Dynamic Tobit models

Andew Harvey [a,*], Yin Liao [b]

[a] *Department of Economics, Cambridge University*
[b] *Business School, Macquarie University, Australia*

ARTICLE INFO

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. Explanatory variables can be included, making static Tobit models a special case. 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.

1. Introduction

The Tobit regression model is widely used 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. (2016) 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, 2021) and Creal et al. (2013). 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

---

* Corresponding author. Tel.: 44-(0) 1223 335228.
E-mail addresses: ach34@cam.ac.uk, Andrew.Harvey@econ.cam.ac.uk (A. Harvey).

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 the 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 heteroscedasticity (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. (2016) and the methods described in Manrique and Shephard (1998) and Varin and Vidoni (2008). 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. Some of the technical details are similar to those in Harvey and Ito (2019), but the motivation of that paper is somewhat different as it is concerned with the construction of models when some of the observations are zero. 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

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

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 $\Pr(y_t = c_U) = 1 - F_x(c_U)$, where F_x is the CDF of x_t.

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

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

For the $t - th$ observation, the score with respect to μ is therefore

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

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

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 $\partial F_x(c_L)/\partial \mu = \exp(\lambda)\partial F_\varepsilon(c_L)/\partial \varepsilon_t \times \partial \varepsilon_t/\partial \mu = -f_x(c_L)$. For the scale $\partial F_x(c_L)/\partial \lambda = -f_x(c_L)(c_L - \mu)$ so

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

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

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

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

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

where u_t is the score, possibly scaled, with respect to location and μ, 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

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

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, $F(x; \upsilon, \xi, \varsigma)$, is a (regularized) incomplete beta function,

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

where $B(z; \xi, \varsigma)$ is the incomplete beta function and $z = (xe^{-\lambda})^{\upsilon}$; 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-distribution 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

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

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 $c_U > \mu$, $\Pr(y_t = c_U) = 1 - F_x(c_U) = [(1 - \beta(z_U; 1/\upsilon, \eta/\upsilon)]/2$.

The score for location in the censored distribution is

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

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.

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. Note that the score is typically scaled by dividing by the information quantity; see, for example, Harvey (2013, p60). In the present context this suggests multiplying the score by $e^{-2\lambda}$.

For scale the score is

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

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

3.2. Exponential generalized beta of the second kind

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

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

where the inverse of the υ parameter in the GB2 distribution has now been replaced 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

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

where $b_t = z_t/(1+z_t)$. For the reasons given in Caivano and Harvey (2014) u_t is better defined by multiplying the above score by a factor proportional to $e^{2\lambda}$.

For scale,

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

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

$$\frac{\partial \ln f(y_t)}{\partial \mu} = -I_L \frac{e^{-\lambda}}{1+z_L} + I_U \frac{e^{-\lambda}z_U}{1+z_U} + (1-I_L-I_U)e^{-\lambda}\frac{z_t-1}{z_t+1}. \tag{12}$$

For scale

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

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

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

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 $\upsilon = e^{-\lambda}$ where λ is the EGB2 scale. The score is

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

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

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

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 for location, (5), and scale,

$$\lambda_{t+1|t} = (1 - \phi_\lambda)\omega_\lambda + \phi_\lambda \lambda_{t|t-1} + \kappa_\lambda u_t, \quad |\phi_\lambda| < 1, \tag{16}$$

can be extended by adding lags as in Harvey (2013, p 63). Other components, including non-stationary trends and seasonals, can also be included. Explanatory variables, $\mathbf{z}_t$, can be added so

$$x_t = \mu_{t|t-1} + \mathbf{z}_t'\boldsymbol{\beta} + \varepsilon_t \exp(\lambda_{t|t-1}), \qquad t = 1, ..., T. \tag{17}$$

where $\boldsymbol{\beta}$ is a vector of parameters. The static Tobit model is then a special case.

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

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

Hence

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

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 $E(x_t \mid x_t \le c_L; x_{t-1}, x_{t-2}, ..)$; compare Zeger and Brookmeyer (1986, p 72). 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 $y_t^\dagger = E(x_t \mid x_t \le c; x_{t-1})$ 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 $y_t = \phi E(x_{t-1} \mid x_{t-1} \le c_L)$.

