Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Journal of Economic Dynamics & Control 27 (2003) 2219–2242
JOURNAL OF Economic Dynamics & Control
www.elsevier.com/locate/econbase

Numerical issues in threshold autoregressive modeling of time series

Jerry Coakley[a,*], Ana-María Fuertes[b], María-Teresa Pérez[c]

[a] *Department of Accounting, Finance and Management, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK*
[b] *Faculty of Finance, Cass Business School, UK*
[c] *Departmento de Matemática Aplicada, ETSII, University of Valladolid, Spain*

Abstract

This paper analyses the contribution of various numerical approaches to making the estimation of threshold autoregressive time series more efficient. It relies on the computational advantages of QR factorizations and proposes Givens transformations to update these factors for sequential LS problems. By showing that the residual sum of squares is a continuous rational function over threshold intervals it develops a new fitting method based on rational interpolation and the standard necessary optimality condition. Taking as benchmark a simple grid search, the paper illustrates via Monte Carlo simulations the efficiency gains of the proposed tools.

JEL classification: C63; C51; C61

Keywords: Band-TAR; QR factorization; Givens rotations; Rational interpolation

1. Introduction

Nonlinear models have been widely applied in recent years to capture asymmetries, limit cycles and jump phenomena in the behavior of economic and financial time series. Among these models, the threshold autoregression (TAR) introduced by Tong and Lim (1980) has received particular attention. This is perhaps the simplest generalization of an AR model which allows for different regimes for the series depending on its past values. TAR models have been successfully applied to model nonlinearities in financial variables by permitting an inner regime of sluggish adjustment for small

* Corresponding author. Tel.: +44-0207-631-6418; fax: +44-0207-631-641.
E-mail address: jcoakley@essex.ac.uk (J. Coakley).

0165-1889/02/$ - see front matter
PII: S0165-1889(02)00123-9

disequilibria— or small deviations from some long run equilibrium path or attractor— and mean reversion in an outer regime comprising large deviations. This nonlinear behavior has been rationalized on the basis of transaction costs or a low signal-to-noise ratio hindering profitable arbitrage opportunities for small disequilibria.

TAR models have also been used successfully to explore asymmetries in macroeconomic variables over the course of the business cycle.[1] There is the question of whether the apparent persistence in a economic time series such as GNP or unemployment provides evidence of asymmetries that standard Gaussian linear (fixed) parameter models cannot accommodate. Thus one aspect of this literature relates to proposals for new unit root tests in a TAR framework—which can be thought of as extensions of existing linear tests—where the alternative hypothesis is stationarity with possible asymmetric adjustment (Enders and Granger, 1998; Berben and van Dijk, 1999; Coakley and Fuertes, 2001a, b).

A practical problem in using TARs is that standard maximum likelihood (ML) estimation algorithms cannot be applied since the log-likelihood function is not continuously differentiable with respect to the threshold parameter. This problem has been commonly tackled by implementing a grid search (GS) over a feasible region of the threshold space. For a given threshold value, the TAR model is piecewise linear in the remaining parameters and thus linear estimation techniques can be applied. The threshold value that maximizes the log-likelihood function over the grid is the ML estimate. Since in principle any point in the continuous threshold space could maximize the log-likelihood, a full or detailed GS with a small step size is preferable to a GS restricted to the order statistics of the threshold variable.

While the latter may deliver inaccurate model parameter estimates for small sample sizes, a practical problem with the detailed GS is that it may prove computationally expensive and especially so for widely dispersed data. Though computation costs may not be an issue in ad hoc fitting of TAR models to single time series, these become relevant in inference applications of TARs using simulation techniques. For instance, exploring the small sample properties of TAR-based tests by Monte Carlo or bootstrap simulation methods and/or estimating response surfaces with a sensible number of replications can become intractable if no attention is paid to estimation time. These problems are aggravated if the model is nonlinear in more than one parameter which effectively implies a high-dimensional grid search.

The purpose of this paper is twofold. First, it explores systematically the value-added of QR factorizations and Givens transformations in TAR fitting. In this sense it seeks to fill an existing gap in investigating numerical aspects of TAR modeling and to provide practical recommendations. Second, by showing that the residual sum of squares of a certain class of TARs is a continuous rational function over threshold intervals, it proposes a novel fitting approach. Its main advantage is allowing for a continuous feasible range for the threshold parameter. Our approach can be considered as

[1] For instance, TARs have been applied to explore the term structure of interest rates by Enders and Granger (1998), the Nelson–Plosser data set by Rothman (1999), unemployment behavior by Caner and Hansen (1998) and Coakley et al. (2002), and to model US output by Pesaran and Potter (1997) and Kapetanios (1999a).

equivalent to a grid search in the limit as the step size becomes increasingly small, while still remaining computationally tractable in contrast to the latter.

The remainder of the paper is organized as follows. In Section 2 we outline the TAR framework and estimation issues. In Section 3 we discuss some numerical tools and combine them in a novel fitting approach which is summarized step-by-step. The proofs of Proposition 1 and Theorems 2 and 3 are deferred to an appendix. In Section 4 we evaluate via Monte Carlo simulation the efficiency gains of these tools and conclude in Section 5.

2. The model

2.1. Band-TAR dynamics

An m-regime TAR model can be written as

$$z_t = \sum_{j=1}^{m} (\phi_0^j + \phi_1^j z_{t-1} + \cdots + \phi_{p_j}^j z_{t-p_j}) I_t(\theta^{j-1} \leqslant v_{t-d} < \theta^j) + \varepsilon_t, \tag{1}$$

where $\varepsilon_t \sim nid(0, \sigma^2)$, $I_t(\cdot)$ is the indicator function, $-\infty = \theta^0 < \theta^1 < \cdots < \theta^m = \infty$ are threshold parameters, p_j and d are the positive integer-valued autoregressive (AR) lag order and threshold delay, respectively. This is a nonlinear model in time but piecewise linear in the threshold space Θ. More specifically, (1) is a discontinuous (in conditional-mean) TAR which partitions the one-dimensional Euclidean space into m linear AR regimes. The specific linear mechanism at any given point in time depends on the values taken by the threshold or switching variable v_{t-d}. The resulting model for $v_{t-d} = z_{t-d}$ is sometimes called a self-exciting TAR (SETAR) to distinguish it from those models where v_{t-d} is exogenous.[2]

Consider the following first-difference reparameterization of (1) for $m = 3$:

$$\begin{aligned} \Delta z_t = {} & A(t,\theta)^- I_t(v_{t-d} < -\theta) + B(t) I_t(-\theta \leqslant v_{t-d} \leqslant \theta) \\ & + A(t,\theta)^+ I_t(v_{t-d} > \theta) + \varepsilon_t \end{aligned} \tag{2}$$

with

$$\begin{aligned} A(t,\theta)^- &= \alpha_1(z_{t-1} + \theta) + \alpha_2(z_{t-2} + \theta) + \cdots + \alpha_p(z_{t-p} + \theta), \\ A(t,\theta)^+ &= \alpha_1(z_{t-1} - \theta) + \alpha_2(z_{t-2} - \theta) + \cdots + \alpha_p(z_{t-p} - \theta), \\ B(t) &= \beta_0 + \beta_1 z_{t-1} + \beta_2 z_{t-2} + \cdots + \beta_q z_{t-q}, \end{aligned}$$

where $\theta > 0$ is an identifying restriction. This is a generalization of the Band-SETAR model introduced by Balke and Fomby (1997) where $p = q = 1$ and $\beta_0 = \beta_1 = 0$. The latter implies random walk behavior in the inner band. An important feature of

[2] (SE)TAR models are special cases of Priestley's (1998) general nonlinear state-dependent models. In the related smooth transition AR (STAR) class of models $I_t(\cdot)$ is replaced by a (continuous) smooth function. See Tong (1983) and Granger and Teräsvirta (1993).

Band-TAR processes is that their stability properties depend on the outer band dynamics only. More specifically, even when the inner band has unit root/explosive behavior, if the roots of the outer band characteristic equation $L^p-(\alpha_1+1)L^{p-1}-\alpha_2 L^{p-2}-\cdots-\alpha_p=0$ lie within the complex unit circle then, whenever $|v_{t-d}|>\theta$, z_t converges to the edges of the band $[-\theta,\theta]$ which act as attractors, and is stationary overall. This Band-TAR scheme has been extensively applied in the recent nonlinear literature to analyse the behavior of (demeaned) financial and economic variables which are expected to exhibit symmetric change-point dynamics around a long run equilibrium path. [3]

A straightforward extension of (2) is an asymmetric Band-TAR with adjustment parameters $\alpha_j^{\mathrm{u}}, j=1,\ldots,p_{\mathrm{u}}$ and $\alpha_j^{\mathrm{l}}, j=1,\ldots,p_{\mathrm{l}}$ for the upper ($v_{t-d}\geqslant\theta^{\mathrm{u}}$) and lower ($v_{t-d}\leqslant-\theta^{\mathrm{l}}$) outer regimes, respectively. Another important related specification is the following continuous (C-)TAR:

$$\Delta z_t=\alpha_1^{\mathrm{u}}(z_{t-1}-\theta)I_t+\alpha_1^{\mathrm{l}}(z_{t-1}-\theta)(1-I_t)+\sum_{j=1}^{p}\alpha_j\Delta z_{t-j}+\varepsilon_t,$$

$$I_t=\begin{cases}1 & \text{if } v_{t-1}\geqslant 0,\\ 0 & \text{otherwise,}\end{cases} \tag{3}$$

where $v_{t-1}=z_{t-1}-\theta$. This model characterizes a process with possibly asymmetric adjustment ($\alpha_1^{\mathrm{u}}\neq\alpha_1^{\mathrm{l}}$) towards the attractor θ. Note that a common feature of (2) and (3) is that θ appears explicitly in the conditional mean of Δz_t. This C-TAR class of models—which has generated an extensive literature [4]—was formally introduced by Chan and Tsay (1998) and proposed by Enders and Granger (1998) as a generalization of the linear augmented Dickey–Fuller regression to test for unit root dynamics.

