A Note on Least Absolute Deviation Estimation of a Threshold Model

Mehmet Caner

Econometric Theory, Vol. 18, No. 3. (Jun., 2002), pp. 800-814.

Econometric Theory, **18**, 2002, 800–814. Printed in the United States of America.
DOI: 10.1017.S0266466602183113

A NOTE ON LEAST ABSOLUTE DEVIATION ESTIMATION OF A THRESHOLD MODEL

MEHMET CANER
University of Pittsburgh

This paper develops the limit law for the least absolute deviation estimator of the threshold parameter in linear regression. In this respect, we extend the literature of threshold models. The existing literature considers only the least squares estimation of the threshold parameter (see Chan, 1993, *Annals of Statistics* 21, 520–533; Hansen, 2000, *Econometrica* 68, 575–605). This result is useful because in the case of heavy-tailed errors there is an efficiency loss resulting from the use of least squares. Also, for the first time in the literature, we derive the limit law for the likelihood ratio test for the threshold parameter using the least absolute deviation technique.

1. INTRODUCTION

In a regression model we are usually interested in whether the regression coefficients are stable or not. We select subsamples to detect this behavior. This selection may be based on continuous variables such as firm size. Rather than using ad hoc models to select subsamples we can use the threshold regression models. This model can be written as

$$y_i = \theta_1' x_i + e_i, \qquad q_i \leq \gamma, \tag{1}$$

$$y_i = \theta_2' x_i + e_i \qquad q_i > \gamma, \tag{2}$$

where q_i is the observed threshold variable. The subsamples are selected according to the value that the threshold variable takes. The random variable e is the regression error.

Hansen (2000) explains different applications of threshold models in the econometric literature. One example is the so-called threshold autoregressive (TAR) model of Tong (1983, 1990). The TAR model is simple and parsimonious and allows for nonlinearities in the conditional expectation function. Thresh-

I thank Donald W.K. Andrews (the co-editor) and two anonymous referees for suggestions that substantially improved the paper. I also thank Bilkent University, Department of Economics, which provided extraordinary research support. Asad Zaman, Shinichi Sakata, and Erdem Basci provided discussion related to technical points in the paper. I also thank the Turkish Academy of Sciences for its financial support. I owe special thanks to Bruce Hansen, who introduced me to threshold models. Address correspondence to: Mehmet Caner, Department of Economics, University of Pittsburgh, W.W. Posvar Hall, Pittsburgh 15260, PA, USA; e-mail: caner@pitt.edu.

  0266-4666/02 $9.50

old models are also special cases of switching models and mixture models. To understand these complex models better, it is useful to analyze the simple threshold model described in (1) and (2). Recently, we have seen these models in the literature; see Tong (1983), Tsay (1997), Chen and Lee (1995), Montgomery, Zarnowitz, Tsay, and Tiao (1996), and Altissimo and Violante (1996).

Estimation of the threshold parameter γ is by least squares (LS) estimation technique in Chan (1993) and Hansen (2000). Robust estimation of the threshold parameter has not been analyzed in the literature. Least absolute deviation (LAD) estimation of a shift is analyzed by Bai (1995). He derives the limit theory for the change-point estimate. The change-point literature considers the model (1) and (2) with $q_i = i$. In the threshold case q_i is observable and may be an element of x_i.

The aim of this paper is to obtain the asymptotic distribution of the threshold parameter γ by using the LAD method. This is the first paper to do so in the literature. We take the approach that is used in the change-point literature. We let the threshold effect $\delta_n = \theta_1 - \theta_2$ converge to zero as the sample size increases. This assumption is used both in the change-point and the threshold literature (see Picard, 1985; Bai, 1995; Hansen, 2000). The limit law for the threshold estimate consists of the functional of two-sided Brownian motion. The limit is similar to what is found in the change-point case of Bai (1995) although the scale factor is different.

We also study the likelihood ratio tests for the threshold parameter. This is not done in the change-point case of Bai (1995). By using the likelihood ratio tests we may construct confidence intervals for the threshold parameter.

When the data are observed from a thick-tailed distribution we show that the LAD estimate of γ is more efficient than the LS estimate. This fact was known in the case of no structural change and the change-point models. We extend this to the threshold regression model.

Section 2 details the method and compares the threshold and change-point models. Section 3 presents the assumptions and the limit law. In Section 4, we form the likelihood ratio test for γ. Section 5 concludes. The Appendix contains all the proofs. The expression $\|.\|$ denotes the Euclidean norm. The symbol $\Rightarrow$ denotes weak convergence with respect to the uniform metric, and $\xrightarrow{d}$ denotes converges in distribution.

2. MODEL

The observed sample is $\{y_i, q_i, x_i\}_{i=1}^n$, where y_i and q_i are real valued and x_i is an m-vector. The threshold variable q_i may be an element of x_i and is assumed to have a continuous distribution. A threshold regression is simply modeled as in (1) and (2). We can rewrite (1) and (2) in a single equation:

$$y_i = x_i'\theta + x_i(\gamma)'\delta_n + e_i, \qquad (3)$$

where $\theta = \theta_2$, $\delta_n = \theta_1 - \theta_2$, $x_i(\gamma) = x_i 1_{\{q_i \le \gamma\}}$, and $1_{\{.\}}$ is an indicator function. The results also generalize to the case where only a subset of parameters switches between regimes and to the case where some regressors only enter in one of the two regimes.

