State-Dependent Threshold STAR Models*

Michael J. Dueker
Russell Investments, U.S.A.

Zacharias Psaradakis
Birkbeck, University of London, U.K.

Martin Sola
Birkbeck, University of London, U.K.
Universidad Torcuato di Tella, Argentina

Fabio Spagnolo
Brunel University, U.K.

March 2010

Abstract

In this paper we consider extensions of smooth transition autoregressive (STAR) models to situations where the threshold is a time-varying function of variables that affect the separation of regimes of the time series under consideration. Our specification is motivated by the observation that unusually high/low values for an economic variable may sometimes be best thought of in relative terms. State-dependent logistic STAR and contemporaneous-threshold STAR models are introduced and discussed. These models are also used to investigate the dynamics of U.S. short-term interest rates, where the threshold is allowed to be a function of past output growth and inflation.

Keywords: Nonlinear autoregressive models; Smooth transition; Threshold; Interest rates.
JEL Classification: C22; E43.

* Part of this paper was written when Martin Sola was visiting the IAE at UAB. He gratefully acknowledges financial support from the Ministerio des Educaion.

1 Introduction

This papers investigates the possibility that the separation of regimes implied by nonlinear regime-switching autoregressive models is better characterized in relative terms rather than being dictated by the (constant) level of some variables. More specifically, we investigate the situation where the regime threshold is a function of variables which potentially affect the evolution of the time series under consideration. Our specification is motivated by the observation that unusually high/small values of an economic variable may sometimes be best thought of in relative terms. For example, interest rates may be considered high or low not in absolute terms but relative to relevant macroeconomic variables that describe the state of the economy. A rate of interest which is considered to be high when the economy is in a state of low inflation and high output growth may be considered too low when the economy in a state of high inflation and low output growth. We propose to model such behavior using variants of smooth transition autoregressive (STAR) models which allow the interest rate threshold to evolve over time as a function of inflation and output growth.

The building blocks of the models introduced here are the logistic STAR (LSTAR) model of Teräsvirta (1994) and the contemporaneous-threshold STAR (C-STAR) model of Dueker et al. (2007). In LSTAR-type models the regime weights are determined by the values of a transition variable relative to a threshold, while in models of the C-STAR type the regime weights depend on the ex ante probability that a latent regime-specific variable exceeds a threshold value.

We discuss a modeling strategy in which the regime threshold is allowed to be a function of relevant exogenous and/or predetermined variables. The resulting state-dependent LSTAR (SD-LSTAR) and state-dependent C-STAR (SDC-STAR) models can be used to describe nonlinear regime-switching behaviour in cases where the threshold between regimes does not remain constant and is affected by state variables. We also introduce a state-dependent C-STAR model with exogenous variables, or SDC-STARX, which nests all other C-STAR and SDC-STAR specifications as special cases. The SDC-STARX model shares with the C-STAR the key feature of having a transition function which depends on all the parameters of the model as well as on the data, including the exogenous variables that affect the state-dependent threshold. Crucially, such models allow us to distinguish between situations where

the threshold is fixed from those where the threshold is time-varying and state-dependent.

The paper is organized as follows. After recalling the definition of LSTAR and C-STAR models in Section 2, the SD-LSTAR, SDC-STAR and the SDC-STARX models are introduced and discussed in Section 3. We examine the stability characteristics of the SDC-STARX model and present the results of simulation experiments that throw light on the small-sample properties of the maximum likelihood (ML) estimator of the parameters of the model. In Section 4 we investigate the dynamics of U.S. short-term interest rates using STAR models with a threshold that depends on inflation and output growth. Our empirical results suggest that constant-threshold models are not a valid reduction of models with state-dependent threshold. Section 5 summarizes and concludes.

