UNIVERSITY OF CAMBRIDGE
Faculty of Economics

Cambridge Working Papers in Economics

Cambridge Working Papers in Economics: 1913

DYNAMIC TOBIT MODELS

Andrew Harvey   Yin Liao

2 February 2019

Score-driven models provide a solution to the problem of modelling time series when the observations are subject to censoring and location and/or scale may change over time. The method applies to generalized-t and EGB2 distributions, as well as to the normal distribution. A set of Monte Carlo experiments show that the score-driven model provides good forecasts even when the true model is parameterdriven. The viability of the new models is illustrated by fitting them to data on Chinese stock returns.

Dynamic Tobit models

Andrew Harvey and Yin Liao
Faculty of Economics, Cambridge University and QUT, Brisbane

February 2, 2019

Abstract

Score-driven models provide a solution to the problem of modeling time series when the observations are subject to censoring and location and/or scale may change over time. The method applies to generalized-t and EGB2 distributions, as well as to the normal distribution. A set of Monte Carlo experiments show that the score-driven model provides good forecasts even when the true model is parameter-driven. The viability of the new models is illustrated by fitting them to data on Chinese stock returns.
KEYWORDS: Censored distributions; dynamic conditional score model; EGARCH models; logistic distribution; generalized t distribution.
JEL classification: C22;C24.

1 Introduction

The Tobit regression model is widely when the dependent variable is subject to censoring; see, for example, Amemiya (1985, ch 10). Here we consider censoring in univariate time series models. A number of researchers, beginning with Zeger and Brookmeyer (1986), have proposed dynamic Tobit models and discussed ways of estimating them. Nearly all these models assume the underlying (uncensored) observations to be Gaussian. Furthermore the dynamics tend to be of the autoregressive or autoregressive-moving average (ARMA) form, as in Park et al (2007) and Wang and Chan (2018), although a recent paper by Allik et al (2015) generalizes to state space models.

The class of dynamic Tobit models proposed here have location given by a filter that depends on the conditional score. Such models, known as Dynamic Conditional Score (DCS) or Generalized Autoregressive Score (GAS) models, have already proved to be highly effective in a number of situations; see, Harvey (2013), Creal et al (2011) and the papers listed in the website http://www.gasmodel.com. In the present context the score has the important feature of automatically solving the problem of how to weight the censored observations. Furthermore the properties of the score mean that the dynamics are driven by a variable which is a martingale difference. Explanatory variables can be included in the models, making static Tobit models a special case.

In the classic Tobit regression model, where location is a linear function of explanatory variables, the uncensored distribution is assumed to be normal. However, whereas for uncensored data Gaussianity leads to least squares in regression and a linear model in time series (and hence the Kalman filter for unobserved components), it no longer yields simple estimation procedures once censoring is introduced. Hence there is no computational disadvantage to adopting other, more flexible, distributions. It is in this spirit that Lewis and McDonald (2014) propose the use of generalized-t and exponential generalized beta of the second kind (EGB2) distributions for censored regression. These distributions may be similarly used as the basis for the dynamic Tobit procedures. The logistic distribution, which is a special case of EGB2, is similar in shape to the normal but when used in a censored DCS model it leads to a particularly simple filter.

Censored score-driven models can also be used when scale is dynamic. For the reasons given in Harvey (2013, ch4) and Harvey and Lange (2018), the preferred specifications are of the exponential generalized autoregressive heteroscdasticity (EGARCH) form.

Score-driven models are observation-driven, rather than parameter-driven. Hence estimation can be carried out relatively simply by maximizing a likelihood function. By contrast, parameter-driven models tend to require computationally intensive techniques, such as Markov chain monte Carlo (MCMC) or particle filters; see the comments in Allik et al (2015). A recent paper by Koopman et al (2016) shows that observation-driven models can provide a good approximation to parameter-driven models, but the converse is not generally true. We find the same here.

The plan of the paper is as follows. Section 2 sets out the basic DCS model for censored observations and Section 3 gives the expressions for the conditional scores for generalized t, EGB2, Gaussian and generalized beta of the second kind (GB2) distributions. Specification and prediction are discussed in Section 4, while in Section 5 a small Monte Carlo study compares the performance of DCS models with the corresponding parameter-driven models. The penultimate section fits DCS models with time-varying scale to censored Chinese stock prices.

2 Censored observations

Let x be a variable with location μ, scale $\varphi = \exp(\lambda)$ and PDF f_x. The observations subject to censoring are defined as

$$y_t = \begin{cases} x_t, & c_L < x_t < c_U, \\ c_L, & x_t \leq c_L, \\ c_U, & x_t \geq c_U \end{cases}, \qquad t = 1, ..., T. \tag{1}$$

For generality we have assumed the possibility of censoring in both tails. The lower and upper bounds, c_L and c_U, are taken to be known. The probabilities of getting censored observations are $\Pr(y_t = c_L) = F_x(c_L)$ and $\Pr(y_t = c_L) = 1 - F_x(c_U)$, where F_x is the CDF of x_t.

Let I_L be an indicator that is one when $y = c_L$ and zero when $y > c_L$. Similarly let I_U be one when $y = c_U$ and zero when $y < c_U$. The distribution of y_t, that is $f(y_t; c_L, c_U)$, is a discrete-continuous mixture and the log-likelihood for the $t - th$ observation is

$$\ln f(y_t; c_L, c_U) = I_L \ln F_x(c_L) + (1 - I_L - I_U) \ln f_x(y_t) + I_U \ln(1 - F_x(c_U)). \tag{2}$$

The score with respect to μ is then

$$\begin{aligned} \frac{\partial \ln f(y_t)}{\partial \mu} &= I_L \frac{\partial \ln F_x(c_L)}{\partial \mu} + (1 - I_L - I_U) \frac{\partial \ln f_x(y_t)}{\partial \mu} + I_U \frac{\partial \ln(1 - F_x(c_U))}{\partial \mu} \\ &= I_L \frac{1}{F_x(c_L)} \frac{\partial F_x(c_L)}{\partial \mu} + (1 - I_L - I_U) \frac{\partial \ln f_x(y_t)}{\partial \mu} - I_U \frac{1}{1 - F_x(c_U)} \frac{\partial F_x(c_U)}{\partial \mu}. \end{aligned} \tag{3}$$

The score for λ takes a similar form.