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 (17), assuming the distribution of ε_t is symmetric, is

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

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 $\Pr(y_{T+1} \ge q_\tau \mid Y_T)$. Hence $q_\tau = F_x^{-1}(\tau)$, where $F_x^{-1}(\tau)$ is the quantile function. The inverse regularized incomplete beta and gamma 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

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

where the $\mu^\dagger$ is the unconditional mean of μ_t and η_t is IID normal with mean zero and variance σ_η^2. 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; see, for example, Durbin and Koopman (2012) or Lee (1999).

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 rolling windows to estimate the models and produce one-step-ahead out-of-sample forecasts. Specifically, we start with the first 2500 observations for model estimation, and re-estimate the models each time a new observation becomes available. One-step-ahead predictions were then computed for the remaining 2500 out-of-sample observations which are also used for forecasting evaluation. 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 model was estimated using the numerically accelerated importance sampling (NAIS) method of Koopman et al. (2016), which is a new efficient method for estimating non-linear state-space models. We adjust the original NAIS method to accommodate censored data.

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 $\mu^\dagger$ 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 the dynamic location by letting η_t have a standard normal distribution. The threshold c was chosen so that the probability of $x_t \leq c$ is 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)

In the second set of experiments, the location is still dynamic, with the same parameters as before, but now the scale in (19) is a stochastic volatility (SV) process, that is

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

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. Second, the differences become 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 state space model is far worse than 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. (2019) 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.

6.1. Data

The data (from the CSMAR database) consist of approximately 1000 daily prices of three A-share stocks from the Shanghai Stock Exchange over a four-year window from Dec 2008 to Jan 2013. The three companies were Baoshan Iron & Steel, Hua Xia Bank and Inner Mongolia Baotou Steel Union and the proportion of trading days in which the stock price hit the limit was 0.041, 0.032 and 0.051. Figure 1 presents the time series plot of the returns for Inner Mongolia Baotou Steel Union, and Figure 2 displays the histogram of the returns.

Table 1
Monte Carlo study: forecasting MSEs relative to those of the DGP model with estimated parameters.