2.2. The estimation problem

Let $\{z_t\}_{t=1}^N$ and $\{v_t\}_{t=1}^N$ be the time series available for estimation of (2). Ordinary LS or, equivalently, conditional ML under Gaussian innovations, lead to the minimization of the following residual sum of squares (RSS) function:

$$RSS(\phi)=\sum_{t}^{n}(\Delta z_t-A(t,\theta)^-)^2 I_t(v_{t-d}<-\theta)+\sum_{t}^{n}(\Delta z_t-B(t))^2 I_t(|v_{t-d}|\leqslant\theta)$$
$$+\sum_{t}^{n}(\Delta z_t-A(t,\theta)^+)^2 I_t(v_{t-d}>\theta)$$

[3] Note that (2) assumes outer regimes with the same dynamics $\alpha(L)$ and symmetric thresholds with respect to zero. For instance, Coakley and Fuertes (2001c) employ this symmetric model to explore the issue of market segmentation in Europe while Obstfeld and Taylor (1997) fit a more restrictive version with $\beta_0=\beta_1=\cdots=\beta_q=0$ to analyse the PPP hypothesis.

[4] For example, Berben and van Dijk (1999) develop a unit root test based on (3) and Coakley and Fuertes (2000) extend (3) to develop tests for AR mean-reversion against sign and amplitude asymmetric adjustment.

with respect to $\phi=(\theta,\alpha',\beta',d,p,q)'$, where $\alpha=(\alpha_1,\ldots,\alpha_p)'$ and $\beta=(\beta_0,\beta_1,\ldots,\beta_q)'$ are the outer and inner AR parameters, respectively, and $n=N-\max(d,p,q)$ the effective sample size.

Let us assume that the lags (d,p,q) are known a priori. Our goal is to estimate the remaining parameters θ, α and β. Since the above objective function is discontinuous in θ, standard gradient-based algorithms cannot be applied. If the threshold space Θ is small, a simple grid search (GS) can be effectively used to find the value $\hat{\theta}\in\Theta$ that minimizes the RSS (or some LS-based criterion) or maximizes the log-likelihood function over a countable set of threshold candidates. The remaining (linear) parameters can be easily estimated by LS conditional on $\hat{\theta}$. Generalizing the latter to unknown d, p and q, Chan (1993) shows that under certain regularity conditions for z_t, including stationarity and geometric ergodicity, and iid but not necessarily Gaussian innovations, this sequential LS approach yields estimators $\hat{\alpha},\hat{\beta},\hat{d},\hat{p}$ and $\hat{q}$ which are strongly consistent at the usual $\sqrt{N}$ rate and asymptotically normal, and an estimator $\hat{\theta}$ which is (super) N-consistent and has a nonstandard distribution. Chan and Tsay (1998) extend this asymptotic result to show that $(\hat{\theta},\hat{\alpha}',\hat{\beta}',\hat{d},\hat{p},\hat{q})$ are strongly $\sqrt{N}$-consistent and asymptotically normal for C-TAR models such as (3).

The threshold space is the continuous region $\Theta\subseteq\mathbb{R}^+$. However, in practice the GS is restricted to a feasible (discrete) range in Θ by fixing a number of threshold candidates which are usually the sample percentiles (or order statistics) of v_{t-d}, that is, $\xi_{(t)}=\{v_{(1)}\leqslant v_{(2)}\leqslant\cdots\leqslant v_{(n)}\}\subset\Theta$. However, since in principle any point in Θ could maximize the log-likelihood, a full or detailed GS using $\xi_\lambda=\bigcup_i\{\theta_i^j\colon\ \theta_i<\theta_i^j<\theta_{i+1},\theta_i^{j+1}=\theta_i^j+\lambda,j=1,2,\ldots\}\cup\xi_{(t)}$ where $\theta_i=v_{(i)},\ i=1,\ldots,n$, is preferable to a GS restricted to $\xi_{(t)}$. While a potential pitfall of using $\xi_{(t)}$ is that it may yield imprecise estimates for small N and widely dispersed data, which imply a large range $v_{(n)}-v_{(1)}$, a practical problem with ξ_λ is that it may prove computationally expensive for small step size λ. In general, the choice of λ depends critically on a trade-off between computation time and threshold bias. This calls for an estimation method capable of handling a continuous threshold range while keeping costs within tractable limits.

For some TARs, threshold values between consecutive order statistics, $v_{(i)}<\theta<v_{(i+1)}$, change neither the partition of the observed data into regimes nor the associated (piecewise linear) LS problem. For these TARs, a sensible range for θ in estimation is $\xi_{(t)}$. However, the latter does not apply to models such as (2) or (3) where the threshold appears explicitly in the conditional mean. This effectively means that varying θ in the range between consecutive order statistics changes the regressors of the upper and lower equations, $\{z_{t-j}-\theta\}$ and $\{z_{t-j}+\theta\}$, $j=1,\ldots,p$, respectively, and hence the LS problem even though the same data partition holds. For these TAR types, using a grid search either with ξ_λ or $\xi_{(t)}$ may yield a suboptimal threshold estimate whose lack of precision will contaminate the distribution of the remaining parameters. These practical issues call for a fitting approach which allows for a continuous feasible range while keeping computation costs within tractable limits. The proposed tools are in this spirit.

3. An efficient estimation approach

3.1. Arranged autoregression and threshold intervals

This section first discusses the ordered-autoregression concept introduced by Tsay (1989) which facilitates efficient estimation of TARs. It then explores how allowing for different threshold values in a given (continuous) threshold interval alters the initial ordered-form setup in the context of model (2).

Without loss of generality, set $v_{t-d} = z_{t-d}$ in (2) and assume d is known a priori. Let L be a plausible maximum lag order, $1 \leqslant (p,q) \leqslant L$. For $p = q = L$ the observed data $\{z_t\}_{t=1}^{N}$ can be represented in a regression framework as $y = f(X) + \varepsilon$, with $X = (x_1, x_2, \ldots, x_L)$ where ε is a disturbance n-vector and y and x_j are data n-vectors of observed data for Δz_t and z_{t-j}, respectively, and $n = N - \max(d, L)$. Each row in this matrix-form setup represents an available *case* for the Band-TAR estimation problem. Let us transform the latter into a change-point problem by rearranging its cases according to the threshold variable v_{t-d}. To facilitate this we augment X with the available observations for the latter as follows:

$$(X|v) = (z_{t-1}, z_{t-2}, \ldots, z_{t-L}|v_{t-d}) = \begin{pmatrix} x_{11} & x_{12} & \ldots & x_{1L} & v_1 \\ x_{21} & x_{22} & \ldots & x_{2L} & v_2 \\ \vdots & \vdots & & \vdots & \vdots \\ x_{n1} & x_{n2} & \ldots & x_{nL} & v_n \end{pmatrix}. \tag{4}$$

Next the rows of y, ε and $(X|v)$ are rearranged following the ordering of v_{t-d}, the last column of $(X|v)$. This yields $y^v = f(X^v) + \varepsilon^v$ with

$$(X^v|v) = (z^v_{t-1}, z^v_{t-2}, \ldots, z^v_{t-L}|v_{(i)}) = \begin{pmatrix} x^v_{11} & x^v_{12} & \ldots & x^v_{1L} & v_{(1)} \\ x^v_{21} & x^v_{22} & \ldots & x^v_{2L} & v_{(2)} \\ \vdots & \vdots & & \vdots & \vdots \\ x^v_{n1} & x^v_{n2} & \ldots & x^v_{nL} & v_{(n)} \end{pmatrix}, \tag{5}$$

where $v_{(i)}$ denotes the ith smallest observation of v_{t-d}. A crucial property of this reformulation (denoted by the superscript v) is that by permuting (rows) cases of the initial matrix-form setup, it preserves the dynamics of z_t.

Let $\theta = \theta_k$ $(\theta_k > 0)$ be a plausible threshold value such that two indexes, k_1 and k_2 $(k_1 < k_2)$, are associated with it satisfying $v_{(i)} < -\theta_k$ for $i = 1, 2 \ldots, k_1$, $v_{(i)} > \theta_k$ for $i = k_2, \ldots, n$, and $-\theta_k \leqslant v_{(i)} \leqslant \theta_k$ for $i = k_1 + 1, \ldots, k_2 - 1$. Using the above ordered-form notation the $s = k_2 - k_1 - 1$ cases classified into the *inner* regime of the Band-TAR

model can be written as $\Delta z_s = Z_s^{\beta}\beta + \varepsilon_s$ where

$$Z_s^{\beta} = \begin{pmatrix} 1 & x^v_{k_1+1,1} & x^v_{k_1+1,2} & \cdots & x^v_{k_1+1,L} \\ 1 & x^v_{k_1+2,1} & x^v_{k_1+2,2} & \cdots & x^v_{k_1+2,L} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x^v_{k_2-1,1} & x^v_{k_2-1,2} & \cdots & x^v_{k_2-1,L} \end{pmatrix}, \tag{6}$$

$\Delta z_s = (y^v_{k_1+1}, y^v_{k_1+2}, \ldots, y^v_{k_2-1})'$ and $\varepsilon_s = (\varepsilon^v_{k_1+1}, \varepsilon^v_{k_1+2}, \ldots, \varepsilon^v_{k_2-1})'$. Likewise the $r = n - (k_2 - k_1 - 1)$ *outer* regime cases can be written as $\Delta z_r = Z_r^{\alpha}(\theta_k)\alpha + \varepsilon_r$ where

$$Z_r^{\alpha}(\theta_k) = \begin{pmatrix} x^v_{11} + \theta_k & x^v_{12} + \theta_k & \cdots & x^v_{1L} + \theta_k \\ \vdots & \vdots & & \vdots \\ x^v_{k_1 1} + \theta_k & x^v_{k_1 2} + \theta_k & \cdots & x^v_{k_1 L} + \theta_k \\ x^v_{k_2 1} - \theta_k & x^v_{k_2 2} - \theta_k & \cdots & x^v_{k_2 L} - \theta_k \\ \vdots & \vdots & & \vdots \\ x^v_{n1} - \theta_k & x^v_{n2} - \theta_k & \cdots & x^v_{nL} - \theta_k \end{pmatrix}, \tag{7}$$

$\Delta z_r = (y^v_1, \ldots, y^v_{k_1}, y^v_{k_2}, \ldots, y^v_n)'$ and $\varepsilon_r = (\varepsilon^v_1, \ldots, \varepsilon^v_{k_1}, \varepsilon^v_{k_2}, \ldots, \varepsilon^v_n)'$. Note that the upper $k_1 \times L$ and lower $(n - k_2 + 1) \times L$ partition matrices of $Z_r^{\alpha}(\theta_k)$ correspond to the $A(t, \theta_k)^-$ and $A(t, \theta_k)^+$ outer AR schemes of (2), respectively.