The derivatives of the CDF can be found by letting $\varepsilon_t = (x_t - \mu)e^{-\lambda}$. Then $\partial F_x(c_L)/\partial \mu = \exp(\lambda) \partial F_\varepsilon(c_L)/\partial \varepsilon_t \times \partial \varepsilon_t / \partial \mu = -f_x$. For the scale $\partial F_x(c_L)/\partial \lambda = -f_x(c_L - \mu)$ so

$$\frac{\partial \ln f(y_t)}{\partial \lambda} = -I_L \frac{f_x(c_L - \mu)}{F_x(c_L)} + (1 - I_L - I_U) \frac{\partial \ln f_x(y_t)}{\partial \lambda} + I_U \frac{f_x(c_U - \mu)}{1 - F_x(c_U)}.$$

The basic idea of the censored DCS Tobit model is to construct a signal-noise model in which location and scale are filters driven by their respective scores. Thus

$$x_t = \mu_{t|t-1} + \varepsilon_t \exp(\lambda_{t|t-1}), \qquad t = 1, ..., T. \tag{4}$$

The distribution of y_t is defined conditional on past values of the observations which appear through the conditional location, $\mu_{t|t-1}$, and/or the logarithm of conditional scale, $\lambda_{t|t-1}$. Alternatively the random variables denoted ε_t can be regarded as independent drawings from a standardized distribution. The stationary first-order model for location is

$$\mu_{t+1|t} = (1-\phi)\omega + \phi\mu_{t|t-1} + \kappa u_t, \quad |\phi| < 1, \tag{5}$$

where u_t is the score with respect to location and ω, ϕ and κ are parameters. A similar equation can be formulated for $\lambda_{t+1|t}$.

3 Distributions and scores

The generalized t and EGB2 distributions are both based on the GB2 distribution, which has PDF

$$f_x(x; \lambda, \upsilon, \xi, \varsigma) = \frac{\upsilon(xe^{-\lambda})^{\upsilon\xi-1}}{\varphi B(\xi,\varsigma)\left[(xe^{-\lambda})^{\upsilon}+1\right]^{\xi+\varsigma}}, \qquad x \geq 0, \qquad \varphi, \upsilon, \xi, \varsigma > 0, \tag{6}$$

where $\exp(\lambda)$ is the scale, υ, ξ and ς are shape parameters and $B(\xi, \varsigma)$ is the beta function. The CDF of x_t, $F(x_t; \upsilon, \xi, \varsigma)$, is a (regularized) incomplete beta function,

$$\beta(z_t; \xi, \varsigma) = B(z_t; \xi, \varsigma)/B(\xi, \varsigma),$$

where $B(z_t; \xi, \varsigma)$ is the incomplete beta function and $z_t = (x_t e^{-\lambda})^{\upsilon}$; see Kleiber and Kotz (2003, p 184). The incomplete beta function can be written in closed form when ς and/or ξ is one. The generalized gamma distribution can be obtained from the GB2 by letting the tail index go to infinity. The CDF of x_t, is then an incomplete gamma function.

We now set out the likelihood functions and scores for censored observations from generalized t and EGB2 distributions. The Gaussian distribution is a special case of both. Finally we return to the GB2 and discuss censored location/scale models for positive observations.

3.1 Generalized t

The generalized t is obtained from GB2 by setting $\upsilon\xi = 1$ and replacing x by $|x-\mu|$. Our preferred parameterization, as in Harvey and Lange (2017), replaces one of the shape parameters, ς, by the tail index, $\eta = \varsigma\upsilon$, and redefines scale by replacing φ by $\varphi\eta^{1/\upsilon}$. The t distribution is given by setting $\upsilon = 2$ in which case the tail index, η, becomes the degrees of freedom. The robustness properties of the score function - or influence function - of location are highlighted by McDonald and Newey (1988). Provided η is finite, it is redescending in that it approaches zero as x moves away from μ. The score function for scale has corresponding robustness features in that it is bounded. The normal distribution is obtained by letting $\eta \to \infty$.

The PDF for the generalized t distribution is

$$f_x(x;\mu,\lambda,\upsilon,\eta) = \frac{\upsilon}{2e^{\lambda}\eta^{1/\upsilon}} \frac{1}{B(1/\upsilon,\eta/\upsilon)} \left(1 + \frac{1}{\eta}\left|\frac{x-\mu}{e^{\lambda}}\right|^{\upsilon}\right)^{-(\eta+1)/\upsilon}, \quad -\infty < x_t < \infty,$$

with CDF

$$F_x(x) = [(1 + sgn(x-\mu)\beta(z;1/\upsilon,\eta/\upsilon)]/2,$$

where $z = (|x-\mu|\, e^{-\lambda})^{\upsilon}/\eta$. Thus, assuming $c_L < \mu$, $\Pr(y_t = c_L) = F_x(c_L) = [(1-\beta(z_L;1/\upsilon,\eta/\upsilon)]/2$ and, assuming $c_U > \mu$, $\Pr(y_t = c_U) = F_x(c_U) = [1+\beta(z_U;1/\upsilon,\eta/\upsilon)]/2$. Hence $1 - F_x(c_U) = [(1-\beta(z_U;1/\upsilon,\eta/\upsilon)]/2$.

The score for location in the censored distribution is

$$\begin{aligned} \frac{\partial \ln f(y_t)}{\partial \mu} &= -I_L \frac{\upsilon}{e^{\lambda}\eta^{1/\upsilon}} \frac{(1-b_L)^{(\eta+1)/\upsilon}}{B(1/\upsilon,\eta/\upsilon)[1-\beta(z_L;1/\upsilon,\eta/\upsilon)]} \\ &\quad + I_U \frac{\upsilon}{e^{\lambda}\eta^{1/\upsilon}} \frac{(1-b_U)^{(\eta+1)/\upsilon}}{B(1/\upsilon,\eta/\upsilon)[1-\beta(z_U;1/\upsilon,\eta/\upsilon)]} \\ &\quad + (1-I_L-I_U)\left[\frac{\eta+1}{\eta e^{\lambda}}(1-b_t)(|y_t-\mu|\, e^{-\lambda})^{\upsilon-1} sgn(y_t-\mu)\right], \end{aligned} \tag{7}$$

where b_t is defined as

$$b_t = \frac{(|y_t-\mu|\, e^{-\lambda})^{\upsilon}/\eta}{(|y_t-\mu|\, e^{-\lambda})^{\upsilon}/\eta + 1} = \frac{z_t}{z_t+1}$$

and b_L and b_U have y_t set to c_L and c_U respectively.

Remark 1. *For the t distribution, $\upsilon = 2$ so $(|y_t-\mu|\, e^{-\lambda})^{\upsilon-1} sgn(y_t-\mu) = (y_t-\mu)e^{-\lambda}$ and the formula for the continuous part of the score becomes much simpler.*

For scale the score is

$$\frac{\partial \ln f(y_t)}{\partial \lambda} = I_L \frac{\upsilon b_L^{1/\upsilon}(1-b_L)^{\eta/\upsilon}}{B(1/\upsilon,\eta/\upsilon)[1-\beta(z_L;1/\upsilon,\eta/\upsilon)]} \tag{8}$$
$$+I_U \frac{\upsilon b_U^{1/\upsilon}(1-b_U)^{\eta/\upsilon}}{B(1/\upsilon,\eta/\upsilon)[1-\beta(z_U;1/\upsilon,\eta/\upsilon)]} + (1-I_L-I_U)[(\eta+1)b_t - 1].$$

The generalized Student-t distribution may be extended to handle skewness and asymmetry as in Harvey and Lange (2017).

