Semiparametric Transition Models

P. Čížek* and C.-H. Koo

CentER, Department of Econometrics & OR, Tilburg University, The Netherlands

A new semiparametric time series model is introduced – the semiparametric transition model – that generalizes the threshold and smooth transition models by letting the transition function to be of an unknown form. The estimation strategy is based on alternating the conditional and unconditional least squares estimation of the transition function and the regression parameters, respectively. The consistency and asymptotic distribution for the regression-coefficient estimator of the semiparametric transition model are derived and shown to be first-order asymptotically independent of the nonparametric transition-function estimates. Monte Carlo simulations demonstrate that the estimation of the semiparametric transition model is more robust to the type of transition between models than the parametric estimators of the threshold and smooth transition models.

JEL codes: C14, C21, C22

Keywords: local linear estimation, nonlinear time series, semiparametric estimation, regime-switching models

* *Corresponding author.* Department of Econometrics & OR, Tilburg University, P.O.Box 90153, 5000LE Tilburg, The Netherlands. *E-mail:* P.Cizek@uvt.nl.

1 Introduction

One class of nonlinear time series models that has been widely applied, for example, in macroeconomics and finance, contains regime-switching models. Among the regime-switching models, the threshold autoregressive (TAR) model of Tong (1983) is a classical one: it was widely studied (see Hansen, 2011, for an overview) and applied (e.g., Potter, 1995a; Rothman, 1998). The TAR model is quite restrictive though in the sense that no gradual change between regimes is allowed.

To overcome this limitation, the smooth transition autoregressive (STAR) model was first introduced by Chan and Tong (1986) and further studied by Teräsvirta (1994); see van Dijk et al. (2002) for a survey. The two-regime STAR model is given by

$$y_t = x_t'\beta_1\{1 - w(z_t;\theta)\} + x_t'\beta_2 w(z_t;\theta) + \varepsilon_t, \qquad t = 1,\ldots,T, \tag{1}$$

where x_t contains lagged values of the response variable y_t, z_t is a continuously distributed transition variable, and $w(\cdot;\theta) : \mathbb{R} \to \mathbb{R}$ is a smooth transition function known up to a finite-dimensional vector θ of parameters. The TAR model would correspond to $w(z;\theta) = I(z > \theta)$ (if the discontinuity is neglected). Among smooth transition functions, a popular choice of $w(z;\theta = (\mu, s)')$ is the logistic distribution function $\Lambda(z;\mu,s) = \{1 + \exp[-s(z-\mu)]\}^{-1}$, which is smooth and monotonic. The corresponding logistic STAR (LSTAR) model has been used to model business cycle asymmetry, for instance, where the regimes correspond to expansions and recessions (Teräsvirta and Anderson, 1992; Skalin and Teräsvirta, 2002). Another practically applied transition function $w(\cdot;\theta)$ is the exponential function $G(z;\mu,s) = 1 - \exp[-s(z-\mu)^2]s$, where the regimes are associated with large and small absolute values of z. This so-called exponential STAR (ESTAR) model has been applied, for example, to real exchange rate data (Taylor et al., 2001; Sarantis, 1999). Finally, recent extensions of the two-regime STAR models (1) include the multiple-regime STAR model (van Dijk and Franses, 1999), the flexible-coefficient STAR model (Medeiros and Veiga, 2003, 2005), the time-varying STAR model (Lundbergh et al., 2003), multivariate STAR (Taylor et al., 2000), and transition models with endogenous explanatory variables (Areosa et al., 2011).

In the STAR model, the transition function $w(\cdot;\theta)$ characterized by parameter θ is assumed to be a known continuous function; typically, it is also bounded between 0 and 1. The assumption that the transition function has a certain parametric form is however hardly justified. Moreover, using a misspecified transition function may lead to inconsistent estimates and thereby wrong inference. Therefore, the present paper introduces a more flexible model in which the transition function is of an unknown form, possibly with a finite set of discontinuities: the semiparametric transition (SETR) model. The SETR model has three main advantages over the STAR model. First, the risk of model misspecification is substantially reduced as the transition function is only assumed to be smooth (up a finite set of discontinuities). Next, even though the estimator of regression coefficients does not rely on any parametric form of the transition function, its rate of convergence to the true values is the same as in the STAR model. Finally, estimates of the transition function in the semiparametric transition model can be used to study important features of the transition between the two regimes (e.g., the size and location of a jump or overshooting behavior of the transition function).

On the one hand, the SETR model nests the TAR, ESTAR, and LSTAR models and even the structural-break model if $z_t = t/T$ is chosen. On the other hand, SETR is a special case of the varying-coefficient model, which was studied by Chen and Tsay (1993) and Hastie and Tibshirani (1993). The varying-coefficient model has the form

$$E[y_t|x_t, z_t] = x_t' m(z_t), \qquad t = 1, \ldots, T, \tag{2}$$

where $m(\cdot) : \mathbb{R} \to \mathbb{R}$ is an unknown vector function and z_t is a scalar index. Recent works on model (2) include Hoover et al. (1998), Wu et al. (1998), and Fan and Zhang (2000) on longitudinal data analysis and Chen and Tsay (1993), Cai et al.

(2000), and Huang and Shen (2004) on nonlinear time series. Moreover, Zhang et al. (2002), Fan and Huang (2005), and Ahmad et al. (2005) studied the partial linearly varying-coefficient model in which some elements of vector function $m(\cdot)$ are constant. Recently, Chen and Hong (2012) designed a test of the STAR models (1) versus the varying-coefficient model (2).

In the varying-coefficient models, the parameters of interest are functions $m(z_t)$ that are estimated nonparametrically. Consequently, they cannot reach the rate of convergence typical for estimators of parametric models such as STAR and require thus larger data sets for sufficiently precise inference. On the contrary, the SETR model applies nonparametric estimation only to the transition function and the regression coefficients of the explanatory variables x_t, which are fixed in each regime, converge to the true values at the same rate as the estimates of the parametric STAR model (1).

The paper is structured as follows. In Section 2, the model and the identification conditions are presented. In Section 3, an estimation method of the semiparametric transition model is proposed. The consistency and asymptotic distribution of the proposed estimator is discussed in Section 4. Finally, a simulation study and real-data application of the SETR estimator are in Sections 5 and 6. All proofs are in the Appendix.

Throughout the paper, the following notation is used. Let $\|x\| = (x'x)^{1/2}$ for any vector $x \in \mathbb{R}^p$ and $\|X\| = \text{tr}(X'X)^{1/2}$ for any $p \times p$ matrix X. For a scalar function $w(z_t)$ of random variable z_t, the (semi)norms used are $\|w\|_\infty = \sup_{z\in\mathbb{R}} |w(z)|$ and $\|w\|_{\infty,\epsilon} = \sup_{f_z(z)>\epsilon} |w(z)|$ for a given $\epsilon > 0$ and the density f_z of z_t. In addition, let $I(\cdot)$ denote the indicator function, $\xrightarrow{P}$ the convergence in probability, and $\xrightarrow{d}$ the convergence in distribution.

2 The semiparametric transition model

Consider the following two-regime semiparametric transition model:

$$y_t = x_t'\beta_1^0\{1 - w^0(z_t)\} + x_t'\beta_2^0 w^0(z_t) + \varepsilon_t, \qquad t = 1, \ldots, T, \tag{3}$$

where y_t is an independent variable, $x_t \in \mathbb{R}^p$ is a vector of covariates, $z_t \in \mathbb{R}$ is a continuous transition variable, and ε_t denotes the error term. The parameters of interest, slopes β_1^0 and β_2^0, are the true vectors of regression coefficients corresponding to the first and second regimes, respectively, and $w^0(\cdot)$ is an unknown piecewise-smooth transition function. When lagged dependent variables are included in the explanatory variables x_t, that is, $x_t = (1, y_{t-1}, y_{t-2}, \ldots, y_{t-p-1})'$, model (3) can be referred to as the semiparametric transition autoregressive model. The transition variable z_t can be exogenous or endogenous. For example in the STAR models, z_t was treated as a lagged dependent variable y_{t-d} in Teräsvirta (1994) and as a linear time trend t/T in Lin and Teräsvirta (1994). Both specifications of z_t fit in this paper, although we concentrate on random z_t rather than a deterministic one here.

The structural-break model, the threshold model, and the smooth transition model are special cases of the SETR model. Suppose $z_t = t/T$ is a linear time trend and the transition function equals $I(z_t \geqslant t_B/T)$ for an unknown break point t_B: then SETR reduces to the structural break model. Similarly, when $w(z_t) = I(z_t \geqslant z_B)$ for a random variable z_t and an unknown threshold z_B, model (3) becomes the threshold model. Finally, assuming that transition function $w^0(z_t)$ has a parametric form $w^0(z_t; \theta)$ characterized by parameter θ yields the smooth transition model (1).

Similarly to many time series models, the estimation method considered here is based on the (nonlinear) least squares (LS). Therefore, the true parameters β_1^0, β_2^0, and w^0 described in model (3) should minimize the expected squared error:

$$\min_{\beta_1,\beta_2,w} E[y_t - x_t'\beta_1\{1 - w(z_t)\} - x_t'\beta_2 w(z_t)]^2 = \min_{\beta_1,\beta_2,w} E[y_t - x_t'\beta_1 - x_t'(\beta_2 - \beta_1)w(z_t)]^2. \tag{4}$$

To motivate and explain the identification conditions, let us write the first-order conditions for β_1^0, β_2^0, and $w^0(z_t)$ corresponding to (4) conditionally on $z_t = z$:

$$E[x_t y_t\{1 - w(z_t)\} - x_t x_t'\beta_1\{1 - w(z_t)\}^2 - x_t x_t'\beta_2\{1 - w(z_t)\}w(z_t)|z_t = z] = 0, \tag{5}$$

$$E[x_t y_t w(z_t) - x_t x_t'\beta_1\{1 - w(z_t)\}w(z_t) - x_t x_t'\beta_2 w(z_t)^2|z_t = z] = 0, \tag{6}$$

$$E[x_t'(\beta_2 - \beta_1)\{y_t - x_t'\beta_1 - x_t'(\beta_2 - \beta_1)w(z_t)\}|z_t = z] = 0. \tag{7}$$

The parameters in (5)–(7) are not identified unless additional assumptions are imposed on the slope parameters and the transition function.

Assumption 1. *Let $\{x_t, z_t, \varepsilon_t\}_{t=1}^{\infty}$ be a sequence of identically distributed random vectors with marginal distributions of z_t and ε_t being absolutely continuous such that*

a) $E[\varepsilon_t|\mathcal{I}_t] = 0$ with $\mathcal{I}_t = \{x_{t-j}, z_{t-j}\}_{j\in\mathbb{N}_0}$;

b) the true slope parameters $\beta^0 = (\beta_1^{0\prime}, \beta_2^{0\prime})'$ are such that $\beta_1^0 \neq \beta_2^0$ and $\beta^0 \in \mathcal{B}$, which is assumed to be a compact subset of $\mathbb{R}^{2p}$;

c) the infimum of eigenvalues of $E[x_t x_t'|z_t \in I_z]$ taken across all intervals $I_z \subseteq \mathbb{R}$, $P(z_t \in I_z) \geqslant \delta$, is positive for any $\delta > 0$ and $E[x_t x_t'|z_t \in I_z]$ is continuous with respect to the bounds of I_z.