Denote the regression parameters by (θ, δ, γ). Let

$$S_n(\theta,\delta,\gamma) = \sum_{i=1}^{n} |y_i - x_i'\theta - x_i(\gamma)'\delta| \quad \textbf{(4)}$$

be the sum of the absolute deviations. By definition, the LAD estimators jointly minimize the objective function in (4). For this minimization γ only takes values in a compact set $\Gamma = [\gamma_l, \gamma_u]$, where γ_l is the lower bound and γ_u is the upper bound.

To obtain $\hat{\gamma}$ we concentrate out γ in (4) and derive conditional LAD estimators of θ, δ. Then, $\hat{\gamma}$ is the value that minimizes this concentrated sum of absolute errors function. Because $S_n(\gamma)$ takes on less than n distinct values $\hat{\gamma}$ can be defined uniquely:

$$\hat{\gamma} = \arg\min_{\gamma \in \Gamma_n} S_n(\gamma), \quad \textbf{(5)}$$

where $\Gamma_n = \Gamma \cap q_1, q_2, \cdots, q_n$. Computing $\hat{\gamma}$ requires at most n function evaluations. The slope estimates can be computed as $\hat{\theta} = \hat{\theta}(\hat{\gamma})$ and $\hat{\delta} = \hat{\delta}(\hat{\gamma})$. As in Hansen (2000), if n is large Γ can be approximated by a grid.

The threshold model in (1) and (2) is similar to the change-point model. In the change-point model $q_i = i$. If the observed values of q_i are distinct, to estimate the parameters in the threshold model, as a first step, we proceed as in the change-point model and sort the data based on q_i. Also when regressors contain the threshold variable q_i, the threshold model is similar to a change-point model with trended data, because sorting the data by q_i induces a trend in the regressors. Even though there are similarities, change-point models are different than the threshold models. Regarding the differences in the distribution we should note that the stochastic process $R_n^*(\gamma) = \sum_{i=1}^{n} x_i \operatorname{sgn}(e_i) 1_{\{q_i \le \gamma\}}$ is not a martingale in the threshold model, which complicates the limit theory for $\hat{\gamma}$. This requires the use of a different set of asymptotic tools compared with the change-point case. The limit law of the estimator of the change point and the threshold parameter in LAD estimation is the same only when the threshold variable is independent from the regressors; this is explained in detail in Section 3.2.

3. THE LIMIT LAW

3.1. Assumptions

Define the moment functionals

$M(\gamma) = E[f(0|x_i)x_i x_i' 1_{\{q_i \le \gamma\}}]$,

$N(\gamma) = E[x_i x_i' 1_{\{q_i \le \gamma\}}]$,

$D(\gamma) = E[f(0|x_i)x_i x_i' | q_i = \gamma]$, and

$$V(\gamma) = E[x_i x_i' | q_i = \gamma],$$

where $f(0|x_i)$ is the conditional density of errors at zero.

Let $g(q)$ denote the density function of q_i and γ_0 denote the true value of γ. Set $D = D(\gamma_0)$, $g = g(\gamma_0)$, and $V = V(\gamma_0)$.

Assumptions.

(a) (x_i, q_i, e_i) is strictly stationary, ergodic, and ρ-mixing sequence with mixing coefficients $\rho_m = O(e^{-\xi m}), \xi > 0$.
(b) The errors e_i admit a positive and continuous density function $f(.)$ in a neighborhood of zero and having zero conditional median.
(c)

$$n^{-1/2} \max_{i \leq i \leq n} \|x_i\| (\log n)^{1/2} = o_p(1).$$

Furthermore

$$E\|x_i\|^4 < \infty.$$

(d) For all $\gamma \in \Gamma$, $E[|x_i|^4 | q_i = \gamma] \leq C$, $E[f(0|x_i)^2 |x_i|^4 | q_i = \gamma] \leq C$, for some $C < \infty$ and $0 < g(\gamma) \leq \bar{g} < \infty$.
(e) $g(\gamma)$, $D(\gamma)$, $V(\gamma)$ are continuous at $\gamma = \gamma_0$.
(f) $\delta = \delta_n = cn^{-\alpha}$ with $c \neq 0$ and $0 < \alpha < \frac{1}{2}$.
(g) $c'Dc > 0$, $c'Vc > 0$.
(h) There exists M, N such that $M > M(\gamma) > 0$, $N > N(\gamma) > 0$ for all $\gamma \in \Gamma$, where Γ is a compact set $[\gamma_l, \gamma_u]$ and γ_l and γ_u are the lower and upper bounds of this set, respectively.

Assumption (a) allows for time-series data. The ρ-mixing assumption controls the degree of time series dependence and is weaker than uniform mixing yet stronger than strong mixing. For further information on the ρ-mixing concept see Peligrad (1982). Assumptions (b) and (c) are standard in the LAD literature (see Pollard, 1991; Weiss, 1991; Bai, 1995). Assumption (d) is standard in the threshold literature (see Hansen, 2000). Assumption (e) imposes a continuous distribution on the threshold variable. Assumption (f) is borrowed from the change-point and the threshold literature; it basically requires that the difference in regression slopes gets smaller with increasing sample size (see Picard, 1985; Bai, 1995; Hansen, 2000). This is needed to get a nuisance-parameter-free limit law. Assumption (g) is a full-rank condition needed to have a nondegenerate asymptotic distribution. Assumption (h) is a full-rank condition that excludes multicollinearity.

3.2. Asymptotic Distribution of the Threshold Parameter

In this section we derive the limit law for the threshold parameter γ.

THEOREM 1. *Under Assumptions (a)–(h),*

$$n^{1-2\alpha}(\hat{\gamma}-\gamma_0) \xrightarrow{d} \omega T,$$

where

$$\omega = \frac{c'Vc}{4(c'Dc)^2 g},$$

$$T = \arg \max_{-\infty<r<\infty} \left[\frac{-1}{2}|r| + W(r) \right],$$

and $W(r)$ is a two-sided Brownian motion.