3.2 EGB2

The EGB2 distribution is formed by taking the logarithm of a GB2; see McDonald and Xu (1995) and Caivano and Harvey (2014). The PDF is

$$f_x(x;\mu,\lambda,\xi,\varsigma) = \frac{e^{-\lambda}\exp\{\xi(x-\mu)e^{-\lambda}\}}{B(\xi,\varsigma)(1+\exp\{(x-\mu)e^{-\lambda}\})^{\xi+\varsigma}}, \quad -\infty < x_t < \infty, \tag{9}$$

where the inverse of the υ parameter in the GB2 distribution has now been repalced by a scale parameter, denoted as $\varphi = \exp(\lambda)$. The distribution is light-tailed and all moments exist. It is positively (negatively) skewed when $\xi > \varsigma$ ($\xi < \varsigma$) and its kurtosis decreases as ξ and ς increase. There is excess kurtosis for finite ξ and/or ς; the maximum kurtosis is nine, but in the symmetric case it is six, obtained when $\xi = \varsigma = 0$, which is the Laplace distribution. The normal distribution is obtained by letting $\varsigma = \xi \to \infty$.

Defining $z_t = \exp((y_t - \mu)e^{-\lambda})$ gives

$$\frac{\partial \ln f(y_t)}{\partial \mu} = -I_L \frac{\xi b_L^{\xi}(1-b_L)^{\varsigma}}{e^{\lambda}B(\xi,\varsigma)\beta(z_L;\xi,\varsigma)} + I_U \frac{\xi b_U^{\xi}(1-b_U)^{\varsigma}}{e^{\lambda}B(\xi,\varsigma)[1-\beta(z_U;\xi,\varsigma)]} \tag{10}$$
$$+(1-I_L-I_U)\frac{(\xi+\varsigma)b_t - \xi}{e^{\lambda}},$$

where $b_t = z_t/(1+z_t)$.

For scale,

$$\frac{\partial F(y_t)}{\partial \lambda} = -I_L \frac{\xi b_L^{\xi}(1-b_L)^{\varsigma}}{B(\xi,\varsigma)\beta(z_L;\xi,\varsigma)}\varepsilon_t + I_U \frac{\xi b_U^{\xi}(1-b_U)^{\varsigma}}{B(\xi,\varsigma)[1-\beta(z_U;\xi,\varsigma)]}\varepsilon_t$$
$$+(1-I_L-I_U)\left[(\xi+\varsigma)\varepsilon_t b_t - \xi\varepsilon_t - 1\right], \tag{11}$$

where $\varepsilon_t = (y_t - \mu)e^{-\lambda}$.

The logistic distribution has $\xi = \varsigma = 1$. Its shape is close to that of the normal but it has heavier tails with an excess kurtosis of 1.2. Whereas the normal leads to linear estimators for uncensored models, it is much more difficult to handle when there is censoring. The fact that the CDF of a logistic distribution has a simple closed form makes it a more natural choice. The formulae above simplify because $\beta(z_t; 1, 1) = b_t$ and $B(1, 1) = 1$. The score for μ is

$$\begin{aligned}\frac{\partial \ln f(y_t)}{\partial \mu} &= -I_L \frac{e^{-\lambda}}{1+z_L} + I_U \frac{e^{-\lambda} z_U}{1+z_U} \\ &\quad + (1 - I_L - I_U) e^{-\lambda} \left[\frac{2z_t}{1+z_t} - 1 \right]. \end{aligned} \tag{12}$$

It may be convenient to multiply by $\exp(\lambda)$ as it is only a scaling factor. For scale

$$\begin{aligned}\frac{\partial \ln f(y_t)}{\partial \lambda} &= -I_L \frac{(c_L - \mu)e^{-\lambda}}{1+z_L} + I_U \frac{(c_U - \mu)e^{-\lambda} z_U}{1+z_U} \\ &\quad + (1 - I_L - I_U) \left[\frac{y_t - \mu}{e^{\lambda}} \frac{2z_t - 1}{1+z_t} - 1 \right]. \end{aligned} \tag{13}$$

3.3 Gaussian distribution

The normal distribution is a special limiting case of both generalized t and EGB2. However, proceeding directly gives

$$\frac{\partial \ln f(y_t)}{\partial \mu} = -I_L \frac{f_x(c_L)}{F_x(c_L)} + I_U \frac{f_x(c_U)}{1 - F_x(c_U)} + (1 - I_L - I_U) \frac{y_t - \mu}{e^{2\lambda}} \tag{14}$$

and

$$\begin{aligned}\frac{\partial \ln f(y_t)}{\partial \lambda} &= -I_L \frac{f_x(c_L)}{F_x(c_L)}(c_L - \mu) + I_U \frac{f_x(c_U)}{1 - F_x(c_U)}(c_U - \mu) \\ &\quad + (1 - I_L - I_U) \left[\frac{(y_t - \mu)^2}{e^{2\lambda}} - 1 \right], \end{aligned} \tag{15}$$

where $\varepsilon_t = (y_t - \mu)/e^{\lambda}$. The CDF is $F_x(c) = (1 + sgn(\varepsilon_c)\gamma(\varepsilon_c^2/2; 1/2))/2$, where $\gamma(.;.)$ is the regularized incomplete gamma function; see Abramowitz and Stegun (1964, p 934).

3.4 Location/scale models

Location/scale models are defined for non-negative variables. When the distribution is GB2, the technical details are similar to those of the EGB2; the location/scale, $\lambda_{t|t-1}$, now plays a similar role to $\mu_{t|t-1}$ in the EGB2 while $\upsilon = \exp(-\lambda)$. The score is

$$\begin{aligned}\frac{\partial \ln f(y_t;\lambda,\upsilon,\xi,\varsigma)}{\partial\lambda} &= -I_L\frac{\xi b_L^{\xi}(1-b_L)^{\varsigma}}{B(\xi,\varsigma)\beta(z_L;\xi,\varsigma)} + I_U\frac{\upsilon\xi b_U^{\xi}(1-b_U)^{\varsigma}}{B(\xi,\varsigma)[1-\beta(z_U;\xi,\varsigma)]}\\ &\quad +(1-I_L-I_U)\upsilon[(\xi+\varsigma)b_t-\xi],\end{aligned}$$

where $z_t = (y_t e^{-\lambda})^{\upsilon}$ and $b_t = z_t/(1+z_t)$. There is some simplification for the Burr distribution, which has $\xi = 1$ and $\beta(z;1,\varsigma) = 1-(1+z)^{-\varsigma}$.

