Journal of Econometrics 172 (2013) 186–194

Contents lists available at SciVerse ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Linear and nonlinear regression with stable errors

John P. Nolan *, Diana Ojeda-Revah

Department of Mathematics and Statistics, American University, United States

ARTICLE INFO

Article history:
Available online 23 August 2012

JEL classification:
C13
C16
C51
G17

Keywords:
Heavy tailed regression
Stable distributions
Score function

ABSTRACT

In this paper we describe methods and evaluate programs for linear regression by maximum likelihood when the errors have a heavy tailed stable distribution. The asymptotic Fisher information matrix for both the regression coefficients and the error distribution parameters are derived, giving large sample confidence intervals for all parameters. Simulated examples are shown where the errors are stably distributed and also where the errors are heavy tailed but are not stable, as well as a real example using financial data. The results are then extended to nonlinear models and to non-homogeneous error terms.

1. Introduction

Ordinary least squares (OLS) is a well established and important procedure for solving regression problems when the errors are normally distributed. Although small departures from normality of the error terms do not affect the regression coefficients greatly, errors with a heavier tailed distribution can result in extreme observations and can significantly affect the estimated OLS regression coefficients. Our focus in this paper is when the error terms have an infinite variance α-stable distribution, possibly with skewness, where OLS can perform very poorly, giving a poor fit and leading to faulty predictions.

Heavy tailed data has been observed in many applications in economics, finance, and engineering. Some basic references are Fama (1963), Mandelbrot (1963), Fama (1965), Nikias and Shao (1995), Adler et al. (1998), Rachev and Mittnik (2000) and Rachev (2003). In these applications, outliers are not mistakes, but an essential part of the error distribution. We are interested in both estimating the regression coefficients and in fitting the error distribution.

The standard linear regression model is

$$y_i = \sum_{j=1}^{k} x_{i,j}\theta_j + \epsilon_i \quad i = 1, \ldots, n, \tag{1}$$

where $x_{i,j}$ are independent variables, y_i are the response variables, θ_j are the coefficients of the regression, to be estimated, and the error terms ϵ_i are i.i.d. random variables. The reasoning used to justify the normal model for the error terms in (1) is that the error is the sum of many unmeasured terms. If the unmeasured terms have a finite variance, then the central limit theorem says that the error terms will be approximately normal. But when the unmeasured terms are heavy tailed, the generalized central limit theorem says that the error terms will be approximately stable.

The main drawback to the use of stable models has been the lack of closed form expressions for densities for all but a few special cases. However, with improved algorithms and increasing computer power, it is now possible to estimate parameters from these distributions numerically, making these models practical and easily applicable to data. In particular, Nolan (2001) implements maximum likelihood (ML) estimation for stable parameters when there is an i.i.d. sample, and gives references to other estimation procedures.

For linear regression when the errors are normally distributed, the OLS solution is the same as the ML solution. When the errors are not normally distributed, these two methods are not the same. Robust estimation methods use some technique that removes or downplays outliers. Here we do not use robust estimation, but instead use robust modeling, as advocated in Lange et al. (1989). In contrast to a robust estimation method, robust modeling is interested in both estimating the regression coefficients, and in fitting the error distribution. Specifically, we use a heavy tailed stable model for the errors, and ML to estimate all the parameters simultaneously.

* Correspondence to: Department of Mathematics and Statistics, American University, 4400 Massachusetts Avenue NW, WA, DC 20016-8050, United States. Tel.: +1 202 885 3140; fax: +1 202885 3155.
E-mail address: jpnolan@american.edu (J.P. Nolan).

0304-4076/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2012.08.008

There are several papers in the literature on this topic. Blattberg and Sargent (1971) proposed two methods for estimating a single regression coefficient in the infinite variance stable case. First, they developed a best linear unbiased estimator (BLUE), assuming symmetric errors and that the index of stability α is known. Those authors credit Wise (1966) with suggesting this approach; we are unable to comment on Wise's contribution because we cannot obtain this source. Second, they applied minimum sum of absolute error (MSAE), also called least absolute deviation or L_1 regression, where the regression coefficients minimize the sum of absolute deviations $\sum_{i=1}^{n} |y_i - (\theta_1 x_{i,1} + \cdots + \theta_k x_{i,k})|$. They showed through simulation, that when $1 < \alpha < 1.7$, BLUE and OLS perform approximately the same, and that MSAE is better than either. In contrast, when $1.7 \leq \alpha < 2$, OLS outperforms BLUE, which slightly outperforms MSAE. Heathcote (1982) discussed functional least squares, using a loss function that depends on the sample characteristic function of the data. These are all robust estimation procedures, and they do not yield a parametric fit to the residuals. El Barmi and Nelson (1997) considered a stable regression model, but they have restrictive assumptions: they assume α is known, β is known and zero, and they do not estimate the scale γ. McCulloch (1998) used ML to estimate the linear regression coefficients when the errors are symmetric stable.

More recently, Samorodnitsky et al. (2007) reconsidered BLUE regression in a more general setting, where the independent variables can also be random, with the independent variables and the dependent variables having any combination of light or heavy tailed distributions. Their analysis is informative and interesting: it shows when the BLUE estimates are consistent and gives optimal convergence rates for BLUE in terms of the tails of both the error distribution and the independent variable distribution. Autoregressive time series models for α-stable models are studied in Andrews et al. (2009). Hallin et al. (2010, in press) use rank based methods to fit linear models in the presence of infinite variance error terms.