A two-sided Brownian motion on the real line is defined as $W(r) = W_1(-r)$ when $r < 0$, $W(r) = W_2(r)$ when $r > 0$, and $W(r) = 0$ when $r = 0$. Here $W_1(-r)$ and $W_2(r)$ are two independent standard Brownian motions on $[0,\infty)$.

Now we consider the similarities and differences of the threshold and the change-point models. We can see that the difference in the limit theory stems from the precision term ω (Bai, 1995, Theorems 3iii and 4iii). First we analyze the similarities between those two models. When x_i and q_i are independent, the limit law for $\hat{\gamma}$ is the same in both models.

To see this point in a simple setting, assume errors e_i are independent and identically distributed (i.i.d.), have a zero median, and admit a positive and continuous density function in a neighborhood of zero. They are independent of the regressors and the threshold variable, and (x_i, q_i) are i.i.d. This replaces our Assumptions (a) and (b). Then the precision term in our Theorem 1 simplifies to

$$\omega = \frac{1}{4f(0)^2 c'E(x_i x_i')c},$$

which is equivalent to the scale term in Theorem 3iii of Bai (1995). The equivalence of the limit laws in the threshold and the change-point cases under the more general Assumptions (a)–(h) can also be established.[1]

The difference between the threshold and the change-point model arises when the threshold variable is a random variable among the regressors. Then the ω term in each model is different in the two models. The precision term in the threshold model is given in Theorem 1 and consists of the ratio of conditional moment matrices. In the change-point model it is given by

$$\omega = \frac{c'E(x_i x_i')c}{4\{c'E[f(0|x_i)x_i x_i']\}^2}.$$

So this is the ratio of the unconditional moment matrices.

This limit is similar to the limit law for the change-point estimator and the threshold estimator in the LS case under the small effect asymptotics. For example in the case of LS estimation of the threshold the limit T is the same but the precision ratio is

$$\omega = \frac{c'E(x_i x_i' e_i^2 | q_i = \gamma_0)c}{\{c'E[x_i x_i' | q_i = \gamma_0]c\}^2 g}$$

by Theorem 1 of Hansen (2000).

There is an advantage of the LAD estimate over the LS estimate of the threshold parameter. To see this point define the rate of efficiency of LAD relative to least squares ARE as their ratio of asymptotic variances (Bai, 1995, p. 415). This is nothing more than the ratio of the precision terms. Then take the case of conditional homoskedasticity, so the precision term for the LS case in the limit law in Theorem 1 of Hansen (2000) is

$$\omega_{\text{LS}} = \frac{\sigma^2}{c'E(x_i x_i' | q_i = \gamma_0)cg},$$

where $Ee_i^2 = \sigma^2$, for all $i = 1, \ldots n$. The precision term ω_{LAD} is

$$\omega_{\text{LAD}} = \frac{1}{[c'E(x_i x_i' | q_i = \gamma_0)cg]4f(0)^2}.$$

So

$$ARE = \frac{\omega_{\text{LS}}}{\omega_{\text{LAD}}} = 4f(0)^2\sigma^2,$$

where this is the ARE of the median to mean in i.i.d. sampling. For example in the case of double exponential distribution if e_i are i.i.d. then $ARE = 2$, indicating LAD is more efficient compared with LS in the case of heavy-tailed distributions.

It is seen that when the threshold effect is large (when α is small in $\delta_n = cn^{-\alpha}$), then the rate of convergence approaches n, so the precision of the estimator is increased. The distribution function for T is given in Bhattacharya and Brockwell (1976):

$$P(T \le x) = 1 + \left(\frac{x}{2\pi}\right)^{1/2} \exp(-x/8) + \frac{3}{2}\exp(x)\Phi\left(-\frac{3x^{1/2}}{2}\right) - \left(\frac{x+5}{2}\right)\Phi\left(-\frac{x}{2}\right),$$

where $\Phi(x)$ denotes the cumulative standard normal distribution function for $x < 0, P(T \le x) = 1 - P(T \le -x)$.

4. LIKELIHOOD RATIO TEST

To test hypotheses about the threshold parameter we benefit from the likelihood ratio statistic. We formulate the null as $H_0: \gamma = \gamma_0$. Using the likelihood ratio test in the LAD framework in Koenker and Bassett (1982) we have

$$LR_n(\gamma) = 4[S_n(\gamma) - S_n(\hat{\gamma})].$$

The likelihood ratio test of H_0 will reject for large values of test statistic $LR_n(\gamma)$. This test is based on the difference between the sum of the absolute residuals in the restricted and the unrestricted models. The test statistic is not the one suggested by Koenker and Bassett (1982); however it reduces to their suggestion when there is conditional homoskedasticity.

Using this framework we have the following theorem.

THEOREM 2. *Under Assumptions (a)–(h),*

$$LR_n(\gamma_0) \xrightarrow{d} \eta^2\, \Xi,$$

where

$$\Xi = \max[2\, W(s) - |s|]$$

and

$$\eta^2 = \frac{c'Vc}{c'Dc}.$$

The distribution function of Ξ is

$$P(\Xi \leq x) = (1 - e^{-x/2})^2.$$

The critical values of the distribution in Theorem 2 are tabulated in Table 1.

If there is conditional homoskedasticity then

$$LR_n(\gamma_0) \xrightarrow{d} \Xi/f(0).$$

So in this case rewriting the likelihood ratio statistic as

$$LR_n^*(\gamma_0) = 4\hat{f}(0)[S_n(\gamma_0) - S_n(\hat{\gamma})] \xrightarrow{d} \Xi,$$

which is free of nuisance parameters. Note that this result depends on the existence of a consistent estimator for the density function for the errors at zero.