The generalized gamma (GG) distribution is

$$f(x;\lambda,\gamma,\upsilon) = \frac{\upsilon}{e^{\lambda}\Gamma(\gamma)}\left(\frac{x}{e^{\lambda}}\right)^{\upsilon\gamma-1}\exp\left(-(xe^{-\lambda})^{\upsilon}\right), \quad 0\le x<\infty,$$

with $\gamma,\upsilon > 0$ and $-\infty < \lambda < \infty$. The gamma distribution sets $\upsilon = 1$, whereas the Weibull has $\gamma = 1$. The exponential distribution has $\upsilon = \gamma = 1$. The CDF of the GG is the regularized incomplete gamma function, $\gamma(z;\gamma)$. The score for the censored GG distribution is

$$\begin{aligned}\frac{\partial \ln f(y_t)}{\partial\lambda} &= -I_L\frac{\upsilon z_L^{\gamma}\exp(-z_L)}{\Gamma(\gamma)\gamma(z_L;\gamma)} + I_U\frac{\upsilon z_U^{\gamma}\exp(-z_U)}{\Gamma(\gamma)(1-\gamma(z_U;\gamma))}\\ &\quad +(1-I_L-I_U)\upsilon(z_t-1).\end{aligned}$$

where $z_t = (y_t e^{-\lambda})^{\upsilon}$. The score for the Weibull distribution simplifies because $F_x = \gamma(z;1) = 1-\exp(-z)$ and so

$$\frac{\partial \ln f(y_t)}{\partial\lambda} = -I_L\frac{\upsilon z_L\exp(-z_L)}{1-\exp(-z_L)} + I_U\upsilon z_U + (1-I_L-I_U)\upsilon(z_t-1).$$

4 Specification and prediction for DCS models

The censored DCS model is set up as in (4), with the appropriate score obtained from Section 2. The first-order dynamics can be extended by adding lags of $\mu_{t|t-1}$ and/or u_t; this is the QARMA model described in Harvey (2013, p 63). Other components, including non-stationary components, such as trends and seasonals, can be included. Explanatory variables, $\mathbf{z}_t$, can also be added so

$$x_t = \mu_{t|t-1} + \mathbf{z}_t'\beta + \varepsilon_t \exp(\lambda_{t|t-1}), \qquad t = 1, ..., T. \tag{16}$$

with $\mu_{t|t-1}$ as in (5) and

$$\lambda_{t+1|t} = (1 - \phi_\lambda)\omega_\lambda + \phi_\lambda \lambda_{t|t-1} + \kappa_\lambda u_t, \quad |\phi| < 1. \tag{17}$$

The static Tobit model is then a special case.

Remark 2. *The use of the DCS filter for location in a Gaussian model is closely related to the approach proposed by Zeger and Brookmeyer (1986). By a standard result for the censored (truncated) normal*

$$E(x \mid x \leq c_L) = \mu - \frac{f_x(c_L)}{F_x(c_L)} \tag{18}$$

Hence

$$-\frac{f_x(c_L)}{F_x(c_L)} = E(x \mid x \leq c) - \mu,$$

which is the score in (14) when $y_t = c_L$. When μ changes over time, it is possible to set up an iterative procedure for an autoregressive model where μ depends on lagged values of x_t based on simply minimizing the sum of squares with lagged censored observations replaced by the conditional expectations $E(x_t \mid x_t \leq c_L;\, x_{t-1}, x_{t-2}, ..)$; compare Zeger and Brookmeyer (1986, p 726). Alternatively the (pseudo) likelihood can be maximized directly. Setting $\kappa = \phi$ in the first-order dynamic score equation, (5), and taking ω there to be zero, gives

$$\mu_{t+1|t} = \phi y_t^\dagger,$$

where $y_t^\dagger = E(x_t \mid x_t \leq c; x_{t-1})$ when $y_t = c_L$ and $y_t^\dagger = y_t$ otherwise. In other words, $y_t = \phi y_{t-1} + e^\lambda \varepsilon_t$ or $y_t = \phi E(x_{t-1} \mid x_{t-1} \leq c_L)$.

Preliminary tests for time-variation can be carried out before a model is fitted. To be specific, a test for serial correlation can be based on the score for the parameter of interest after fitting a static censored (regression) model. Such tests are Lagrange multiplier (LM) tests; see Harvey (2013, section 2.5). Other things being equal, the choice of a logistic distribution is an attractive option; see Lewis and McDonald (2014). After a model has been fitted, the scores can again be used as a diagnostic, as in Harvey and Thiele (2016).

The one-step ahead predictive distribution gives the probability of a censored observation. For example the probability of hitting the lower bound at time t with a t_η-distribution is $F_x(c_L) = [(1 - \beta(z_L; 1/2, \eta/2)]/2$, where $z_L = (c_L - \mu_{t|t-1})^2 e^{-2\lambda_{t|t-1}}/\eta$.

The conditional mean for x_{T+1} in (16), assuming the distribution of ε_t is symmetric, is

$$E(x_{T+1} \mid Y_T) = \mu_{T+1|T} + \mathbf{z}'_{T+1}\beta.$$

The prediction for the mean of the censored variable, y_t, is different, but can be calculated if needed.

Conditional quantiles can be obtained from the full predictive distribution. The $\tau - th$ conditional quantile is defined as $\Pr(y_{T+1} \geq q_\tau \mid Y_T)$. Hence $q_\tau = F_x^{-1}(\tau)$, where $F_x^{-1}(\tau)$ is the quantile function. The inverse regularized incomplete beta and ganna functions are to be found in Abramowitz and Stegun (1964, p 944-5).

Multi-step forecasts of the $l - step$ ahead distribution can be made by simulation. Values of $\mu_{T+\ell|T}$ and $\lambda_{T+\ell|T}$ are obtained by simulating beta variates, and hence the score, ℓ times. A value of $x_{T+\ell}$ is then simulated, again from a beta variate (which converts into a generalized t or EGB2). For Gaussian and GG distributions gamma variables are simulated.

5 Parameter-driven models

It could be argued that a parameter-driven unobserved components models is more plausible than the DCS model. The basic UC model with constant scale is

$$x_t = \mu_t + \varepsilon_t \exp(\lambda), \qquad t = 1, ..., T, \tag{19}$$

with

$$\mu_{t+1} = (1 - \phi)\mu + \phi\mu_t + \eta_t, \quad |\phi| < 1, \tag{20}$$

where η_t is IID normal with mean zero and variance σ_η^2 and μ is the unconditional mean of μ_t. Throughout this article we assume η_t and ε_t are independent. Without the censoring the model could be handled by the Kalman filter when ε_t is Gaussian. With censoring, or indeed with non-Gaussian disturbances, it becomes a nonlinear state space model and estimation must be carried out by a computer intensive method such as MCMC, particle filtering or simulated EM[1]; see, for example, Durbin and Koopman (2012). A simpler (approximate) approach is described in Allik et al (2015).