Type	Distribution	I: Probability of censored data is 0.05 ($c = 0.05$)			II: Probability of censored data is 0.25 ($c = 0.25$)		
		State Space	DCS Tobit	DCS	State Space	DCS Tobit	DCS
		Panel A: State space model as DGP					
		High persistence ($\phi = 0.9$)					
CV model	normal	1.000	1.016	1.067	1.000	1.021	1.134
	t(4)	1.000	1.015	1.065	1.000	1.020	1.127
	t (normal)	1.000	1.019	1.068	1.000	1.024	1.145
SV model	normal	1.000	1.020	1.071	1.000	1.025	1.139
	t(4)	1.000	1.019	1.069	1.000	1.023	1.135
	t (normal)	1.000	1.023	1.074	1.000	1.028	1.152
		Low persistence ($\phi = 0.5$)					
CV model	normal	1.000	1.010	1.056	1.000	1.029	1.115
	t(4)	1.000	1.012	1.054	1.000	1.023	1.117
	t (normal)	1.000	1.015	1.058	1.000	1.028	1.120
SV model	normal	1.000	1.017	1.060	1.000	1.034	1.215
	t(4)	1.000	1.014	1.058	1.000	1.030	1.206
	t (normal)	1.000	1.020	1.063	1.000	1.038	1.220
		Panel B: DCS Tobit model as DGP					
		High persistence ($\phi = 0.9$)					
CV model	normal	1.567	1.000	1.024	1.643	1.000	1.034
	t(4)	1.763	1.000	1.037	1.899	1.000	1.047
	t (normal)	1.854	1.000	1.037	1.902	1.000	1.047
SV model	normal	1.989	1.000	1.020	1.931	1.000	1.030
	t(4)	2.010	1.000	1.032	2.013	1.000	1.042
	t (normal)	2.014	1.000	1.039	2.109	1.000	1.049
		Low persistence ($\phi = 0.5$)					
CV model	normal	1.781	1.000	1.015	1.853	1.000	1.021
	t(4)	1.833	1.000	1.025	1.902	1.000	1.029
	t (normal)	1.910	1.000	1.027	1.945	1.000	1.032
SV model	normal	1.945	1.000	1.019	2.010	1.000	1.025
	t(4)	1.977	1.000	1.029	2.113	1.000	1.030
	t (normal)	2.012	1.000	1.032	2.195	1.000	1.034
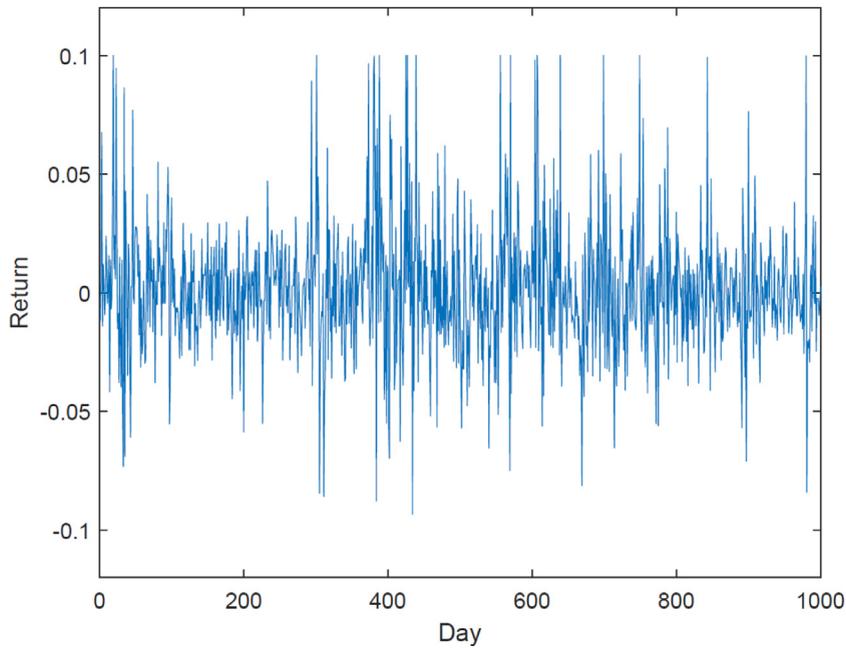
Fig. 1. Daily Inner Mongolia Baotou Steel Union stock returns from December 11, 2008 to January 21, 2013
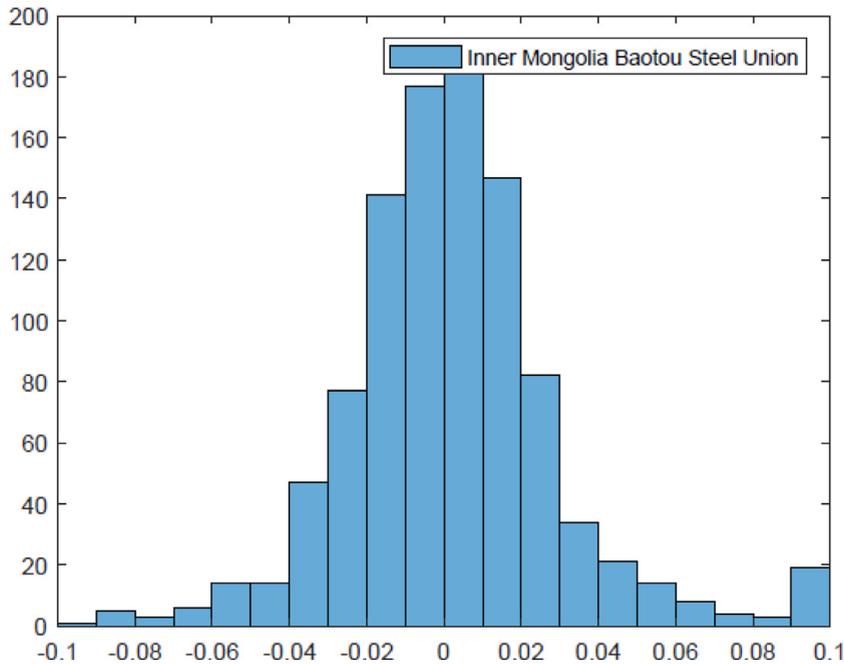
Fig. 2. Empirical Distribution of Inner Mongolia Baotou Steel Union stock returns.