TABLE 1. Critical values in Theorem 2

$P(\Xi \leq x) =$	.80	.85	.90	.95	.99
$c_{xi}(\beta)$	4.5	5.10	5.94	7.35	10.59

Any kernel estimator such as the Epanechnikov kernel can be used for estimation of this density (Hardle and Linton, 1994), or a histogram estimator can be used as in Buchinsky (1994).

This test is not carried out in the change-point case. Bai (1995) only studies the estimation of the change-point model. This is the first time such a test also is carried out in the threshold framework. Careful analysis of the Bai (1995) paper shows us that, in the case of x_i being independent from q_i, the limit law for the test statistic is the same. Take the very simple case of i.i.d. regressors and the threshold variable with the added assumption of conditional homoskedasticity (these are simple cases of our Assumptions (a) and (b)). Using the test statistic $LR_n^*(\gamma_0)$ benefiting from the proof of Theorem 3iii of Bai (1995) in combination with the proof of Theorem 2 in Hansen (2000) we have, for the change-point case,

$$\eta^2 = \frac{c'E(x_i x_i')c}{c'E(x_i x_i')c} = 1.$$

In the threshold model when x_i and q_i are independent, using the discussion after Theorem 2, $\eta^2 = 1$.

This equivalence can also be shown using our Assumptions (a)–(h) in Theorem 3(iii) of Bai (1995). The same difference between the threshold and the change-point models in the case of Theorem 1 carries over to the Theorem 2 results also.

Note that we may use the likelihood ratio test to establish confidence intervals for γ. These are described in Theorem 3 of Hansen (2000). Details are given in Caner (1999).

Because the limit in Theorem 2 depends on the nuisance parameter η^2 we need a consistent estimate for that term. Along the lines of Hansen (2000, Sect. 3.4), let $r_{1i} = (\delta_n' x_i)^2$, $r_{2i} = f(0|x_i)(\delta_n' x_i)^2$. Then

$$\eta^2 = \frac{E(r_{1i}|q_i = \gamma_0)}{E(r_{2i}|q_i = \gamma_0)}$$

is the ratio of the conditional expectations. We have to use the sample counterparts of r_{1i}, r_{2i}. Let $\hat{r}_{1i} = (\hat{\delta}' x_i)^2$ and $\hat{r}_{2i} = K_{h,d}(\hat{e}_i)(\hat{\delta}' x_i)^2$, where $K_{h,d}(\hat{e}_i)$ is the following kernel suggested by Weiss (1991) to estimate conditional densities with dependent data:

$$K_{h,d}(\hat{e}_i) = h^{-1} K(\hat{e}_i/h).$$

The term $K(.)$ can be the simple kernel

$$K(\hat{e}_i/h) = 1_{\{|\hat{e}_i/h| \leq\}}/2,$$

where $\hat{e}_i$ is the LAD regression residual.

We can use kernel regression. Following Weiss (1991, Sect. 5) and Hansen (2000), the Nadaraya–Watson kernel estimator is

$$\hat{\eta}^2 = \frac{\sum_{i=1}^{n} K_h(\hat{\gamma} - q_i)\hat{r_{1i}}}{\sum_{i=1}^{n} K_h(\hat{\gamma} - q_i)\hat{r_{2i}}} = \frac{\sum_{i=1}^{n} K_h(\hat{\gamma} - q_i)(\hat{\delta}'x_i)^2}{\sum_{i=1}^{n} K_h(\hat{\gamma} - q_i)K_{h,d}(\hat{e}_i)(\hat{\delta}'x_i)^2},$$

where $K_h(u)$ in the preceding formula can be the Epanechnikov kernel. The bandwidth can be selected according to a minimum square error criterion (see Weiss, 1991; Hardle and Linton, 1994).

The slope coefficients can be obtained as in Hansen (2000), as if the threshold parameter is fixed. This is analyzed in Caner (1999).

5. CONCLUSION

This paper derives the limit law for the threshold parameter using a LAD estimator. We find that the LAD estimator is more efficient than the LS estimator in the case of thick-tailed distributions.

An interesting topic for future research may be analyzing the multiregime threshold models using LAD technique. Another worthwhile consideration is the behavior of LAD threshold estimators in a nonstationary framework.

NOTE

1. The precision term in Theorem 4iii of Bai (1995) is different given his Assumptions A.8 and A.9. This is because he benefits from a functional central limit theorem for dependent variables in his case (Wooldridge and White, 1988). However in our threshold case basically we can only benefit from Theorem 16.1 of Billingsley (1968). For details regarding this last point see the proofs of Lemmas A.3, A.4, and A.11 of Hansen (2000).

REFERENCES

Altissimo, F. & G. Violante (1996) Persistence and Non-linearity in US GNP and Unemployment: An Endogenous Delay Threshold VAR. Working paper, University of Pennsylvania, Department of Economics.

Bai, J. (1995) Least absolute deviation estimation of a shift. *Econometric Theory* 11, 403–436.

Bhattacharya, P.K. & P.J. Brockwell (1976) The minimum of an additive process with applications to signal estimation and storage theory. *Probability Theory and Related Fields* 37, 51–75.

Billingsley, P. (1968) *Convergence of Probability Measures.* New York: Wiley.

Buchinsky, M. (1994) Changes in the US wage structure 1963–1987: Application of quantile regression. *Econometrica* 62, 405–458.

Caner, M. (1999) Least Absolute Deviation Estimation of a Threshold Model. Mimeo, Bilkent University, Department of Economics.