2 Constant-Threshold STAR Models

A two-regime (conditionally heteroskedastic) STAR model of order $p \geq 1$ for the univariate time series $\{y_t\}$ may be written in the form

$$y_t = G(\mathbf{z}_{t-1})y_{1t} + \{1 - G(\mathbf{z}_{t-1})\}y_{2t}, \qquad t \geq p+1, \tag{1}$$

where $G(\mathbf{z}_{t-1})$ is a continuous function of a vector of exogenous and/or predetermined variables $\mathbf{z}_{t-1}$, with $0 \leq G(\mathbf{z}_{t-1}) \leq 1$, and

$$y_{it} = \mu_i + \sum_{j=1}^{p} \alpha_j^{(i)} y_{t-j} + \sigma_i u_{it}, \qquad i = 1, 2. \tag{2}$$

In (2), μ_i and $\alpha_j^{(i)}$ are real constants, σ_i is a positive constant, and $\{u_{it}\}$ are independent and identically distributed (i.i.d.) random variables (for each i) such that $\mathbb{E}(u_{it}) = \mathbb{E}(u_{it}^2 - 1) = 0$, $\{u_{1t}\}$ and $\{u_{2t}\}$ are mutually independent, and u_{it} is independent of the initial values $y_1, \ldots, y_p$ for all i and t.

STAR models like (1)–(2) have been used extensively in the analysis of economic and financial data. The main feature that differentiates alternative STAR specifications is the choice of the mixing (or transition) function $G(\cdot)$ and the transition variables $\mathbf{z}_{t-1}$ (cf. Teräsvirta, 1998; van Dijk et al., 2002). A popular choice for $G(\cdot)$ in (1) is the logistic specification

$$G(y_{t-1}) = \frac{\exp\{-\gamma(y_{t-1} - y^*)\}}{1 + \exp\{-\gamma(y_{t-1} - y^*)\}}, \qquad \gamma > 0, \tag{3}$$

which gives rise to the pth-order LSTAR model, or LSTAR(p) (Teräsvirta, 1994). The location parameter y^* in (3) may be interpreted as the threshold between the two regimes associated with the limiting values $\lim_{y_{t-1}\to\infty} G(y_{t-1}) = 0$ and $\lim_{y_{t-1}\to-\infty} G(y_{t-1}) = 1$, while the slope parameter γ determines the smoothness of the transitions between the two regimes.

An alternative STAR model was proposed recently by Dueker et al (2007). The two-regime, pth-order C-STAR model, or C-STAR(p), can be obtained from (1)–(2) by putting $\mathbf{y}_{t-1} = (y_{t-1}, ..., y_{t-p})'$ and specifying the mixing function $G(\cdot)$ as

$$G(\mathbf{y}_{t-1}) = \frac{F_1(\{y^* - \mu_1 - \boldsymbol{\alpha}_1'\mathbf{y}_{t-1}\}/\sigma_1)}{F_1(\{y^* - \mu_1 - \boldsymbol{\alpha}_1'\mathbf{y}_{t-1}\}/\sigma_1) + 1 - F_2(\{y^* - \mu_2 - \boldsymbol{\alpha}_2'\mathbf{y}_{t-1}\}/\sigma_2)}. \tag{4}$$

Here, $F_i(\cdot)$ is the cumulative distribution function of u_{it} $(i = 1, 2)$ (which is assumed to possess a continuous and positive density), y^* is a threshold parameter, and $\boldsymbol{\alpha}_i = (\alpha_1^{(i)}, \ldots, \alpha_p^{(i)})'$ $(i = 1, 2)$. Thus, under the C-STAR(p) specification, we have

$$y_t = \frac{\mathbb{P}(y_{1t} < y^*|\mathbf{y}_{t-1})y_{1t} + \mathbb{P}(y_{2t} \geq y^*|\mathbf{y}_{t-1})y_{2t}}{\mathbb{P}(y_{1t} < y^*|\mathbf{y}_{t-1}) + \mathbb{P}(y_{2t} \geq y^*|\mathbf{y}_{t-1})}.$$

