Stat Papers (2014) 55:627–642
DOI 10.1007/s00362-013-0509-y

REGULAR ARTICLE

Finite mixture modeling of censored regression models

Maria Karlsson · Thomas Laitila

Received: 14 March 2012 / Revised: 7 February 2013 / Published online: 9 March 2013

Abstract A finite mixture of Tobit models is suggested for estimation of regression models with a censored response variable. A mixture of models is not primarily adapted due to a true component structure in the population; the flexibility of the mixture is suggested as a way of avoiding non-robust parametrically specified models. The new estimator has several interesting features. One is its potential to yield valid estimates in cases with a high degree of censoring. The estimator is in a Monte Carlo simulation compared with earlier suggestions of estimators based on semi-parametric censored regression models. Simulation results are partly in favor of the proposed estimator and indicate potentials for further improvements.

Keywords Finite mixture models · Censoring · Tobit · EM-algorithm

1 Introduction

A frequent problem in regression analysis is the occurrence of censored observations of the dependent variable. In failure and survival time studies, the termination of a study may leave units without observed failures or deaths. For these units, the data information at hand are the eventual occurrence of failures and deaths after the time of closure of the study. In econometric applications, regression analysis based on censored observations are applied in cases where observations are limited to

M. Karlsson (✉)
Department of Statistics, USBE, Umeå University, Umeå, Sweden
e-mail: maria.karlsson@stat.umu.se

T. Laitila
Department of Statistics, Örebro University, Örebro, Sweden

T. Laitila
Statistics Sweden, Örebro, Sweden

non-negative values, providing a large fraction of zeros representing unobservable negative values.

Maximum likelihood (ML) estimation was an early suggestion as an alternative to the least squares estimator which is biased and inconsistent under censoring. Tobin (1958) and Glasser (1965) proposed ML estimation under a normal distribution assumption, where the model studied by Tobin (1958) is generally known as the "Tobit model". Amemiya (1973) derived results on consistency and asymptotic normality of the ML estimator under normality. A related estimator is the two-step estimator developed by Heckman (1976) under normality.

A general problem with the ML estimator is its sensitivity to model misspecification (e.g., White 1982). The potential inconsistency of the ML estimator due to non-normality has encouraged researchers to develop estimators based on less restrictive assumptions. There is a rich variety of suggestions of estimators for semi- and non-parametric censored regression models (Miller 1976; Buckley and James 1979; Powell 1984, 1986a,b; Horowitz 1986, 1988; Lee 1992; Honoré and Powell 1994; Khan and Powell 2001; Lewbel and Linton 2002; Karlsson 2006; Huang et al. 2007; Abarin and Wang 2009; Moral-Arce et al. 2011).

For the purpose of detecting market segments of customers, Jedidi et al. (1993) consider a finite mixture of censored regression models based on a normal distribution assumption. The idea of a finite mixing of distributions dates back to Pearson (1894) who tried to establish evidence of two species of crabs in one data set of measurements. However, one problem facing Pearson was the difficulty in distinguishing between a finite mixture of symmetric distributions and a single asymmetric distribution (Pearson 1894). This result implies that a finite mixture of distributions can be used as an approximation of an unknown distribution. Or as stated by McLachlan and Peel (2000, p. 1), "But as any continuous distribution can be approximated arbitrary well by a finite mixture of normal densities with common variance (...), mixture models provide a convenient semiparamatric framework in which to model unknown distributional shapes, whatever the objective..."

Focus in this paper is placed on the potentials of using the model by Jedidi et al. (1993) for estimation of censored regressions models. The mixture is utilized as a means for modeling an unknown distribution, and is not primarily motivated by data containing observations from different populations. Adapting the model by Jedidi et al. (1993) for estimation of censored regression models extends the partially adaptive estimator suggested by Caudill (2012), where the distribution of the disturbance term, with the constant term added, is modeled by a mixture of normal distributions. Here this model is extended to also include a mixture of slope coefficients.

This paper presents results from Monte Carlo simulations comparing estimates derived from a finite mixture of Tobit models with those derived from estimators compared by Moon (1989) and Honoré and Powell (1994), and the partially adaptive estimator of Caudill (2012). The simulation study also includes comparisons of estimators derived from finite mixture models extended with skedastic functions to address potential heteroskedasticity.

Censored regression models and the finite mixture of Tobit models proposed by Jedidi et al. (1993) are described in the next section. Section 3 contains a description of the EM-algorithm used for estimation of the finite mixture of Tobit models. Setups and results of the Monte Carlo studies performed are contained in Sect. 4. An empirical illustration is given in Sect. 5, while the final section includes a discussion on results and suggestions for further research.

2 Finite mixture of Tobit models

Let (y_i, x_i^T), $(i = 1, \ldots, n)$, denote n pairs of observations generated from n independent distributions where y_i is an observation of the scalar, real valued, random variable $Y_i = max(0, Y_i^*)$, x_i is an observation of the p-dimensional, real valued, random vector X_i, and $Y_i^* | (X_i = x_i) \sim N(x_i^T \beta_0, \sigma_0^2)$. Here the vector of parameters β_0 and the variance σ_0^2 are unknown population quantities. The marginal distribution of the vector X_i is assumed not to depend on β_0 or σ_0^2, and due to the normal assumption, $\phi((y_i^* - x_i^T \beta_0)/\sigma_0)$ is the density for Y_i^* conditionally on $X_i = x_i$.

This model is the Tobit model (Tobin 1958), where observations of the dependent variable in the classical linear regression model are left censored at zero. The Tobit model can be extended to represent cases with right censoring and/or different censoring points by appropriate transformations where the censoring points are assumed fixed and known.