The DCS model can be regarded as an approximation to (19) in much the same way as the pseudo-likelihood approach of Zeger and Brookmeyer (1986) is an approximation to the likelihood for a model in which x_t is an AR process. This raises the question of how well the DCS model performs when the true data generating process (DGP) is the UC model. To answer this question, we carried out a set of Monte Carlo experiments. In each simulation, we generated 5000 observations and used the first 2500 observations for estimation. One-step-ahead predictions were then compute for the remaining 2500 out-of-sample observations. This procedure was repeated 1000 times, and hence we have 1000 estimates for each parameter, and 2500×1000 forecasts. The DCS Tobit model was estimated by ML, as the likelihood function of the model is available in closed form, whereas the state-space models was estimated by using the numerically accelerated importance sampling (NAIS) method of Koopman et al. (2016), which is a newly proposed efficient method for estimating the non-linear and non-Gaussian state-space model. We adjust the original NAIS method to accommodate censored data; the detailed procedure is set out in an online Appendix.

Having compared the performance of the DCS model for a UC DGP, we reversed the experiment and estimated the UC model for a DCS DGP. The details are as follows.

Experiment 1: Constant Volatility (CV).
The model is (19) and (20). The constant μ is set to zero, as it is in all subsequent cases. We used a standard normal and a Student t-distribution with four degrees of freedom for ε_t, and generated η_t from a standard normal distribution. The threshold c was chosen so that the probability of $x_t \leq c = 0.05$ or 0.25. To consider both low and high persistent data processes, we set ϕ to be either 0.5 or 0.9. Lastly, we set λ to be zero, so the signal-to-noise ratio is one. When data were generated from the DCS model, we set $\kappa = 0.5$.

Experiment 2: Stochastic Volatility (SV) Model
In the second set of experiments, the location is still dynamic, with the same parameters as before, but now the scale in (19) follows a stochastic volatility

[1] Lee (1999) proposes estimating a dynamic Tobit model with autoregressive disturbances and GARCH effects by simulated ML.

(SV) process as well, that is

$$\lambda_{t+1} = (1 - \phi_\lambda)\omega + \phi_\lambda \lambda_t + \eta_t^\lambda, \quad |\phi_\lambda| < 1,$$

where η_t^λ is standard normal. In the DCS scale model, $\kappa_\lambda = 0.1$.

Panel A of Table 1 presents the results for the state-space DGPs. There are three noteworthy findings. First, the differences in forecasting accuracy between the state space models and the corresponding DCS Tobit models are small and are smaller for the more persistent observations. The relative MSE of the DCS Tobit model ranges from 1.023 to 1.038 when $\phi = 0.9$ and from 1.019 to 1.028 when $\phi = 0.5$. Second, the differences in forecasting accuracy between the state space models and the DCS models are smaller when the true DGP has a $t-$distribution. Third, both models perform better when there is no stochastic volatility.

Panel B of Table 1 presents the results when the data are generated by the DCS Tobit model. The main conclusion is that the state space model is far worse thatn the DCS model. This finding is consistent with the results reported by Koopman et al (2016) for other score-driven models.

Finally we compare the relative performance of the standard DCS model, in which censored observations are taken at face value, and the DCS Tobit model. The Tobit model is always better, irrespective of whether the DGP is a state space model or the DCS. However, the gains are bigger when the true model is the state space model. Even though the gains are not large when the censored DCS is the true model, it is reassuring to know that the censored DCS model works well. It is worth re-iterating that the censored model also predicts the probability of censoring, so there are reasons for using it above and beyond simply predicting the next observation.

6 Application to the Chinese stock market

Daily price limit rules are widely used by stock markets across the globe, and these rules are particularly popular in emerging markets with a large fraction of inexperienced investors. China's equity market, which is the second largest in the world by capitalization, imposes daily price limits of 10% on regular stocks and 5% on special treatment (ST) stocks. A recent paper by Chen et al (2018) studies the impact of the price limit rule on destructive trading behavior in China. Here, we focus on modeling Chinese stock return dynamics to illustrate our DCS Tobit model.

Table 1: Forecasting results from Monte Carlo study (Relative MSEs). This table provides one-step-ahead forecasting results for time-varying location (and scale) parameters. Panel A and B present respectively the results when the state space and DCS Tobit model are used as DGPs. We report the forecasting MSEs relative to those of the DGP model with estimated parameters.

		I: The Probability of censored data is 0.05 ($c = 0.05$)				II: The Probability of censored data is 0.25 ($c = 0.25$)			
Type	Distribution	State Space		DCS Tobit	DCS	State Space		DCS Tobit	DCS
Panel A: State space model as DGP									
High persistence ($\phi = 0.9$)									
		True	Estimated	Estimated	Estimated	True	Estimated	Estimated	Estimated
CV model	normal	0.987	1.000	1.016	1.067	0.985	1.000	1.021	1.134
	t(4)	0.912	1.000	1.015	1.065	0.987	1.000	1.020	1.127
	t (normal)	0.981	1.000	1.019	1.068	0.980	1.000	1.024	1.145
SV model	normal	0.974	1.000	1.020	1.071	0.972	1.000	1.025	1.139
	t(4)	0.977	1.000	1.019	1.069	0.967	1.000	1.023	1.135
	t (normal)	0.966	1.000	1.023	1.074	0.965	1.000	1.028	1.152
Low persistence ($\phi = 0.5$)									
		True	Estimated	Estimated	Estimated	True	Estimated	Estimated	Estimated
CV model	normal	0.989	1.000	1.010	1.056	0.990	1.000	1.029	1.115
	t(4)	0.990	1.000	1.012	1.054	0.991	1.000	1.023	1.117
	t (normal)	0.987	1.000	1.015	1.058	0.985	1.000	1.028	1.120
SV model	normal	0.978	1.000	1.017	1.060	0.980	1.000	1.034	1.215
	t(4)	0.971	1.000	1.014	1.058	0.973	1.000	1.030	1.206
	t (normal)	0.965	1.000	1.020	1.063	0.969	1.000	1.038	1.220
Panel B: DCS Tobit model as DGP									
High persistence ($\phi = 0.9$)									
		True	Estimated	Estimated	Estimated	True	Estimated	Estimated	Estimated
CV model	normal	0.994	1.567	1.000	1.024	0.992	1.643	1.000	1.034
	t(4)	0.996	1.763	1.000	1.037	0.995	1.899	1.000	1.047
	t (normal)	0.993	1.854	1.000	1.037	0.991	1.902	1.000	1.047
SV model	normal	0.995	1.989	1.000	1.020	0.993	1.931	1.000	1.030
	t(4)	0.995	2.010	1.000	1.032	0.992	2.013	1.000	1.042
	t (normal)	0.996	2.014	1.000	1.039	0.993	2.109	1.000	1.049
Low persistence ($\phi = 0.5$)									
		True	Estimated	Estimated	Estimated	True	Estimated	Estimated	Estimated
CV model	normal	0.997	1.781	1.000	1.015	0.986	1.853	1.000	1.021
	t(4)	0.998	1.833	1.000	1.025	0.989	1.902	1.000	1.029
	t (normal)	0.996	1.910	1.000	1.027	0.985	1.945	1.000	1.032
SV model	normal	0.998	1.945	1.000	1.019	0.988	2.010	1.000	1.025
	t(4)	0.997	1.977	1.000	1.029	0.990	2.113	1.000	1.030
	t (normal)	0.998	2.012	1.000	1.032	0.992	2.195	1.000	1.034