Since the mixing weights involve the probability that the contemporaneous value of the latent variable y_{1t} (y_{2t}) is smaller (greater) than the threshold level y^*, the model is called a contemporaneous-threshold STAR model. As with conventional STAR models, the C-STAR model may be thought of as a regime-switching model that allows for two regimes associated with the latent variables y_{1t} and y_{2t}. Alternatively, the model may be thought of as allowing for a continuum of regimes, each of which is associated with a different value of $G(\mathbf{y}_{t-1})$.[1]

[1] It is perhaps worth noting here that realizations of y_{1t} and y_{2t} such that $y_{1t} \geq y^*$ and $y_{2t} < y^*$ are not precluded by the C-STAR model. To illustrate the point numerically, suppose that $y_{1t} = -0.5 + 0.6y_{t-1} + 3u_t$ and $y_{2t} = -0.5 + 0.9y_{t-1} + 3u_t$, with $u_t \sim \mathcal{N}(0, 1)$; assume further that $y_{t-1} = 5$ and $y^* = 10$. Then, the mixing weights are $\mathbb{P}(y_{1t} < y^*|y_{t-1}) = \mathbb{P}(3u_t < y^* + 0.5 - 0.6y_{t-1}|y_{t-1}) = \mathbb{P}(u_t < 2.5) = 0.994$ and $\mathbb{P}(y_{2t} \geq y^*|y_{t-1}) = \mathbb{P}(3u_t \geq y^* + 0.5 - 0.9y_{t-1}|y_{t-1}) = \mathbb{P}(u_t \geq 1.6666667) = 0.0478$, so that $G(y_{t-1}) = 0.9541$. Hence, conditionally on $y_{t-1} = 5$, the C-STAR(1) model assigns a large weight to the regime associated with y_{1t}, so that most of the regime-specific area in this regime is below the threshold and very little of that generated by the other regime is above the threshold. It is not, therefore, against the logic of the model to obtain a realization such as $y_{2t} < y^*$ (which is very likely to happen); the identifying conditions of the model imply that the weight to the regime associated with y_{2t} is going to be small whenever it is likely that the realizations of y_{2t} are such that $y_{2t} < y^*$.

3 State-Dependent Threshold STAR Models

3.1 Definitions

In this section we generalize the standard LSTAR and C-STAR specifications to allow for state dependency of the threshold. In the case of the LSTAR, this is achieved by replacing the transition function in (3) with

$$G(y_{t-1}) = \frac{\exp\{-\gamma(y_{t-1} - y^*_{t-1})\}}{1 + \exp\{-\gamma(y_{t-1} - y^*_{t-1})\}}, \tag{5}$$

where y^*_{t-1} is a time-varying threshold. The latter is specified as a linear combination of the elements of a k-dimensional vector $\mathbf{x}_{t-1} = (x_{1,t-1}, \ldots, x_{k,t-1})'$ of observable exogenous and/or predetermined variables, that is,

$$y^*_{t-1} = y^* + \boldsymbol{\delta}'\mathbf{x}_{t-1}, \tag{6}$$

where y^* is an unknown threshold intercept and $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_k)'$ is a vector of unknown parameters. We call the model defined by (1), (2), (5) and (6) the state-dependent pth-order LSTAR model, or SD-LSTAR(p). Note that the SD-LSTAR(p) nests the conventional LSTAR(p) model. Hence, one may evaluate whether the latter model is a valid reduction of the SD-LSTAR(p) by testing the hypothesis that $\boldsymbol{\delta} = \mathbf{0}$.

To generalize the C-STAR(p) specification to allow for a state-dependent threshold, the mixing weights in (4) are replaced with

$$G(\mathbf{z}_{t-1}) = \frac{F_1(\{y^*_{t-1} - \mu_1 - \boldsymbol{\alpha}_1'\mathbf{y}_{t-1}\}/\sigma_1)}{F_1(\{y^*_{t-1} - \mu_1 - \boldsymbol{\alpha}_1'\mathbf{y}_{t-1}\}/\sigma_1) + 1 - F_2(\{y^*_{t-1} - \mu_2 - \boldsymbol{\alpha}_2'\mathbf{y}_{t-1}\}/\sigma_2)}, \tag{7}$$

where y^*_{t-1} is a time-varying threshold depending on the observable exogenous and/or predetermined variables $\mathbf{x}_{t-1}$, and $\mathbf{z}_{t-1} = (\mathbf{y}'_{t-1}, \mathbf{x}'_{t-1})'$. As before, the threshold y^*_{t-1} is specified to be the linear combination of the elements of $\mathbf{x}_{t-1}$ given in (6). We call the model defined by (1), (2), (6) and (7) the state-dependent pth-order C-STAR model, or SDC-STAR(p).