The order statistics of the moduli of v_{t-d}, which are denoted by $\theta_1 \leqslant \theta_2 \leqslant \cdots \leqslant \theta_n$, are used to define the continuous threshold space Θ. To guarantee that each regime contains enough observations (cases) for the submodels to be estimable an asymptotic-theory based rule defines Θ such that for some $\kappa \geqslant 0$, and any θ, $\lim_{n\to\infty} r(n, \theta)/n \geqslant \kappa$, and $\lim_{n\to\infty} s(n, \theta)/n \geqslant \kappa$ where r and s are the outer and inner regime cases, respectively. A usual choice is $\kappa = 0.15$. After filtering out possible repeated θ_i values, the threshold space is eventually defined as $\Theta = \{\bigcup_i [\theta_i, \theta_{i+1})\} \subseteq \mathbb{R}^+$ where $[\theta_i, \theta_{i+1})$, $i = \tau_0, \tau_0 + 1, \ldots, \tau_1 - 1$, is a countable number of continuous nonoverlapping intervals (threshold intervals hereafter) and $\kappa n \leqslant \tau_0, (1 - \kappa)n \geqslant \tau_1$. Note that $\tau_0 = \kappa n$ and $\tau_1 = (1 - \kappa)n$ only if there are no repeated order statistics below and above the κ- and $(1 - \kappa)$-quantiles of v_{t-d}, respectively.

Without loss of generality, let us start the iterations from the extreme right-hand interval $[\theta_{\tau_1 - 1}, \theta_{\tau_1})$ in Θ and allow $\theta = \theta_{\tau_1 - 1}$ as the first threshold candidate. The latter defines the inner and outer regressor matrices, Z_s^{β} and $Z_r^{\alpha}(\theta)$, respectively, where $s = k_2 - k_1 - 1$, $r = k_1 + (n - k_2 + 1)$, and k_1 and $n - k_2 + 1$ are the number of cases from the top and bottom of X^v, respectively, classified as outer cases. The outer regressor

matrix can be rewritten as $Z_r^{\alpha}(\theta) = Z0_r + U_r^{\theta}$ where

$$Z0_r = \begin{pmatrix} x_{11}^v & x_{12}^v & \dots & x_{1L}^v \\ \vdots & \vdots & & \vdots \\ x_{k_1 1}^v & x_{k_1 2}^v & \dots & x_{k_1 L}^v \\ x_{k_2 1}^v & x_{k_2 2}^v & \dots & x_{k_2 L}^v \\ \vdots & \vdots & & \vdots \\ x_{n1}^v & x_{n2}^v & \dots & x_{nL}^v \end{pmatrix} \tag{8}$$

and $U_r^{\theta} = u_r u_{\theta}' = (1,\dots,1,-1,\dots,-1)'(\theta,\dots,\theta,\theta,\dots,\theta)$ is a rank-one matrix with u_r an r-vector whose first k_1 components are $1s$ and the remaining $(n-k_2+1)$ components are $-1s$, and u_{θ} is a p-vector. Thus $Z_r^{\alpha}(\theta)$ can be seen as a rank-one correction of $Z0_r$. We call $Z0_r$ and U_r^{θ}, basis and correction components, respectively. For threshold values in a given interval, say $\theta_{\tau_1-1} \leqslant \theta < \theta_{\tau_1}$, the same data partition holds and thus the basis is invariant, whereas the correction changes due to the explicit dependence on θ. The net result is that for within-interval threshold variations $Z_r^{\alpha}(\theta)$ changes, whereas Z_s^{β} is invariant since it does not explicitly depend on θ.

3.2. QR Approach to LS solving

Given a general LS problem

$$\min_{\gamma} \|X\gamma - y\|_2, \tag{9}$$

where $X \in \mathbb{R}^{n\times m}$ $(n \geqslant m)$, $y \in \mathbb{R}^n$, $\gamma \in \mathbb{R}^m$ and $\|.\|_2$ represents the Euclidean length or 2-norm in $\mathbb{R}^n$, its solution can be written in terms of the so-called pseudoinverse or Moore–Penrose inverse $X^{\perp}$ (Björck, 1996) as

$$\hat{\gamma} = X^{\perp} y. \tag{10}$$

However, rather than calculating $X^{\perp}$ explicitly, decompositions of X or methods to solve the normal equations $(X'X\gamma = X'y)$ associated with (9) are commonly used.

When X is a full-column rank matrix the vector $\hat{\gamma} \in R^m$ that solves (9), called the LS estimator, is unique and given by[5]

$$\hat{\gamma} = (X'X)^{-1}X'y, \tag{11}$$

where $(X'X)^{-1}X'$ is the Moore–Penrose inverse of a full-column rank matrix. The latter can be efficiently calculated via the QR approach which has the merit of being relatively cheap in terms of number of operations and computationally stable. The following Theorem defines the QR factorization.

[5] A full-column rank regressor matrix implies no exact collinearity between the regressors. This condition is guaranteed in an AR(p) model since $X_{n\times(p+1)} = (1, x_{t-1}, x_{t-2}, \dots, x_{t-p})$.

Theorem 1 (QR factorization). *Let $X \in \mathbb{R}^{n\times m}$ with $n \geqslant m$. Then there is an orthogonal matrix $Q \in \mathbb{R}^{n\times n}$, such that*

$$X = Q\begin{pmatrix} R \\ 0 \end{pmatrix}, \tag{12}$$

where $R \in \mathbb{R}^{m\times m}$ is upper triangular with nonnegative diagonal elements. This is called the QR factorization of X.

Since orthogonal transformations preserve Euclidean length (9) is equivalent to

$$\min_{\gamma} \|Q'(X\gamma - y)\|_2, \tag{13}$$

where

$$\|Q'(X\gamma - y)\|_2^2 = \left\| \begin{pmatrix} R \\ 0 \end{pmatrix} \gamma - \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} \right\|_2^2 = \|R\gamma - d_1\|_2^2 + \|d_2\|_2^2 \tag{14}$$

for any $\gamma \in \mathbb{R}^m$. Since $rank(X) = rank(R) = m$, the LS estimator $\hat{\gamma}$ can be obtained by solving the upper triangular linear system $R\gamma = d_1$.[6] The residuals, $\hat{e} = y - X\hat{\gamma}$, can be calculated efficiently by $\hat{e} = Q(0, d_2)'$ and the RSS by $\|\hat{e}\|_2^2 = \|d_2\|_2^2$.

Another important advantage of the QR approach is that when X is altered in particular ways, the factorization of the resulting matrix $\tilde{X}$ needs not be calculated anew but can be easily updated from previous factors using, for instance, Givens rotations. The latter is a computationally stable approach which requires considerably fewer operations than a new factorization.[7] Two scenarios are of particular interest for TAR modeling: when $\tilde{X}$ is obtained by adding rows to X or as a rank-one correction of X. Section 3.4 discusses in detail how Givens rotations are integrated in the proposed fitting approach.

3.3. Locally continuous $RSS(\theta)$ functions

Let us consider the linear regression model $y = X(\theta)\gamma + \varepsilon$ and associated LS problem

$$\min_{\gamma} \|X(\theta)\gamma - y\|, \tag{15}$$

where the $n \times m$ $(n \geqslant m)$ regressor matrix depends explicitly on a parameter θ. Following our discussion in Section 3.2, the LS solution of (15) can be written as

$$\hat{\gamma}^{\theta} = X(\theta)^{\perp} y \tag{16}$$

and the *RSS*, defined by $\|\hat{e}^{\theta}\|_2^2 = (y - X(\theta)\hat{\gamma}^{\theta})'(y - X(\theta)\hat{\gamma}^{\theta})$, turns out to be

$$\|\hat{e}^{\theta}\|_2^2 = y' P^{\perp}_{R(X)} y = y'(I - X(\theta)X(\theta)^{\perp})y, \tag{17}$$

[6] This factorization can be modified using pivoting techniques, such as the rank-revealing QR approach, to solve LS problems where the regressor matrix is rank deficient.

[7] See Stoer and Bulirsch (1992) and Golub and Van Loan (1996). Schlittgen (1997) proposes a fast estimation algorithm for a different type of TAR models—where θ does not appear explicitly in the conditional mean—which also uses Givens updates of the QR decomposition.

where $P^{\perp}_{R(X)}$ is the orthogonal projection of R^n onto the range of $X(\theta)$ and I is the identity matrix.

If $X(\theta)$ is full-column rank, then the second-moment matrix $X(\theta)'X(\theta)$ is nonsingular and (17) can be computed by

$$\|\hat{e}^{\theta}\|_2^2 = y'(I - X(\theta)(X(\theta)'X(\theta))^{-1}X(\theta)')y. \tag{18}$$

Let $X(\theta)$ denote a first degree polynomial matrix, that is, a matrix whose entries are first degree polynomials. Then there exist constant matrices X_0 and X_1 such that

$$X(\theta) = X_0 + X_1\theta, \tag{19}$$

which implies a second degree polynomial moment matrix

$$X(\theta)'X(\theta) = X_0'X_0 + (X_0'X_1 + X_1'X_0)\theta + X_1'X_1\theta^2. \tag{20}$$

Our estimation approach builds upon the next theorem.

Theorem 2. *If the $n \times m$ $(n \geqslant m)$ regressor matrix $X(\theta)$ in (15) is a first degree polynomial matrix, $X(\theta) = X_0 + X_1\theta$, with X_1 of rank one and whose ith row is $(c_i, \ldots, c_i)$ with c_i constant, then the sum of squared residuals $\|\hat{e}^{\theta}\|_2^2$ is a rational function of degree type $(4,2)$ provided $X(\theta)$ is a full-column rank matrix.*

The result in Theorem 2 is based on the following Proposition.

