NESTED SUB-SAMPLE SEARCH ALGORITHM FOR ESTIMATION OF THRESHOLD STOCHASTIC REGRESSION MODELS

by Dong Li (Tsinghua University) and Howell Tong
(London School of Economics & Political Science)

Threshold stochastic regression models have attracted much attention and been widely used to model nonlinear phenomena in diverse areas, such as economics, finance, ecology and others. Their success is partly due to their simple fitting and often clear interpretation. Threshold models are typically characterized by piecewise linearization via partitioning a complex system into regimes by some threshold (or covariate) variable, thereby providing a relatively easy-to-handle approximation of a complex system. When the model within each regime is a linear regression, we have the well-known two-phase regression of Quandt (1958). On the other hand, when the model within each regime is a linear autoregression, we have the well-known threshold autoregressive (TAR) model of Tong (1978), including the self-exciting threshold autoregressive model and its smooth cousin, the smooth threshold (or transition) autoregressive model, as special cases. See also Tong and Lim (1980), Chan and Tong (1986), Tong (1990), and the references therein. Recently Hansen (2011) has provided a fairly comprehensive review of the impacts of TAR models on econometrics and economics by reference to 75 influential papers published in the literature. Chen et al. (2011) has provided a similar review of the impacts on finance. More recently, Chan et al. (2014) has adopted the LASSO method to estimate TAR models with multiple thresholds, with promising results. A concise overview of the history and prospects of threshold models is given by Tong (2011).

As far as theoretical results are concerned, much progress has been in two-phase regression since Quandt (1958) and TAR models since Tong (1978). For the former, see, e.g., Bacon and Watts (1971), Goldfeld and Quandt (1972), Maddala (1977), Quandt (1984) and others. For the latter, see, e.g., Chan (1993), who first showed that the least squares estimator (LSE) of the threshold parameter is super-consistent and obtained its limiting distribution theoretically; Hansen (1997, 2000), who presented an alternative approximation to the limiting distribution of the estimated threshold when the threshold effect diminishes as the sample size increases; Gonzalo and Pitarakis (2002), who developed a sequential estimation approach that makes the estimation of multiple threshold models computationally feasible and formally discussed the large sample properties; Li and Ling (2012), who established the asymptotic theory of LSE in multiple threshold models and proposed a resampling method for implementing the limiting distribution of the estimated threshold directly when the threshold effect is fixed. Other significant results related to threshold models include Tsay (1989, 1998), Hansen (1996), Caner and Hansen (2001), Gonzalo and Wolf (2005), Seo and Linton (2007), and Yu (2012) among others.

Despite the theoretical progress in threshold models, computational issues are somewhat lacking behind, which hinders wider practical applications. A key issue is computational cost.

A commonly used approach to fit a threshold model is the (conditional) least squares method. When the threshold is known, the threshold model is piecewise linear in the remaining parameters and thus linear estimation techniques can be applied. However, when the threshold is unknown, the ordinary least squares method for linear regression cannot be

applied immediately since the threshold parameter lies in an indicator function. This issue has been commonly tackled by using the single grid search (SGS) algorithm over a feasible threshold space; see Tong and Lim (1980), Chan (1993), Hansen (1997, 2000), Gonzalo and Pitarakis (2002), Li and Ling (2012), Yu (2012), and others. The SGS algorithm requires least squares operations of order O(n) for single threshold models, where n is the sample size. If n is small, the SGS algorithm can be effectively used to search for the estimate of the threshold over a set of threshold candidates by enumeration. However, when n is large, this algorithm can be rather time-consuming.
The situation is worse when we wish to fit threshold models to a panel of observations. Gonzalo and Wolf (2005) considered subsampling inference of threshold models and massive computations are needed in the choice of the block size. Similarly, massive computations are also needed in bootstrap
estimation of single threshold models in Seijo and Sen (2011). In practice, conventional numerical approach for threshold modelling incurs inevitably high cost for large sample. For example, about np^3 least squares operations are needed when fitting a threshold model with p covariates to data with sample size n. To illustrate, if n is 1000 and p is 10, then we need about one million least squares operations. Thus, it is crucially important to find ways to reduce the computational cost when fitting a threshold model for large n.