The conditional probability of the event $Y_i = 0$ equals $\Phi(-x_i^T \beta_0/\sigma_0)$. The pdf of the conditional distribution for Y_i can then be written as

$$f(y_i : \beta_0, \sigma_0^2) = \phi((y_i - x_i^T \beta_0)/\sigma_0)^{d_i} \Phi(-x_i^T \beta_0/\sigma_0)^{(1-d_i)}$$

where $d_i = 1$ if $y_i > 0$ and $d_i = 0$ if $y_i = 0$. The log-likelihood based on the set of n observations can then be written as

$$\log L_n(\beta, \sigma^2) = \sum_{i=1}^{n} \log f(y_i : \beta, \sigma^2)$$

In the approach suggested by Jedidi et al. (1993) to define a finite mixture Tobit (FMT) model, the conditional distribution of Y_i is assumed to be defined by a mixture of $K > 1$ distributions with pdfs of the form of $f(.)$. To handle this, let $B = \{\beta_1, \beta_2, \ldots, \beta_K\}$ denote a set of K different parameter vectors of lengths p, and let $\Sigma = \{\sigma_1^2, \sigma_2^2, \ldots, \sigma_K^2\}$ be a set of variances. Also, let $\Lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_K\}$ denote a set of scalar weights satisfying $0 < \lambda_k < 1$ and $\sum_{k=1}^{K} \lambda_k = 1$. These sets are combined into the set $\Psi = \{B, \Sigma, \Lambda\}$ defined for a parameter space $\Psi \in \Xi$.

Now, let $\Psi_0 \in \Xi$ be a specific point in the parameter space and let $f(y : \beta_{k0}, \sigma_{k0}^2)$ denote the pdf of the k:th component in the mixture. Then the conditional pdf for Y_i in the FMT model is obtained as

$$f(y : \Psi_0) = \sum_{k=1}^{K} \lambda_{k0} f(y : \beta_{k0}, \sigma_{k0}^2)$$

With this FMT model, the parameter vector β_0 in the censored regression model is defined by the weighted sum

$$\beta_0 = \sum_{k=1}^{K} \lambda_{k0}\beta_{k0}.$$

The ML estimator is defined as the value $\hat{\Psi} \in \Xi$ which maximizes the log-likelihood function

$$\log L(\Psi) = \sum_{i=1}^{n} \log \sum_{k=1}^{K} \lambda_k f(y : \beta_k, \sigma_k^2) \tag{1}$$

The ML estimator of the parameter vector β_0 in the censored regression model is defined as

$$\hat{\beta} = \sum_{k=1}^{K} \hat{\lambda}_k \hat{\beta}_k \tag{2}$$

3 Estimation

The inclusion of a log of a sum in the definition of the log-likelihood function (1) makes it difficult to numerically solve for the ML estimates. A trick to overcome the computational complexities is to assume that the population actually consists of K subpopulations and each observation is obtained from one of these subpopulations. The information on the subpopulation from which the observation is obtained is missing.

Let W_i be a component membership vector, i.e., its elements W_{ik} are defined to be either 1 or 0 if the component of origin of Y_i is K or not. W_i cannot be observed but has a multinomial distribution consisting of one draw from K categories with probabilities Λ. The complete-data log-likelihood is then

$$\begin{aligned}\log L_{compl}(\Psi) = \sum_{i=1}^{n}\sum_{k=1}^{K} w_{ik}\Bigg[&\log\lambda_k + d_i \log\left(\frac{1}{\sigma_k}\phi\left(\frac{y_i - x_i^T\beta_k}{\sigma_k}\right)\right)\\ &+(1-d_i)\log\left(\Phi\left(\frac{-x_i^T\beta_k}{\sigma_k}\right)\right)\Bigg]\end{aligned} \tag{3}$$

Taking the conditional expectation of the complete-data log-likelihood given the observed data and the current estimate (or guess) of Ψ is the E-step of the EM-algorithm. Let $\Psi^{(0)}$ be the initial guess for Ψ and let

$$Q(\Psi|\Psi^{(s)}) = E_{\Psi^{(s)}}(\log L_{compl}(\Psi))$$

denote the conditional expectation of (3) in the $(s + 1)$th iteration. Because the complete-data log-likelihood is linear in w_{ik},

$$
\begin{aligned}
E_{\Psi^{(s)}}(W_{ik}|y) = P(W_{ik} = 1|y) &= \tau_k^{(s)}(y_i|\Psi^{(s)}) \\
&= \lambda_k^{(s)} \frac{f_k(y_i|\beta_k^{(s)}, \sigma_k^{(s)})}{\sum_{k=1}^K \lambda_k^{(s)} f_k(y_i|\beta_k^{(s)}, \sigma_k^{(s)})}
\end{aligned} \tag{4}
$$

(for details see McLachlan and Peel 2000, Section 2.8) and the E-step is completed by replacing w_{ij} in (3) with its conditional expectation (4) so that

$$
\begin{aligned}
Q(\Psi|\Psi^{(s)}) = \sum_{i=1}^n \sum_{k=1}^K \tau_k^{(s)}(y_i|\Psi^{(s)}) \Bigg[&\log \lambda_k + d_i \log\left(\frac{1}{\sigma_k} \phi\left(\frac{y_i - x_i^T \beta_k}{\sigma_k} \right)\right) \\
&+ (1 - d_i) \log\left(\Phi\left(\frac{-x_i^T \beta_k}{\sigma_k} \right)\right)\Bigg]
\end{aligned} \tag{5}
$$

Note that, with an assumption of a population divided into components (subpopulations), $\tau_k(y_i)$ can be interpreted as the estimated probability of an observation belonging to the kth component of the population.