Table 2: Summary statistics for the three Chinese stock returns

	Baoshan Iron & Steel	Hua Xia Bank	Inner Mongolia Baotou Steel Union
Mean	0.002	0.002	0.001
Std	0.026	0.023	0.022
Skewness	0.595	0.388	0.434
Kurtosis	7.661	6.471	7.137
Percentage of price limit hitting(%)	4.060	3.227	5.098

6.1 Data

The data (from the CSMAR database) consist of daily prices of three A-share stocks from the Shanghai Stock Exchange, namely Baoshan Iron & Steel, Hua Xia Bank and Inner Mongolia Baotou Steel Union. The sample period is a four-year window from Dec 2008 to Jan 2013 (around 1000 trading days), including around 4% of trading days when the stock price hits the limit. Table 2 gives summary statistics; note that there is excess kurtosis despite the censoring. Figure 1 displays the histogram of the returns Inner Mongolia Baotou Steel Union.

6.2 Model Specification

We specify the dynamics of scale for the three Chinese stock returns using a number of DCS models with Gaussian, logistic and student t conditional distributions. The conditional mean in (4) is assumed to be constant. We first fit the time-varying volatility model (17) and then generalize so as to have two components for volatility with leverage effects, as in Harvey and Lange (2018), that is

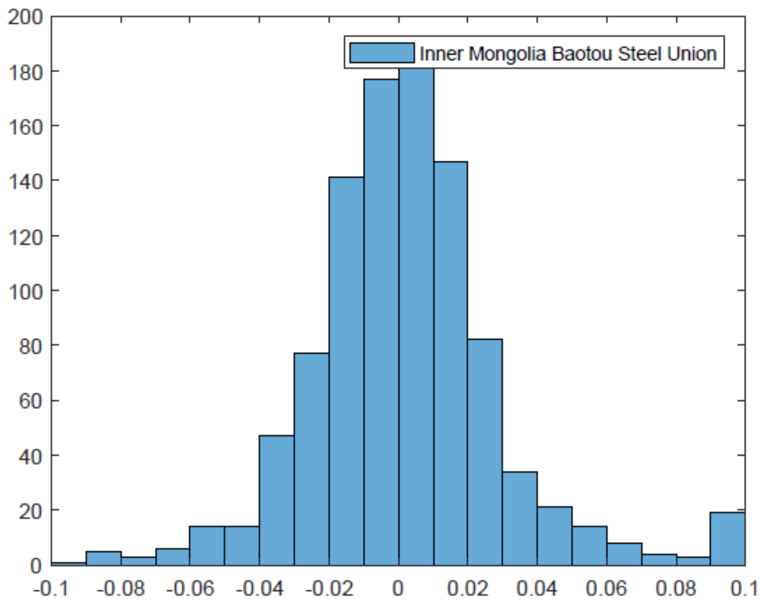
$$\begin{aligned} \lambda_{t|t-1} &= \omega + \lambda_{1,t|t-1} + \lambda_{2,t|t-1}, \\ \lambda_{i,t+1|t} &= \phi_{\lambda,i}\, \lambda_{i,t|t-1} + \kappa_{\lambda,i}\, u_t + \kappa_i^*\, sgn(-\varepsilon_t)\,(u_t+1), \qquad i = 1, 2. \end{aligned}$$

Other features of the data, such as skewness, could be accommodated by fitting more general distributions such as the skewed, asymmetric generalized-t described in Harvey and Lange (2018).

6.3 Empirical Findings

Table 3 reports the estimation results of the standard and censored DCS models for the three Chinese stocks. The estimated parameters turn out

Figure 1: Empirical Distribution of Inner Mongolia Baotou Steel Union stock returns.



to be similar to the ones found by Harvey and Lange (2018) for US and Japanese stock indices. The two component models always outperform the corresponding one component models and the leverage effect is much stronger in the short run. Overall, the t-distribution gives the best fit, but the logistic is not far behind. It should be borne in mind that the t-distribution requires a shape parameter to be estimated, whereas the logistic does not. Furthermore the logistic score takes a relatively simple form.

Figure 2 shows the ACF plot of scale scores from the best model, the two-component Student t, for a typical stock. There is no indication of significant serial correlation in the scores and other diagnostics were satisfactory.

The most important result in the present context is that in all cases the censored DCS models achieve higher likelihood value than the corresponding basic DCS models.

Figure 3 shows the estimated dynamic probability of hitting the stock price limit for the Inner Mongolia Baotou Steel Union stock returns during the whole sample period. The probability is computed for the best fitting model, namely the censored DCS two component model with Student t distribution. The average of the estimated dynamic probabilities is very close to the unconditional probability of hitting the price limit calculated from the original returns. The estimated volatility for the censored DCS model is shown in Figure 4. There are some differences as compared with the standard DCS model, but no general conclusions can be drawn.

7 Conclusion

Score-driven models provide a relatively straightforward solution to the problem of modeling time series when the observations are subject to censoring and location and/or scale may change over time. The method applies to generalized-t and EGB2 distributions, as well as to the normal distribution. A particularly appealing model is obtained for the logistic distribution.

A set of Monte Carlo experiments show that the score-driven model provides good forecasts even when the true model is parameter-driven. The converse appears not to be the case. When dynamic censored models are fitted to data on Chinese stock returns, they give a better fit than models that do not take account of censoring. Furthermore they yield an estimated of the probability that the next observations will be subject to censoring.

Table 3: In-sample Estimation Results. This table reports in-sample estimation results using the 1000 return observations for the three Chinese stocks.