Further, let $\mathcal{W}$ denotes the space of measurable functions $w : \mathbb{R} \to \mathbb{R}$ that are continuous up to a finite number of points $s_1, \ldots, s_J \in \mathbb{R}$, are uniformly bounded, $\sup_{z\in\mathbb{R}} |w(z)| < M < +\infty$, and are differentiable (from left or right at points $s_1, \ldots, s_J$) with derivatives uniformly bounded by M such that all $w \in \mathcal{W}$ satisfy

d) there exist intervals (a_1, b_1), $P(z_t \in (a_1, b_1)) > 0$, and (a_2, b_2), $P(z_t \in (a_2, b_2)) > 0$, such that $w(z_t) = 0$ for $z_t \in (a_1, b_1)$ and $w(z_t) = 1$ for $z_t \in (a_2, b_2)$.

Assumption 1.a claims that $\{\varepsilon_t\}_{t=1}^{\infty}$ is a martingale difference sequence with respect to the σ-field $\mathcal{I}_t$ generated by the current and past values of (x_t, z_t). This condition guarantees the conditional mean of y_t is correctly represented by the regression function in model (3). Condition 1.b requires the slope coefficients to be different in the two regimes: otherwise, it is not possible to distinguish the regimes and to identify the transition function (i.e., (7) would always equal zero if $\beta_1 = \beta_2$). The full-rank Assumption 1.c is similar to usual assumptions in the threshold and structural-break models for identification (e.g., Assumption A2 in Bai and Perron, 1998) and it can be seen as a weaker form of Assumption 1.7 in Hansen (2000), for instance: $E(x_t x_t' | z_t = z) > 0$, which is sufficient for solving (7) and reduces to $E(x_t x_t') > 0$ if x_t is independent of z_t. (Note that only $(\beta_2^0 - \beta_1^0)' E(x_t x_t' | z_t \in I_z)(\beta_2^0 - \beta_1^0) > 0$ is strictly necessary, see (7), but we impose the positive definiteness as β_1^0 and β_2^0 are generally unknown.) The full-rank condition is imposed for any interval I_z with non-zero probability of $z_t \in I_z$ to identify the transition function $w^0(z_t)$ almost everywhere. If the aim is to identify only slopes β_1 and β_2, much weaker assumption has to hold: two matrices $E[x_t x_t' | z_t \in (a_1, b_1)]$ and $E[x_t x_t' | z_t \in (a_2, b_2)]$ have to be non-singular, where the intervals are defined in Assumption 1.d.

Assumption 1 also defines the space of functions $\mathcal{W}$, where the transition function is searched for. Although we assume differentiability of the functions, which will be necessary later to derive the asymptotic distribution, assuming that functions w are Lipschitz (within the intervals of continuity) uniformly on $\mathcal{W}$ would be sufficient. Moreover, note that – without left or right continuity (or differentiability) of functions at the points of discontinuity – the identification of w^0 would not be possible at those points.

Finally, Assumption 1.d ensures that the system described by model (3) is with a positive probability in the first regime described by β_1^0 (when $z_t \in (a_1, b_1)$) and in the second regime defined by β_2^0 (when $z_t \in (a_2, b_2)$). On the one hand, this is an identification assumption for w: if there are instead constants $c_1 < c_2$ such that $w(z_t) = c_1$ for $z_t \in (a_1, b_1)$ and $w(z_t) = c_2$ for $z_t \in (a_2, b_2)$ in model (3), Assumption 1.d is satisfied in model (3) for parameter vectors $\widetilde{\beta}_1^0 = \beta_1^0(1 - c_1) + \beta_2^0 c_1$ and $\widetilde{\beta}_2^0 = \beta_1^0(1-c_2) + \beta_2^0 c_2$ instead of original β_1^0 and β_2^0. On the other hand, Assumption 1.d is essential for identification of the slope parameters β_1 and β_2 because they are not identifiable by using other values of z_t alone due to further unspecified $w(z_t)$. More specifically, for $\{z_t : w(z_t) \neq 0 \text{ or } 1\}$, the first-order condition (5) with respect to β_1 is equal to the first-order condition (6) taken with respect to β_2 multiplied by a scalar factor $\{1 - w(z_t)\}/w(z_t)$. Although practical difference is likely negligible, this assumptions excludes the LSTAR and ESTAR models as their transition functions never reach 0 and 1. The SETR analog of LSTAR would be based on the assumption that $w(z_t) = 0$ if $z_t < b_1$, $P(z_t \in (-\infty, b_1)) > 0$, and $w(z_t) = 1$ if $z_t > a_2$, $P(z_t \in (a_2, +\infty)) > 0$. (Analogously to common practice in the structural-break estimation, one could thus set that z_t below its αth quantile and above its $(1-\alpha)$th quantile correspond to the first and second regime, respectively.) Similarly, the SETR analog of ESTAR would hinge on the assumption that $w(z_t) = 0$ if $|z_t| < b_1$, $P(z_t \in (-b_1, b_1)) > 0$, and $w(z_t) = 1$ if $|z_t| > a_2$, $P(z_t \in (-\infty, -a_2) \cup (a_2, +\infty)) > 0$.

The identification result is stated in the following theorem. Note that the transition function is identified only up to a set with $f_z(z) = 0$ (f_z being the density of z_t), that is, the minimum of the LS criterion (4) is attained at β^0 and any function w such that $\|w - w^0\|_{\infty,\epsilon} = 0$ for any $\epsilon > 0$.

Theorem 1. *If $\{y_t, x_t, z_t\}$ follow model (3) and Assumption 1 is satisfied, then (β^0, w^0) are uniquely identified in $\mathcal{B} \times \mathcal{W}$ (up to a set with zero density in the case of w^0): it holds for any $\delta > 0$ and $\epsilon > 0$ that*

$$\inf_{\|\beta-\beta^0\|>\delta \text{ or } \|w-w^0\|_{\infty,\epsilon}>\delta} E[y_t - x_t'\beta_1 - x_t'(\beta_2-\beta_1)w(z_t)]^2 > E[y_t - x_t'\beta_1^0 - x_t'(\beta_2^0-\beta_1^0)w^0(z_t)]^2, \tag{8}$$

where $\beta \in \mathcal{B}$ and $w \in \mathcal{W}$.

Although Theorem 1 establishes that the slopes and transition function can be found by minimizing the (nonlinear) least squares criterion, the joint minimization with respect to $\beta = (\beta_1', \beta_2')'$ and w is computationally cumbersome (see Section 3 for details). We therefore design an iterative algorithm that requires only linear least squares estimation. Let us introduce the basic notation and concepts for this algorithm.

First, given some parameter values $\beta \in \mathbb{R}^{2p}$, the LS criterion (4) can be minimized with respect to w or the first-order condition (7) can be solved to obtain value $w(z_t)$ at $z_t = z$. Although we do not assume $E(x_t x_t' | z_t = z) > 0$, Assumption 1.c guarantees $E(x_t x_t' | z_t \in I_z) > 0$ for any interval $I_z \ni z$ with length $|I_z| > 0$. Equation (7) can be thus used conditionally on $z_t \in I_z$ (instead of $z_t = z$) to solve for $w(z)$ if $|I_z| \to 0$ and $w(z)$ is continuous in I_z (the derivatives of $w(z)$ are uniformly bounded). This solution of (7) for a given β will be denoted

$$w(z, \beta) = \lim_{|I_z| \to 0} \frac{E[x_t'(\beta_2 - \beta_1)(y_t - x_t'\beta_1) | z_t \in I_z]}{E[\{x_t'(\beta_2 - \beta_1)\}^2 | z_t \in I_z]}. \tag{9}$$

On the other hand, given some transition function w, the slope estimates of parameters β can be estimated by minimizing the LS criterion (4) with respect to β only or solving the unconditional counterpart of (5)–(6) for β. Considering a given w and using abbreviated notation $\omega_t = [1 - w(z_t), w(z_t)]'$, the LS estimate of β given w minimizes $E[y_t - (\omega_t \otimes x_t)'\beta]^2$ and it can be denoted and expressed as

$$\beta(w) = \{E[(\omega_t \otimes x_t)(\omega_t \otimes x_t)' | z_t]\}^{-1} E[(\omega_t \otimes x_t) y_t | z_t] \tag{10}$$

since $x_t'\beta_1\{1 - w(z_t)\} - x_t'\beta_2 w(z_t) = (\omega_t \otimes x_t)'\beta$. According to Theorem 1, it holds that $\beta^0 = \beta(w^0)$ and $\|w^0(z) - w(z, \beta^0)\|_{\infty,\epsilon} = 0$ for any $\epsilon > 0$.

3 Estimation

Before discussing the estimation method, let $\widehat{\beta}_T$ and $\widehat{w}_T(\cdot)$ denote the unconditional estimators of β^0 and $w^0(\cdot)$ that minimize the sum of squared residuals ($\beta = (\beta_1', \beta_2')'$):

$$\min_{\beta, w} \sum_{t=1}^{T} \left\{ y_t - x_t'\beta_1 - x_t'(\beta_2 - \beta_1) w(z_t) \right\}^2 . \tag{11}$$

Similarly, let $\widehat{\beta}_T(w)$ and $\widehat{w}_T(\cdot, \beta)$ denote the conditional estimators of $\beta(w)$ in (10) and $w(\cdot, \beta)$ in (9) given a fixed w and a fixed β, respectively.

Estimating the slope coefficients β and transition function w through direct minimization in (11) is intractable in practice. One common strategy in regime-switching models is concentration (e.g., see Hansen, 2000, for the TAR model and Leybourne et al., 1998, for the STAR model). Given fixed β, the semiparametric transition model in (3) can be viewed as a varying-coefficient model. Applying a nonparametric estimator of the varying-coefficient literature (see Fan and Zhang, 2008, for a review) yields the conditional estimators $\widehat{w}_T(z_1, \beta), \ldots \widehat{w}_T(z_T, \beta)$. The $2p$ slope coefficients are then estimated via minimizing the concentrated sum of squared residuals:

$$\widehat{\beta}_T = \arg\min_{\beta} \sum_{t=1}^{T} \left\{ y_t - x_t'\beta_1 - x_t'(\beta_2 - \beta_1) \widehat{w}_T(z_t, \beta) \right\}^2 . \tag{12}$$

This is however computationally demanding and could be difficult if p is large.