The M-step on the $(s + 1)$th iteration requires calculation of the global maximum of (5) with respect to Ψ to give the updated estimate $\Psi^{(s+1)}$. The updated estimates of the mixing weights are calculated independently of the other updated estimates of B and Σ. Thus, the λ_k on the $(s + 1)$th iteration is

$$
\lambda_k^{(s+1)} = \frac{1}{n} \sum_{i=1}^n \tau_k^{(s)}(y_i|\Psi^{(s)}).
$$

The updated estimates of B and Σ can be computed using a optimization algorithm, maximizing

$$
\sum_{i=1}^n \sum_{k=1}^K \tau_k^{(s)}(y_i|\Psi^{(s)}) \log f_k(y_i|\beta_k, \sigma_k)
$$

with respect to (B, Σ) or by using another EM-algorithm. This latter approach is utilized by Jedidi et al. (1993) by recognizing that the non-positive y_i^* are unobservable and using a "minor EM-algorithm" within the M-step of the EM-algorithm described above.

One property of the EM-algorithm is its monotone increase in the log-likelihood value over a sequence of iterations. This implies that the EM-algorithm will converge to the ML estimate if the likelihood is strictly concave within a neighborhood of the ML estimate and the starting point of the algorithm is within this neighborhood. Unfortunately the likelihood function of a finite mixture of Tobit models is not globally concave over the parameter space and the EM iterations may end up in undesirable points as local maxima. The suggested remedy for this is to try out several different starting points (Wu 1983).

The existence of local maxima for a finite mixture of Tobit models is implied by the results of Amemiya (1973), who showed that the log-likelihood of the Tobit model is not globally concave. Finite mixtures of normal distributions also presents the problem of an unbounded log-likelihood function (McLachlan and Peel 2000; Caudill 2012), which is made possible by variances not bounded away from zero. The suggested way of avoiding this problem is to restrict the parameter space for the variances. This might also have implications for the identification of the ML estimate within the defined parameter space, since Amemiya (1973) shows the Tobit log-likelihood to be globally concave if it is defined conditionally on a specified value of the variance. This topic and the properties of the EM-algorithm deserve further attention but are out of the scope of the present paper.

4 Simulation

Honoré and Powell (1994) compared several estimators of the censored regression model under some different designs. Here, a selection of the their designs are used to evaluate the performance of using a finite mixture of two Tobit models, i.e., a FMT with $K = 2$, and estimating β_0 via the EM-algorithm and weighting the estimated component regression coefficient as described in (2). For the M-step maximization the Nelder–Mead algorithm included in the package `optim` in `R` (R Core Team 2012) was used.

For comparison, the symmetrically censored least squares (SCLS) estimator (Powell 1986b) and the censored least absolute deviations (CLAD) estimator (Powell 1984) are included. The SCLS is based on symmetric censoring of the upper tail of the distribution of the response variable to compensate for the censoring in the lower tail. The CLAD estimator minimizes the sum of absolute deviations between the observed responses and $\max(0, x_i^T \beta)$. Both estimators are consistent and asymptotically normal for a wide class of distributions of the error term and also robust to heteroskedasticity. However, for consistency of the SCLS estimator the error term distribution needs to be symmetric. The optimizations required to compute these estimates were also done by the Nelder–Mead algortithm in `R`.

The simulation also includes the partially adaptive estimator for the censored regression model (PAM), suggested by Caudill (2012). In terms of the FMT estimator, the PAM estimator equals the FMT with the restriction of equal slope coefficients over mixture components, i.e., if β_k^s denotes the slope coefficients in β_k, then $\beta_k^s = \beta_l^s$ for all $k, l \in \{1, 2, \ldots, K\}$.

Data is generated from the following model

$$Y_i = \max(0, X_i^T \beta_0 + \varepsilon), \tag{6}$$

where $X_i = (1, X_{1i})$ and X_{1i} is uniformly distributed on $[0, 20]$, the slope coefficient of β_0 equals 1. The error term ε is generated according to the five designs listed below and the intercept is chosen such that the censoring rate is either 25, 50, or 70 %.

The designs considered are:

- Design 1: Laplace. The errors term, ε, is Laplace distributed with mean 0 and variance 100.
- Design 2: Normal mixture (right skewed). The error term, ε, is normally distributed with mean 12 and variance 220 with probability 0.2 and normally distributed with mean -3 and variance 25 with probability 0.8. This yields a right skewed distribution.
- Design 3: Gumbel (right skewed). The error term, ε, is Gumbel distributed with location $\sqrt{600/\pi^2}\gamma$ and scale $\sqrt{600/\pi^2}$, where γ is the Euler-Mascheroni constant $= 0.57721\ldots$ This means that the mean of ε is 0 and variance is 100 as it is in the other models.
- Design 4: Heteroskedastic errors ($\varepsilon = \sqrt{(ce^x)} \cdot N(0, 1)$). The error term $\varepsilon = \sqrt{(ce^x)} \cdot N(0, 1)$ with $c = 4.122307 \times 10^{-6}$ chosen so that the average variance of ε is 100 as it is in the other models.
- Design 5: Heteroskedastic errors ($\varepsilon = \sqrt{(\alpha_0 + \alpha_1 \cdot x)} \cdot N(0, 1)$). The error term $\varepsilon = \sqrt{(\alpha_0 + \alpha_1 \cdot x)} \cdot N(0, 1)$ where $\alpha_0 = 50$ and $\alpha_1 = 5$ chosen so that the average variance of ε is 100 as it is in the other models.

Similar to Honoré and Powell (1994), sample sizes of 200 and 800 are included and, in addition, $n = 2,000$ is also considered. Another difference in the study design compared to the design of Honoré and Powell (1994) is the inclusion of the higher censoring rate, i.e., 70 % censoring. This is a common situation when measuring the time to a rare event where many individuals will not experience the event before the end of the study period, or when measuring consumer desired spendings on a particular brand where a majority do not purchase the brand at all (e.g., Marell et al. 2004; Jedidi et al. 1993; Fack and Landais 2010).