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 EGARCH volatility model, (16), and then generalize so as to have two components for volatility with leverage effects, as in Harvey and Lange (2018), that is

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

By fitting a model that takes account of the censoring, we are able to provide estimates of the underlying level of volatility.

6.3. Empirical Findings

Table 2 and Table 3 report the estimation results of the standard and censored DCS models for the three Chinese stocks using one and two component models. The estimated parameters turn out 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. (Recall that the logistic score takes a relatively simple form and that, unlike the t-distribution, there is no shape parameter to be estimated.)

Figure 3 shows the ACF plot of scale scores from the best model, the censored two-component Student t, for Inner Mongolia Baoshan Steel Union. There is no indication of significant serial correlation and other diagnostics were satisfactory. The same was true for the other two stocks. Figure 4 shows the estimated two components of the scale. The corresponding figure for the logistic distribution was similar.
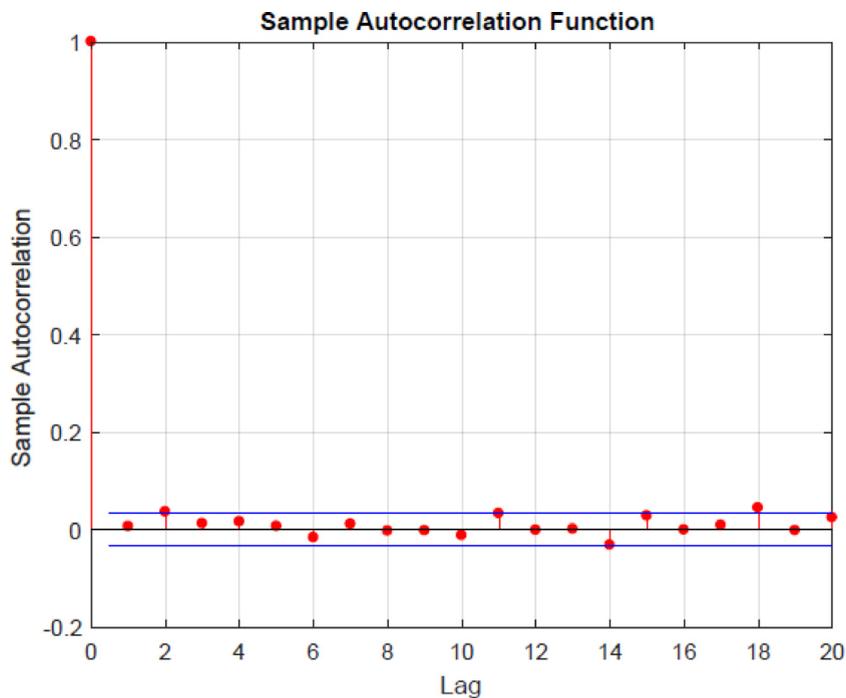
Figure 5 shows the estimated dynamic probability of hitting the stock price limit for the Inner Mongolia Baotou Steel Union stock returns during the whole sample period. The probability is computed from the model used to construct Figures 3 and 4. The average of the estimated dynamic probabilities is very close to the unconditional probability of hitting the price limit calculated from the original returns.

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

Table 2

Estimation results using the 1000 observations on returns for three Chinese stocks: one component model