Proposition 3. *If an $n \times n$ second degree polynomial matrix is obtained as $A(\theta) = (B + C\theta)'(B + C\theta)$, where C is a rank-one matrix whose ith row is a vector of the form $(c_i, \ldots, c_i)$ with c_i constant, then* $\det A(\theta)$ *is a second degree polynomial.*

The proofs of these results can be found in the Appendix. Theorem 2 allows us to consider a continuous range of values in the identification of the nonlinear parameter θ while keeping computation costs within tractable limits. As shown in Section 3.1, for $\theta \in [\theta_i, \theta_{i+1})$ the outer regime LS problem can be written as (15) with $X(\theta) = Z_r^{\alpha}(\theta)$. Since the latter is a first degree polynomial matrix, from Theorem 2 it follows that the RSS of the Band-TAR model is a rational function of θ of degree type (4,2) over each threshold interval, that is $RSS(\theta) \equiv \Psi^{4,2}(\theta)$ for $\theta \in [\theta_i, \theta_{i+1})$. Since $Z_r^{\alpha}(\theta)'Z_r^{\alpha}(\theta)$ is nonsingular, the denominator of $\Psi^{4,2}(\theta)$ never vanishes in the interval and thus $RSS(\theta)$ is continuous over $[\theta_i, \theta_{i+1})$. It follows that a global minimum exists:

$$\theta^* = \arg \min_{[\theta_i, \theta_{i+1})} \Psi^{4,2}(\theta), \tag{21}$$

which can be found by applying the necessary optimality condition. Since $\Psi^{4,2}(\theta)$ is a relatively low degree rational function, its coefficients can be cheaply identified in each interval via rational interpolation using just seven (arbitrary) support points. [8]

[8] We implement a simple Neville–Aitken type approach which generates a tableau of values of intermediate rational functions $\Psi^{\mu,\nu}(\theta)$ following the zigzag $(0,0) \to (1,0) \to (2,0) \to (3,0) \to (3,1) \to (4,1) \to (4,2)$ in the (μ, ν)-diagram.

3.4. QR updating: Givens transformations

This section discusses, first, how Givens rotations can be used to iterate efficiently within and across thershold intervals and, second, how the lags p, q and d can be identified.

For the outer regime, the algorithm starts by considering the submodel for $p = L$. The outer-regime LS problem for any $\theta \in [\theta_i, \theta_{i+1})$ can be written as

$$\min_{\alpha} \|Z_r^{\alpha}(\theta)\alpha - \Delta z_r\|_2, \tag{22}$$

where, as shown in Section 3.1, its $r \times L$ regressor matrix has two components, the basis $Z0_r$ and the correction U_r^{θ}, $Z_r^{\alpha}(\theta) = Z0_r + U_r^{\theta}$. Since within-interval variations of θ affect only U_r^{θ}, it follows that the LS problems associated with different candidates in $[\theta_i, \theta_{i+1})$ can be solved readily by simply updating the QR factors of $Z_r^{\alpha}(\theta)$ for the different U_r^{θ}. Moreover, since $Z_r^{\alpha}(\theta)$ is just a rank-one correction of $Z0_r$, these within-interval updates can be cheaply obtained via Givens rotations.

Now let us consider across-interval variations of θ, that is, the algorithm moves to the next (say, contiguous to the left) threshold interval, $[\theta_{i-1}, \theta_i) \subset \Theta$. In contrast to the within-interval variations, not only does the correction U_r^{θ} change but also c ($c \geqslant 1$) new cases are classified into the outer regime which defines a new basis $Z0_r$ with $r = r + c$. Since the new $Z0_r$ is just the previous interval's basis with c additional rows, its QR factorization can be cheaply obtained via Givens rotations.

Consider now the augmented matrix $(Z_r^{\alpha}(\theta)|\Delta z_r)_L$ and its QR factorization

$$(Z_r^{\alpha}(\theta)|\Delta z_r)_L = Q_r \begin{pmatrix} R_r & d_r \\ 0 & s_r \\ 0 & 0 \end{pmatrix}, \tag{23}$$

where Q_r and R_r are $r \times r$ and $L \times L$ matrices, respectively, d_r is a L-dimensional vector and s_r is an scalar and, for simplicity, we have dropped θ in the right-hand side of the equation. It follows that

$$Z_r^{\alpha}(\theta) = Q_r \begin{pmatrix} R_r \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \Delta z_r = Q_r \begin{pmatrix} d_r \\ s_r \\ 0 \end{pmatrix} \tag{24}$$

and $\|Q_r'(Z_r^{\alpha}(\theta)\alpha - \Delta z_r)\|_2^2 = \|R_r\alpha - d_r\|_2^2 + \|s_r\|_2^2$. It follows that the LS estimator $\hat{\alpha}$ can be efficiently calculated by solving (back substitution) the upper triangular system $R_r\hat{\alpha} = d_r$, that is

$$\begin{pmatrix} r_{11} & r_{12} & \dots & r_{1L-1} & r_{1L} \\ 0 & r_{22} & \dots & r_{2L-1} & r_{2L} \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & r_{L-1L-1} & r_{L-1L} \\ 0 & 0 & \dots & 0 & r_{LL} \end{pmatrix} \hat{\alpha} = \begin{pmatrix} d_{r1} \\ d_{r2} \\ \vdots \\ d_{rL-1} \\ d_{rL} \end{pmatrix} \tag{25}$$

and the RSS estimator is given by $RSS_r^{\alpha}(\theta, L) = \|s_r\|_2^2 = s_r^2$.

If for the same given θ in $[\theta_i, \theta_{i+1})$ we consider the model for $p = L-1$, the associated outer matrix $(Z_r^{\alpha}(\theta)|\Delta z_r)_{L-1}$ is just $(Z_r^{\alpha}(\theta)|\Delta z_r)_L$ with the last column of $Z_r^{\alpha}(\theta)$ deleted.[9] This has minimal impact on the earlier QR factors in (23), that is

$$(Z_r^{\alpha}(\theta)|\Delta z_r)_{L-1} = Q_r \begin{pmatrix} \tilde{R}_r & d_r \\ 0 & s_r \\ 0 & 0 \end{pmatrix}, \tag{26}$$

where $\tilde{R}_r$ is just R_r with its last column deleted, giving $RSS_r^{\alpha}(\theta, L-1) = d_{rL}^2 + s_r^2 = d_{rL}^2 + RSS_r^{\alpha}(\theta, L)$. This can be generalized to any order $p < L$ as $RSS_r^{\alpha}(\theta, p) = d_{r,\,p+1}^2 + RSS_r^{\alpha}(\theta, p+1) = d_{r,\,p+1}^2 + d_{r,\,p+2}^2 \cdots + d_{rL}^2 + s_r^2$. Once the iterations for $p = L, L-1, \ldots, 1$ are completed the best order (conditional on d and θ) can be obtained by minimizing an Akaike Information Criterion (AIC) following Tong (1983):

$$\tilde{p}_r(\theta, d) = \arg \min_{1 \leqslant p \leqslant L} AIC_{\alpha}(\theta, d, p), \tag{27}$$

where $AIC_{\alpha}(\theta, d, p) = r \ln(RSS_r^{\alpha}(\theta, p)/r) + 2p$ and $AIC_{\alpha}(\theta, d, \tilde{p}_r)$ represents the minimal criterion associated with d and a specific θ in $[\theta_i, \theta_{i+1})$.[10] For each interval, our fitting procedure finds first the optimal θ_p^* associated with each $p = L, L-1, \ldots, 1$, by minimizing a continuous rational function as discussed in Section 3.2, and then minimizes the resulting $AIC_{\alpha}(\theta_p^*, d, p)$ sequence to find $\tilde{p}_r(\theta_{\tilde{p}_r}^*, d)$.

For the inner regime, the algorithm proceeds analogously to calculate RSS_s^{β} and AIC_{β} sequentially for $q = L, L-1, \ldots, 1$. However, since the inner-regime LS problem is within-interval invariant, a unique best-fit $\tilde{q}_s$ and minimal $AIC_{\beta}(d, \tilde{q}_s)$ are associated with all $\theta \in [\theta_i, \theta_{i+1})$. For the Band-TAR, a minimal AIC for each interval is calculated as $AIC_{\mathrm{TAR}}(\theta_{\tilde{p}_r}^*, d, \tilde{p}_r, \tilde{q}_s) = AIC_{\alpha}(\theta_{\tilde{p}_r}^*, d, \tilde{p}_r) + AIC_{\beta}(d, \tilde{q}_s)$, which gives an AIC sequence. The values (θ, p, q) that minimize this sequence conditional on a given d, that is

$$(\tilde{\theta}, \tilde{p}, \tilde{q}) = \arg \min_{[\theta_i, \theta_{i+1}) \subset \Theta} AIC_{\mathrm{TAR}}(\theta_{\tilde{p}_r}^*, d, \tilde{p}_r, \tilde{q}_s) \tag{28}$$

are the LS estimates and $AIC_{\mathrm{TAR}}(\tilde{\theta}, d, \tilde{p}, \tilde{q})$ is the minimal AIC of the Band-TAR model for a given d.[11]

[9] Using for order $p < L$ the same $Z_r^{\theta}(\theta)$ matrix as for order L with its last $L-p$ columns deleted is computationally efficient but implies that $L-p$ available cases (rows) for $Z_r^{\theta}(\theta)$ are not used. Though in this sense the LS solution for the $p < L$ models is suboptimal, this is tolerated since in general (parsimony principle) L is not too large and this simplification delivers a much faster algorithm. Nevertheless, once (θ, d, p, q) are identified the algorithm estimates α and β using all available data.

[10] Alternatively one could adopt the more parsimonious Schwarz Bayesian criterion (SBC) or the Hannan–Quinn criterion (HQC) which lies somewhere between the AIC and HQC. Discussing the relative adequacy of these and other criteria goes beyond the scope of this paper. See Kapetanios (1999b).

[11] Note that only RSS_r^{α} and RSS_s^{β}, but not $\hat{\alpha}$ and $\hat{\beta}$, are required for the identification of (θ, d, p, q). A number of operations can be avoided by updating only the Q and R matrices involved in the former's estimation. In particular, the interval-by-interval iterations for the outer regime require updates of both the Q and R factors of $Z0_r^{\alpha}$ to obtain Z_r^{α}'s factorization (due to the rank-one correction U_r^{α}), whereas the inner-regime iterations only require updating the R factor of Z_s^{β}.