Instead of this traditional concentration approach, we propose an iterative estimation algorithm. Based on Assumption 1.d, an initial consistent slope estimator $\widehat{\beta}_T^{(0)}$ is constructed by using the data that are purely from the first and second regimes. Then the sum of squared residuals given $\beta = \widehat{\beta}_T^{(0)}$ is minimized locally (in neighborhoods of points $z_1, \ldots, z_T$) to obtain the corresponding initial estimator $\widehat{w}_T^{(0)} = \widehat{w}_T(\cdot, \widehat{\beta}_T^{(0)})$ of the transition function. Next, the slope estimate is updated to $\widehat{\beta}_T^{(1)} = \widehat{\beta}_T(\widehat{w}_T^{(0)})$ by minimizing the sum of squared residuals given the initial estimate $w = \widehat{w}_T^{(0)}$, and similarly, the transition-function estimate can be updated to

$\widehat{w}_T^{(1)} = \widehat{w}_T(\cdot, \widehat{\beta}_T^{(1)})$. The procedure can be iterated by estimating $\widehat{\beta}_T^{(k)} = \widehat{\beta}_T(\widehat{w}_T^{(k-1)})$ and $\widehat{w}_T^{(k)} = \widehat{w}_T(\cdot, \widehat{\beta}_T^{(k)})$ for $k = 2, 3, \ldots, K$. In practice, we used $K = 2$: given that the initial estimates $\widehat{\beta}_T^{(0)}$ and $\widehat{w}_T^{(0)}$ are not very precise, $\widehat{\beta}_T^{(2)}$ is the first slope estimate based on an iterated and presumably more precise estimate $\widehat{w}_T^{(1)}$ of the transition function. This delivers fast estimation and consistent and asymptotically normal estimator as shown later in Section 4.

In the rest of Section 3, we discuss first the choice of the initial slope estimator $\widehat{\beta}_T^{(0)}$ in Section 3.1, then the local nonparametric estimation of $\widehat{w}_T(\cdot, \beta)$ in Section 3.2, and finally iterated LS estimator $\widehat{\beta}_T(w)$ in Section 3.3.

3.1 Initial estimator of β

As the regions of the first and second regimes are assumed to be known, simple consistent initial estimators $\widehat{\beta}_{1,T}^{(0)}$ and $\widehat{\beta}_{2,T}^{(0)}$ of β_1 and β_2 can be obtained by employing the ordinary LS method in the regions of the first and second regimes, respectively. For example, a researcher can assume the observations with $z_t < q_z(\alpha)$ and $z_t > q_z(1-\alpha)$ follow purely the first and second regimes, respectively, where $q_z(\alpha)$ denotes the αth quantile of the z_t distribution. As the researcher might be willing to assume this only for a rather small α to avoid misspecification and there would thus be only small numbers of observations in each regime, the initial estimators would be very imprecise. In general, the same argument holds for any choice of intervals (a_1, b_1) and (a_2, b_2) in Assumption 1.d that are assumed to be very short.

Given Theorem 1, we suggest the following improvement of the simple initial estimator $\widehat{\beta}_{1,T}^{(0)}$ and $\widehat{\beta}_{2,T}^{(0)}$ described in the previous paragraph. Starting from short intervals $(a_1^0, b_1^0) \subseteq (a_1, b_1)$ and $(a_2^0, b_2^0) \subseteq (a_2, b_2)$, construct increasing sequences of intervals $(a_j^0, b_j^0) \subset (a_j^1, b_j^1) \subset \ldots \subset (a_j^\kappa, b_j^\kappa)$ for $j = 1, 2$. For each pair of intervals (a_1^k, b_1^k) and (a_2^k, b_2^k), $k = 1, \ldots, \kappa$, estimate $\widehat{\beta}_{1,T}^{(0,k)}$ and $\widehat{\beta}_{2,T}^{(0,k)}$, forming estimate $\widehat{\beta}_T^{(0,k)}$, compute the transition function $\widehat{w}_T^{(0,k)} = \widehat{w}_T(\cdot, \widehat{\beta}_T^{(0,k)})$, and evaluate the sum S_k^2 of least squares (11). Then define the initial estimate by $\widehat{\beta}_{1,T}^{(0)} = \widehat{\beta}_{1,T}^{(0,\hat{k})}$ and $\widehat{\beta}_{2,T}^{(0)} = \widehat{\beta}_{2,T}^{(0,\hat{k})}$

for $\widehat{k} = \arg\min_{k=0,\ldots,\kappa} S_k^2$, that is, the estimate minimizing the unconditional LS criterion.

The benefit of the described procedure is that the estimation becomes insensitive to the choice of the initial intervals (a_1^0, b_1^0) and (a_2^0, b_2^0). On the one hand, choosing too short initial intervals $(a_1^0, b_1^0) \subsetneq (a_1, b_1)$ and $(a_2^0, b_2^0) \subsetneq (a_2, b_2)$, where (a_1, b_1) and (a_2, b_2) are the longest intervals satisfying Assumption 1.d, does not affect the estimate precision much since longer intervals (a_1^k, b_1^k) and (a_2^k, b_2^k), $k > 1$, are considered as well and the best fit is chosen. On the other hand, including long intervals that do not satisfy Assumption 1.d, $(a_j^\kappa, b_j^\kappa) \supsetneq (a_j, b_j)$, $j = 1, 2$, does not affect the consistency of this procedure as is verified later in Theorem 2 in Section 4.

3.2 Local linear estimator of $w(\cdot, \beta)$

Given $\beta = (\beta_1', \beta_2')'$ with $\beta_1 \neq \beta_2$, the semiparametric transition model (3) can be reformulated as a varying-coefficient model with a single covariate and no intercept:

$$\widetilde{y}_t = y_t - x_t'\beta_1 = x_t'(\beta_2 - \beta_1)w(z_t, \beta) = \widetilde{x}_t m(z_t) + \varepsilon_t, \tag{13}$$

where $\widetilde{y}_t = y_t - x_t'\beta_1$, $\widetilde{x}_t = x_t'(\beta_2 - \beta_1)$, and $m(z_t) = w(z_t, \beta)$.

In the case of a smooth varying-coefficient function $m(\cdot)$, a number of estimators are described in the literature. There are three main approaches to estimate smooth $m(\cdot)$: kernel local polynomial smoothing (e.g., Wu et al., 1998; Fan and Zhang, 1999), polynomial splines (e.g., Huang et al., 2002, 2004), or spline smoothing (e.g., Hoover et al., 1998). In this paper, we opt for the local constant and local linear smoothing method. The local linear estimator $\widehat{m}_T(z)$ of $m(z)$ is the a-minimizer of

$$\min_{a\in\mathbb{R}, b\in\mathbb{R}} \sum_{t=1}^{T} [\widetilde{y}_t - \widetilde{x}_t\{a + b(z_t - z)\}]^2 K_h(z_t - z), \tag{14}$$

where $K_h(v) = K(v/h_T)/h_T$, $K(v)$ is a symmetric kernel function, and h_T is the bandwidth. The local constant estimator corresponds to (14) without term $b(z_t - z)$.

Solving the first-order conditions of (14) leads to

$$\widehat{m}_T(z) = (1,0)\left\{\sum_{t=1}^{T} \dot{x}_t \dot{x}_t' K_h(z_t - z)\right\}^{-1} \sum_{t=1}^{T} \dot{x}_t \widetilde{y}_t K_h(z_t - z) \tag{15}$$

with vector $\dot{x}_t = [\widetilde{x}_t, \widetilde{x}_t(z_t - z)]'$. Analogously, the local constant estimator can be expressed in the form (15) using $\widetilde{x}_t$ instead of $\dot{x}_t$.

Although the local linear smoother is sufficient for consistent estimation of the slope parameters even if the transition function is discontinuous at a finite number of points (see Section 4), the estimation of the transition function will possibly suffer. Unfortunately, there is a rather limited research on the nonparametric estimation of piecewise continuous functions with jumps in the context of varying-coefficient models. In this work, we employ the generalization of the nonparametric estimation procedure for discontinuous function that was originally designed for nonparametric regression by Gijbels et al. (2007) and that was generalized to the varying-coefficient models by Čížek and Koo (2014). Its short description follows.

Let the conventional kernel function be $K^c(v) = K(v)$, where $K(v)$ is a symmetric kernel with support $[-1, 1]$, and the left-side and right-side kernels be $K^l(v) = K(v) \cdot I(v \in [-1, 0))$ and $K^r(v) = K(v) \cdot I(v \in (0, 1])$, respectively. Using these three kernels, there are three sets of local linear (or local constant) estimates of $m(z)$ and their derivatives $m'(z)$:

$$[\widehat{a}^j(z), \widehat{b}^j(z)] = \arg\min_{a,b} \sum_{t=1}^{T} [\widetilde{y}_t - \widetilde{x}_t\{a + b(z_t - z)\}]^2 K_h^j(z_t - z), \quad j \in \{l, r, c\}, \tag{16}$$

where superscripts l, r, c indicate whether the left-, right-, or two-sided h_T-neighborhood of z is used. The goodness of fit of the three estimates can be measured by their weighted residual mean squares (WRMSs) defined by

$$\text{WRMS}^j(z) = \frac{\sum_{t=1}^{T} [\widetilde{y}_t - \widetilde{x}_t\{\widehat{a}^j + \widehat{b}^j(z_t - z)\}]^2 K_h^j(z_t - z)}{\sum_{t=1}^{T} K_h^j(z_t - z)}, \quad j \in \{l, r, c\}. \tag{17}$$

If $m(z_t)$ is continuous at $z_t = z$, all three WRMSs are consistent estimates of $E[\varepsilon_t^2|\widetilde{x}_t, z_t]$, while WRMS^l is the only consistent estimate in the left h_T-neighbourhood of a jump point and WRMS^r is the only consistent estimator in the right h_T-neighborhood of a jump point (cf. Proposition 2.2 in Gijbels et al., 2007, and Theorem 3 in Čížek and Koo, 2014). With this idea, the estimator of the varying coefficient $m(\cdot)$ is defined by

$$\widehat{m}_T(z) = \begin{cases} \widehat{a}^c(z), & \text{if diff}(z) \leqslant u_T, \\ \widehat{a}^l(z), & \text{if diff}(z) > u_T \text{and WRMS}^l(z) < \text{WRMS}^r(z), \\ \widehat{a}^r(z), & \text{if diff}(z) > u_T \text{and WRMS}^l(z) > \text{WRMS}^r(z), \\ \dfrac{\widehat{a}^l(z) + \widehat{a}^r(z)}{2}, & \text{if diff}(z) > u_T \text{and WRMS}^l(z) = \text{WRMS}^r(z), \end{cases} \tag{18}$$

where $\text{diff}(z) = \text{WRMS}^c(z) - \min\{\text{WRMS}^l(z), \text{WRMS}^r(z)\}$ and the threshold value $u_T > 0$ is such that $u_T \to 0$ as $T \to \infty$. Parameter u_T can be determined along with h_T, for example, by the least-squares cross-validation (e.g., Yao and Tong, 1998).

3.3 Least squares estimator of $\beta(w)$

Given some transition function w, the semiparametric transition model is linear in the slope parameter β. Hence, the ordinary LS estimation can be directly applied. Denote $\omega_t = [1 - w(z_t), w(z_t)]'$. Similarly to (10), the sum of squared residuals $T^{-1} \sum_{t=1}^{T} \{y_t - (\omega_t \otimes x_t)'\beta\}^2$ is minimized with respect to β (with w fixed), which yields the conditional LS estimator

$$\widehat{\beta}_T(w) = \left\{ \frac{1}{T} \sum_{t=1}^{T} (\omega_t \otimes x_t)(\omega_t \otimes x_t)' \right\}^{-1} \frac{1}{T} \sum_{t=1}^{T} (\omega_t \otimes x_t) y_t. \tag{19}$$

4 Asymptotic properties

In the asymptotic analysis, we consider absolutely regular time series and transition functions from $\mathcal{W}$ constrained to piecewise smooth functions.