	Baoshan Iron & Steel		Hua Xia Bank		Inner Mongolia Baotou Steel Union	
	Panel A: SV model with Gaussian Distribution					
	DCS	Censored DCS	DCS	Censored DCS	DCS	Censored DCS
DCS	Censored DCS	DCS	Censored DCS			
μ 0.0014	0.0014	0.0009	0.0010	0.0003	0.0003	
ω	-3.6239	-3.6140	-4.0320	-4.0197	-3.6738	-3.9556
ϕ_λ	0.9183	0.9198	0.9750	0.9600	0.9900	0.9895
κ_λ	0.0401	0.0400	0.0282	0.0254	0.0106	0.0072
log likelihood	-2256.9470	-2241.1552	-2643.6784	-2645.8744	-2839.4584	-2839.4039
BIC	4525.8939	4494.3104	5299.3568	5297.7488	5690.9168	5670.8077
	Panel B: SV model with Logistic Distribution					
	DCS	Censored DCS	DCS	Censored DCS	DCS	Censored DCS
μ	0.0008	0.0008	0.0006	0.0006	0.0004	0.0004
ω	-4.5122	-4.5132	-4.8804	-4.8804	-5.3341	-5.3379
ϕ_λ	0.9409	0.9433	0.9588	0.9588	0.9767	0.9782
κ_λ	0.0287	0.0338	0.0186	0.0186	0.0273	0.0267
log likelihood	-2248.8988	-2230.1476	-2639.7811	-2639.7811	-2824.3777	-2818.7886
BIC	4509.7967	4472.2951	5291.5621	5291.5621	5660.7554	5649.5773
	Panel C: SV model with Student t Distribution ($v = 2$)					
	DCS	Censored DCS	DCS	Censored DCS	DCS	Censored DCS
μ	0.0002	0.0001	0.0003	0.0003	0.0000	0.0000
ω	-3.9033	-3.8449	-4.2255	-4.1574	-3.6678	-3.6599
ϕ_λ	0.9418	0.9475	0.9852	0.9941	1.0000	0.9999
κ_λ	0.0672	0.0713	0.0315	0.0306	0.0321	0.0283
ν	4.1592	3.9478	6.2385	6.0512	4.5174	4.4792
log likelihood	-2236.2685	-2224.3694	-2623.6669	-2615.3882	-2805.9347	-2784.9099
BIC	4487.5371	4463.7389	5262.3338	5245.7765	5626.8694	5584.8199
	Panel D: two component SV model with Gaussian Distribution					
	DCS	Censored DCS	DCS	Censored DCS	DCS	Censored DCS
μ	0.0015	0.0014	0.0009	0.0010	0.0021	0.0014
ω	-4.0018	-3.6147	-4.0306	-4.0191	-3.8142	-3.9101
$\phi_{\lambda,1}$	0.9912	0.9903	0.9825	0.9736	0.9946	0.9932
$\phi_{\lambda,2}$	0.7953	0.7873	0.8830	0.8447	0.8473	0.8806
$\kappa_{\lambda,1}$	0.0089	0.0347	0.0195	0.0146	0.0054	0.0060
$\kappa_{\lambda,2}$	0.0067	0.0056	0.0093	0.0108	0.0054	0.0051
κ_1^*	0.0674	0.0858	0.0021	0.0704	0.0154	0.0172
κ_2^*	0.1125	0.1309	0.0342	0.1360	0.0292	0.0386
log likelihood	-2209.5083	-2189.5540	-2610.2233	-2602.1798	-2797.0238	-2770.6155
BIC	4443.0165	4403.1079	5244.4467	5228.3595	5618.0476	5565.2311
	Panel E: two component SV model with Logistic Distribution					
	DCS	Censored DCS	DCS	Censored DCS	DCS	Censored DCS
μ	0.0009	0.0014	0.0006	0.0006	0.0000	0.0004
ω	-4.2339	-4.6147	-4.6309	-4.8692	-4.7476	-5.3451
$\phi_{\lambda,1}$	0.9994	0.9903	0.9978	0.9497	0.9901	0.9819
$\phi_{\lambda,2}$	0.8765	0.8873	0.8783	0.8427	0.8319	0.8194
$\kappa_{\lambda,1}$	0.0193	0.0360	0.0179	0.0228	0.0364	0.0218
$\kappa_{\lambda,2}$	0.0073	0.0046	0.0192	0.0174	0.0185	0.0097
κ_1^*	0.0184	0.0858	0.0177	0.0175	0.0217	0.0430
κ_2^*	0.0264	0.1309	0.0239	0.0275	0.0317	0.0644
log likelihood	-2204.4196	-2172.5540	-2605.0996	-2590.5984	-2727.2849	-2719.1177
BIC	4432.8392	4369.1079	5234.1992	5205.1968	5478.5698	5462.2355
	Panel F: two component SV model with student t Distribution ($v = 2$)					
	DCS	Censored DCS	DCS	Censored DCS	DCS	Censored DCS
μ	0.0003	0.0002	0.0003	0.0002	0.0002	0.0001
ω	-3.9153	-3.8777	-4.1334	-3.5777	-3.8418	-3.7488
$\phi_{\lambda,1}$	0.9907	0.9897	0.9733	0.9731	0.9889	0.9887
$\phi_{\lambda,2}$	0.8785	0.8921	0.8651	0.8620	0.8873	0.8742
$\kappa_{\lambda,1}$	0.0179	0.0153	0.0223	0.0215	0.0110	0.0121
$\kappa_{\lambda,2}$	0.0017	0.0015	0.0104	0.0116	0.0021	0.0019
κ_1^*	0.0055	0.0048	0.0032	0.0028	0.0029	0.0030
κ_2^*	0.0488	0.0472	0.0371	0.0342	0.0330	0.0314
ν	4.4216	4.4631	5.7148	4.4831	4.6440	5.0274
log likelihood	-2201.3214	-2164.4532	-2601.3231	-2583.8890	-2718.2125	-2707.3345
BIC	4429.6428	4355.9064	5229.6462	5194.7780	5463.4250	5441.6690

Figure 2: Correlogram of scale scores from censored DCS two component model with student t distribution fitted to Inner Mongolia Baotou Steel Union stock returns

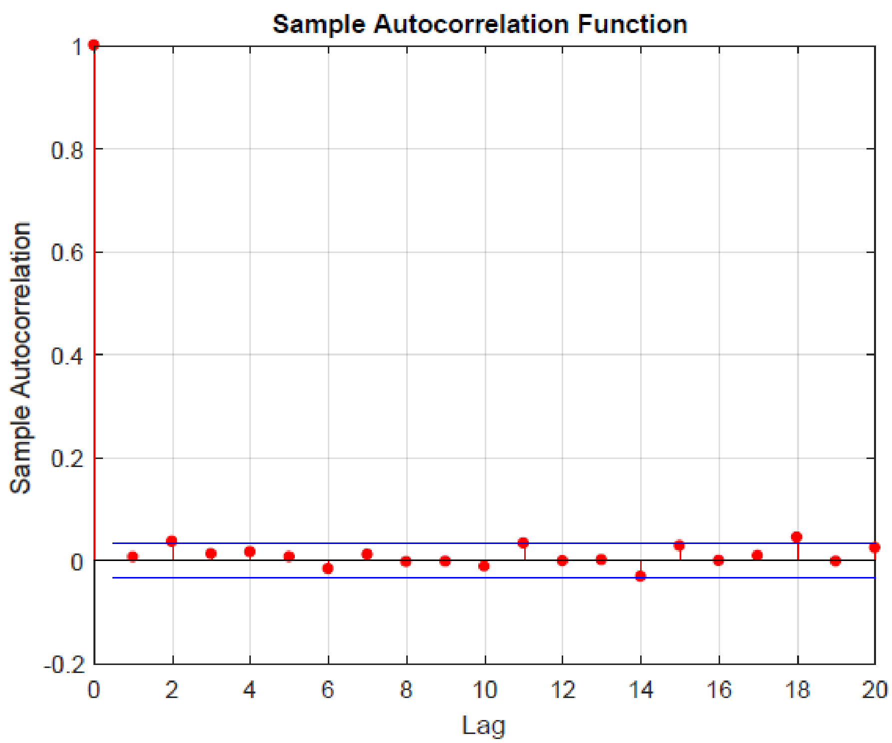


Figure 3: Estimated probability of price limit hitting for Inner Mongolia Baotou Steel Union stock returns from two component model with Student t distribution. The horizontal line shows the unconditional probability of price limit hitting.