If d is not known a priori—as is common in practice—the above procedure is repeated for different $d \in \{1,\ldots,D\}$ where D is some maximum plausible threshold delay. This gives a sequence of $AIC_{\mathrm{TAR}}(\tilde{\theta}, d, \tilde{p}, \tilde{q})$ values. The LS estimates of (θ, d, p, q) are obtained by minimizing this sequence, that is

$$(\hat{\theta}, \hat{d}, \hat{p}, \hat{q}) = \arg \min_{1 \leqslant d \leqslant D} AIC_{\mathrm{TAR}}(\tilde{\theta}, d, \tilde{p}, \tilde{q}) \tag{29}$$

or alternatively, following Tong and Lim (1980), the normalized AIC (NAIC) sequence

$$(\hat{\theta}, \hat{d}, \hat{p}, \hat{q}) = \arg \min_{1 \leqslant d \leqslant D} AIC_{\mathrm{TAR}}(\tilde{\theta}, d, \tilde{p}, \tilde{q})/(N - \max(d, L)) \tag{30}$$

where $N - \max(d, L)$ is the effective sample size. Once $\hat{\theta}, \hat{d}$, $\hat{p}$ and $\hat{q}$, are identified the model becomes (piecewise) linear in the remaining parameters, α and β, which can be easily estimated by LS.

3.5. The proposed fitting approach

The following steps can be followed to identify and estimate the Band-TAR parameters:

(i) Fix D and L, some maximum feasible values for the threshold delay and AR lag orders, respectively. Set the minimum number of observations required in each regime to $m = \kappa N$ where N is the sample size. A usual choice is $\kappa = 0.15$.

(ii) For each $d \in \{1, 2, \ldots, D\}$ repeat the following:

- Arrange the data in ordered autoregressive form (y^v, X^v) as described in Section 3.1.
- Define the threshold space $\Theta = \{\bigcup_i [\theta_i, \theta_{i+1}), i = \tau_0, \tau_0 + 1, \ldots, \tau_1 - 1\} \subset \mathbb{R}^+$ by means of continuous nonoverlapping thershold intervals as described in Section 3.1.
- Repeat the following operations for the *outer regime*:
  (1) Use as initial threshold interval $[\theta_i, \theta_{i+1})$, the extreme right-hand interval in Θ. Filter out the outer-regime cases in (y^v, X^v) to form the (augmented) basis matrix $(Z0_r^{\alpha} | \Delta z_r)_p$ for order $p = L$. Compute the QR factorization of the latter.
  (2) Choose 7 arbitrary (e.g. equally spaced) thresholds $\theta_i^j \in [\theta_i, \theta_{i+1}), j = 1, \ldots, 7$.
  (3) Factorize the augmented regressor matrix $(Z_r^{\alpha}(\theta_i^j) | \Delta z_r)_p, j = 1, \ldots, 7$, by means of (rank-one-correction) Givens updates of the QR factorization of $(Z0_r^{\alpha} | \Delta z_r)_p$.
  (4) Use the updated R factor to compute $RSS_r^{\alpha}(\theta_i^j, p)$, $j = 1, \ldots, 7$.
  (5) Identify $\Psi_p^{4,2}(\theta)$ via rational interpolation with support points $(\theta_i^j, RSS_r^{\alpha}(\theta_i^j, p)), j = 1, \ldots 7$.
  (6) Find θ_p^*, the value that minimizes $\Psi_p^{4,2}(\theta)$ in $[\theta_i, \theta_{i+1})$ using the necessary optimality condition. Compute the associated minimal $RSS_r^{\alpha}(\theta_p^*, p)$ and $AIC_{\alpha}(\theta_p^*, d, p)$.

(7) Set $p=p-1$. Apply Givens rotations to the QR factorization of $(Z0_r^{\alpha}|\Delta z_r)_{p+1}$ to re-factor the new basis matrix, $(Z0_r^{\alpha}|\Delta z_r)_p$, which is just the previous basis with the last column of $Z0_r^{\alpha}$ deleted.
(8) Repeat steps 2–7 until $p=1$ is completed.
(9) Find $\tilde{p}_r$ (and associated optimal $\theta^*_{\tilde{p}_r}$), the best-fit order for the active threshold interval as the value that minimizes the $AIC_{\alpha}(\theta^*_p, d, p)$ sequence

$$\tilde{p}_r = \arg \min_{1 \leqslant p \leqslant L} AIC_{\alpha}(\theta^*_p, d, p)$$

and compute the minimal $AIC_{\alpha}(\theta^*_{\tilde{p}_r}, d, \tilde{p}_r)$.
(10) Move to the next (contiguous to the left) threshold interval in Θ. Re-factor the new basis $(Z0_r^{\alpha}|\Delta z_r)_p$ for $p=L$ by applying Givens rotations to the QR decomposition of the previous interval analogous basis (these two matrices differ only in that the former has c additional row(s) or $r=r+c, c \geqslant 1$).
(11) Repeat steps 2–10 until the last threshold interval in Θ (extreme left-hand interval) has been considered.

- Repeat the following operations for the *inner regime*:
(1) Use as starting interval $[\theta_i, \theta_{i+1})$ the last interval used in the outer iterations. Filter out the inner-regime cases in (y^v, X^v) to form the augmented regressor matrix, $(Z_s^{\beta}|\Delta z_s)_q$ for order $q=L$. Calculate its QR decomposition.
(2) Use the R factor of $(Z_s^{\beta}|\Delta z_s)_q$ to compute $RSS_s^{\beta}(q)$ and the associated $AIC_{\beta}(d,q) = s\ln(RSS_s^{\beta}/s) + 2(q+1)$.
(3) Set $q=q-1$ and apply Givens rotations to the R factor of $(Z_s^{\beta}|\Delta z_s)_{q+1}$ to obtain the R factor of the new $(Z_s^{\beta}|\Delta z_s)_q$ which is just the former matrix with the last column of Z_s^{β} deleted.
(4) Repeat steps 2 and 3 until $q=1$.
(5) Determine $\tilde{q}_s$, the best-fit order for the active threshold interval as the value that minimizes the $AIC_{\beta}(d,q)$ sequence

$$\tilde{q}_s = \arg \min_{1 \leqslant q \leqslant L} AIC_{\beta}(d,q)$$

and compute the minimal $AIC_{\beta}(d, \tilde{q}_s)$.
(6) Move to the next (contiguous to the right) interval in Θ. Obtain the R factor of the new $(Z_s^{\beta}|\Delta z_s)_q$ for $q=L$ by applying Givens rotations to the R factor of the analogous matrix for the previous interval (these two matrices differ only in that the new matrix has additional row(s) or $s=s+c, c \geqslant 1$).
(7) Repeat steps 2–6 until the last interval in Θ (right-hand extreme interval) has been considered.

- For each threshold interval compute an overall AIC from the (outer and inner) AIC obtained in steps 9 and 5, *i.e.* $AIC_{\mathrm{TAR}}(\theta^*_{\tilde{p}_r}, d, \tilde{p}_r, \tilde{q}_s) = AIC_{\alpha}(\theta^*_{\tilde{p}_r}, d, \tilde{p}_r) + AIC_{\beta}(d, \tilde{q}_s)$. Minimize the latter sequence, conditional on d, across intervals to find:

$$(\tilde{\theta}, \tilde{p}, \tilde{q}) = \arg \min_{[\theta_i, \theta_{i+1}) \subset \Theta} AIC_{\mathrm{TAR}}(\theta^*_{\tilde{p}_r}, d, \tilde{p}_r, \tilde{q}_s)$$

and calculate the associated minimal NAIC as

$$NAIC_{\mathrm{TAR}}(\tilde{\theta}, d, \tilde{p}, \tilde{q}) = AIC_{\mathrm{TAR}}(\tilde{\theta}, d, \tilde{p}, \tilde{q})/(N - \max(d, L)).$$

(iii) The LS estimates $\hat{\theta}, \hat{d}, \hat{p}$ and $\hat{q}$ are obtained by minimizing the NAIC sequence:

$$(\hat{\theta}, \hat{d}, \hat{p}, \hat{q}) = \min_{1 \leqslant d \leqslant D} NAIC_{\mathrm{TAR}}(\tilde{\theta}, d, \tilde{p}, \tilde{q}).$$

(iv) Finally, estimate the adjustment parameters of the Band-TAR model, α and β, conditional on $(\hat{\theta}, \hat{d}, \hat{p}, \hat{q})$ by ordinary LS using all available data.

4. Simulation analysis

This section investigates via Monte Carlo simulation the merits of the proposed numerical tools in the context of Band-TAR fitting.

4.1. Monte Carlo design

Three particular cases of Band-SETAR model (2) are considered in the experiments: [12]

(I) $q = 2, p = 2, d = 1, \theta = 0.35, \beta' = \{0.5, -0.55, -0.75\}, \alpha' = \{-0.8, -0.75\}$,
(II) $q = 1, p = 3, d = 2, \theta = 0.92, \beta' = \{0.4, -1.0\}, \alpha' = \{-0.5, -0.73, -0.35\}$,
(III) $q = 3, p = 5, d = 1, \theta = 0.18, \ \beta' = \{-0.95, -1.65, 0.8, 0.45\}$,

$$\alpha' = \{-1.8, 0.35, 0.4, -0.6, -0.75\}.$$

The error term ε_t is generated as $nid(0, \sigma_\varepsilon^2)$ with $\sigma_\varepsilon^2 = \{0.2, 0.4, 0.9\}$. Thus effectively nine different data generating processes (DGPs) are used. The sample sizes are $N = \{100, 200\}$ after discarding the initial 200 observations. In all experiments we replicate the simulations 500 times. Four fitting approaches are compared:

- The (continuous) approach summarized step-by-step in Section 3.5 which uses rational interpolation together with QR factorization and Givens updating [F1 hereafter].
- A grid search (GS) which uses QR factorization and Givens updating [F2].
- A GS using QR factorization (no Givens updating) [F3].
- A simple GS (no QR factorization or Givens updating) [F4].

The maximum lags considered are $D = 4$ and $L = \{4, 7\}$. [13] The minimum number of observations per regime is $m = \tau N$ with $\tau = 0.15$ and the step size of the GS is $\lambda = 10^{-1}$. Since once the change-point $\hat{\theta}$ is identified the Band-TAR model becomes piecewise linear in the remaining parameters, our comparative analysis focuses on bias and efficiency of the former and on computation costs. The following descriptive measures are employed to summarize the results: mean bias (B_μ), root mean squared

[12] The simulations were programmed in GAUSS 3.2.26 and run on a 500MHz Pentium III.
[13] To keep the analysis simple we use $L = 4$ for DGP I and DGP II and $L = 7$ for DGP III.