Chan, K.S. (1993) Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Annals of Statistics* 21, 520–533.

Chen, C.W. & J.C. Lee (1995) Bayesian inference of threshold autoregressive models. *Journal of Time Series Analysis* 16, 483–492.
Gyorfi, L., W. Hardle, P. Sarda, & P. Vieu (1990) *Nonparametric Curve Estimation from Time Series*. New York: Springer-Verlag.
Hansen, B.E. (2000) Sample splitting and threshold estimation. *Econometrica* 68, 575–605.
Hardle, W. & O. Linton (1994) Applied nonparametric methods. In R.F. Engle & D. McFadden (eds.), *Handbook of Econometrics*, vol. 4, pp. 2295–2339. Amsterdam: Elsevier Science.
Koenker, R. & G. Bassett (1982) Tests of linear hypotheses and l_1 estimation. *Econometrica* 50, 1577–1583.
Montgomery, A.L., V. Zarnowitz, R. Tsay, & G.C. Tiao (1996) Non-linearities in Modeling and Forecasting the US Unemployment Rate. Working paper, University of Chicago Business School.
Newey, W.K. & D. McFadden (1994) Large sample estimation and hypotheses testing. In R.F. Engle & D. McFadden (eds.), *Handbook of Econometrics*, vol. 4, pp. 2113–2245. Amsterdam: Elsevier Science.
Peligrad, M. (1982) Invariance principles for mixing sequences of random variables. *Annals of Probability* 10, 968–981.
Picard, D. (1985) Testing and estimating change-points in time series. *Advances in Applied Probability* 17, 841–867.
Pollard, D. (1991) Asymptotics for least absolute deviation regression estimators. *Econometric Theory* 7, 186–199.
Tong, H. (1983) *Threshold Models in Non-linear Time Series Analysis*. Lecture Notes in Statistics 21, Berlin: Springer.
Tong, H. (1990) *Non-Linear Time Series: A Dynamical System Approach*. Oxford: Oxford University Press.
Tsay, R.S. (1997) Unit Root Tests with Threshold Innovations. Working paper, University of Chicago Business School.
Weiss, A. (1991) Estimating nonlinear dynamic models using least absolute error estimates. *Econometric Theory* 7, 46–68.
Wooldridge, J.M. & H. White (1988) Some invariance principles and central limit theorems for dependent and heterogenous processes. *Econometric Theory* 4, 210–230.

APPENDIX

We first begin with the consistency result for the threshold estimate $\hat{\gamma}$ (Theorem A.1). However the following lemma is needed for the consistency proof.

LEMMA A.1. *Under Assumptions (a)–(c) and (f) we have*

$$\inf_\gamma \inf_{\theta,\delta} \sum_{i=1}^{n} |e_i - n^{-1/2}x_i'\theta - n^{-1/2}x_i'(\gamma)\delta - \Delta x_i'(\gamma)\delta_n| - |e_i - \Delta x_i'(\gamma)\delta_n|$$

$$= O_p(\log n),$$

where $\Delta x_i(\gamma) = x_i(\gamma) - x_i(\gamma_0)$.

Proof of Lemma A.1. First, substitute $e_i - \Delta x_i(\gamma)'\delta_n$ for e_i in Lemma A.1ii of Bai (1995). Then the proof proceeds as in Lemma A.1ii of Bai (1995). However there is one change that has to be made. Instead of Lemma A.2 of Bai (1995), we use the exponential inequality for ρ-mixing random variables (see Gyorfi, Hardle, Sarda, and Vieu, 1990, Theorem 2.2.2) given our Assumptions (a) and (b). ∎

THEOREM A.1. *Under Assumptions (a)–(h),*

$$\hat{\gamma} \xrightarrow{p} \gamma_0.$$

Proof. Conditional on γ, write up the objective function

$$S_n(\gamma) = \sum_{i=1}^{n} |y_i - x_i'\theta - x_i'(\gamma)\delta|. \tag{A.1}$$

As in equation (5) of Bai (1995), the following reparametrized objective function is easier to work with:

$$S_n(\gamma) = \sum_{i=1}^{n} |y_i - x_i'(\theta_0 + n^{-1/2}\theta) - x_i'(\gamma)(\delta_n + n^{-1/2}\delta)|. \tag{A.2}$$

Note that $(\hat{\theta}, \hat{\delta})$ minimizes (A.1) given γ and $(\tilde{\theta}, \tilde{\delta})$ minimizes (A.2) given γ. Then see that $\tilde{\theta} = n^{1/2}(\hat{\theta} - \theta_0)$, $\tilde{\delta} = n^{1/2}(\hat{\delta} - \delta_n)$.

We can rewrite (A.2) using (3) and Assumptions (a) and (b) and adding and subtracting $x_i(\gamma)'\delta_n$:

$$S_n(\gamma) = \sum_{i=1}^{n} |e_i - n^{-1/2}x_i'\theta - n^{-1/2}x_i(\gamma)'\delta - \delta_n'\Delta x_i(\gamma)|. \tag{A.3}$$

Similar arguments and details regarding parametrization of the objective function can be found in Bai (1995, p. 408).