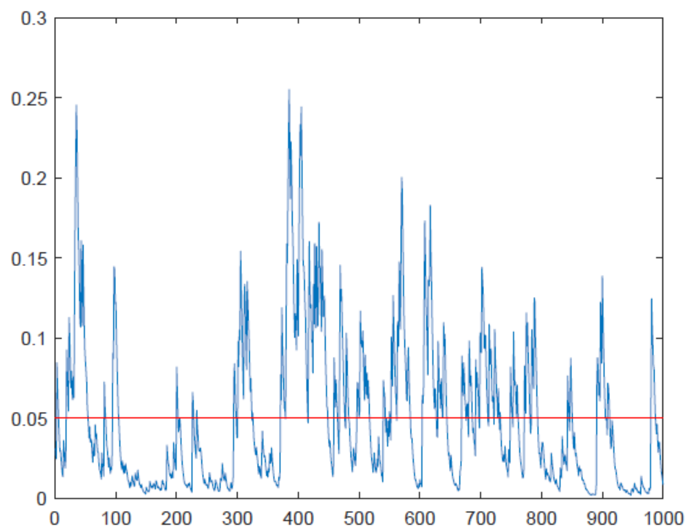
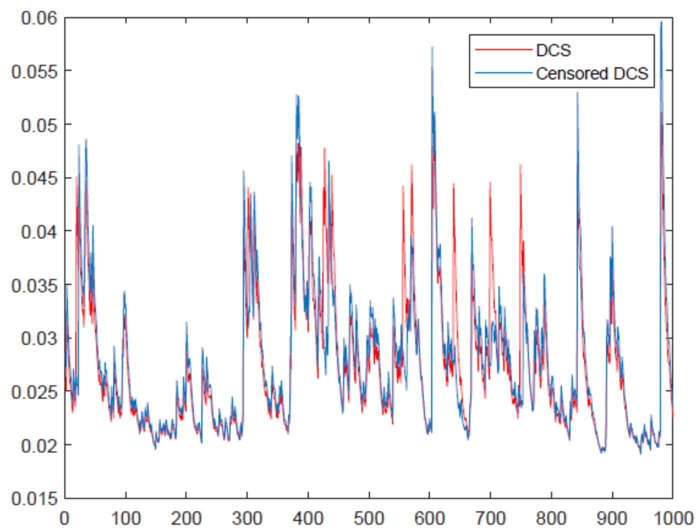


Figure 4: Comparison of estimated volatility from two component DCS and censored DCS model with student t distribution for Baotou Steel Union stock returns.



REFERENCES

Amemiya, T. (eds.), 1985, *Advanced Econometrics*, Harvard University Press.

Abramowitz, M. and I. A. Stegun, (eds.), 1964, *Handbook of Mathematical Functions*, New York: Dover Publications Inc.

Allik, B., Miller, C., Piovoso, M. J. and R. Zurakowski (2015) The Tobit Kalman Filter: An Estimator for Censored Measurements. IEEE Transactions on Control Systems Technology (forthcoming)

Blasques, F., Koopman, S. J. and A. Lucas (2014). Maximum likelihood estimation for generalized autoregressive score models. Tinbergen Institute Discussion Paper, TI 2014-029/III, Amsterdam.

Caivano, M. and A.C. Harvey (2014) Time series models with an EGB2 conditional distribution. *Journal of Time Series Analysis*, 34, 558-71.

Chen, T., Gao, Z. Y., He, J. B., Jiang, W. X. and Xiong, W (2018). Daily price limits and destructive market behavior. *Journal of Econometrics (In press)*

Creal, D., Koopman, S. J. and A. Lucas (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics,* 28, 777-95.

Durbin, J. and S. J. Koopman (2012). *Time Series Analysis by State Space Methods*. 2nd edition. Oxford: Oxford Statistical Science Series.

Harvey, A.C. (2013). *Dynamic Models for Volatility and Heavy Tails: with applications to financial and economic time series.* Econometric Society Monograph, Cambridge University Press.

Harvey, A.C. and R-J. Lange (2016). Volatility Modeling with a Generalized t-Distribution, *Journal of Time Series Analysis,* 38, 175-90.

Harvey, A. C. and R-J. Lange (2018). Modeling the Interactions between Volatility and Returns using EGARCH-M, *Journal of Time Series Analysis,* 39, 909-919.

Harvey, A.C. and S. Thiele (2016) Testing against changing correlation. *Journal of Empirical Finance*, 38, 575-89.

Harvey, A.C. and R. Ito (2018). Modeling time series when some observations are zero. *Journal of Econometrics,(to appear).*

Kleiber, C. and S. Kotz (2003). *Statistical Size Distributions in Economics and Actuarial Sciences.* New York: Wiley.

Koopman, S. J., Lucas, A. and M. Scharth (2016). Predicting time-varying parameters with parameter-driven and observation-driven models, *Review of Economics and Statistics*, 98, 97-110.

Lee, L. F. (1999) Estimation of dynamic and ARCH Tobit models, *Journal of Econometrics, 92, 355-390.*

Lewis, R.A and J.B. McDonald (2014). Partially adaptive estimation of censored regression models. *Econometric Reviews*, 33, 732-50.

McDonald, J.B. and W.K. Newey (1988). Partially Adaptive Estimation of Regression Models via the Generalized t Distribution, *Econometric Theory*, 4, 428-457.

McDonald J. B. and Xu, Y. J. (1995). A generalization of the beta distribution with applications. *Journal of Econometrics* 66, 133-152.

Park, J. W., Genton, M. G., and Ghosh, S. K. (2007), Censored time series analysis with autoregressive moving average models, Canadian Journal of Statistics, 35, 151–168.

Wang, C. and K-S. Chan (2018) Quasi-Likelihood Estimation of a Censored Autoregressive Model With Exogenous Variables, Journal of the American Statistical Association, 113, 1135-1145.

Zeger, S. L. and Brookmeyer, R. (1986), Regression analysis with censored autocorrelated data, Journal of the American Statistical Association, 81, 722–729.