Table 1
Simulation results for fitting procedure F1

DGP	N	σ_ε^2	$\sigma_{\hat\theta}^2$	B_μ	t_μ(min)	$RMSE$	B_τ	t_τ(min)	MAD
I	100	0.2	0.00038	−0.00028	0.2425	0.01949	0.00013	0.2335	0.01023
I	100	0.4	0.00108	0.00060	0.2332	0.02940	−0.00105	0.2341	0.01440
I	100	0.9	0.26110	0.24211	0.2319	0.56500	0.04996	0.2325	0.07308
I	200	0.2	0.00045	0.00015	1.043	0.02084	−0.00102	1.043	0.00896
I	200	0.4	0.00173	0.00372	1.058	0.04014	−0.00026	1.051	0.01501
I	200	0.9	0.02059	0.04181	1.044	0.10943	0.00715	1.047	0.02957
II	100	0.2	0.03862	−0.12870	0.2351	0.23480	−0.04672	0.2365	0.06494
II	100	0.4	0.03635	−0.04014	0.2357	0.17770	−0.00126	0.2368	0.04246
II	100	0.9	0.04400	0.00578	0.2359	0.20891	0.00623	0.2370	0.05006
II	200	0.2	0.00793	−0.03527	1.041	0.09572	−0.00924	1.046	0.01970
II	200	0.4	0.00299	−0.00062	1.046	0.05807	−0.00111	1.047	0.01642
II	200	0.9	0.00463	0.00281	1.040	0.06839	−0.00026	1.040	0.01881
III	100	0.2	0.03246	0.09917	0.2658	0.20280	0.06753	0.2666	0.06753
III	100	0.4	0.21313	0.30622	0.2659	0.54164	0.13130	0.2667	0.13130
III	100	0.9	1.1013	0.91536	0.2664	1.3917	0.28592	0.2670	0.28601
III	200	0.2	0.00207	0.06007	1.287	0.07535	0.05349	1.288	0.05349
III	200	0.4	0.03413	0.12804	1.301	0.22463	0.10799	1.300	0.10799
III	200	0.9	0.69216	0.54177	1.297	0.99211	0.22343	1.297	0.22343

Note: $\sigma_{\hat\theta}^2$ is the sample variance of $\hat\theta$; $B_\mu = \Sigma(\hat\theta - \theta)/M$; $RMSE = \sqrt{\Sigma(\hat\theta - \theta)^2/M}$; $MAD = median(|\hat\theta - \theta|)$; $B_\tau = median(\hat\theta - \theta)$; $t_\mu = \Sigma\hat t/M$; $t_\tau = median(\hat t)$; $\hat t$ denotes computation time in min.

error (RMSE), median bias (B_τ), mean absolute deviation (MAD), sample variance ($\sigma_{\hat\theta}^2$), mean computation time (t_μ) and median computation time (t_τ).

4.2. Monte Carlo results

The estimation results, summarized in Tables 1–4, indicate that the threshold parameter estimates ($\hat\theta$) do not appear to suffer from systematic upward or downward bias. The different bias measures for $\hat\theta$ from procedure F1 are generally smaller than those from any of the GS methods (F2–F4) despite the relatively small step size employed. For instance, for DGP II (with $N = 100$, $\sigma_\varepsilon^2 = 0.2$) the RMSE and MAD from F1 are only 28% and 44% that from F2, respectively. We checked whether reducing the step size from $\lambda = 10^{-1}$ to $\lambda = 10^{-3}$ overturns this outcome. Focusing on the GS procedure F2 the results suggest that, though the biases fall the estimator $\hat\theta$ from F1 still remains less biased. For instance, the RMSE for DGP I and DGP II with $N = 100$ and $\sigma_\varepsilon^2 = 0.2$ falls from 0.06960 to 0.02210 and from 0.24739 to 0.23592, respectively. The downside of such step size reduction is that computation costs appear multiplied by a factor of some 3 to 13 times depending on the specifications (Coakley et al., 2001).

Table 2
Simulation results for fitting procedure F2

DGP	N	σ_ε^2	$\sigma_{\hat{\theta}}^2$	B_μ	t_μ(min)	$RMSE$	B_τ	t_τ(min)	MAD
I	100	0.2	0.00485	0.00098	0.0592	0.06960	−0.00543	0.0592	0.02287
I	100	0.4	0.01684	0.01107	0.0593	0.13012	−0.00513	0.0593	0.03243
I	100	0.9	0.28533	0.25132	0.0598	0.58985	0.04611	0.0598	0.07943
I	200	0.2	0.00051	−0.00192	0.3659	0.02254	−0.00280	0.3659	0.01037
I	200	0.4	0.00168	−0.00495	0.3659	0.04125	−0.00423	0.3660	0.01772
I	200	0.9	0.02431	0.04267	0.3659	0.16149	0.00788	0.3658	0.03132
II	100	0.2	0.03947	−0.14770	0.0586	0.24739	−0.06622	0.0584	0.07105
II	100	0.4	0.02909	−0.04311	0.0584	0.20121	−0.01117	0.0584	0.05067
II	100	0.9	0.05414	−0.00738	0.0591	0.23268	0.00291	0.0591	0.06890
II	200	0.2	0.02559	−0.03774	0.3660	0.18606	−0.01618	0.3660	0.02109
II	200	0.4	0.00311	−0.00780	0.3659	0.06629	−0.00441	0.3658	0.01632
II	200	0.9	0.00625	−0.00412	0.3662	0.07909	−0.00486	0.3660	0.02003
III	100	0.2	0.03795	0.09348	0.0637	0.21840	0.06558	0.0637	0.06558
III	100	0.4	0.22717	0.26022	0.0639	0.54261	0.13932	0.0638	0.13932
III	100	0.9	1.2808	0.98141	0.0650	1.4972	0.28545	0.0648	0.28545
III	200	0.2	0.00215	0.06485	0.4103	0.07956	0.05966	0.4103	0.05966
III	200	0.4	0.02246	0.11452	0.4113	0.18849	0.09823	0.4107	0.09823
III	200	0.9	0.73273	0.54422	0.4111	1.0136	0.21697	0.4108	0.21697

With the exception of DGP III, the threshold dispersion ($\sigma_{\hat{\theta}}^2$) suggests that the estimates from F1 are generally more efficient. Hence, these results provide prima facie evidence that permitting a continuous range of variation for the threshold in Band-TAR fitting can help reduce small-sample biases and increase the efficiency of the threshold estimator.[14] This is important given that the remaining (linear) parameters in the TAR model are estimated conditional on $\hat{\theta}$. Finally, the GS procedures produce virtually identical results in terms of bias and efficiency, as expected, since they define the same feasible set of threshold candidates.[15]

In terms of computation costs, a ceteris paribus comparison of F3 and F4 across specifications indicates that by using the QR approach to LS solving, TAR estimation time—as measured by t_μ and t_τ—falls by a factor of 1.5 on average for the different DGPs. These time savings are likely to increase sharply in TAR-based inference analysis using simulations. For instance, in a Monte Carlo analysis of the small sample properties (such as size and power) of a Band-TAR unit root test—the null is unit root behavior and the alternative Band-TAR stationarity—with $N = 100$, $L = 7$ and $D = 4$

[14] This conclusion can be extended to other TAR schemes where the threshold also appears explicitly in the conditional mean, such as C-TAR model (3).

[15] We also compared the detailed GS approaches F2–F4 with a GS restricted to the order statistics of the simulated z_t. The results reveal that the threshold estimates from the latter suffer more often from 'suboptimality' in the sense that they are more biased and less efficient on average. See Coakley et al. (2000).

Table 3
Simulation results for fitting procedure F3

DGP	N	σ_{ε}^2	$\sigma_{\hat{\theta}}^2$	B_{μ}	t_{μ}(min)	$RMSE$	B_{τ}	t_{τ}(min)	MAD
I	100	0.2	0.00449	0.00122	0.0841	0.06699	−0.00453	0.0838	0.02121
I	100	0.4	0.03547	0.03357	0.0848	0.19111	−0.00599	0.0845	0.03633
I	100	0.9	0.18521	0.21141	0.0872	0.47909	0.07603	0.0868	0.08679
I	200	0.2	0.00050	−0.00345	0.3162	0.02251	−0.00334	0.3162	0.00965
I	200	0.4	0.00361	0.00015	0.3313	0.06045	−0.00374	0.3310	0.01800
I	200	0.9	0.05206	0.04977	0.3178	0.23330	0.00628	0.3172	0.03186
II	100	0.2	0.04131	−0.39739	0.0674	0.44626	−0.42112	0.0706	0.42112
II	100	0.4	0.07537	−0.18965	0.0790	0.33344	−0.07572	0.0833	0.11401
II	100	0.9	0.04350	−0.02784	0.0878	0.21020	−0.01035	0.0878	0.05149
II	200	0.2	0.03163	−0.22267	0.2683	0.28486	−0.20523	0.2731	0.20523
II	200	0.4	0.00498	−0.01379	0.3128	0.07183	−0.00801	0.3158	0.01979
II	200	0.9	0.00668	−0.00733	0.3313	0.08171	−0.00787	0.3313	0.02115
III	100	0.2	0.03239	0.10851	0.3110	0.21001	0.07437	0.3105	0.07437
III	100	0.4	0.24141	0.30478	0.3163	0.57776	0.13939	0.3155	0.13939
III	100	0.9	1.1949	0.94876	0.3539	1.4466	0.29352	0.3508	0.29352
III	200	0.2	0.00188	0.06268	1.2815	0.07621	0.05491	1.2810	0.05491
III	200	0.4	0.05126	0.14308	1.2816	0.26763	0.11208	1.2808	0.11208
III	200	0.9	0.78740	0.62651	1.4050	1.0855	0.23885	1.4047	0.23885

and a sensible number of replications, say 10,000, the difference between employing F3 or F4 is some 1500 min. ($\simeq$ 25 h.). A comparison of F2 and F3 reveals that using Givens rotations to update the QR factors (rather than calculating them anew) reduces computation time by a ratio of 1.5 when $L = D = 4$ (DGP I and DGP II) and a ratio of some 5.5 times when $L = 7$ and $D = 4$ (DGP III). These ratios are magnified for the larger sample size $N = 200$. Therefore, Givens rotations can reduce the costs of TAR fitting, and more so the larger is $L(D)$ which increases the number of potential models.