Notice that

$$G(\mathbf{z}_{t-1}) = \frac{\mathbb{P}(y_{1t} < y^*_{t-1}|\mathbf{z}_{t-1})}{\mathbb{P}(y_{1t} < y^*_{t-1}|\mathbf{z}_{t-1}) + \mathbb{P}(y_{2t} \geq y^*_{t-1}|\mathbf{z}_{t-1})}$$

and

$$1 - G(\mathbf{z}_{t-1}) = \frac{\mathbb{P}(y_{2t} \geq y^*_{t-1}|\mathbf{z}_{t-1})}{\mathbb{P}(y_{1t} < y^*_{t-1}|\mathbf{z}_{t-1}) + \mathbb{P}(y_{2t} \geq y^*_{t-1}|\mathbf{z}_{t-1})}.$$

Hence, under the assumptions in (6)–(7), the SDC-STAR(p) model may be rewritten as

$$y_t = \frac{\mathbb{P}(y_{1t} < y^*_{t-1}|\mathbf{z}_{t-1})y_{1t} + \mathbb{P}(y_{2t} \geq y^*_{t-1}|\mathbf{z}_{t-1})y_{2t}}{\mathbb{P}(y_{1t} < y^*_{t-1}|\mathbf{z}_{t-1}) + \mathbb{P}(y_{2t} \geq y^*_{t-1}|\mathbf{z}_{t-1})}. \tag{8}$$

As in the case of the C-STAR, the mixing weights involve the probability that the contemporaneous value of y_{1t} (y_{2t}) is smaller (greater) than some threshold level y^*_{t-1}. The SDC-STAR(p) model reduces to a C-STAR(p) under the restriction $\boldsymbol{\delta} = \mathbf{0}$.

A related model can be obtained by including the exogenous and/or predetermined variables $\mathbf{x}_{t-1}$ in the equation that describes the dynamics of the latent variables y_{1t} and y_{2t}. More specifically, let

$$y_{it} = \mu_i + \boldsymbol{\alpha}^{(i)\prime}\mathbf{y}_{t-1} + \boldsymbol{\delta}^{(i)\prime}\mathbf{x}_{t-1} + \sigma_i u_{it}, \qquad i = 1, 2, \tag{9}$$

where $\boldsymbol{\delta}^{(i)} = (\delta^{(i)}_1, \ldots, \delta^{(i)}_k)'$ are unknown parameters, and take $G(\cdot)$ to have the form

$$G(\mathbf{z}_{t-1}) = \frac{F_1(\{y^*_{t-1} - \mu_1 - \boldsymbol{\beta}_1'\mathbf{z}_{t-1}\}/\sigma_1)}{F_1(\{y^*_{t-1} - \mu_1 - \boldsymbol{\beta}_1'\mathbf{z}_{t-1}\}/\sigma_1) + 1 - F_2(\{y^*_{t-1} - \mu_2 - \boldsymbol{\beta}_2'\mathbf{z}_{t-1}\}/\sigma_2)}, \tag{10}$$

with $\boldsymbol{\beta}_i = (\boldsymbol{\alpha}_i', \boldsymbol{\delta}_i')'$ ($i = 1, 2$). Equations (1), (6), (9) and (10) define a pth-order state-dependent C-STAR model with exogenous variables, or SDC-STARX(p).