The results in the current paper are based on dissertation work in Ojeda (2001). The ML approach described here extends previous work in several ways. It allows for non-symmetric error distributions, allows a larger range for α than McCulloch (1998), simultaneously estimates multiple regression coefficients $\theta_1, \ldots, \theta_k$ and the parameters of the stable error distribution, increases the accuracy and speed of the computations, gives asymptotic joint confidence regions for all the parameters with normal distributions for the estimates and $n^{-1/2}$ rate of convergence, and extends the ML approach to nonlinear regression with stable error terms.

We close this section with a discussion on parameterizations of stable distributions. Then in Section 2 a ML estimation procedure for linear regression with stable errors is given. We also describe three methods of computing confidence interval estimates for the coefficients, compute the score function for the location, and assess the accuracy of the estimation methods by simulations. In Section 3 we extend these results to nonlinear regression problems, allowing non-homogeneous errors. The paper ends with a short conclusion section.

Parameterizations of stable distributions

Stable distributions are a four parameter family, with index of stability $\alpha \in (0, 2]$, skewness $\beta \in [-1, 1]$, scale $\gamma > 0$, and location $\delta \in (-\infty, \infty)$. Since there is no known formula for the density or distribution function of a general stable random variable, these distributions are specified by their characteristic function. A random variable X is *stable* if $X \stackrel{d}{=} aZ + b$, where $a \neq 0$, $b \in \mathbb{R}$ and $Z = Z(\alpha, \beta)$ is a random variable with characteristic function

$$E \exp(iuZ) = \begin{cases} \exp\left(-|u|^{\alpha}\left[1 + i\beta(\operatorname{sign} u)\tan\frac{\pi\alpha}{2}(|u|^{-\alpha-1} - 1)\right]\right) & \alpha \neq 1 \\ \exp\left(-|u|\left[1 + i\beta(\operatorname{sign} u)\frac{2}{\pi}\log|u|\right]\right) & \alpha = 1. \end{cases}$$

This is the simplest parameterization of a standardized stable random variable that is jointly continuous in (α, β); it is the (M) parameterization on page 11 of Zolotarev (1986), and discussed on page 7 of Samorodnitsky and Taqqu (1994). It is important for numerical purposes to have a continuous parameterization, otherwise small changes in the data can result in arbitrarily large changes in the parameters.

There are numerous meanings of the location and scale parameters in the literature ((Zolotarev, 1986) gives six, (Nolan, in preparation) lists another five). We describe three parameterizations, which we denote by $S(\alpha, \beta, \gamma, \delta; k)$, $k = 0, 1, 2$. The user can select any of these three parameterizations to use for the estimates of the parameters $(\alpha, \beta, \gamma, \delta)$ in the program described below. See the comments in the next section for details on the connection between the parameterization and the intercept.

We will say

$$X \sim S(\alpha, \beta, \gamma, \delta; 0) \quad \text{if } X = \gamma Z(\alpha, \beta) + \delta.$$

This is the simplest parameterization that is continuous in all four parameters and is a scale-location family. A second parameterization is

$$X \sim S(\alpha, \beta, \gamma, \delta; 1)$$
$$\text{if } X = \begin{cases} \gamma\left(Z(\alpha, \beta) + \beta\tan\frac{\pi\alpha}{2}\right) + \delta & \alpha \neq 1 \\ \gamma\left(Z(1, \beta) + \beta\frac{2}{\pi}\log\gamma\right) + \delta & \alpha = 1. \end{cases}$$

This is the most commonly used parameterization in the modern literature, e.g. Samorodnitsky and Taqqu (1994). But it is not continuous in the parameters: as $\alpha \to 1$, the term $\beta\tan\frac{\pi\alpha}{2}$ diverges when $\beta \neq 0$, leading to an arbitrarily large shift in the center of the distribution. Also, it is not a scale-location family when $\alpha = 1$. Both of these facts make this a poor choice for numerical work. Finally, we will define

$$X \sim S(\alpha, \beta, \gamma, \delta; 2) \quad \text{if} \quad X = \gamma\alpha^{1/\alpha}(Z(\alpha, \beta) - m(\alpha, \beta)) + \delta,$$

where $m(\alpha, \beta)$ is the mode of $Z(\alpha, \beta)$. This parameterization is continuous in all parameters, it is a scale-location family, with the mode of X at the location parameter δ, and it has the property that $S(\alpha, \beta, \gamma, \delta; 2)$ converges to a normal distribution with mean δ and standard deviation γ as $\alpha \to 2$. This is not the case in the other parameterizations, where the limiting standard deviation is $\sqrt{2}\gamma$. More information on the parameterizations can be found in Nolan (1998), where values of $m(\alpha, \beta)$ have been computed numerically.

2. Linear regression

The regression model (1) can be written in matrix form as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\epsilon},$$

where $\mathbf{X} = (x_{i,j})_{n\times k}$ is the design matrix, $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)^T$ are the regression coefficients, and $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)^T$ are the errors. We will assume that there is a constant term in the model, say $x_{i,1} = 1$ for all i. Then there are $k + 3$ parameters to be