Our minimization problem is

$$\hat{\gamma} = \operatorname{argmin}_{\gamma \in \Gamma} S_n(\gamma).$$

However this is equivalent to

$$\hat{\gamma} = \operatorname{argmax}_{\gamma \in \Gamma} Q_n(\gamma), \tag{A.4}$$

where $Q_n(\gamma) = S_n(\gamma_0) - S_n(\gamma)$. That can be written as

$$Q_n(\gamma) = \sum_{i=1}^{n} |e_i - n^{-1/2}x_i'\theta - n^{-1/2}x_i'(\gamma_0)\delta| - \sum_{i=1}^{n} |e_i - n^{-1/2}x_i'\theta - n^{-1/2}x_i'(\gamma)\delta - \Delta x_i'(\gamma)\delta_n|. \tag{A.5}$$

Next, add and subtract $\sum_{i=1}^{n}|e_i - \Delta x_i'(\gamma)\delta_n|$ and $\sum_{i=1}^{n}|e_i|$ to and from (A.5) and multiply each side of (A.5) by $n^{2\alpha-1}$ to have

$$
\begin{aligned}
n^{2\alpha-1}&Q_n(\gamma)\\
&= -n^{2\alpha-1}\left(\sum_{i=1}^{n}|e_i - \Delta x_i'(\gamma)\delta_n| - |e_i|\right)\\
&\quad + n^{2\alpha-1}\left(\sum_{i=1}^{n}|e_i - n^{-1/2}x_i'\theta - n^{-1/2}x_i'(\gamma_0)\delta| - |e_i|\right)\\
&\quad - n^{2\alpha-1}\left(\sum_{i=1}^{n}|e_i - n^{-1/2}x_i'\theta - n^{-1/2}x_i'(\gamma)\delta - \Delta x_i'(\gamma)\delta_n| - |e_i - \Delta x_i(\gamma)'\delta_n|\right).
\end{aligned}
\tag{A.6}
$$

To obtain the consistency of $\hat{\gamma}$, we need to analyze the three terms on the right-hand side of (A.6) separately. First, it can be shown by using Pollard (1991) or equation (A.41) of Bai (1995) that

$$
\begin{aligned}
n^{2\alpha-1}&\sum_{i=1}^{n}|e_i - \Delta x_i'(\gamma)\delta_n| - |e_i|\\
&= n^{2\alpha-1}\delta_n'\sum_{i=1}^{n}\Delta x_i(\gamma)D_i\\
&\quad + n^{2\alpha-1}\delta_n'\left(\sum_{i=1}^{n}f(0|x_i)\Delta x_i(\gamma)\Delta x_i(\gamma)'\right)\delta_n + o_p(1),
\end{aligned}
\tag{A.7}
$$

where $D_i = \operatorname{sgn}(e_i)$ and sgn denotes the sign function.

We need the following result to consider the first-term on the right-hand side of (A.7):

$$
n^{-1/2}\sum_{i=1}^{n}x_i(\gamma)\operatorname{sgn}(e_i) = O_p(1). \tag{A.8}
$$

Equation (A.8) can be obtained by using the same arguments as in Lemma A.4 of Hansen (2000) by replacing e_i there with $\operatorname{sgn}(e_i)$ and using Assumptions (a) and (b). The same result in (A.8) holds if we replace γ with γ_0. Because $\alpha < \frac{1}{2}$ and $\Delta x_i(\gamma) = x_i 1_{\{q_i \le \gamma\}} - x_i 1_{\{q_i \le \gamma_0\}}$ by (A.8) we have

$$
n^{\alpha-1}\sum_{i=1}^{n}x_i 1_{\{q_i\le\gamma\}}\operatorname{sgn}(e_i) - n^{\alpha-1}\sum_{i=1}^{n}x_i 1_{\{q_i\le\gamma_0\}}\operatorname{sgn}(e_i) = o_p(1), \tag{A.9}
$$

uniformly over γ.

Next, note that Lemma 1 of Hansen (1996) shows that Assumptions (a) and (b) are sufficient for

$$
n^{-1}\sum f(0|x_i)x_i(\gamma)x_i(\gamma)' \xrightarrow{p} M(\gamma) = E(f(0|x_i)x_i x_i' 1_{\{q_i\le\gamma\}}).
$$

Without losing any generality, assuming $\gamma_0 = \min(\gamma, \gamma_0)$ and using Assumption (f) we see that the second term in (A.7) weakly converges to

$$n^{-1}c'\left(\sum f(0|x_i)\Delta x_i(\gamma)\Delta x_i(\gamma)'\right)cf(0) \Rightarrow c'\Delta M(\gamma)c, \tag{A.10}$$

uniformly over $\gamma \in [\gamma_0, \gamma_u]$, where $\Delta M(\gamma) = M(\gamma) - M(\gamma_0)$.

Now we can use Lemma A.1 to have the third term on the right-hand side of (A.6) as

$$n^{2\alpha-1}\sum |e_i - n^{-1/2}x_i'\theta - n^{-1/2}x_i(\gamma)'\delta - \Delta x_i(\gamma)'\delta_n| - |e_i - \Delta x_i(\gamma)'\delta_n| = o_p(1), \tag{A.11}$$

uniformly over γ.

By Lemma A.1ii of Bai (1995) and Assumption (f), the second term on the right-hand side of (A.6) converges in probability to zero.

Then use (A.7)–(A.11) in combination with (A.6) to have

$$n^{2\alpha-1}Q_n(\gamma) \Rightarrow -c'\Delta M(\gamma)c \equiv b(\gamma).$$

Set $\gamma > \gamma_0$, taking partial derivatives

$$\frac{d}{\partial\gamma}b(\gamma) = -c'D(\gamma)g(\gamma)c \leq 0,$$

by Assumption (e). So $b(\gamma)$ is continuous and weakly decreasing on $[\gamma_0, \gamma_u]$. Also when $\gamma \to \gamma_0$, $b(\gamma) = 0$, so γ_0 is the maximum point. Moreover,

$$\frac{d}{\partial\gamma}b(\gamma)|_{\gamma=\gamma_0} = -c'Dcg < 0,$$

by Assumption (g). So $b(\gamma)$ is a continuous function that is uniquely maximized at γ_0 on $[\gamma_0, \gamma_u]$.