A comparison of F2 and F4 suggests that both the QR factorization and Givens rotations jointly reduce estimation costs by some 2.2 to 8.2 times depending on the specifications. The latter translate into a time difference of some 4320 min. ($\simeq$ 72 h.) in a Monte Carlo analysis of the small sample properties of a Band-TAR test with $N = 100$, $L = 7$, $D = 4$ and 10,000 replications. Finally, as expected from the way the feasible threshold range is defined, computation costs increase with innovation volatility for the GS methods F2–F4 whereas they are invariant to data dispersion for F1 and depend only on sample size. This difference may be relevant when fitting Band-TARs to highly volatile time series such as those involving financial variables. [16]

[16] As an illustration of Band-TAR modeling, the algorithm outlined above is applied to US dollar exchange rate data of six core members of the ERM (Belgium, Denmark, France, Germany, Italy and the Netherlands) plus the UK. See Coakley et al. (2000).

Table 4
Simulation results for fitting procedure F4

DGP	N	σ_ε^2	$\sigma_{\hat{\theta}}^2$	B_μ	t_μ(min)	$RMSE$	B_τ	t_τ(min)	MAD
I	100	0.2	0.00319	−0.00218	0.1261	0.05646	−0.00442	0.1260	0.02009
I	100	0.4	0.02356	0.03199	0.1268	0.15666	−0.00149	0.1267	0.03178
I	100	0.9	0.24187	0.24309	0.1276	0.58816	0.05634	0.1271	0.07822
I	200	0.2	0.00047	−0.00102	0.4558	0.02168	−0.00215	0.4558	0.00960
I	200	0.4	0.00216	−0.00119	0.4561	0.04641	−0.00408	0.4559	0.01820
I	200	0.9	0.07688	0.06923	0.4577	0.28552	0.01025	0.4572	0.03192
II	100	0.2	0.04214	−0.38885	0.1096	0.43961	−0.42031	0.1124	0.42031
II	100	0.4	0.03849	−0.08029	0.1373	0.21182	−0.01053	0.1437	0.45778
II	100	0.9	0.04498	−0.03942	0.1449	0.21551	−0.00836	0.1446	0.05100
II	200	0.2	0.03119	−0.21916	0.3947	0.28135	−0.21112	0.4096	0.21112
II	200	0.4	0.00399	−0.00697	0.4562	0.06605	−0.00638	0.4568	0.01829
II	200	0.9	0.00559	−0.00455	0.4599	0.07485	−0.00414	0.4588	0.01970
III	100	0.2	0.05707	0.12185	0.4376	0.26796	0.07536	0.4361	0.07536
III	100	0.4	0.28301	0.28402	0.4443	0.63731	0.14287	0.4423	0.14287
III	100	0.9	1.2430	1.0356	0.5328	1.5209	0.3099	0.5348	0.3099
III	200	0.2	0.00174	0.06087	1.8149	0.07375	0.05744	1.8150	0.05744
III	200	0.4	0.07061	0.15025	1.9549	0.30504	0.11622	1.9548	0.11622
III	200	0.9	0.86391	0.62951	1.8256	1.12181	0.23258	1.8252	0.23258

5. Conclusions

This paper investigates numerical aspects of TAR fitting. In this context it evaluates the computational advantages of the QR approach and Givens transformations. In addition it shows that, for a particular class of TARs, the model residual sum of squares is a continuous rational function over particular threshold intervals. Building on this result we propose a novel algorithm which, in contrast to a grid search, allows for a continuum of values for the threshold parameter while keeping computational costs within tractable limits. While the latter is discussed in the context of Band-TARs it can be easily generalized to other TAR models. Monte Carlo experiments are conducted to compare the efficiency of different fitting procedures.

Our simulation results suggest that QR factorizations and Givens updates significantly reduce (up to eight times for the specifications used) the computation costs of the sequential conditional LS involved in TAR fitting, and especially so the larger the maximum plausible lag order and delay parameter. If threshold parameter accuracy is important, the continuous-threshold-range method is the preferred approach while, if speed is more relevant, a fast GS using QR factorizations and Givens updates is recommended. Finally, when the objective is a trade-off between threshold accuracy and computation costs, we suggest a mixed approach. The latter consists of a GS which uses the rational function component in long threshold intervals.

Issues for future research include improving the rational interpolation algorithm in terms of computation time and stability and investigating further the properties of the residual sum of squares rational functions.

Acknowledgements

We thank Michel Juillard, Miguel Angel Martín, an anonymous referee and participants at the Society for Computational Economics 6th CEF International Conference, Universitat Pompeu Fabra, Barcelona, July 2000, for helpful comments.

Appendix.

This appendix includes the proof of the main results in Section 3.3. It starts by proving the following more general result.

Theorem 4. *Given an $n \times n$ polynomial matrix of degree r, $A(\theta)=A_0+\theta A_1 \cdots+\theta^r A_r$, where A_i for $i=1,\ldots,r$ are rank-one matrices, then* $\det A(\theta)$ *is a polynomial of degree* $r(r+1)/2$ *if* $n \geqslant r$ *or* $nr-n(n-1)/2$ *if* $n<r$.

Proof. If $A_i=(\mathbf{a}_1^i,\ldots,\mathbf{a}_n^i)'$ and we denote by $\mathbf{p}_i(\theta)$ the polynomial vector $\mathbf{a}_i^0+\theta\mathbf{a}_i^1+\cdots+\theta^r\mathbf{a}_i^r$ then the polynomial matrix can be written as $A(\theta)=(\mathbf{p}_1(\theta),\mathbf{p}_2(\theta)\ldots\mathbf{p}_n(\theta))'$. Using the properties of the determinant

$$\det A(\theta)=\det\begin{pmatrix}(\mathbf{a}_1^0)'\\ \mathbf{p}_2(\theta)'\\ \cdots\\ \mathbf{p}_n(\theta)'\end{pmatrix}+\theta\det\begin{pmatrix}(\mathbf{a}_1^1)'\\ \mathbf{p}_2(\theta)'\\ \cdots\\ \mathbf{p}_n(\theta)'\end{pmatrix}+\cdots+\theta^r\det\begin{pmatrix}(\mathbf{a}_1^r)'\\ \mathbf{p}_2(\theta)'\\ \cdots\\ \mathbf{p}_n(\theta)'\end{pmatrix}. \tag{31}$$

Analogously, it can be shown that

$$\det(\mathbf{a}_1^i,\mathbf{p}_2(\theta),\ldots,\mathbf{p}_n(\theta))'=\det P_{(i,0)}+\theta\det P_{(i,1)}+\cdots+\theta^r\det P_{(i,r)},$$

where $P_{(i,j)}$ denotes the matrix $(\mathbf{a}_1^i,\mathbf{a}_2^j,\mathbf{p}_3(\theta),\ldots,\mathbf{p}_n(\theta))'$.

Substituting the above expression in (31) and factoring out θ^k gives

$$\begin{aligned}\det A(\theta)=&\det P_{(0,0)}+\theta[\det P_{(0,1)}+\det P_{(1,0)}]\\ &+\theta^2[\det P_{(0,2)}+\det P_{(1,1)}+\det P_{(2,0)}]\\ &+\cdots+\theta^r[\det P_{(0,r)}+\det P_{(1,r-1)}+\cdots+\det P_{(r,0)}]\\ &+\theta^{r+1}[\det P_{(1,r)}+\cdots+\det P_{(r,1)}]+\cdots+\theta^{2r}\det P_{(r,r)}.\end{aligned}$$

Note that in the above formula every power θ^k multiplies a sum of determinants, $\det P_{(i,j)}$, with the common property that the sum of their indices i and j (upper indices of the first two—constant—rows) equals power k. Moreover, those determinants for

which $i=j$, for $i,j\neq 0$ vanish. This last property is due to the hypothesis of rank unity for A_i $i=1,\ldots,r$.

This process is continued recursively until all matrices appearing in the formula for $\det A(\theta)$ are constant. It follows that the determinant of the matrix whose row upper indices add up to a maximum possible value, s, (without nonzero repeated indices) establishes the highest attainable power, θ^s, in $\det A(\theta)$. Then, the latter matrix must contain rows $\mathbf{a}_{i_1}^r$, $\mathbf{a}_{i_2}^{r-1},\ldots,\mathbf{a}_{i_n}^{r-n+1}$ if $n<r$ or $\mathbf{a}_{i_1}^r$, $\mathbf{a}_{i_2}^{r-1},\ldots,\mathbf{a}_{i_r}^1$ if $n\geqslant r$, giving $s=r+(r-1)+\cdots+(r-n+1)=nr-n(n-1)/2$ for the former and $s=r+(r-1)+\cdots+1=r(r+1)/2$ for the latter. □

The above proof provides a formula for $\det A(\theta)$. As an example, let us consider the particular case $r=2$. In this case, the degree of this polynomial is $r(r+1)/2=3$ and

$$\det A(\theta)=\det A_0+\theta\left(\sum_{i=1}^{n}\det A_{0(1)}^{i}\right)+\theta^2\left(\sum_{i=1}^{n}\det A_{0(2)}^{i}\right)+\theta^3\left(\sum_{i,j=1\ i\neq j}^{n}\det A_{0(1,2)}^{i,j}\right), \tag{32}$$

where

$$A_{0(1)}^{i}=(\mathbf{a}_1^0,\ldots,\mathbf{a}_{i-1}^0,\mathbf{a}_i^1,\mathbf{a}_{i+1}^0,\ldots,\mathbf{a}_n^0)',$$

$$A_{0(2)}^{i}=(\mathbf{a}_1^0,\ldots,\mathbf{a}_{i-1}^0,\mathbf{a}_i^2,\mathbf{a}_{i+1}^0,\ldots,\mathbf{a}_n^0)',$$

$$A_{0(1,2)}^{i,j}=(\mathbf{a}_1^0,\ldots,\mathbf{a}_{i-1}^0,\mathbf{a}_i^1,\mathbf{a}_{i+1}^0,\ldots,\mathbf{a}_{j-1}^0,\mathbf{a}_j^2,\mathbf{a}_{j+1}^0,\ldots,\mathbf{a}_n^0)'.$$

Proceeding analogously, results can be established for polynomial matrices in which the A_i components have rank different from one. We do not include them here for space considerations. We are particularly interested in the case where $r=2$ and A_1 and A_2 have rank two and one, respectively. For the latter, the analogous formula to (32) is

$$\det A(\theta)=\det A_0+\theta\left(\sum_{i=1}^{n}\det A_{0(1)}^{i}\right)+\theta^2\left(\sum_{i=1}^{n}\det A_{0(2)}^{i}+\sum_{i,j=1}^{n}\det A_{0(1,1)}^{i,j}\right)+\theta^3\left(\sum_{i,j=1}^{n}\det A_{0(1,2)}^{i,j}\right)+\theta^4\left(\sum_{i,j,k=1}^{n}\det A_{0(1,1,2)}^{i,j,k}\right), \tag{33}$$

where indexes i,j, and k in the same sum are always different. This formula will be used to prove Proposition 1.