In the time series literature, Tong (1983, Appendix A10) proposed and later Tsay (1989) re-discovered the SGS approach based on the rearranged technique, which essentially turns the threshold estimation into a change-point problem of the associated order statistics obtained from the observations. See also Ertel and Fowlkes (1976). This method is now available by calling the function \texttt{tar} in the package \texttt{TSA} in \textsf{R}; see Chan and Ripley (2012). For threshold regression models,
the SGS algorithm is available by calling the program in \textsf{R} developed by Hansen (2000) on the website:\par\qquadhttp://www.ssc.wisc.edu/\~{}bhansen/progs/ecnmt$_{-}$00.html.\\Wu and Chang (2002) proposed a genetic algorithm for TAR models. However, this algorithm has many limitations, as recognised by the above authors, so it is not widely used in practice. Coakley, Fuertes and P\'{e}rez (2003) presented an algorithm based on the QR decomposition of matrices for a particular class of TAR models (called the band-type TAR model). For general threshold models, it is fair to say that the SGS algorithm remains to-date the most commonly adopted technique in practice due to its simplicity and reliability, although it is time-consuming for large n.

In this paper, we propose a novel algorithm, namely the nested sub-sample search algorithm, or the NeSS algorithm for brevity, to produce a much faster search that is reliable in the context of threshold estimation. Compared with existing algorithms, the NeSS algorithm reduces the computational cost drastically from O(n) down to O(log n) least squares operations for large sample size. The idea is simple. We shrink the nested feasible set step by step and finally maximize over a small feasible set by enumeration
so that it is expected to save computational costs. The performance of our method is evaluated via Monte Carlo simulation studies in finite samples.

The method can be generalized to cover maximum likelihood estimation and other areas.

References:

Bacon, B. W. and Watts, D. G. (1971). Estimating the transition between two intersecting straight lines. \textit{Biometrika} \textbf{58}, 525--534.

Caner, M. (2002). A note on least absolute deviation estimation of a threshold model. \textit{Econometric Theory} \textbf{18}, 800--814.

Caner, M. and Hansen, B. E. (2001). Threshold autoregression with a unit root. \textit{Econometrica} \textbf{69},
1555--1596.

Chan, K. S. (1990). Testing for threshold autoregression. \textit{Ann. Statist.} \textbf{18},
1886--1894.

Chan, K. S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model.
\textit{Ann. Statist.} \textbf{21}, 520--533.

Chan, K.S., Li, D., Ling, S. and Tong, H. (2014). On conditionally heteroscedastic AR models with thresholds.
\textit{Statist. Sinica} \textbf{24}, 625--652.

Chan, K.S. and Ripley, B. (2012). TSA: Time Series Analysis. R package version 1.01, URL http://cran.r-project.org/web/packages/TSA/.

Chan, K.S. and Tong, H. (1986). On estimating thresholds in autoregressive models. {\it J. Time Series Anal.} {\bf 7}, 179--190.

Chan, N.H., Yau, C.Y. and Zhang, R.M. (2015). LASSO estimation of threshold autoregressive models. {\it J. Econometrics} (to appear).

Chen, C.W.S., So, M.K.P. and Liu, F.C. (2011). A review of threshold time series models in finance. {\it Stat. and Its Interface} {\bf 4}, 167--182.

Coakley, J., Fuertes, A.-M. and P\'{e}rez, M.-T. (2003).
Numerical issues in threshold autoregressive modeling of time series.
\textit{J. Econom. Dynam. Control} \textbf{27}, 2219--2242.

Ertel, J. E. and Fowlkes, E. B. (1976). Some algorithms for linear spline and piecewise multiple linear regression. \textit{J. Amer. Statist. Assoc.} \textbf{71}, 640--648.