4.1 Results

Samples from designs 1–3 is replicated 1,000 times and the results in terms of average and median bias, root mean square error (RMSE), and median absolute deviation (MAD) of the estimators of the slope coefficient are found in Tables 1, 2, and 3. A general result for all estimators is decreasing bias, RMSE, and MAD with increasing sample size and decreasing censoring rate.

Apparent is that when the censoring rate is high the FMT and PAM estimators are better than their semi-parametric competitors. When the censoring rate is 70 % they out-perform the two other estimators in terms of bias, RMSE, and MAD. The SCLS estimator does worst and is associated with large bias estimates, especially when the sample size is small.

At the lower censoring rate of 50 %, the FMT and PAM are generally associated with better results then the SCLS and CLAD estimators in terms of RMSE and MAD. Again the SCLS estimator does worst while, e.g., bias estimates are notably smaller than in the 70 % censoring case. In some cells, the biases of the SCLS are on par with those of the other estimators. In the Laplace and Gumbel distribution cases, the CLAD estimator has the smallest bias estimates (average and median bias) among all the four

Table 1 Average and median bias, RMSE, and MAD of the estimators of the slope coefficient in a Monte Carlo simulation of 1,000 replicates from Design 1 (Laplace).

	Estimator	Average bias	RMSE	Median bias	MAD
25 % censoring, $n = 200$	FMT	0.025	0.169	0.020	0.101
	SCLS	0.051	0.253	0.003	0.131
	CLAD	0.018	0.147	0.003	0.091
	PAM	0.032	0.127	0.028	0.081
25 % censoring, $n = 800$	FMT	0.015	0.075	0.012	0.050
	SCLS	0.012	0.101	0.002	0.063
	CLAD	0.007	0.073	0.002	0.046
	PAM	0.026	0.070	0.023	0.045
25 % censoring, $n = 2,000$	FMT	0.009	0.048	0.006	0.031
	SCLS	−0.000	0.058	−0.003	0.040
	CLAD	−0.001	0.046	−0.002	0.028
	PAM	0.015	0.044	0.012	0.028
50 % censoring, $n = 200$	FMT	0.042	0.206	0.028	0.124
	SCLS	0.189	0.770	0.038	0.231
	CLAD	0.049	0.295	−0.009	0.151
	PAM	0.051	0.167	0.040	0.105
50 % censoring, $n = 800$	FMT	0.021	0.088	0.017	0.056
	SCLS	0.032	0.194	−0.003	0.108
	CLAD	0.015	0.139	−0.006	0.078
	PAM	0.038	0.097	0.032	0.061
50 % censoring, $n = 2,000$	FMT	0.014	0.055	0.015	0.012
	SCLS	0.005	0.109	−0.007	0.073
	CLAD	−0.001	0.077	−0.005	0.052
	PAM	0.021	0.065	0.013	0.038
70 % censoring, $n = 200$	FMT	0.072	0.347	0.042	0.150
	SCLS	2.630	19.67	0.102	0.418
	CLAD	0.193	0.913	−0.003	0.242
	PAM	0.063	0.215	0.045	0.139
70 % censoring, $n = 800$	FMT	0.021	0.105	0.020	0.069
	SCLS	0.151	0.611	0.013	0.206
	CLAD	0.039	0.273	−0.020	0.137
	PAM	0.040	0.122	0.035	0.086
70 % censoring, $n = 2,000$	FMT	0.010	0.063	0.009	0.041
	SCLS	0.049	0.261	0.010	0.135
	CLAD	0.016	0.173	−0.013	0.099
	PAM	0.017	0.080	0.012	0.049

estimators for the larger sample sizes. Results of the FMT and PAM are close where the average bias and RMSE tend to be smaller for the FMT for the larger sample sizes.

In the cases of a 25 % censoring degree, results for the FMT, PAM, and CLAD estimators are comparable and better than those of the SCLS estimator over all designs and sample sizes. Bias and RMSE estimates are all small in general.

Simulation results obtained under the heteroskedasticity designs 4 and 5 are depicted in Tables 4 and 5. The relative performance of estimators according to the results are somewhat different from those indicated by Tables 1, 2, and 3. Large bias and RMSE estimates are observed for the FMT and PAM estimators when the variance function is of a multiplicative form (Table 4), while results in general associates small bias and RMSE for the CLAD estimator. Small bias and RMSE estimates are also obtained for the SCLS estimator in the 25 % censoring case. With an additive

Table 2 Average and median bias, RMSE, and MAD of the estimators of the slope coefficient in a Monte Carlo simulation of 1,000 replicates from Design 2 (Normal mixture).