Proof of Proposition 1. The matrix $A(\theta)$ obtained as $(B+C\theta)'(B+C\theta)$ is the following second degree polynomial matrix:

$$A(\theta)=A_0+\theta A_1+\theta^2 A_2, \tag{34}$$

where $A_1 = B'C + C'B$ and $A_2 = C'C$. Taking into account the special structure of C we have

$$A_1 = \begin{pmatrix} d_1 + d_1 & d_1 + d_2 & \dots & d_1 + d_n \\ d_2 + d_1 & d_2 + d_2 & \dots & d_2 + d_n \\ \vdots & \vdots & & \vdots \\ d_n + d_1 & d_n + d_2 & \dots & d_n + d_n \end{pmatrix}, \quad A_2 = \begin{pmatrix} \kappa & \dots & \kappa \\ \vdots & & \vdots \\ \kappa & \dots & \kappa \end{pmatrix}, \tag{35}$$

where $d_i = \sum_{j=1}^{n} (c_j b_{ji})$ and $\kappa = \sum_{i=1}^{n} c_i^2$. A_2 is clearly a rank one matrix and we prove next that A_1 has rank two. To see this, let us take any 3×3 submatrix from A_1 and calculate its determinant

$$\begin{aligned} \det(\mathbf{d}_1 + \mathbf{d}_j, \mathbf{d}_2 + \mathbf{d}_j, \mathbf{d}_3 + \mathbf{d}_j)' &= \det(\mathbf{d}_1, \mathbf{d}_2 + \mathbf{d}_j, \mathbf{d}_3 + \mathbf{d}_j)' \\ &\quad + \det(\mathbf{d}_j, \mathbf{d}_2 + \mathbf{d}_j, \mathbf{d}_3 + \mathbf{d}_j)' \\ &= \det(\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3 + \mathbf{d}_j)' + \det(\mathbf{d}_1, \mathbf{d}_j, \mathbf{d}_3 + \mathbf{d}_j)' \\ &\quad + \det(\mathbf{d}_j, \mathbf{d}_2, \mathbf{d}_3 + \mathbf{d}_j)' + \det(\mathbf{d}_j, \mathbf{d}_j, \mathbf{d}_3 + \mathbf{d}_j)', \end{aligned}$$

where $\mathbf{d}_j = (d_{j_1}, d_{j_2}, d_{j_3})'$, $\mathbf{d}_k = d_{i_k}(1,1,1)'$ with $k = 1,2,3$. The first and fourth terms (determinants) of the last equality are clearly zero. Writing the remaining two terms as sums of determinants they are also seen to vanish.

Hence formula (33) can be applied yielding that $\det A(\theta)$ is a fourth degree or lower polynomial. However, notice that in the resulting formula every determinant appearing in the sum multiplying θ^4 vanishes because

$$\begin{aligned} \det A^{i,j,k}_{0(1,1,2)} &= \det(\mathbf{a}_1^0 \dots, \kappa\mathbf{1}, \dots, d_j\mathbf{1} + \mathbf{d}, \dots, d_k\mathbf{1} + \mathbf{d}, \dots, \mathbf{a}_0^n)' \\ &= \det(\dots, \kappa\mathbf{1}, \dots, d_j\mathbf{1}, \dots, d_k\mathbf{1} + \mathbf{d}, \dots)' \\ &\quad + \det(\dots, \kappa\mathbf{1}, \dots, \mathbf{d}, \dots, d_k\mathbf{1} + \mathbf{d}, \dots)' \\ &= \det(\dots, \kappa\mathbf{1}, \dots, \mathbf{d}, \dots, d_k\mathbf{1}, \dots)' + \det(\dots, \kappa\mathbf{1}, \dots, \mathbf{d}, \dots, \mathbf{d}, \dots)' \\ &= 0 \end{aligned}$$

where $\mathbf{1}$ and $\mathbf{d}$ denote the vectors $(1, \dots, 1)'$ and $(d_1, \dots, d_n)$, respectively. Besides, the coefficient of θ^3 is also zero because

$$\det A^{i,j}_{0(1,2)} = -\det A^{j,i}_{0(1,2)}. \tag{36}$$

This last equality follows from

$$\begin{aligned} \det A^{i,j}_{0(1,2)} &= \det(\mathbf{a}^0, \dots, d_i\mathbf{1} + \mathbf{d}, \dots, \kappa\mathbf{1}, \dots, \mathbf{a}_n^0)' \\ &= \det(\mathbf{a}^0, \dots, \mathbf{d}, \dots, \kappa\mathbf{1}, \dots, \mathbf{a}_n^0)'. \end{aligned} \tag{37}$$

Then $\det A(\theta)$ has degree two since the coefficients of both θ^4 and θ^3 are zero. □

Proof of Theorem 2. Taking into account that

$$(X(\theta)'X(\theta))^{-1} = \frac{1}{\det(X(\theta)'X(\theta))}\,\mathrm{adj}(X(\theta)'X(\theta)), \tag{38}$$

we conclude from Proposition 1 that the entries of the matrix $(X(\theta)'X(\theta))^{-1}$ are rational functions of degree (2,2). The result of this theorem is a straightforward consequence of the latter and formula (18). □

References

Balke, N.S., Fomby, T.B., 1997. Threshold cointegration. International Economic Review 38, 627–645.

Berben, R.P., van Dijk, D., 1999. Unit root tests and asymmetric adjustment: a ressesment. Econometric Institute Research Report 9902/A, Erasmus University, Rotterdam.

Björck, A., 1996. Numerical Methods for Least Squares Problems. SIAM, Philadelphia.

Caner, M., Hansen, B.E., 1998. Threshold autoregressions with a unit root. Working Papers in Economics No. 381, Boston College.

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

Chan, K.S., Tsay, R.S., 1998. Limiting properties of the least squares estimator of a continuous threshold autoregressive model. Biometrika 85, 413–426.

Coakley, J., Fuertes, A.M., 2000. Bootstrap LR tests of sign and amplitude asymmetry. Discussion Paper in Economics No. 4/2000, Birkbeck College, University of London.

Coakley, J., Fuertes, A.M., 2001a. Asymmetric dynamics in UK real interest rates. Applied Financial Economics 12, 379–387.

Coakley, J., Fuertes, A.M., 2001b. A nonlinear analysis of excess foreign exchange returns. Manchester School 69, 623–642.

Coakley, J., Fuertes, A.M., 2001c. Border costs and real exchange rate dynamics in Europe. Journal of Policy Modeling 23, 669–676.

Coakley, J., Fuertes, A.M., Pérez, M.T., 2000. Numerical issues in threshold autoregressive modeling of time series. Discussion Paper in Economics No. 12/2000, Birkbeck College, University of London.

Coakley, J., Fuertes, A.M., Pérez, M.T., 2001. An efficient rational interpolation approach to least squares estimation for Band-TARs. In: Vulvov, L., Waśniewski, J., Yalamov, P. (Eds.), Lecture Notes in Computer Science, Vol. 1988. Springer, Berlin, pp. 198–206.

Coakley, J., Fuertes, A.M., Zoega, Z., 2002. Evaluating the persistence and structuralist theories of unemployment from a nonlinear perspective. Studies in Nonlinear Dynamics and Econometrics 5, 179–202.

Enders, W., Granger, C.W.J., 1998. Unit root tests and asymmetric adjustment. Journal of Business and Economic Statistics 16, 304–311.

Golub, G.H., Van Loan, C.F., 1996. Matrix Computations. John Hopkins University Press, London.

Granger, C.W.J., Teräsvirta, T., 1993. Modelling Nonlinear Economic Relationships. Oxford University Press, Oxford.

Kapetanios, G., 1999a. Threshold models for trended time series. Working Paper, Department of Applied Economics, University of Cambridge.

Kapetanios, G., 1999b. Model selection in threshold models. Working Paper, Department of Applied Economics, University of Cambridge.

Obstfeld, M., Taylor, A.M., 1997. Non-linear aspects of goods-market arbitrage and adjustment: Heckscher's commodity points revisited. Journal of the Japanese and International Economies 11, 441–479.

Pesaran, M.H., Potter, S., 1997. A floor and ceiling model of US output. Journal of Economic Dynamics and Control 21, 661–696.

Priestley, M.B., 1998. Nonlinear and Nonstationary Time Series Analysis. Academic Press, London.

Rothman, P., 1999. Time series evidence on whether adjustment to long run equilibrium is asymmetric. Paper presented at the SNDE 7th Annual Symposium, New York University.

Schlittgen, R., 1997. Fitting of threshold models for time series. Discussion Paper No. 97-13, University of California, San Diego.

Stoer, J.F., Bulirsch, R., 1992. Introduction to Numerical Analysis. Springer, Berlin.

Tong, H., 1983. Threshold Models in Non-linear Time Series Analysis. Springer, Berlin.

Tong, H., Lim, K.S., 1980. Threshold autoregression, limit cycles and cyclical data. Journal of the Royal Statistical Society B 42, 245–292.

Tsay, R.S., 1989. Testing and modelling threshold autoregressive processes. Journal of the American Statistical Association 84, 231–240.