The same analysis applies when γ is in $[\gamma_l, \gamma_0]$. Then we use Theorem 2.1 of Newey and McFadden (1994) to have the consistency result. ■

Now we present the rate of convergence result. Set $a_n = O(n^{1-2\alpha})$.

LEMMA A.2. *Under Assumptions (a)–(h),*

$$a_n(\hat{\gamma} - \gamma_0) = O_p(1).$$

Note that the proof of this lemma is similar to Lemma A.9 of Hansen (2000) and thus is omitted. This is written as a technical appendix. Interested readers can obtain the proof from the author on request.

For the subsequent results the following notation is useful. Set

$$\Delta_i(v) = 1_{\{q_i \leq \gamma_0 + v/a_n\}} - 1_{\{q_i \leq \gamma_0\}} = d_i(\gamma_0 + v/a_n) - d_i(\gamma_0),$$

$$\Delta x_i(v) = x_i 1_{\{q_i \leq \gamma_0 + v/a_n\}} - x_i 1_{\{q_i \leq \gamma_0\}} = x_i \Delta_i(v).$$

Let $v \in \Psi$ be any compact subset of R and $\|\phi\| \leq \overline{M}$ as in Bai (1995, p. 432) where $\phi = (\theta, \delta)$. We need the following lemma to derive the limit law for the threshold estimate.

LEMMA A.3. *Under Assumptions (a)–(h),*

$$\sum_{i=1}^{n} |e_i - n^{-1/2}x_i'\theta - n^{-1/2}x_i(\gamma_0)'\delta| - |e_i|$$

$$- \sum_{i=1}^{n} |e_i - n^{-1/2}x_i'\theta - n^{-1/2}x_i(v)'\delta - \Delta x_i(v)'\delta_n| - |e_i - \Delta x_i(v)'\delta_n|$$

$$= o_p(1),$$

uniformly on any compact set Ψ.

Note that the proof of this lemma is simple but tedious and thus is omitted. This is written as a technical appendix. Interested readers can obtain the proof from the author on request.

The limit behavior of $\hat{\gamma}$ can be obtained by analyzing the local behavior of the objective function in (A.6):

$$Q_n(v) = \left(\sum_{i=1}^{n} |e_i - n^{-1/2}x_i'\theta - n^{-1/2}x_i(\gamma_0)'\delta| - |e_i| \right)$$

$$- \left(\sum_{i=1}^{n} |e_i - n^{-1/2}x_i'\theta - n^{-1/2}x_i(\gamma_0)'\delta - \Delta x_i(v)'\delta_n| - |e_i - \Delta x_i(v)'\delta_n| \right)$$

$$- \left(\sum_{i=1}^{n} |e_i - \Delta x_i(v)'\delta_n| - |e_i| \right). \tag{A.12}$$

In (A.12) the third term is analyzed via Pollard (1991) or Bai (1995, equation (A.41)).

$$\sum_{i=1}^{n} |e_i - \Delta x_i'(v)\delta_n| - |e_i| = \delta_n' \sum \Delta x_i(v)\operatorname{sgn}(e_i)$$

$$+ \delta_n' \left(\sum f(0|x_i)\Delta x_i(v)\Delta x_i(v)' \right)\delta_n + o_p(1), \tag{A.13}$$

uniformly on any given compact set Ψ.

We use Lemma A.10 of Hansen (2000) to have

$$\frac{a_n}{n^{1-2\alpha}} \delta_n' \left(\sum f(0|x_i)\Delta x_i(v)\Delta x_i(v)' \right)\delta_n = \frac{a_n}{n} c' \left(\sum f(0|x_i)\Delta x_i(v)\Delta x_i(v)' \right) c$$

$$\xrightarrow{p} \mu|v|, \tag{A.14}$$

uniformly on any compact set Ψ, where $\mu = c'Dcg$.

Then use Lemma A.11 of Hansen (2000) to have, uniformly on any compact set Ψ,

$$\left(\frac{a_n}{n} \right)^{1/2} \sum_{i=1}^{n} x_i \operatorname{sgn}(e_i)\Delta_i(v) \Rightarrow B(v), \tag{A.15}$$

where $B(v)$ is a vector Brownian motion with covariance matrix $EB(1)B(1)' = Vg$. Using $\Delta x_i(v) = x_i \Delta_i(v)$,

$$\left(\frac{a_n}{n^{1-2\alpha}}\right)\delta_n' \sum \Delta x_i(v)\operatorname{sgn}(e_i) = \left(\frac{a_n}{n}\right)^{1/2} c' \sum_{i=1}^{n} x_i \operatorname{sgn}(e_i)\Delta_i(v)$$

$$\Rightarrow c'B(v), \tag{A.16}$$

with $c'B(v)$ a Brownian motion with variance $\lambda = c'Vcg$.

By combining (A.13)–(A.16) and using Assumption (g) we obtain

$$\frac{a_n}{n^{1-2\alpha}}\left(\sum_{i=1}^{n} |e_i - \Delta x_i(v)\delta_n'| - |e_i|\right) \Rightarrow \lambda^{1/2}W(v) + \mu|v|, \tag{A.17}$$

where $W(v)$ is a standard Brownian motion. Then using (A.17), Lemma A.3, and (A.12) we have

$$\frac{a_n}{n^{1-2\alpha}} Q_n(v) \Rightarrow -\lambda^{1/2}W(v) - \mu|v| + o_p(1) = Q(v), \tag{A.18}$$

uniformly on any compact set Ψ.