First, the definition of an absolutely regular (or β-mixing) process is given. Consider a strictly stationary process $\{X_t\}_{t=1}^{\infty}$ and let $\mathcal{F}_k^l$ be the σ-algebra generated by $\{X_t\}_{t=k}^{l}$. The β-mixing coefficients are defined by

$$\beta(m) = \sup_{t \in \mathbb{N}} E[\sup_{A \in \mathcal{F}_{t+m}^{\infty}} |P(A|\mathcal{F}_1^t) - P(A)|].$$

If $\lim_{m \to \infty} \beta(m) = 0$, the process $\{X_t\}_{t=1}^{\infty}$ is called β-mixing or absolutely regular.

Next, we define the class of smooth functions $C_M^{\gamma}(\mathcal{X})$ on a bounded set $\mathcal{X} \subset \mathbb{R}^d$ following van der Vaart and Wellner (1996, p. 154); see also Ichimura and Lee (2010). Let $\underline{\gamma}$ be the largest integer smaller than γ, and for any vector $k = (k_1, \ldots, k_d) \in \mathbb{N}^d$, let the differential operator $D^k = \frac{\partial^{|k|}}{\partial x_1^{k_1} \ldots \partial x_d^{k_d}}$ for $|k| = \sum_{i=1}^{d} k_i$. Additionally, define the function norm

$$\|f\|_{\gamma} = \max_{|k| \leqslant \underline{\gamma}} \sup_{x} |D^k f(x)| + \max_{|k| = \underline{\gamma}} \sup_{x \neq x'} \frac{|D^k f(x) - D^k f(x')|}{\|x - x'\|^{\gamma - \underline{\gamma}}},$$

where the suprema are taken over all x and x' in the interior of $\mathcal{X}$. Then $C_M^{\gamma}(\mathcal{X})$ is the set of all continuous functions $f : \mathcal{X} \mapsto \mathbb{R}$ with $\|f\|_{\gamma} \leqslant M$.

To show the consistency of the estimators proposed in Section 3, the following assumptions are introduced.

Assumption 2. *Let the random variables x_t, z_t, ε_t and random vectors $v_t = (v_{1t}, v_{2t}, z_t)'$ with v_{1t} and v_{2t} representing any element of vectors x_t and $(x_t', \varepsilon_t)'$, respectively, satisfy the following conditions:*

a) process $\{x_t, z_t, \varepsilon_t\}_{t=1}^{T}$ is strictly stationary and absolutely regular with β-mixing coefficients $\beta(m)$, $m \in \mathbb{N}$, such that $\beta(m) = o(m^{-(2+\xi)/\xi})$ as $m \to \infty$ for some $\xi > 0$.

b) the following moments are finite: $E\|x_t x_t^\top\|^{2+\xi} < \infty$, $E\|\varepsilon_t x_t\|^{2+\xi} < \infty$, $E|z_t|^{2+\xi} < \infty$, *and* $E|\varepsilon_t|^{2+\xi} < \infty$, *where* ξ *is given in 2.a.*

c) assuming that the support $\mathcal{Z}$ *of* z_t *is partitioned,* $\mathcal{Z} = \bigcup_{j=1}^{\infty} I_j$, *into bounded, convex sets with nonempty interiors, the space* $\mathcal{W}$ *of transition functions contains only piecewise continuous functions such that, after restricting them to* I_j, $\mathcal{W}_{|I_j}$ *belongs to* $C_M^{\gamma}(I_j)$ *for some* $\gamma > 3$ *and* $j \in \mathbb{N}$.

d) finally, let $\sum_{j,k,l=1}^{\infty} \max\{\lambda(I_{jkl}^3), 1\} \cdot \max I_{jkl}^3 \cdot Q^{[(1+\delta)(3+\xi)]^{-1}}(I_{jkl}^3)$ *be finite for some* $\delta > 0$, *where the partition of* $\mathbb{R}^3 = \bigcup_{j,k,l=1}^{\infty} I_{jkl}^3$ *is defined by* $I_{jkl}^3 = I_j \times [k, k+1) \times [l, l+1)$, $\lambda(I_{jkl}^3)$ *denotes the Lebesque measure of* I_{jkl}^3, $\max I_{jkl}^3 = \sup_{v=(v_1,v_2,v_3)' \in I_{jkl}^3} \max\{|v_1|, |v_2|, |v_3|\}$, *and* $Q(I_{jkl}^3) = P(v_t \in I_{jkl}^3)$.

If $\{x_t, z_t, \varepsilon_t\}_{t=1}^T$ is a series of independent random vectors, Assumption 2.a is automatically fulfilled. Under dependence, the stationarity condition in Assumption 2.a excludes time trends and integrated processes. Additionally, the mixing condition in Assumption 2.a controls the degree of dependence in sequence $\{x_t, z_t, \varepsilon_t\}_{t=1}^T$ and is a standard assumption to guarantee the validity of the stochastic limit theorems. Sufficient conditions such that the nonlinear autoregressive models (which contain the TAR, STAR, and the semiparametric transition model for many transition functions w) are geometrically ergodic and thus β-mixing under Assumption 2.b can be found in Chen and Tsay (1993) and Meitz and Saikkonen (2010), for instance.

Furthermore, Assumption 2.b imposes that a sufficient number of moments exists. Assumption 2.b together with 2.a and 1.b are essential to guarantee the validity of the law of large numbers (LLN) and the central limit theorem (CLT) for dependent sequences (e.g., Arcones and Yu, 1994 and Davidson 1994, Section 24.4). Assumption 2.c defines a class of functions such that LLN can be applied uniformly to this class of functions (cf. van der Vaart and Wellner, 1996, Sections 2.7 and 2.8). The transition functions have to be piecewise smooth and at least three times differentiable in the continuity regions. Finally, Assumption 2.d is a technical assumption used again for the uniform LLN. It does not restrict variables with a bounded support, which are commonly used (or imposed by means of trimming) in semiparametric literature. For variables with infinite support, it requires that the probability of observing large values are small. To facilitate an easy understanding, consider the univariate equivalent of Assumption 2.d: $\sum_{j=1}^{\infty} \max\{\lambda(I_j), 1\} \cdot \max I_j \cdot Q^{[\delta(3+\xi)]^{-1}}(I_j)$. As intervals I_j can be chosen of the maximum length 1 without loss of generality, the sum is bounded by $\sum_{j=1}^{\infty} |j+1| \cdot \{Q^{[(1+\delta)(3+\xi)]^{-1}}([j, +\infty)) + Q^{[(1+\delta)(3+\xi)]^{-1}}([-\infty, -j])\}$. Considering case of small $\xi > 0$ so that $(1+\delta)(3+\xi) < 3.5$, this bound is finite if the distribution of random variable v_t has tails decreasing to zero proportionally to or faster than $1/j^7$, for instance. This assumption can be further weakened (along with the order of differentiability) if the error term ε_t is independent of transition variable z_t.

The following theorem establishes the consistency of the unconditional estimators. This guarantees that minimizing the LS criterion (11) with respect to both β and w leads to consistent estimates.

Theorem 2. *Under Assumptions 1 and 2, it holds that* $\widehat{\beta}_T - \beta^0 \xrightarrow{P} 0$, $\|\widehat{w}_T - w^0\|_{\infty,\epsilon} \to 0$ *for any* $\epsilon > 0$, *and* $E\{\widehat{w}_T(z_t) - w^0(z_t)\}^2 \to 0$ *as* $T \to +\infty$.

Since the estimation procedure suggested in Section 3 estimates the regression coefficients β given an estimate of the transition function w and vice versa, it is necessary to impose some conditions on the nonparametric estimator of $w(\cdot, \beta)$ in (13) in order to derive the asymptotic distribution of the slope parameters.

Assumption 3. *Let* $\zeta_T > 0$ *such that* $\zeta_T \to 0$ *as* $T \to +\infty$, $\mathcal{Z}_T^c$ *be a subset of the support* $\mathcal{Z}$ *of* z_t *excluding all* ζ_T*-neighborhoods of discontinuities* $\{s_j\}_{j=1}^J$, $\mathcal{Z}_T^c = \mathcal{Z} \setminus \bigcup_{j=1}^J [s_j - \zeta_T, s_j + \zeta_T]$, *and* $U(\beta^0, \delta) = \{\beta \in \mathcal{B} : \|\beta - \beta^0\| < \delta\}$. *It is assumed that there exist* $\delta > 0$ *such that, for all* $\beta \in U(\beta^0, \delta)$ *and any* $0 < \widetilde{\delta} < \delta$,

a) $P\{\widehat{w}_T(z, \beta) \in \mathcal{W}\} \to 1$ *as* $T \to +\infty$;

b) estimator $\widehat{w}_T(z,\beta)$ is uniformly bounded on $\mathcal{Z}\times\mathcal{B}$ and uniformly consistent on $\mathcal{Z}_T^c$: $\sup_{z\in\mathcal{Z}_T^c}|\widehat{w}_T(z,\beta)-w_T(z,\beta)|\xrightarrow{P}0$ as $T\to+\infty$ for any $\beta\in U(\beta^0,\delta)$;

c) estimator $\widehat{w}_T(z,\beta)$ is stochastically equicontinuous in β on $\mathcal{Z}_T^c$: $\sup_{z\in\mathcal{Z}_T^c}\sup_{\beta\in U(\beta^0,\delta)}\sup_{\widetilde{\beta}\in U(\beta,\widetilde{\delta})}|\widehat{w}_T(z,\beta)-\widehat{w}_T(z,\widetilde{\beta})|\xrightarrow{P}0$ as $T\to+\infty$;

d) function $w(z,\beta)$ has a uniformly bounded derivative with respect to $\beta\in U(\beta^0,\delta)$: $\sup_{z\in\mathcal{Z}_T^c}\sup_{\beta\in U(\beta^0,\delta)}\|\partial w(z,\beta)/\partial\beta\|<\infty$;

e) the density of z_t is bounded on $\mathcal{Z}$.

While Assumptions 3.d and 3.e are additional regularity conditions, Assumptions 3.a–3.c are relevant to the properties of the conditional estimator of the transition estimator. As mentioned in Section 3, general nonparametric estimators $\widehat{w}_T(\cdot,\beta)$ of univariate varying-coefficient model (13) are considered, where the response variable $\widetilde{y}_t = y_t - x_t'\beta_1$ and explanatory variables $\widetilde{x}_t = x_t'(\beta_2-\beta_1)$ for fixed β_1 and β_2. First, the estimates are supposed to converge to a function from the function space $\mathcal{W}$ in Assumption 3.a as is common in semiparametric literature (e.g., Ichimura and Lee, 2010). Next, Assumptions 3.b requires the nonparametric estimator to be uniformly consistent. This condition is typically satisfied on compact subsets of $\mathbb{R}$, but can be extended to $\mathbb{R}$ for bounded functions. For the jump-preserving varying-coefficient estimator introduced in Section 3.2, Assumption 3.b is verified by Čížek and Koo (2014, Theorem 4). Finally, the nonparametric estimator $\widehat{w}_T(\cdot,\beta)$ is required to be stochastically equicontinuous by Assumption 3.c similarly to Ichimura and Lee (2010), who argue that this restriction holds for estimators continuously differentiable in $\beta\in U(\beta^0,\delta)$. Note that the estimator depends on β only via linear transformations $\widetilde{y}_t$ and $\widetilde{x}_t$.