	Estimator	Average bias	RMSE	Median bias	MAD
25 % censoring, $n = 200$	FMT	0.002	0.129	0.001	0.084
	SCLS	0.118	0.254	0.082	0.125
	CLAD	0.012	0.157	−0.002	0.096
	PAM	0.001	0.093	0.002	0.0193
25 % censoring, $n = 800$	FMT	0.004	0.059	0.001	0.040
	SCLS	0.103	0.141	0.095	0.096
	CLAD	0.008	0.075	0.002	0.048
	PAM	0.005	0.046	0.002	0.031
25 % censoring, $n = 2,000$	FMT	−0.002	0.036	−0.001	0.023
	SCLS	0.091	0.109	0.089	0.089
	CLAD	0.000	0.045	−0.003	0.030
	PAM	0.001	0.029	0.002	0.019
50 % censoring, $n = 200$	FMT	0.007	0.175	−0.003	0.107
	SCLS	0.390	2.281	0.025	0.294
	CLAD	0.039	0.423	−0.032	0.202
	PAM	0.008	0.166	−0.007	0.105
50 % censoring, $n = 800$	FMT	0.007	0.078	0.006	0.051
	SCLS	0.126	0.327	0.078	0.150
	CLAD	0.026	0.200	−0.009	0.119
	PAM	0.009	0.075	0.006	0.050
50 % censoring, $n = 2,000$	FMT	−0.001	0.046	−0.002	0.030
	SCLS	0.082	0.171	0.070	0.103
	CLAD	0.009	0.119	−0.003	0.076
	PAM	0.002	0.046	0.001	0.031
70 % censoring, $n = 200$	FMT	0.043	0.347	0.001	0.180
	SCLS	6.287	35.21	−0.429	6.288
	CLAD	−0.021	2.370	−0.266	1.524
	PAM	0.027	0.307	−0.006	0.183
70 % censoring, $n = 800$	FMT	0.016	0.136	0.004	0.082
	SCLS	2.583	43.47	−0.194	1.524
	CLAD	−0.041	1.535	−0.140	0.481
	PAM	0.016	0.135	0.010	0.084
70 % censoring, $n = 2,000$	FMT	0.002	0.079	0.001	0.049
	SCLS	0.437	7.931	−0.165	0.803
	CLAD	−0.074	0.935	−0.109	0.301
	PAM	0.005	0.083	0.005	0.054

variance function (Table 5) and 25 % censoring, bias and RMSE estimates are small for all estimators. For 50 % censoring, only CLAD has small bias and RMSE estimates. None of the estimators works well under 70 % censoring according to the results in Table 5.

4.2 Heteroskedastic components

In an effort to improve the performance of the FMT and PAM estimators in the heteroskedastic cases considered, a skedastic variance function is included in the regression models. The function used is $\sigma_k = \exp(\alpha_{0k} + \alpha_{1k}x)$ in (1). The results in terms of bias and RMSE of the regression slope estimators for the simulation designs 4 and 5

Table 3 Average and median bias, RMSE, and MAD of the estimators of the slope coefficient in a Monte Carlo simulation of 1,000 replicates from Design 3 (Gumbel).

	Estimator	Average bias	RMSE	Median bias	MAD
25 % censoring, $n = 200$	FMT[1]	0.046	0.170	0.037	0.105
	SCLS	0.152	0.335	0.091	0.150
	CLAD	0.028	0.213	−0.007	0.128
	PAM	0.053	0.155	0.046	0.100
25 % censoring, $n = 800$	FMT	0.040	0.087	0.036	0.056
	SCLS	0.103	0.162	0.089	0.096
	CLAD	0.012	0.106	0.005	0.070
	PAM	0.044	0.092	0.039	0.058
25 % censoring, $n = 2,000$	FMT	0.036	0.061	0.032	0.040
	SCLS	0.087	0.112	0.084	0.085
	CLAD	0.001	0.068	−0.002	0.043
	PAM	0.027	0.062	0.020	0.037
50 % censoring, $n = 200$	FMT[1]	0.062	0.233	0.039	0.136
	SCLS	0.395	1.253	0.135	0.302
	CLAD	0.084	0.429	−0.028	0.217
	PAM	0.060	0.205	0.050	0.134
50 % censoring, $n = 800$	FMT	0.037	0.100	0.032	0.066
	SCLS	0.134	0.328	0.074	0.147
	CLAD	0.029	0.212	−0.010	0.123
	PAM	0.053	0.122	0.048	0.081
50 % censoring, $n = 2,000$	FMT	0.031	0.067	0.028	0.043
	SCLS	0.085	0.178	0.067	0.101
	CLAD	0.001	0.121	−0.009	0.083
	PAM	0.036	0.086	0.029	0.053
70 % censoring, $n = 200$	FMT[1]	0.089	0.511	0.047	0.178
	SCLS	5.321	34.52	0.122	0.717
	CLAD	0.504	4.217	0.002	0.441
	PAM	0.064	0.246	0.045	0.158
70 % censoring, $n = 800$	FMT	0.035	0.128	0.034	0.087
	SCLS	0.977	5.643	0.070	0.403
	CLAD	0.204	0.807	−0.040	0.237
	PAM	0.055	0.142	0.051	0.093
70 % censoring, $n = 2,000$	FMT	0.023	0.079	0.023	0.054
	SCLS	0.525	3.004	0.077	0.299
	CLAD	0.096	0.470	−0.016	0.191
	PAM	0.041	0.098	0.037	0.066

[1] For sample size $n = 200$ the FMT did not converge in a few replicates of the simulation study and the results are based on 999, 999, and 997 replicates for the three different censoring levels respectively

are reported in Tables 6 and 7 respectively. Estimators defined by the heteroskedastic mixtures are denoted FMT.vf and PAM.vf, respectively.

As is indicated by the results, the modification of the FMT and PAM estimators drastically improves their performance under heteroskedasticity. Average biases are small and in comparison with the CLAD estimator, results are similar to those in Tables 1, 2, and 3. The FMT.vf and PAM.vf estimators works better than the CLAD estimator for large censoring, and are generally associated with smaller RMSE and MAD values. The results for the FMT.vf and PAM.vf estimators are comparable with one exception. In Table 6 under 70 % censoring, the PAM.vf has large bias and RMSE estimates, while the FMT.vf estimator has small bias and RMSE estimates.

Table 4 Average and median bias, RMSE, and MAD of the estimators of the slope coefficient in a Monte Carlo simulation of 1,000 replicates from Design 4 (Heteroskedastic errors, $\varepsilon = \sqrt{(ce^x)} \cdot N(0, 1)$).