The SDC-STARX(p) model nests several specifications. If $\boldsymbol{\delta}^{(1)} = \boldsymbol{\delta}^{(2)} = \boldsymbol{\delta} = \mathbf{0}$, we obtain a restricted SDC-STARX(p) model, or RSDC-STARX(p), the mixing weights of which are the same as those of a SDC-STAR(p) model with threshold $y^*_{t-1} = y^* - \boldsymbol{\delta}^{(1)\prime}\mathbf{x}_{t-1}$; the two specifications are not, however, equivalent because they imply different conditional distributions for the latent variables y_{it}. If $\boldsymbol{\delta}^{(1)} = \boldsymbol{\delta}^{(2)} = \mathbf{0}$, the SDC-STARX($p$) model becomes a SDC-STAR(p). If $\boldsymbol{\delta} = \mathbf{0}$, we obtain a pth-order C-STAR model with exogenous variables, or C-STARX(p), and constant threshold y^*. Finally, if $\boldsymbol{\delta}^{(1)} = \boldsymbol{\delta}^{(2)} = \boldsymbol{\delta} = \mathbf{0}$, the SDC-STARX($p$) reduces to a C-STAR($p$) with constant threshold y^*.

3.2 Stability

As discussed in Tong (1990), the stability properties of a nonlinear autoregressive model may be assessed by examining the noiseless part, or skeleton, of the model. In this subsection, we analyze the stability characteristics of the SDC-STARX model by considering the properties of its skeleton.[2] The stability properties of SDC-STAR and C-STARX models follow as

[2] Since the properties of models of the LSTAR and C-STAR type have been discussed extensively in the literature, the SDC-STAR and SDC-STARX models will be the focus of much of the discussion in the remainder of this paper.

special cases of the SDC-STARX, while the stability properties of the SD-LSTAR may be analyzed in a similar fashion.

For simplicity and clarity of exposition, let us consider the case of a SDC-STARX(1) model with a threshold which depends on a single exogenous variable x_{t-1} ($k = 1$). The skeleton of such a model is defined as

$$Y_t = S(Y_{t-1}, X_{t-1}), \tag{11}$$

where

$$\begin{aligned} S(Y_{t-1}, X_{t-1}) &= G(Y_{t-1}, X_{t-1})\{\mu_1 + \alpha_1^{(1)} Y_{t-1} + \delta_1^{(1)} X_{t-1}\} \\ &\quad + \{1 - G(Y_{t-1}, X_{t-1})\}\{\mu_2 + \alpha_1^{(2)} Y_{t-1} + \delta_1^{(2)} X_{t-1}\}, \end{aligned} \tag{12}$$

and $G(Y_{t-1}, X_{t-1})$ is given by (10) with $\mathbf{z}_{t-1} = (Y_{t-1}, X_{t-1})'$ and $y^*_{t-1} = y^* + \delta x_{t-1}$. Assuming $\{x_t\}$ is stationary, a fixed point of the skeleton is any value Y_e which satisfies the equation

$$Y_e = S(Y_e, X_e), \tag{13}$$

where $X_e = \mathbb{E}(x_t)$. The value Y_e is said to be an equilibrium point of the SDC-STARX(1) model and, since the model is nonlinear, there may be one, several or no equilibrium points satisfying (13).

An examination of the local stability of each of the equilibrium points can be carried out by considering a first-order Taylor expansion about the fixed point,

$$Y_t - Y_e \approx \lambda(Y_e, X_e)(Y_{t-1} - Y_e), \tag{14}$$

where

$$\lambda(Y_e, X_e) = \left. \frac{\partial S(Y_{t-1}, X_e)}{\partial Y_{t-1}} \right|_{Y_{t-1} = Y_e}. \tag{15}$$