In the following theorems, the consistency and asymptotic distribution of the estimator proposed in Section 3 will be derived. The estimation starts with an estimate $\widehat{\beta}_T^{(0)}$, which is consistent either by Assumption 1.d if one pair of intervals is used or by Theorem 2 otherwise. Based on a consistent estimator $\breve{\beta}_T$ such as $\widehat{\beta}_T^{(0)}$

or any subsequent iterations $\widehat{\beta}_T^{(k)}$, the transition function is estimated by $\widehat{w}_T(\cdot, \breve{\beta}_T)$, which is shown to be asymptotically equivalent to infeasible $\widehat{w}_T(\cdot, \beta^0)$.

Theorem 3. *If Assumptions 1–3 hold and* $\breve{\beta}_T \xrightarrow{P} \beta^0$, *then* $\sup_{z \in \mathcal{Z}_T^c} |\widehat{w}_T(z, \breve{\beta}_T) - \widehat{w}_T(z, \beta^0)| \xrightarrow{P} 0$ *and* $E[\widehat{w}_T(z, \breve{\beta}_T) - \widehat{w}_T(z, \beta^0)]^2 \to 0$ *as* $T \to +\infty$.

An immediate consequence of Theorem 3 and Assumption 3.b is the weak consistency of $\widehat{w}_T(\cdot, \breve{\beta}_T)$. Note that the convergence in probability is in this case equivalent to the convergence in mean due to uniformly bounded functions w and $\widehat{w}_T$.

Corollary 1. *If Assumptions 1–3 hold and* $\breve{\beta}_T \xrightarrow{P} \beta^0$, $\sup_{z \in \mathcal{Z}_T^c} |\widehat{w}_T(z, \widehat{\beta}_T) - w(z, \beta^0)| \xrightarrow{P} 0$ *and* $E[\widehat{w}_T(z_t, \widehat{\beta}_T) - w(z_t, \beta^0)]^2 \to 0$ *as* $T \to +\infty$.

The next step of the estimation procedure is based on a consistent estimate $\breve{w}_T$ of the transition function such as $\widehat{w}_T^{(0)} = \widehat{w}_T(\cdot, \widehat{\beta}_T^{(0)})$ or later iterations $\widehat{w}_T^{(k)} = \widehat{w}_T(\cdot, \widehat{\beta}_T^{(k)})$: given the transition function, the slope parameters are estimated. To derive their consistency and limiting distribution, the matrices entering the asymptotic variance of the estimator have to introduced.

Assumption 4. *Let the covariance matrices*

$$Q^0 = E[(\omega_t^0 \otimes x_t)(\omega_t^0 \otimes x_t)'] \quad \textit{and} \quad V^0 = E[\varepsilon_t^2(\omega_t^0 \otimes x_t)(\omega_t^0 \otimes x_t)']$$

with $\omega_t^0 = [1 - w^0(z_t), w^0(z_t)]'$. *We assume* Q^0 *and* V^0 *to be finite and positive definite.*

Assumption 4 corresponds to the usual full-rank condition. With Assumptions 1–4, we first claim – similarly to Theorem 3 – that the difference between the slope estimator $\widehat{\beta}_T(\breve{w}_T)$ and the infeasible estimator $\widehat{\beta}_T(w^0)$ based on the true transition function w^0 converges to zero in probability at a rate faster than $T^{-1/2}$.

Theorem 4. *If Assumptions 1–4 hold and estimator* $\breve{w}_T$ *satisfies* $E[\breve{w}_T(z_t) - w^0(z_t)]^2 \to$

0, *then it holds for* $T \to +\infty$ *that*

$$\sqrt{T}(\widehat{\beta}_T(\breve{w}_T) - \widehat{\beta}_T(w^0)) \xrightarrow{P} 0.$$

Finally, the limiting distribution of the infeasible estimator $\widehat{\beta}_T(w^0)$ (assuming known w^0) is derived in Theorem 5, and by Theorem 4, this distribution describes asymptotically also the feasible estimator $\widehat{\beta}_T(\breve{w}_T)$.

Theorem 5. *Under Assumptions 1–4,*

$$\sqrt{T}\{\widehat{\beta}_T(w^0) - \beta^0\} \xrightarrow{d} N(0, Q^{0^{-1}}\ V^0 Q^{0^{-1}}).$$

The asymptotic variance of the infeasible and feasible estimators thus corresponds to the variance of the linear least-squares estimator of model (3) with a known transition w^0. In practice, the asymptotic variance in Theorem 5 can be estimated directly by taking the finite sample equivalents of Q^0 and V^0 since a consistent estimate of w^0 is obtained as a part of the estimation procedure. In particular, if the estimation stops after κ iterations, one can define $\widehat{\omega}_t = [1 - \widehat{w}_T(z_t, \widehat{\beta}_T^{(\kappa)}), \widehat{w}_T(z_t, \widehat{\beta}_T^{(\kappa)})]'$ and $\widehat{\varepsilon}_t = y_t - (\widehat{\omega}_t \otimes x_t)'\widehat{\beta}_T^{(\kappa)}$ and estimate Q^0 and V^0 by $\widehat{Q}_T = \frac{1}{T}\sum_{t=1}^T (\widehat{\omega}_t \otimes x_t)(\widehat{\omega}_t \otimes x_t)'$ and $\widehat{V}_T = \frac{1}{T}\sum_{t=1}^T \widehat{\varepsilon}_t^2 (\breve{\omega}_t \otimes x_t)(\breve{\omega}_t \otimes x_t)'$.

5 Simulation study

In this section, the performance of the proposed estimator of the semiparametric transition (SETR) model is evaluated by Monte Carlo simulations. Furthermore, these simulations provide a comparison with the existing parametric estimators of the TAR and LSTAR models.

Four different data generating processes (DGPs) are considered. All DGPs are based on the semiparametric transition model (3) and an autoregressive model of order 2:

$$y_t = [\beta_{1;0} + \beta_{1;1}y_{t-1} + \beta_{1;2}y_{t-2}]\{1 - w(z_t)\} + [\beta_{2;0} + \beta_{2;1}y_{t-1} + \beta_{2;2}y_{t-2}]w(z_t) + \varepsilon_t,$$

where errors $\varepsilon_t \sim N(0,1)$ are independent and identically distributed and the true values of the regression coefficients used in the simulation are $\beta_{1;0} = -0.25, \beta_{1;1} = 0.4, \beta_{1;2} = -0.6$ and $\beta_{2;0} = 0.25, \beta_{2;1} = -0.8, \beta_{2;2} = 0.2$. The functional forms of the weighting function $w(z_t)$ and their arguments are listed below ($U(0,1)$ denotes the uniform distribution on interval $[0,1]$):

DGP1a $w(z) = I(z > \tau)$ with $\tau = 0.4$ and $z_t = y_{t-2}$;

DGP1b $w(z) = I(z > \tau)$ with $\tau = 0.4$ and $z_t = t/T$, where $t = 1,\ldots,T$;

DGP2 $w(z) = [1 + \exp\{-\nu(z-\tau)\}]^{-1}$ with $\nu = 2$, $\tau = 0.4$, and $z_t = y_{t-2}$;

DGP3 $w(z) = 0.5[1 - \cos\{4\pi(z-0.1)\}]I(z \in [0.1, 0.85]) + I(z > 0.85)$ and $z_t \sim U(0,1)$ are independent and identically distributed;

DGP4 $w(z) = (z^{-1/2} - 1)I(z \in [0.2, 0.7]) + I(z > 0.7)$ and $z_t \sim U(0,1)$ are independent and identically distributed.

The DGP1a is a TAR model, where the transition function is piecewise constant with discontinuity at 0.4. Although the case of deterministic transition variable z_t is not in the focus of this paper, DGP1b replicates DGP1a for the case of z_t being time. The DGP2 corresponds to the standard LSTAR model, where the shape parameter $\nu = 2$ so that the logistic function is flat enough to be distinguished from the indicator function of DGP1. While DGP1 and DGP2 use the lagged dependent variable in the role of the transition variable, the last two DGP3 and DGP4 rely on a uniformly distributed transition variable independent of ε_s and y_{s-1}, $s \leqslant t$, and moreover, they are not nested in neither TAR, nor LSTAR models. The transition function in DGP3 is continuous and reaches both regimes two times (see Figure

3), whereas the transition function in DGP4 is discontinuous with two jumps (see Figure 4). In all cases, the order of the baseline autoregressive process is 2 and is assumed to be known.

For each data-generating process, 1000 samples of sizes $T = 200$, 400, and 800 are generated and estimated by the TAR, LSTAR, and the semiparametric transition procedure (SETR), where the weighting function is estimated by the local-constant estimator of varying-coefficient model (13) assuming continuity of w (SETR/C) or by the jump-preserving local-constant estimator of (13) designed by Čížek and Koo (2014) for piecewise smooth functions w with jumps (SETR/J). In both cases, the quartic kernel is used and the bandwidth h_T and parameter u_T in (18) are determined by least squares leave-one-out cross-validation. The proposed SETR estimation uses 4 initial estimators (for each of the two regimes), which are based on the data below the αth quantile and above the $(1-\alpha)$th quantile of the transition variable z_t for $\alpha = 0.05, 0.10, 0.20$, and 0.40. Furthermore, the estimation involves two iterations: (i) based on the initial estimates $\widehat{\beta}_T^{(0)}$, the initial weighting function $\widehat{w}_T^{(0)}$ is estimated; (ii) the LS estimate $\widehat{\beta}_T^{(1)}$ corresponding to $\widehat{w}_T^{(0)}$ is obtained and $\widehat{w}_T^{(1)}$ is computed given $\widehat{\beta}_T^{(1)}$; as the initial estimators $\widehat{\beta}_T^{(0)}$ are typically rather imprecise, the procedure is repeated again so that (iii) based on the estimates $\widehat{\beta}_T^{(1)}$ and $\widehat{w}_T^{(1)}$, the corresponding LS estimate $\widehat{\beta}_T^{(2)}$ and the weighting function $\widehat{w}_T^{(2)}$ are estimated and reported (see Section 3 for details). Regarding the TAR and LSTAR models, the transition parameters τ and ν are determined by a grid search. All estimates are summarized by means of their bias and mean squared error (MSE).

5.1 TAR results

The estimation results for the TAR model are summarized in Tables 1 and 2 for DGP1a and DGP1b, respectively; sample sizes cover $T = 200, 400$, and 800. The TAR and LSTAR estimates provide best and precise estimates as both correspond to the specified DGP: the grid for the transition parameter ν was reaching up to

Table 1: Biases and MSEs of all estimator for DGP1a and $T = 200, 400$, and 800.