	Estimator	Average bias	RMSE	Median bias	MAD
25 % censoring, $n = 200$	FMT	0.531	0.607	0.531	0.531
	SCLS	0.076	0.361	0.012	0.136
	CLAD	0.001	0.032	0.001	0.018
	PAM	0.090	0.117	0.087	0.087
25 % censoring, $n = 800$	FMT	0.552	0.569	0.569	0.569
	SCLS	0.007	0.104	−0.004	0.068
	CLAD	−0.000	0.013	−0.000	0.009
	PAM	0.140	0.151	0.132	0.132
25 % censoring, $n = 2,000$	FMT	0.560	0.565	0.563	0.563
	SCLS	0.001	0.062	−0.002	0.040
	CLAD	0.000	0.008	−0.000	0.005
	PAM	0.168	0.174	0.165	0.165
50 % censoring, $n = 200$	FMT	1.218	1.430	1.124	1.124
	SCLS	1.664	11.06	0.024	0.271
	CLAD	0.022	0.187	0.001	0.099
	PAM	0.284	0.357	0.262	0.262
50 % censoring, $n = 800$	FMT	1.238	1.268	1.263	1.263
	SCLS	0.070	1.153	−0.006	0.129
	CLAD	0.001	0.077	−0.001	0.052
	PAM	0.349	0.371	0.341	0.341
50 % censoring, $n = 2,000$	FMT	1.260	1.269	1.262	1.262
	SCLS	0.009	0.125	−0.004	0.078
	CLAD	0.001	0.046	−0.003	0.030
	PAM	0.418	0.432	0.407	0.407
70 % censoring, $n = 200$	FMT	3.113	4.112	2.402	2.402
	SCLS	10.65	117.9	0.062	0.661
	CLAD	0.286	1.244	0.015	0.310
	PAM	1.461	1.637	1.394	1.394
70 % censoring, $n = 800$	FMT	2.936	3.114	3.038	3.038
	SCLS	1.610	10.79	0.008	0.308
	CLAD	0.040	0.282	0.005	0.167
	PAM	1.254	1.301	1.232	1.232
70 % censoring, $n = 2,000$	FMT	3.020	3.121	3.233	3.233
	SCLS	0.185	1.085	−0.007	0.189
	CLAD	0.009	0.160	−0.005	0.106
	PAM	1.253	1.272	1.250	1.250

In case of homoskedasticity, the FMT.vf and PAM.vf estimators are oveparameterized which might affect their properties. In Table 8, simulation results for these estimators under designs 1–3 and $n = 200$ is presented. Results shows on large bias and RMSE results for the estimators.

5 An empirical example: the Mroz data

To illustrate the usefulness of the FMT estimator, the FMT model is applied to data on annual work hours of married women studied by Mroz (1987), which is available, e.g., in the package `Ecdat` (Croissant 2011) in `R` (R Core Team 2012).

Table 5 Average and median bias, RMSE, and MAD of the estimators of the slope coefficient in a Monte Carlo simulation of 1,000 replicates from Design 5 (Heteroskedastic errors, $\varepsilon = \sqrt{(}\alpha_0 + \alpha_1 \cdot x) \cdot N(0, 1)$).

	Estimator	Average bias	RMSE	Median bias	MAD
25 % censoring, $n = 200$	FMT	0.060	0.148	0.061	0.098
	SCLS	0.055	0.251	0.009	0.128
	CLAD	0.027	0.193	0.008	0.102
	PAM	0.080	0.157	0.082	0.105
25 % censoring, $n = 800$	FMT	0.052	0.086	0.051	0.061
	SCLS	0.009	0.098	0.003	0.063
	CLAD	0.009	0.087	0.004	0.057
	PAM	0.076	0.102	0.074	0.077
25 % censoring, $n = 2,000$	FMT	0.053	0.068	0.051	0.052
	SCLS	0.004	0.060	0.002	0.040
	CLAD	0.004	0.056	−0.000	0.038
	PAM	0.078	0.089	0.077	0.077
50 % censoring, $n = 200$	FMT	0.156	0.247	0.147	0.159
	SCLS	0.518	2.830	0.064	0.325
	CLAD	0.116	0.492	0.008	0.208
	PAM	0.189	0.251	0.185	0.189
50 % censoring, $n = 800$	FMT	0.147	0.170	0.145	0.145
	SCLS	0.070	0.327	0.000	0.155
	CLAD	0.036	0.225	0.008	0.125
	PAM	0.186	0.203	0.182	0.182
50 % censoring, $n = 2,000$	FMT	0.157	0.168	0.161	0.161
	SCLS	0.019	0.166	−0.002	0.101
	CLAD	0.015	0.136	0.001	0.084
	PAM	0.188	0.195	0.188	0.188
70 % censoring, $n = 200$	FMT	0.268	0.568	0.227	0.234
	SCLS	8.354	44.79	0.370	1.612
	CLAD	1.502	6.159	0.189	0.674
	PAM	0.299	0.371	0.291	0.291
70 % censoring, $n = 800$	FMT	0.256	0.283	0.257	0.257
	SCLS	2.626	11.83	0.167	0.567
	CLAD	0.492	1.943	0.055	0.331
	PAM	0.297	0.316	0.294	0.294
70 % censoring, $n = 2,000$	FMT	0.278	0.286	0.276	0.276
	SCLS	1.380	8.981	0.079	0.425
	CLAD	0.226	1.022	−0.008	0.249
	PAM	0.299	0.306	0.298	0.298

The data set contains data on hours worked outside the home for 753 married women, of whom 325 worked zero hours, and data on several other characteristics of the women. In Caudill (2012), the PAM estimator and the Tobit estimator are both applied to the Mroz data. However, the results presented below and in Caudill (2012) are not directly comparable because in the data set in `Ecdat` there is no data on the "nonwife income", i.e., family income exclusive of the wifes income, as was included in the model of Caudill (2012).