|	Baoshan Iron & Steel	Hua Xia Bank	Inner Mongolia Baotou Steel Union
	Panel A: SV model with Gaussian Distribution		
μ	0.0014	0.0010	0.0003
	(0.0004)	(0.0003)	(0.0001)
ω	-3.6140	-4.0197	-3.9556
	(1.0213)	(0.7862)	(1.0089)
ϕ_λ	0.9198	0.9600	0.9895
	(0.2315)	(0.3421)	(0.3336)
κ_λ	0.0400	0.0254	0.0072
	(0.0120)	(0.0121)	(0.0034)
log likelihood	-2241.1552	-2645.8744	-2839.4039
BIC	4494.3104	5297.7488	5670.8077
	Panel B: SV model with Logistic Distribution		
μ	0.0008	0.0006	0.0004
	(0.0003)	(0.0003)	(0.0002)
ω	-4.5132	-4.8804	-5.3379
	(1.2321)	(1.0987)	(1.3236)
ϕ_λ	0.9433	0.9588	0.9782
	(0.2554)	(0.3008)	(0.2978)
κ_λ	0.0338	0.0186	0.0267
	(0.0123)	(0.0032)	(0.0114)
log likelihood	-2230.1476	-2639.7811	-2818.7886
BIC	4472.2951	5291.5621	5649.5773
	Panel C: SV model with Student t Distribution		
μ	0.0001	0.0003	0.0000
	(0.0000)	(0.0001)	(0.0000)
ω	-3.8449	-4.1574	-3.6599
	(0.9832)	(1.0081)	(1.1342)
ϕ_λ	0.9475	0.9941	0.9999
	(0.3426)	(0.4089)	(0.4231)
κ_λ	0.0713	0.0306	0.0283
	(0.0227)	(0.0104)	(0.0097)
ν	3.9478	6.0512	4.4792
	(1.0987)	(2.7867)	(1.7765)
log likelihood	-2224.3694	-2615.3882	-2784.9099
BIC	4463.7389	5245.7765	5584.8199

Fig. 3. Correlogram of scale scores from censored DCS two component model with Student t distribution fitted to Inner Mongolia Baotou Steel Union stock returns.

Table 3

Estimation results using the 1000 observations on returns for three Chinese stocks: two component model