		TAR		LSTAR		SETR/C		SETR/J	
T		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
200	$\widehat{\beta}_{1,0}$	0.001	0.142	-0.008	0.149	-0.007	0.257	0.016	0.201
	$\widehat{\beta}_{1,1}$	-0.009	0.078	-0.003	0.079	0.041	0.142	0.012	0.133
	$\widehat{\beta}_{1,2}$	-0.004	0.133	-0.006	0.137	0.038	0.189	0.029	0.165
	$\widehat{\beta}_{2,0}$	0.002	0.215	0.018	0.227	0.124	0.399	0.046	0.338
	$\widehat{\beta}_{2,1}$	0.005	0.072	0.000	0.073	-0.010	0.123	0.005	0.128
	$\widehat{\beta}_{2,2}$	-0.004	0.124	-0.010	0.127	-0.052	0.168	-0.023	0.148
400	$\widehat{\beta}_{1,0}$	-0.004	0.093	-0.009	0.095	-0.024	0.162	0.010	0.115
	$\widehat{\beta}_{1,1}$	-0.005	0.055	-0.002	0.055	0.049	0.110	0.007	0.088
	$\widehat{\beta}_{1,2}$	-0.004	0.091	-0.005	0.091	0.025	0.125	0.016	0.103
	$\widehat{\beta}_{2,0}$	0.008	0.149	0.014	0.150	0.136	0.287	0.029	0.214
	$\widehat{\beta}_{2,1}$	0.005	0.052	0.002	0.052	-0.017	0.090	0.003	0.083
	$\widehat{\beta}_{2,2}$	-0.004	0.083	-0.007	0.084	-0.048	0.118	-0.011	0.097
800	$\widehat{\beta}_{1,0}$	-0.001	0.066	-0.003	0.066	-0.027	0.110	0.012	0.075
	$\widehat{\beta}_{1,1}$	-0.001	0.038	0.000	0.038	0.045	0.090	-0.001	0.064
	$\widehat{\beta}_{1,2}$	-0.002	0.063	-0.003	0.063	0.017	0.084	0.010	0.068
	$\widehat{\beta}_{2,0}$	-0.003	0.102	-0.000	0.103	0.123	0.224	0.005	0.149
	$\widehat{\beta}_{2,1}$	0.002	0.035	0.001	0.034	-0.020	0.068	0.002	0.058
	$\widehat{\beta}_{2,2}$	-0.001	0.058	-0.001	0.058	-0.042	0.084	-0.002	0.066

Table 2: Biases and MSE of all estimator for DGP1b and $T = 200, 400$, and 800.

		TAR		LSTAR		SETR/C		SETR/J	
T		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
200	$\widehat{\beta}_{1,0}$	-0.001	0.116	-0.004	0.117	-0.002	0.129	0.007	0.127
	$\widehat{\beta}_{1,1}$	-0.005	0.077	0.001	0.079	0.000	0.136	-0.026	0.135
	$\widehat{\beta}_{1,2}$	-0.002	0.070	-0.006	0.071	-0.010	0.101	0.006	0.096
	$\widehat{\beta}_{2,0}$	0.003	0.100	0.005	0.101	-0.024	0.103	-0.007	0.102
	$\widehat{\beta}_{2,1}$	-0.009	0.089	-0.012	0.090	0.052	0.127	0.022	0.115
	$\widehat{\beta}_{2,2}$	-0.024	0.091	-0.027	0.092	0.012	0.105	-0.008	0.101
400	$\widehat{\beta}_{1,0}$	0.000	0.082	-0.002	0.082	-0.006	0.094	0.001	0.092
	$\widehat{\beta}_{1,1}$	-0.004	0.055	-0.001	0.056	0.007	0.100	-0.010	0.095
	$\widehat{\beta}_{1,2}$	0.002	0.050	0.000	0.050	-0.007	0.073	0.003	0.069
	$\widehat{\beta}_{2,0}$	0.007	0.071	0.008	0.071	-0.014	0.070	0.002	0.071
	$\widehat{\beta}_{2,1}$	-0.004	0.065	-0.005	0.065	0.041	0.088	0.013	0.075
	$\widehat{\beta}_{2,2}$	-0.012	0.066	-0.014	0.066	0.015	0.072	-0.004	0.069
800	$\widehat{\beta}_{1,0}$	-0.001	0.055	-0.002	0.056	-0.006	0.064	-0.001	0.063
	$\widehat{\beta}_{1,1}$	-0.001	0.040	0.001	0.040	0.009	0.072	-0.006	0.072
	$\widehat{\beta}_{1,2}$	0.000	0.036	-0.001	0.036	-0.008	0.053	0.001	0.052
	$\widehat{\beta}_{2,0}$	0.000	0.046	0.001	0.046	-0.017	0.048	-0.002	0.046
	$\widehat{\beta}_{2,1}$	-0.002	0.046	-0.003	0.046	0.031	0.061	0.007	0.050
	$\widehat{\beta}_{2,2}$	-0.007	0.046	-0.007	0.046	0.015	0.051	-0.002	0.048

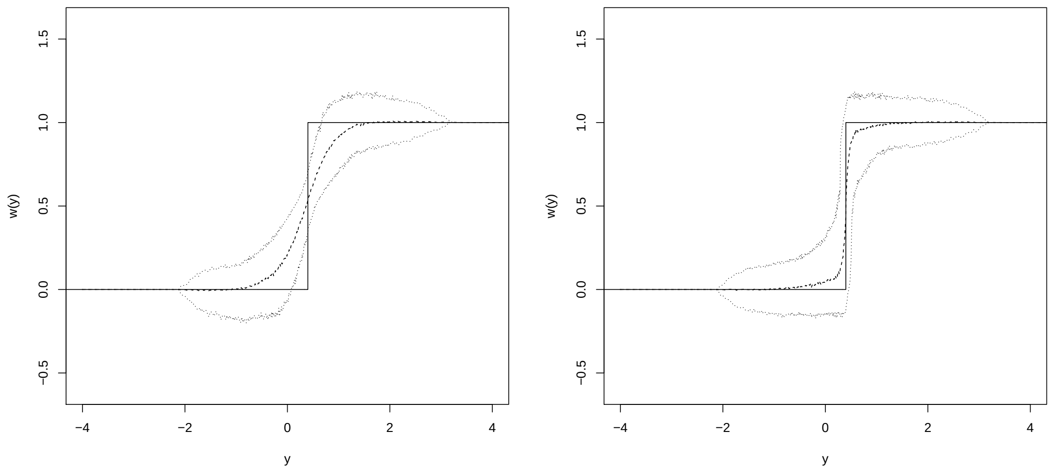

Figure 1: The mean estimates (dashed line) and 5% and 95% quantiles (dotted lines) of the transition function in DGP1a with $T = 400$; the solid line depicts the true transition function. The left and right panels correspond to SETR/C and SETR/J estimates, respectively.

$\nu = 1000$ and the logistic transition function can thus became numerically identical to the discontinuous transition of TAR. Regarding the SETR estimation, both SETR/C and SETR/J provide consistent estimates in the sense that the biases and mean squared errors (MSE) decrease with an increasing sample size; the MSEs even support the $\sqrt{n}$ convergence rate of the semiparametric estimators in that the MSEs at $n = 800$ are approximately half of the MSEs at $n = 200$. It is however noticeable that the SETR/J, which accounts for the discontinuity of the transition function, exhibits much smaller biases than the SETR/C. The source of the SETR/C bias is visible on Figure 1, where the average of estimated weight functions is presented along with the corresponding 90% confidence bands. Whereas SETR/C estimates are significantly biased, SETR/J exhibits much smaller bias and its confidence band includes the true transition function.

Comparing SETR/J to the parametric TAR and LSTAR estimates, the parametric estimates are more precise: the overall MSE of SETR (across the full vector of parameters) is approximately 10%–30% higher depending on the model and sample size; the difference is most likely related to the nonparametric estimation of a dis-

Table 3: Biases and MSE of all estimator for DGP2 and $T = 400$.

	TAR		LSTAR		SETR/C		SETR/J	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\widehat{\beta}_{1,0}$	0.081	0.153	0.027	0.262	0.044	0.265	0.062	0.287
$\widehat{\beta}_{1,1}$	-0.158	0.179	-0.004	0.118	0.024	0.113	0.021	0.115
$\widehat{\beta}_{1,2}$	0.013	0.121	0.013	0.171	0.043	0.182	0.053	0.193
$\widehat{\beta}_{2,0}$	-0.356	0.395	-0.004	0.451	-0.027	0.390	-0.057	0.423
$\widehat{\beta}_{2,1}$	0.177	0.203	0.009	0.117	0.003	0.105	0.013	0.108
$\widehat{\beta}_{2,2}$	0.102	0.143	0.007	0.171	0.011	0.162	0.020	0.174

continuous function. One can also note that the estimates are overall more precise in the case of DGP1b with the deterministic transition variable than in the case of DGP1a with the lagged dependent variable acting as the transition variable.

5.2 LSTAR results

The estimation results for the LSTAR model are summarized in Tables 3, from now on only for $T = 400$. The LSTAR model and estimator provides now correct parametric specification and provide thus best results in terms of very small bias and MSE. On the other hand, TAR is misspecified, which manifests itself by relatively large bias of some parameter estimates. Further, both SETR/C and SETR/J provide consistent estimates with relatively small biases and MSEs, which are surprisingly close to those of LSTAR: the precision of the parametric and semiparametric estimation is on the same level. Since the transition function is now smooth, SETR/C is more precise than SETR/J, which accounts for the possible discontinuities of the transition function and provides thus slightly more noisy estimates of the transition function. The difference is not very large though as can be seen from the transition function estimates on Figure 2.

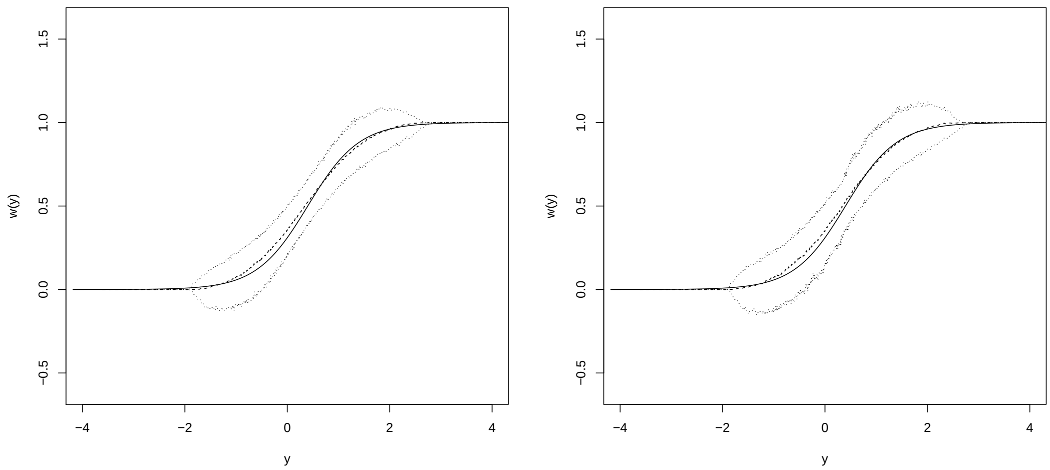

Figure 2: The mean estimates (dashed line) and 5% and 95% quantiles (dotted lines) of the transition function in DGP2 with $T = 400$; the solid line depicts the true transition function. The left and right panels correspond to SETR/C and SETR/J estimates, respectively.