Here the explantory variables included in the model are the wife's education in years (*educw*), the wife's previous labor market experience (*exper*), the wife's previous labor market experience squared (*exper**exper*) and the wife's age (*agew*). The response variable is the wife's annual work hours outside the home. A two component finite

Table 6 Results for the FMT.vf and PAM.vf estimators of the slope coefficient in a Monte Carlo simulation of 1,000 replicates from Design 4 (Heteroskedastic errors, $\varepsilon = \sqrt{(ce^x)} \cdot N(0, 1)$).

	FMT.vf		PAM.vf	
	Average bias	RMSE	Average bias	RMSE
25 % censoring, $n = 200$	−0.005	0.010	0.000	0.006
25 % censoring, $n = 800$	−0.007	0.008	−0.000	0.003
25 % censoring, $n = 2,000$	−0.007	0.008	0.000	0.002
50 % censoring, $n = 200$	0.040	0.061	0.002	0.050
50 % censoring, $n = 800$	0.040	0.046	−0.001	0.027
50 % censoring, $n = 2,000$	0.041	0.044	−0.002	0.020
70 % censoring, $n = 200$	0.058	0.193	0.165	0.284
70 % censoring, $n = 800$	0.033	0.093	0.131	0.184
70 % censoring, $n = 2,000$	0.030	0.060	0.124	0.161

Table 7 Results for the FMT.vf and PAM.vf estimators of the slope coefficient in a Monte Carlo simulation of 1,000 replicates from Design 5 PAM=OK(Heteroskedastic errors, $\varepsilon = \sqrt{(\alpha_0 + \alpha_1 \cdot x)} \cdot N(0, 1)$.

	FMT.vf		PAM.vf	
	Average bias	RMSE	Average bias	RMSE
25 % censoring, $n = 200$	0.012	0.130	0.016	0.134
25 % censoring, $n = 800$	0.006	0.066	0.009	0.067
25 % censoring, $n = 2,000$	0.009	0.041	0.012	0.043
50 % censoring, $n = 200$	0.028	0.184	0.042	0.192
50 % censoring, $n = 800$	0.016	0.089	0.029	0.095
50 % censoring, $n = 2,000$	0.019	0.058	0.031	0.066
70 % censoring, $n = 200$	0.027	0.320	0.062	0.320
70 % censoring, $n = 800$	0.008	0.146	0.044	0.158
70 % censoring, $n = 2,000$	0.018	0.092	0.048	0.110

Table 8 Results for the FMT.vf and PAM.vf estimators of the slope coefficient in a Monte Carlo simulation of 1,000 replicates from models with homoskedastic variance, i.e., Designs 1–3. Sample size, $n = 200$.

		FMT.vf		PAM.vf	
	%	Average bias	RMSE	Average bias	RMSE
Design 1	25	0.123	0.173	0.095	0.168
	50	0.288	0.335	0.238	0.313
	70	0.448	0.509	0.388	0.482
Design 2	25	−0.039	0.098	0.252	0.299
	50	−0.420	0.465	0.711	0.785
	70	0.564	0.658	0.997	1.182
Design 3	25	0.196	0.235	0.154	0.219
	50	0.383	0.428	0.260	0.343
	70	0.575	0.651	0.353	0.497

mixture of Tobit models is used to compute both the FMT and PAM estimates, the latter by restricting the slope coefficients to be the same in the two components. Also the usual (single component) Tobit models are estimated using the package

Table 9 Estimation results for the Mroz data

	FMT	SE	PAM	SE	Tobit	SE
constant	77.767	27.65	358.147	15.10	−55.738	403.46
educw	49.564	10.65	39.182	6.39	53.880	21.16
exper	134.104	14.32	151.341	8.37	151.558	18.01
exper*exper	−1.917	0.51	−2.561	0.24	−2.279	0.56
agew	−35.222	3.78	−29.469	2.15	−35.460	6.70
σ	946.948		1051.411		1184.418	
λ_1	0.879		0.793		–	
Loglikelihood value	−3819.95		−3843.80		−3855.79	
AIC	7665.90		7705.60		7723.58	
CAIC	7726.03		7747.22		7751.34	
BIC	7726.01		7747.21		7751.33	
ABIC	7684.75		7718.65		7732.28	

`censReg` (Henningsen 2012) in `R`. The standard errors of the FMT and PAM estimates are estimated by delete-one jackknife. The standard errors of the Tobit estimates are estimated by taking the square root of the diagonal elements of minus the inverse of the hessian matrix recieved from `censReg`.

The parameter estimates (FMT, PAM, and Tobit) are reported in Table 9 and differ quite substantially between estimators. The maximized value of the log-likelihood function is highest for the FMT model, −3,819.95. The values of the PAM and Tobit models are −3,843.80 and −3,855.79 respectively. In Table 9 the informations criteria Akaikes Information Criterion (AIC), consistent AIC (CAIC), Bayesian Information Criterion (BIC), and adjusted BIC (ABIC) are given. The FMT model is the preferred model according to all four of these model selection statistics. The Tobit model has the highest AIC, CAIC, BIC and ABIC values of the three models considered.

6 Discussion

This paper is concerned with the estimation of censored regression models through estimation of a finite mixture of Tobit models. Estimates of the parameters in the censored regression model are defined as weighted sums of the corresponding component estimates. The properties of this FMT estimator are studied by means of simulation where earlier suggested semiparametric estimators, the CLAD and STLS estimators, are included for comparison. Another estimator included is the PAM estimator, which is based on a censored regression model with a finite mixture of normal distributions for the disturbance term.