	Baoshan Iron & Steel	Hua Xia Bank	Baotou Steel Union
Panel A: two component SV model with Gaussian Distribution			
μ	0.0014	0.0010	0.0014
	(0.0007)	(0.0008)	(0.0010)
ω	-3.6147	-4.0191	-3.9101
	(1.1897)	(1.8765)	(1.9098)
$\phi_{\lambda,1}$	0.9903	0.9736	0.9932
	(0.4241)	(0.3998)	(0.4056)
$\phi_{\lambda,2}$	0.7873	0.8447	0.8806
	(0.3034)	(0.2128)	(0.2225)
$\kappa_{\lambda,1}$	0.0347	0.0146	0.0060
	(0.0114)	(0.0067)	(0.0023)
$\kappa_{\lambda,2}$	0.0056	0.0108	0.0051
	(0.0011)	(0.0043)	(0.0024)
κ_1^*	0.0858	0.0704	0.0172
	(0.0233)	(0.0267)	(0.0085)
κ_2^*	0.1309	0.1360	0.0386
	(0.0457)	(0.0378)	(0.0112)
log likelihood	-2189.5540	-2602.1798	-2770.6155
BIC	4403.1079	5228.3595	5565.2311
Panel B: two component SV model with Logistic Distribution			
μ	0.0014	0.0006	0.0004
	(0.0071)	(0.0004)	(0.0002)
ω	-4.6147	-4.8692	-5.3451
	(1.0987)	(1.1214)	(2.0908)
$\phi_{\lambda,1}$	0.9903	0.9497	0.9819
	(0.4543)	(0.3326)	(0.2928)
$\phi_{\lambda,2}$	0.8873	0.8427	0.8194
	(0.2321)	(0.2897)	(0.3336)
$\kappa_{\lambda,1}$	0.0360	0.0228	0.0218
	(0.0124)	(0.0178)	(0.0109)
$\kappa_{\lambda,2}$	0.0046	0.0174	0.0097
	(0.0015)	(0.0043)	(0.0022)
κ_1^*	0.0858	0.0175	0.0430
	(0.0321)	(0.0054)	(0.0156)
κ_2^*	0.1309	0.0275	0.0644
	(0.0531)	(0.0113)	(0.0328)
log likelihood	-2172.5540	-2590.5984	-2719.1177
BIC	4369.1079	5205.1968	5462.2355
Panel C: two component SV model with Student t Distribution			
μ	0.0002	0.0002	0.0001
	(0.0001)	(0.0001)	(0.0001)
ω	-3.8777	-3.5777	-3.7488
	(-1.2245)	(-1.7865)	(-1.5045)
$\phi_{\lambda,1}$	0.9897	0.9731	0.9887
	(0.3321)	(0.4376)	(0.3089)
$\phi_{\lambda,2}$	0.8921	0.8620	0.8742
	(0.2248)	(0.3435)	(0.4007)
$\kappa_{\lambda,1}$	0.0153	0.0215	0.0121
	(0.0032)	(0.0046)	(0.0055)
$\kappa_{\lambda,2}$	0.0015	0.0116	0.0019
	(0.0007)	(0.0048)	(0.0005)
κ_1^*	0.0048	0.0028	0.0030
	(0.0022)	(0.0013)	(0.0011)
κ_2^*	0.0472	0.0342	0.0314
	(0.0201)	(0.0112)	(0.0108)
ν	4.4631	4.4831	5.0274
	(1.7862)	(2.2001)	(2.6754)
log likelihood	-2164.4532	-2583.8890	-2707.3345
BIC	4355.9064	5194.7780	5441.6690
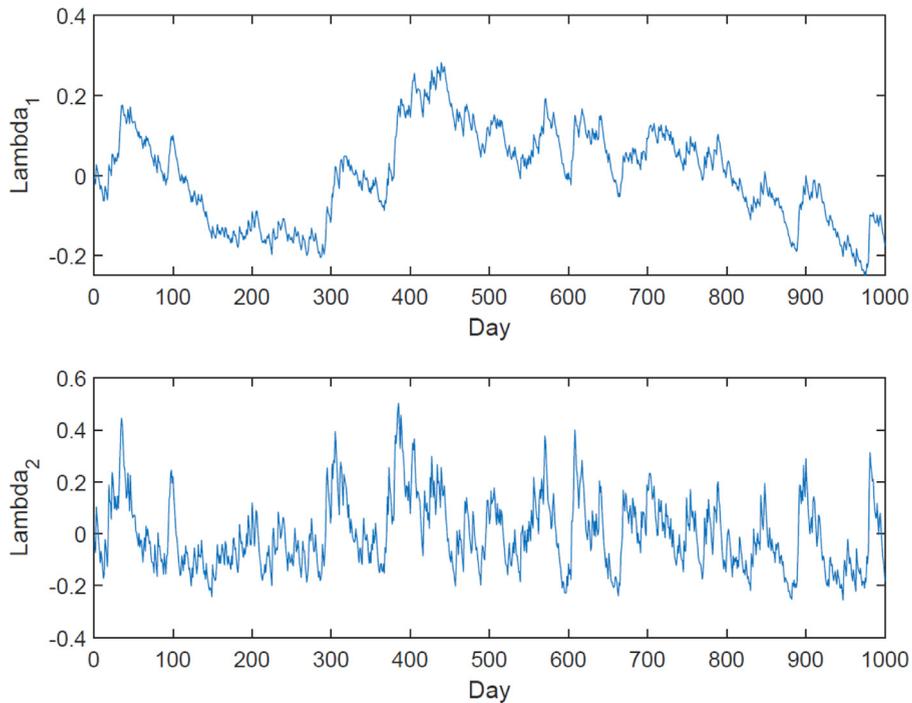
Fig. 4. Estimated components of scale, λ_1 and λ_2, from a censored two-component Student t model fitted to Inner Mongolia Baotou Steel Union stock returns. The top panel is for the long-run component (λ_1).
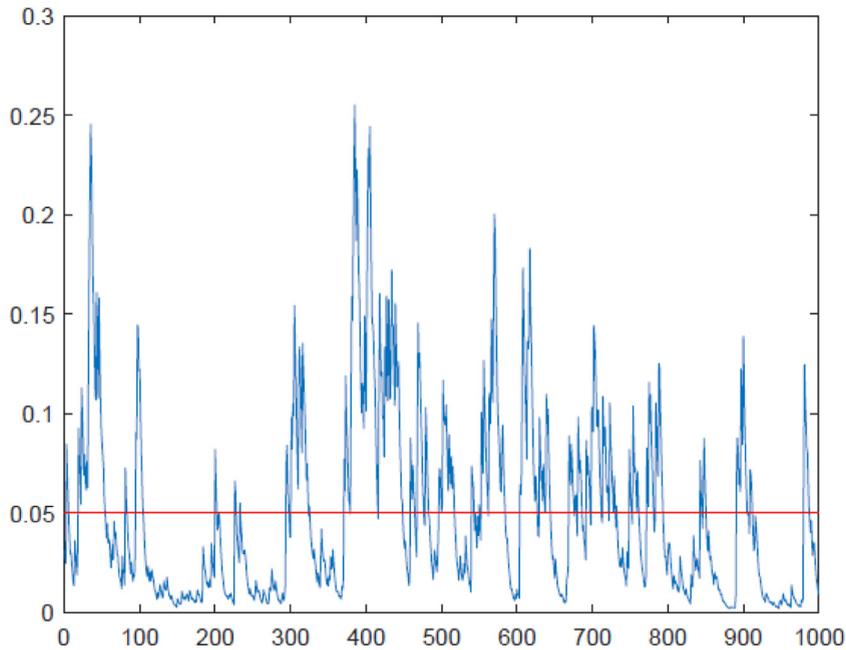
Fig. 5. Estimated probability of price limit hitting for Inner Mongolia Baotou Steel Union stock returns from two component model with Student t distribution. The horizontal line shows the unconditional probability of price limit hitting.