5.3 Cosinus function

Another example of model with a continuous transition function is DGP3 with the corresponding estimation results in Tables 4 and the transition function estimates on Figure 3 (again for $T = 400$). In this case, both parametric models – TAR and LSTAR – are misspecified, which leads to substantial biases in both cases. On the other hand, both SETR/C and SETR/J provide consistent estimates with relatively small biases and the smallest MSEs. Since the transition function is again smooth, SETR/C should be more precise than SETR/J, but the difference between the two methods seems negligible.

5.4 Two-jump function

Finally, we present the results for DGP4, which includes two jumps with a smooth transition between them, see Figure 4. Also in this case, both parametric models, TAR and LSTAR, are misspecified, which leads to substantial biases in both cases – see Table 5 for the simulation results ($T = 400$). The semiparametric transition

Table 4: Biases and MSE of all estimator for DGP3 and $T = 400$.

	TAR		LSTAR		SETR/C		SETR/J	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\widehat{\beta}_{1,0}$	0.130	0.178	0.122	0.177	-0.003	0.095	-0.002	0.096
$\widehat{\beta}_{1,1}$	-0.307	0.372	-0.282	0.364	0.033	0.124	0.031	0.126
$\widehat{\beta}_{1,2}$	0.201	0.250	0.185	0.245	-0.023	0.096	-0.022	0.096
$\widehat{\beta}_{2,0}$	-0.060	0.136	-0.053	0.137	0.008	0.091	0.007	0.092
$\widehat{\beta}_{2,1}$	0.153	0.247	0.130	0.246	-0.036	0.122	-0.033	0.125
$\widehat{\beta}_{2,2}$	-0.104	0.175	-0.089	0.176	0.022	0.096	0.020	0.097

Table 5: Biases and MSE of all estimator for DGP4 and $T = 400$.

	TAR		LSTAR		SETR/C		SETR/J	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\widehat{\beta}_{1,0}$	0.064	0.158	0.062	0.166	0.005	0.102	0.004	0.105
$\widehat{\beta}_{1,1}$	-0.162	0.291	-0.158	0.314	0.021	0.132	0.010	0.125
$\widehat{\beta}_{1,2}$	0.106	0.198	0.103	0.215	-0.013	0.102	-0.006	0.098
$\widehat{\beta}_{2,0}$	-0.085	0.134	-0.082	0.138	-0.005	0.087	-0.002	0.086
$\widehat{\beta}_{2,1}$	0.202	0.268	0.196	0.277	-0.017	0.122	-0.012	0.118
$\widehat{\beta}_{2,2}$	-0.133	0.185	-0.128	0.190	0.010	0.090	0.007	0.088

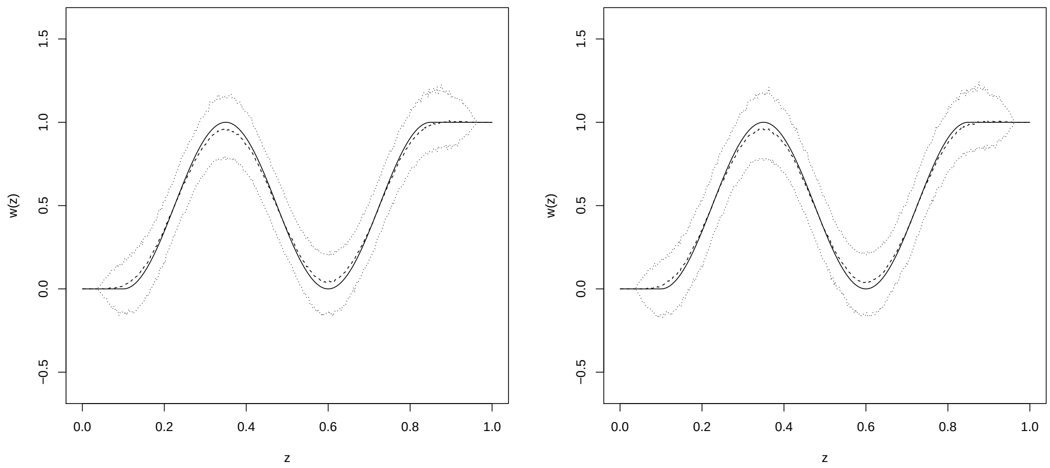

Figure 3: The mean estimates (dashed line) and 5% and 95% quantiles (dotted lines) of the transition function in DGP3 with $T = 400$; the solid line depicts the true transition function. The left and right panels correspond to SETR/C and SETR/J estimates, respectively.

methods SETR/C and SETR/J provide consistent estimates with relatively small biases and the smallest MSEs. Due to discontinuities of the transition function, SETR/J is slightly better than SETR/C. The difference is not very large though as the biases of the transition function estimates are similar in both cases (see Figure 4). The reason behind this seemingly surprising results, especially in comparison to DGP1a and DGP1b, is the bandwidth choice: the cross-validation selects for SETR/C a smaller bandwidth in the presence of two breaks than in the case of a constant function with one break only, which leads to a more precise approximation of the discontinuous weight function.

To sum up, the estimation of the semi-parametric transition model performs well in all cases. Obviously, the MSEs of the estimates from the semiparametric estimation are larger than those from the parametric estimations, when the DGPs are correctly specified in the case of TAR or LSTAR. But the gap is relatively small in the case of TAR and practically negligible in the case of LSTAR and the semiparametric procedure offers extra flexibility in modeling the transition function.

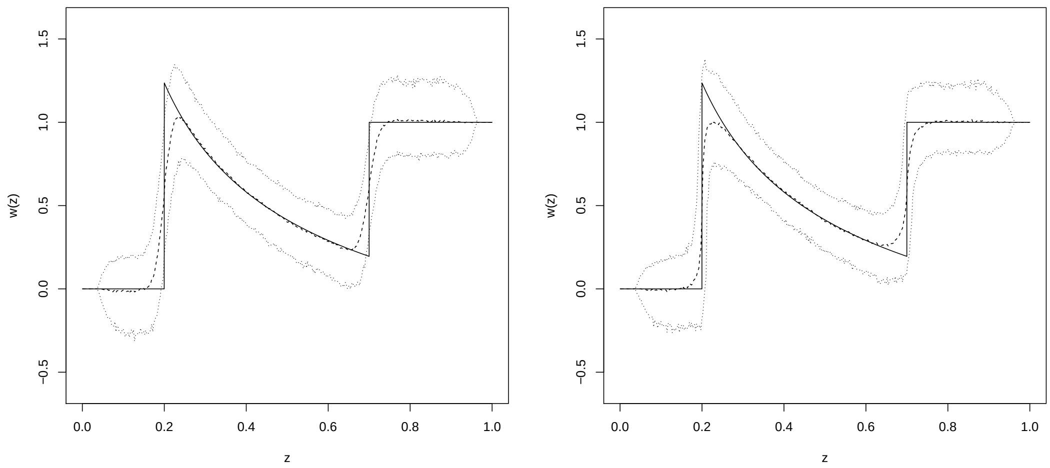

Figure 4: The mean estimates (dashed line) and 5% and 95% quantiles (dotted lines) of the transition function in DGP4 with $T = 400$; the solid line depicts the true transition function. The left and right panels correspond to SETR/C and SETR/J estimates, respectively.

6 Application to GDP

To demonstrate the use of the proposed semiparametric transition model, we analyze the quarterly GDP of the USA in years 1948–2007. The GDP and GNP series have been analyzed in the context of threshold autoregression or multiple regime models by many authors, for example, by Potter (1995b) or Tiao and Tsay (1994); see Hansen (2011) for an overview of this line of research. In particular, we consider the logarithm of the growth of quaterly GDP in two time periods (similarly to Clements and Krolzig, 1998): from 1948–1990 and from 1960–2007 as some authors suspect that the post-war behavior was characterized by a different dynamics than later at the end of the 20th century. (Although the proposed model can be theoretically extended to multiple regimes and even structural breaks, estimating a more complex model is not feasible due to a small sample size.) As in Potter (1995b), the baseline model is AR(5) without the third and fourth autoregressive terms (although their omission does not affect results much). This model led to more stable results than the AR(2) model used in some works concerning the GNP and GDP series in the

Table 6: Coefficient estimates for the TAR and SETR model of US GDP based on AR(5).

		1948–1990		1960–2007	
		TAR	SETR	TAR	SETR
	AR(1)	0.210	0.392	0.736	0.380
Regime 1	AR(2)	-0.859	-1.222	-2.231	-0.107
	AR(5)	-0.069	0.374	1.166	0.621
	AR(1)	0.326	0.274	0.256	0.204
Regime 2	AR(2)	-0.006	-0.057	0.167	0.135
	AR(5)	-0.175	-0.257	-0.155	-0.472
	Threshold	-0.187	—	-0.692	—

USA in the sense that the estimation results were not overly sensitive to changes in the time span or the bandwidth parameter. The transition variable z_t is chosen as the second lag of the dependent variable in agreement with practically all papers analyzing these series.

The estimation was performed by the algorithm described in Section 3, where we assume that observations with the values of the transition variable below its 5% quantile or above its 95% quantile are completely in regime 1 or regime 2. Recall that this constraint is also imposed on the estimates of the transition function $w(z_t)$. The estimation was performed by the jump-preserving local-constant estimator of Čížek and Koo (2014), see Section 3. Its bandwidth was fixed to $h = 1.5$ for easier comparison across time periods (the cross-validated bandwidth ranges from 1.1 to 2.0 depending on exact time interval), but the threshold value u_T was chosen by leave-one-out cross-validation on a grid from 0.1 to 1.0. Estimation employes the quartic kernel.