The overall picture of the simulation study is that the results are promising and the idea of mixing Tobit models can yield estimators with better properties than those of previously suggested estimators. There are however some further developments needed for the theory to work in practice. Our observation is that a finite mixture of Tobit models performs generally better than the other estimators when the censored regression model estimated has homoskedastic disturbances. In case of heteroskedasticity, the approach does not work. However, extending the Tobit models in the mixture with skedastic functions for the variances again yield an estimator with better results than the other estimators. Unfortunately, the skedastic function extended Tobit models do not work under homoskedasticity.

In a comparison of the FMT and the PAM estimators, the results are not clearcut in favour of the FMT estimator. However, it seems as if the extra parameters in a mixture of Tobit models yields an extra flexibility which improves estimator properties. This is conditional on a correct specification with respect to assumed homoskedasticity or heteroskedasticity.

In further developments of the FMT estimator, it is of interest to define an approach which encompasses both homoskedasticty and heteroskedasticity. One direct option is to start with an FMT model including skedastic functions for potential heteroskedasticity. A model specification test, e.g., a likelihood ratio (LR) test, can then be used for discriminating between the FMT and the FMT.vf estimators. Several aspects have to be considered here. One is the anticipated increase in the variance of the estimator, which might rule out the advantages of the FMT estimator observed in this paper. Second, the LR test would be valid if the true model is an FMT mixture. As the mixture of FMT is here suggested as an approximation, the properties of the LR test have to be considered. A third aspect is the form of the skedastic function. Here an exponential function has been used and results imply robustness against a misspecification of the form of heteroskedasticity. However, other forms of skedastic functions can be considered, as well as studying the performance under other forms of heteroskedasticity.

The number of components to include in a mixture of Tobit models has not been addressed in this paper. Suggestions of criteria for choosing the number of components are found in e.g., McLachlan and Peel (2000, Ch. 6). One important aspect here is the use of the finite mixture for approximating an unknown distribution. It is therefore of interest to assess the sufficient number of components rather than the true number of components. An expected shortcoming of an estimator defined with a selection criteria for the number of components is increased variance. Our simulation results indicate that a fixed two components model can be sufficient for defining estimators with good properties in terms of low bias and variance. Similar findings for non-censored data are found in Bartolucci and Scaccia (2005).

References

Abarin T, Wang L (2009) Second-order least squares estimation of censored regression models. J Stat Plan Inference 139:125–135

Amemiya T (1973) Regression analysis when the dependent variable is truncated normal. Econometrica 41:997–1016

Bartolucci F, Scaccia L (2005) The use of mixtures for dealing with non-normal regression errors. Comput Stat Data Anal 48:821–834

Buckley J, James I (1979) Linear regression with censored data. Biometrika 66:429–436

Caudill SB (2012) A partially adaptive estimator for the censored regression model based on a mixture of normal distributions. Stat Methods Appl 21:121–137

Croissant Y (2011) Ecdat: data sets for econometrics. R package version 0.1-6.1
Fack G, Landais C (2010) Are tax incentives for charitable giving efficient? Evidence from France. Am Econ J Econ Policy 2:117–141
Glasser M (1965) Regression analysis with dependent variable censored. Biometrics 21:300–307
Heckman JJ (1976) The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. Ann Econ Social Meas 5:475–492
Henningsen A (2012) censReg: Censored Regression (Tobit) Models. R package version 0.5-16.
Honoré BE, Powell JL (1994) Pairwise difference estimators of censored and truncated regression models. J Econ 64:241–278
Horowitz JL (1986) A distribution-free least squares estimator for censored linear regression models. J Econ 32:59–84
Horowitz JL (1988) Semiparametric M-estimation of censored linear regression models. Adv Econ 7:45–83
Huang J, Ma S, Xie H (2007) Least absolute deviations estimation for the accelerated failure time model. Statistica Sinica 17:1533–1548
Jedidi K, Ramaswamy V, DeSarbo WS (1993) A maximum likelihood method for latent class regression involving a censored dependent variable. Psychometrika 58:375–394
Karlsson M (2006) Estimators of regression parameters for truncated and censored data. Metrika 63: 329–341
Khan S, Powell JL (2001) Two-step estimation of semiparametric censored regression models. J Econ 103:73–110
Lee MJ (1992) Winsorized mean estimator for censored regression. Econ Theory 8:368–382
Lewbel A, Linton O (2002) Nonparametric censored and truncated regression. Econometrica 70:765–779
Marell A, Davidsson P, Gärling T, Laitila T (2004) Direct and indirect effects on households' intentions to replace the old car. J Retail Consumer Serv 11:1–8
McLachlan G, Peel D (2000) Finite mixture models. Wiley, New York
Miller RG (1976) Least squares regression with censored data. Biometrika 63:449–464
Moon C-G (1989) A Monte Carlo comparison of semiparametric Tobit estimators. J Appl Econ 4:361–382
Moral-Arce I, Rodríguez-Póo JM, Sperlich S (2011) Low dimensional semiparametric estimation in a censored regression model. J Multivar Anal 102:118–129
Mroz TA (1987) The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. Econometrica 55:765–799
Pearson K (1894) Contributions to the mathematical theory of evolution. Philos Trans R Soc Lond A 185:71–110
Powell JL (1984) Least absolute deviations estimation for the censored regression model. J Econ 25:303–325
Powell JL (1986a) Censored regression quantiles. J Econ 32:143–155
Powell JL (1986b) Symmetrically trimmed least squares estimation for Tobit models. Econometrica 54:1435–1460
R Core Team (2012) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
Tobin J (1958) Estimation of relationships for limited dependent variables. Econometrica 26:24–36
White H (1982) Maximum likelihood estimation of misspecified models. Econometrica 50:1–25
Wu CF (1983) On the convergence properties of the EM algorithm. Ann Stat 11:95–103