Goldfeld, S.M. and Quandt, R.E. (1972). \textit{Nonlinear Methods in Econometrics}, North-Holland, Amsterdam.

Gonzalo, J. and Pitarakis, J. Y. (2002). Estimation and model selection based inference in single and multiple threshold
models. \textit{J. Econometrics} \textbf{110}, 319--352.

Gonzalo, J. and Wolf, M. (2005). Subsampling inference in threshold autoregressive models. \textit{J. Econometrics} \textbf{127}, 201--224.

Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. \textit{Econometrica} \textbf{64}, 413--430.

Hansen, B. E. (1997). Inference in TAR models. \textit{Stud. Nonlinear Dyn. Econom.} \textbf{2}, 1--14.

Hansen, B. E. (2000). Sample splitting and threshold estimation. \textit{Econometrica} \textbf{68}, 575--603.

\item
Hansen, B. E. (2011). Threshold autoregression in economics. \textit{Stat. and Its Interface} \textbf{4}, 123--127.

\item
Li, D. and Ling, S. (2012). On the least squares estimation of multiple-regime threshold autoregressive models.
\textit{J. Econometrics} \textbf{167}, 240--253.

Liu, J., Wu, S. and Zidek, J. V. (1997). On segmented multivariate regression. \textit{Statist. Sinica} \textbf{7}, 497--525.

Maddala, G. S. (1977). \textit{Econometrics}, McGraw-Hill, New York.

Quandt, R. E. (1958). The estimation of the parameters of a linear regression system obeying two separate regimes. \textit{J. Amer. Stat. Ass.} \textbf{53}, 873--880.

Quandt, R. E. (1984). Computational problems and methods, \textit{Handbook of Econometrics} Vol. 1, ed. Z. Griliches and
M.D. Intriligator, North-Holland, Amsterdam, 701--764.

Samia, N.I. and Chan, K.S.(2011). Maximum likelihood estimation of a generalized threshold stochastic regression model. \textit{Biometrika} \textbf{98}, 433--448.

Seijo, E. and Sen, B. (2011). Change-point in stochastic design regression and the bootstrap. \textit{Ann. Statist.} \textbf{39}, 1580--1607.

Seo, M. H. and Linton, O. (2007). A smoothed least squares estimator for threshold regression models. \textit{J. Econometrics} \textbf{141}, 704--735.

Tong, H. (1978). On a Threshold Model. In: \textit{Pattern Recognition and Signal Processing}, ed. C.H. Chen, Sijthoff and Noordhoff, Amsterdam, 575--586.

Tong, H. (1983). \textit{Threshold models in nonlinear time series analysis.} Lecture Notes in Statistics, \textbf{21}. Springer-Verlag, New York.

\item
Tong, H. (1990). \textit{Nonlinear time series. A dynamical system approach}. Oxford University Press, New York.

Tong, H. (2011). Threshold models in time series analysis --- 30 years on. \textit{Stat. Interface} \textbf{4}, 107--118.

Tong, H. and Lim, K. S. (1980). Threshold autoregression, limit cycles and cyclical data (with discussion). \textit{J. R. Stat. Soc.} B \textbf{42}, 245--292.

Tsay, R. S. (1989). Testing and modeling threshold autoregressive processes. \textit{J. Amer. Statist. Assoc.} \textbf{84}, 231--240.

Tsay, R. S. (1998). Testing and modeling multivariate threshold models. \textit{J. Amer. Statist. Assoc.} \textbf{93}, 1188--1202.

Wu, B. and Chang, C.-L. (2002). Using genetic algorithms to parameters (d,r) estimation for threshold autoregressive models. \textit{Comput. Statist. Data Anal.} \textbf{38}, 315--330.

Yu, P. (2012). Likelihood estimation and inference in threshold regression. \textit{J. Econometrics} \textbf{167}, 274--294.