The estimation results are reported in Table 6 along with the TAR estimates traditionally used for this kind of analysis. Although the magnitude of the coefficients cannot be directly compared as the SETR model involves a general weighting function, both TAR and SETR estimates exhibit common patterns: similarly to

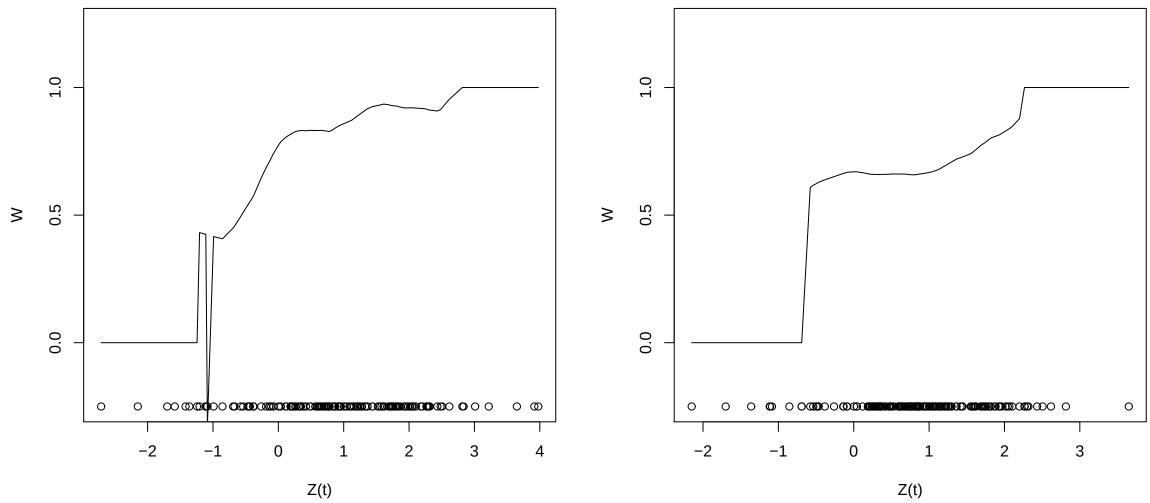

Figure 5: Transition function estimates for the semiparametric transition model of US GDP based on AR(5): 1948–1990 in the left panel and 1960–2007 in the right panel. The circles indicate the values of the transition variable observed in the data set.

Potter (1995b), for instance, the AR(1) coefficients are positive in all regimes, but the AR(2) coefficients are negative in regime 1, which corresponds to small values of z_t (below threshold in TAR), that is, to recession. In regime 2, which corresponds to large values of z_t (above threshold in TAR), the AR(2) coefficients are close to zero or positive depending on the time period used. (Note that the substantially negative AR(2) coefficient of the TAR model for data 1960–2007 is likely due to a highly imprecise estimate of regime 1 as there are only 8 observations below the threshold and the baseline model has 4 parameters).

The estimates of the transition function $w(z_t)$ for both periods are in Figure 5. In both cases, one can notice a discontinuity in the weighting functions at or above -1, which is also a feature of the TAR model. However as z_t increases, the transition function tends to gradually increase towards 1 for large values of z_t. Note that these characteristics of the transition function are not specific to the particular choice of bandwidth. Further, the oscillation of the estimates for years 1948–1990 around $z_t = -1$ is caused by the lack of data in that area, which leads both to volatile estimates and a large uncertainity in the selection of the right-, left-, or symmetric-estimates, see (18). Altogether, these results provide some evidence in favor of the semiparametric transition model by demonstrating that, for example, TAR might be too restrictive in some situations, even though a formal rejection of TAR would have to be based on confidence bands and, due to their likely width, a larger sample size.

7 Conclusion

The traditional TAR and STAR models both rely on the parametric form of the transition function. When the transition function differs from what these models assume, the estimation results often become biased and inconsistent. To remedy this problem, we develop the semiparametric transition model that generalizes the two-regime (smooth) transition model by assuming an unknown transition function. We propose an iterative estimation procedure for the semiparametric transition model which is based on the straightforward application of (local) least squares. Practically any consistent estimator discussed in the varying-coefficient literature can be used to estimate the conditional transition function as long as it is stochastically equicontinuous in its dependent variable and regressors. The consistency and asymptotic normality for the regression-coefficients estimator are derived in the paper, while the transition-function estimates are only shown to be consistent.

The simulation study using different types of transition functions indicates that the slope estimators from the parametric estimations of the TAR and STAR models are sensitive to the choice of the transition functions. On the other hand, the estimation of the proposed SETR function performs similarly to the parametric procedures (with a correctly specified transition function) if the transition function is smooth. Hence, the semiparametric transition model is a practically applicable alternative even in the parametric settings such as STAR.

In this paper, only a single transition variable and a two-regimes case are considered. Similar to the STAR model, the SETR model can be extended to a linear combination of several transition variables and to multiple regimes scenarios. Moreover, the asymptotic properties of the estimator of the transition function should be further investigated. Finally, asymptotic distribution and tests can be developed in future research for studying the features of the transition function (e.g., overshooting behaviour).

References

Ahmad, I., Leelahanon, S., Li, Q., 2005. Efficient estimation of a semiparametric partially linear varying coefficient model. The Annals of Statistics 33, 258–283.

Arcones, M. A., Yu, B., 1994. Central limit theorems for empirical and u-processes of stationary mixing sequences. Journal of Theoretical Probability 7, 47–72.

Areosa, W. D., McAleer, M., Medeiros, M. C., 2011. Moment-based estimation of smooth transition regression models with endogenous variables. Journal of Econometrics 165, 100–111.

Bai, J., Perron, P., 1998. Estimating and testing linear models with multiple structural changes. Econometrica 66 (1), 47–78.

Cai, Z., Fan, J., Yao, Q., 2000. Functional-coefficient regression models for nonlinear time series. Journal of the American Statistical Association 95 (451), 941–956.

Chan, K. S., Tong, H., 1986. On estimating thresholds in autoregressive models. Journal of Time Series Analysis 7, 179–190.

Chen, B., Hong, Y., 2012. Testing for smooth structural changes in time series models via nonparametric regression. Econometrica 80 (3), 1157–1183.

Chen, R., Tsay, R. S., 1993. Functional-coefficient autoregressive models. Journal of the American Statistical Association 88 (421), 298–308.

Čížek, P., Koo, C.-H., 2014. Jump-preserving varying-coefficient models, mimeo.

Clements, M. P., Krolzig, H.-M., 1998. A comparison of the forecast performance of markov-switching and threshold autoregressive models of us gnp. Econometrics Journal 1, C47–C75.

Davidson, J., 1994. Stochastic Limit Theory. Oxford University Press.

Fan, J., Huang, T., 2005. Profile likelihood inferences on semiparametric varying-coefficient partially linear models. Bernoulli 11 (6), 1031–1057.

Fan, J., Zhang, J., 2000. Two-step estimation of functional linear models with applications to longitudinal data. Journal of the Royal Statistical Society Series B 62 (2), 303–322.

Fan, J., Zhang, W., 1999. Statistical estimation in varying-coefficient models. The Annals of Statistics 27, 1491–1518.

Fan, J., Zhang, W., 2008. Statistical methods with varying coefficient models. Statistics and Its Interface 1, 179–195.

Gijbels, I., Lambert, A., Qiu, P., 2007. Jump-preserving regression and smoothing using local linear fitting: A compromise. Annals of the Institute of Statistical Mathematics 59 (2), 235–272.

Hansen, B. E., 2000. Sample splitting and threshold estimation. Econometrica 68, 575–603.

Hansen, B. E., 2011. Threshold autoregression in economics. Statistics and Its Interface 4, 123–127.

Hastie, T. J., Tibshirani, R. J., 1993. Varying-coefficient models. Journal of the Royal Statistical Society Series B 55, 757–796.

Hoover, D. R., Rice, J. A., Wu, C. O., Yang, L.-P., 1998. Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. Biometrika 85, 809–822.

Huang, J. Z., Shen, H., 2004. Functional coefficient regression models for non-linear time series: A polynomial spline approach. Scandinavian Journal of Statistics 31 (4), 515–534.

Huang, J. Z., Wu, C. O., Zhou, L., 2002. Varying-coefficient models and basis function approximations for the analysis of repeated measurements. Biometrika 89, 111–128.

Huang, J. Z., Wu, C. O., Zhou, L., 2004. Polynomial spline estimation and inference for varying coefficient models with longitudinal data. Statistics Sinica 14, 763–788.

Ichimura, H., Lee, S., 2010. Characterization of the asymptotic distribution of semiparametric m-estimators. Journal of Econometrics 159, 252–266.

Leybourne, S., Newbold, P., Vougas, D., 1998. Unit roots and smooth transitions. Journal of Time Series Analysis 19, 83–97.

Lin, C.-F. J., Teräsvirta, T., 1994. Testing the constancy of regression parameters against continuous structural change. Journal of Econometrics 62 (2), 211–228.

Lundbergh, S., Teräsvirta, T., van Dijk, D., 2003. Time-varying smooth transition autoregressive models. Journal of Business & Economic Statistics 21 (1), 104–21.

Medeiros, M. C., Veiga, A., 2003. Diagnostic checking in a flexible nonlinear time series model. Journal of Time Series Analysis 24 (4), 461–482.

Medeiros, M. C., Veiga, A., 2005. A flexible coefficient smooth transition time series model. IEEE Transactions on Neural Networks 16, 97–113.

Meitz, M., Saikkonen, P., 2010. A note on the geometric ergodicity of a nonlinear ar-arch model. Statistics & Probability Letters 80 (7-8), 631–638.

Potter, S. M., 1995a. A nonlinear approach to us gnp. Journal of Applied Econometrics 10 (2), 109–25.

Potter, S. M., 1995b. A nonlinear approach to us gnp. Journal of Applied Econometrics 10 (2), 109–125.

Rothman, P., 1998. Forecasting asymmetric unemployment rates. Review of Economics and Statistics 80, 164–168.

Sarantis, N., 1999. Modeling non-linearities in real effective exchange rates. Journal of International Money and Finance 18 (1), 27–45.

Skalin, J., Teräsvirta, T., 2002. Modeling asymmetries and moving equilibria in unemployment rates. Macroeconomic Dynamics 6 (2), 202–241.

Taylor, M. P., Peel, D. A., Sarno, L., 2001. Nonlinear mean-reversion in real exchange rates: Toward a solution to the purchasing power parity puzzles. International Economic Review 42 (4), 1015–42.

Taylor, N., van Dijk, D., Franses, P. H., Lucas, A., 2000. Sets, arbitrage activity, and stock price dynamics. Journal of Banking & Finance 24 (8), 1289–1306.

Teräsvirta, T., 1994. Specification, estimation, and evaluation of smooth transition autoregressive models. Journal of the American Statistical Association 89, 208–218.

Teräsvirta, T., Anderson, H. M., 1992. Characterizing nonlinearities in business cycles using smooth transition autoregressive models. Journal of Applied Econometrics 7 (S), S119–36.

Tiao, G. C., Tsay, R. S., 1994. Some advances in non-linear and adaptive modelling in time-series. Journal of Forecasting 13, 109–131.

Tong, H., 1983. Threshold Models in Non-Linear Time Series Analysis: Lecture Notes in Statistics. Springer, Berlin.

van der Vaart, A. W., Wellner, J. A., 1996. Weak Convergence and Empirical Processes. Springer-Verlag, New York.

van Dijk, D., Franses, P. H., 1999. Modeling multiple regimes in the business cycle. Macroeconomic Dynamics 3, 311–340.

van Dijk, D., Teräsvirta, T., Franses, P. H., 2002. Smooth transition autoregressive models – a survey of recent developments. Econometric Reviews 21, 1–47.

Wu, C. O., Chiang, C. T., Hoover, D. R., 1998. Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. Journal of the American Statistical Association 93, 1388–1402.

Yao, Q., Tong, H., 1998. Cross-validatory bandwidth selections for regression estimation based on dependent data. Journal of Statistical Planning and Inference 68, 387–415.

Zhang, W., Lee, S.-Y., Song, X., 2002. Local polynomial fitting in semivarying coefficient model. Journal of Multivariate Analysis 82 (1), 166 – 188.