Proof of Theorem 1. Making the change of variables $v = (\lambda/4\mu^2)r$, using Lemma A.2 and (A.18), and following the proof of Theorem 1 in Hansen (2000) we have the desired result. ∎

Proof of Theorem 2. This simply follows from the proof of Theorem 2 in Hansen (2000) after making the change of the variables $v = \lambda r/(4\mu^2)$ given in (A.17) and (A.18). ∎

LINKED CITATIONS

You have printed the following article:

A Note on Least Absolute Deviation Estimation of a Threshold Model
Mehmet Caner
Econometric Theory, Vol. 18, No. 3. (Jun., 2002), pp. 800-814.
Stable URL:
http://links.jstor.org/sici?sici=0266-4666%28200206%2918%3A3%3C800%3AANOLAD%3E2.0.CO%3B2-R

This article references the following linked citations. If you are trying to access articles from an off-campus location, you may be required to first logon via your library web site to access JSTOR. Please visit your library's website or contact a librarian to learn about options for remote access to JSTOR.

Notes

[1] **Least Absolute Deviation Estimation of a Shift**
Jushan Bai
Econometric Theory, Vol. 11, No. 3. (Aug., 1995), pp. 403-436.
Stable URL:
http://links.jstor.org/sici?sici=0266-4666%28199508%2911%3A3%3C403%3ALADEOA%3E2.0.CO%3B2-3

[1] **Some Invariance Principles and Central Limit Theorems for Dependent Heterogeneous Processes**
Jeffrey M. Wooldridge; Halbert White
Econometric Theory, Vol. 4, No. 2. (Aug., 1988), pp. 210-230.
Stable URL:
http://links.jstor.org/sici?sici=0266-4666%28198808%294%3A2%3C210%3ASIPACL%3E2.0.CO%3B2-H

[1] **Sample Splitting and Threshold Estimation**
Bruce E. Hansen
Econometrica, Vol. 68, No. 3. (May, 2000), pp. 575-603.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28200005%2968%3A3%3C575%3ASSATE%3E2.0.CO%3B2-C

References

NOTE: *The reference numbering from the original has been maintained in this citation list.*

LINKED CITATIONS

Least Absolute Deviation Estimation of a Shift
Jushan Bai
Econometric Theory, Vol. 11, No. 3. (Aug., 1995), pp. 403-436.
Stable URL:
http://links.jstor.org/sici?sici=0266-4666%28199508%2911%3A3%3C403%3ALADEOA%3E2.0.CO%3B2-3

Changes in the U.S. Wage Structure 1963-1987: Application of Quantile Regression
Moshe Buchinsky
Econometrica, Vol. 62, No. 2. (Mar., 1994), pp. 405-458.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28199403%2962%3A2%3C405%3ACITUWS%3E2.0.CO%3B2-H

Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model
K. S. Chan
The Annals of Statistics, Vol. 21, No. 1. (Mar., 1993), pp. 520-533.
Stable URL:
http://links.jstor.org/sici?sici=0090-5364%28199303%2921%3A1%3C520%3ACALDOT%3E2.0.CO%3B2-Q

Sample Splitting and Threshold Estimation
Bruce E. Hansen
Econometrica, Vol. 68, No. 3. (May, 2000), pp. 575-603.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28200005%2968%3A3%3C575%3ASSATE%3E2.0.CO%3B2-C

Tests of Linear Hypotheses and l"1 Estimation
Roger Koenker; Gilbert Bassett
Econometrica, Vol. 50, No. 6. (Nov., 1982), pp. 1577-1584.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28198211%2950%3A6%3C1577%3ATOLHAL%3E2.0.CO%3B2-4

Invariance Principles for Mixing Sequences of Random Variables
Magda Peligrad
The Annals of Probability, Vol. 10, No. 4. (Nov., 1982), pp. 968-981.
Stable URL:
http://links.jstor.org/sici?sici=0091-1798%28198211%2910%3A4%3C968%3AIPFMSO%3E2.0.CO%3B2-H

NOTE: *The reference numbering from the original has been maintained in this citation list.*

LINKED CITATIONS

Testing and Estimating Change-Points in Time Series
Dominique Picard
Advances in Applied Probability, Vol. 17, No. 4. (Dec., 1985), pp. 841-867.
Stable URL:
http://links.jstor.org/sici?sici=0001-8678%28198512%2917%3A4%3C841%3ATAECIT%3E2.0.CO%3B2-U

Asymptotics for Least Absolute Deviation Regression Estimators
David Pollard
Econometric Theory, Vol. 7, No. 2. (Jun., 1991), pp. 186-199.
Stable URL:
http://links.jstor.org/sici?sici=0266-4666%28199106%297%3A2%3C186%3AAFLADR%3E2.0.CO%3B2-6

Estimating Nonlinear Dynamic Models Using Least Absolute Error Estimation
Andrew A. Weiss
Econometric Theory, Vol. 7, No. 1. (Mar., 1991), pp. 46-68.
Stable URL:
http://links.jstor.org/sici?sici=0266-4666%28199103%297%3A1%3C46%3AENDMUL%3E2.0.CO%3B2-E

Some Invariance Principles and Central Limit Theorems for Dependent Heterogeneous Processes
Jeffrey M. Wooldridge; Halbert White
Econometric Theory, Vol. 4, No. 2. (Aug., 1988), pp. 210-230.
Stable URL:
http://links.jstor.org/sici?sici=0266-4666%28198808%294%3A2%3C210%3ASIPACL%3E2.0.CO%3B2-H

NOTE: *The reference numbering from the original has been maintained in this citation list.*