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. The viability of models when scale is dynamic is illustrated with data on Chinese stock returns. A censored two-component Student t EGARCH model fits well and produces an estimate of the probability that the next observations will be subject to censoring.

Acknowledgment

References

Abramowitz, M., Stegun, I.A. (Eds.), 1964. Handbook of mathematical functions. New York: Dover Publications Inc.

Allik, B., Miller, C., Piovoso, M.J., Zurakowski, R., 2016. The tobit kalman filter: an estimator for censored measurements. IEEE Transactions on Control Systems Technology 24, 365–371.

Amemiya, T., 1985. Advanced Econometrics. Harvard University Press.

Caivano, M., Harvey, A.C., 2014. Time series models with an EGB2 conditional distribution. Journal of Time Series Analysis 34, 558–571.

Chen, T., Gao, Z.Y., He, J.B., Jiang, W.X., Xiong, W., 2019. Daily price limits and destructive market behavior. Journal of Econometrics 208, 249–264.

Creal, D., Koopman, S.J., Lucas, A., 2013. Generalized autoregressive score models with applications. Journal of Applied Econometrics 28, 777–795.

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

Harvey, A., 2021. Score-driven time series models. Forthcoming in Annual Review of Statistics and Its Application 94. doi:10.1146/annurev-statistics-040120-021023.

Harvey, A.C., 2013. Dynamic models for volatility and heavy tails: with applications to financial and economic time series. In: Econometric society monograph. Cambridge University Press.

Harvey, A.C., Ito, R., 2019. Modeling time series when some observations are zero. Journal of Econometrics 14, 33–43.

Harvey, A.C., Lange, R.J., 2017. Volatility modeling with a generalized t-distribution. Journal of Time Series Analysis 38, 175–190.

Harvey, A.C., Lange, R.J., 2018. Modeling the interactions between volatility and returns using EGARCH-M. Journal of Time Series Analysis 39, 909–919.

Harvey, A.C., Thiele, S., 2016. Testing against changing correlation. Journal of Empirical Finance 38, 575–589.

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

Koopman, S.J., Lucas, A., Scharth, M., 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., McDonald, J.B., 2014. Partially adaptive estimation of censored regression models. Econometric Reviews 33, 732–750.

Manrique, A., Shephard, N., 1998. Simulation-based likelihood inference for limited dependent processes. The Econometrics Journal 1, 174–202.

McDonald, J., Newey, W.K., 1988. Partially adaptive estimation of regression models via the generalized t distribution. Econometric Theory 4, 428–457.

McDonald, J.B., Xu, Y.J., 1995. A generalization of the beta distribution with applications. Journal of Econometrics 66, 133–152.

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

Varin, C., Vidoni, P., 2008. Pairwise likelihood inference for general state space models. Econometric Reviews 28, 170–185.

Wang, C., Chan, K.S., 2018. Quasi-likelihood estimation of a censored autoregressive model with exogenous variables. Journal of the American Statistical Association 113, 1135–1145.

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