If $|\lambda(Y_e, X_e)| < 1$, then the equilibrium is locally stable and Y_t is a contraction in the neighbourhood of (Y_e, X_e). It is straightforward to verify that

$$\begin{aligned} \frac{\partial S(Y_{t-1}, X_e)}{\partial Y_{t-1}} &= \alpha_1^{(2)} + \{\alpha_1^{(1)} - \alpha_1^{(2)}\} G(Y_{t-1}, X_e) \\ &\quad + \left(\mu_1 - \mu_2 + \{\alpha_1^{(1)} - \alpha_1^{(2)}\} Y_{t-1} + \{\delta_1^{(1)} - \delta_1^{(2)}\} X_e\right) \frac{\partial G(Y_{t-1}, X_e)}{\partial Y_{t-1}}, \end{aligned} \tag{16}$$

where

$$\frac{\partial G(Y_{t-1}, X_e)}{\partial Y_{t-1}} = -\frac{\sigma_1^{-1} \alpha_1^{(1)} f_1(w_1)[1 - F_2(w_2)] + \sigma_2^{-1} \alpha_1^{(2)} f_2(w_2) F_1(w_1)}{\{F_1(w_1) + 1 - F_2(w_2)\}^2}, \tag{17}$$

$f_i(\cdot)$ is the density of $F_i(\cdot)$ $(i = 1, 2)$, and

$$w_i = \sigma_i^{-1}\left(y^* - \mu_i - \alpha_1^{(i)} Y_{t-1} + \{\delta - \delta_1^{(i)}\} X_e\right), \qquad i = 1, 2.$$

As a numerical illustration, we consider a SDC-STARX(1) model in which the noise variables u_{it} $(i = 1, 2)$ have the Student-t distribution with ν_i degrees of freedom (re-scaled to have unit variance). The data-generating process (DGP) for the threshold-determining variables $\{x_t\}$ is the autoregressive model

$$x_t = 0.9 x_{t-1} + \varepsilon_t,$$

where $\{\varepsilon_t\}$ are i.i.d. random variables, independent of $\{u_{1t}\}$ and $\{u_{2t}\}$, having the standardized Student-t distribution with 5 degrees of freedom. The parameters of the model take the following values:

$$\mu_1 = -0.5, \quad \mu_2 = 0.5, \quad \alpha_1^{(1)} = 0.9, \quad \alpha_1^{(2)} = 0.9, 0.99, \quad \sigma_1 = 3, \quad \sigma_2 = 2,$$
$$\nu_1 = 3, \quad \nu_2 = 4, \quad y^* = 0, \quad \delta = 0, 0.3, \quad \delta_1^{(1)} = 0, 0.1, \quad \delta_1^{(2)} = 0, -0.1.$$

For each parameter configuration, we use a grid of starting values to solve equation (13) numerically and find the equilibrium points; the local stability of each equilibrium point is then examined by considering the expansion in (14)–(17).

When $\delta = 0.3$, $\delta_1^{(1)} = 0.1$, $\delta_1^{(2)} = -0.1$, and $\alpha_1^{(2)} = 0.9$, we find a unique locally stable equilibrium $Y_e = -0.67$ with $\lambda(Y_e, X_e) = 0.93$. When $\delta = 0$ (so that the SDC-STARX model reduces to a C-STARX), $\delta_1^{(1)} = 0.1$, $\delta_1^{(2)} = -0.1$ and $\alpha_1^{(2)} = 0.9$, there is also a unique locally stable equilibrium point $Y_e = 0.10$, for which $\lambda(Y_e, X_e) = 0.93$. Figure 1 shows plots of the skeleton and of the data generated under the latter scenario. The top left shows the simulated data and the skeleton, using $y_0 = 10$ and $x_0 = 0$ as initial values. It can be seen that the skeleton converges very quickly to the fixed point $Y_e = -0.67$. The stability of this fixed point can be also inferred from the the lower right panel, where changes in the skeleton $(\Delta Y_t = Y_t - Y_{t-1})$ are plotted against past values of the skeleton (Y_{t-1}). These changes are positive to the left of $Y_e = -0.67$ (so that Y_t tends to increase towards the equilibrium value Y_e), but negative to the right of $Y_e = -0.67$ (so that Y_t tends to decrease towards Y_e). The values of the mixing weights $G(\cdot)$ for the SDC-STARX and C-STARX models, which are shown in the top right panel, differ substantially in several parts of the artificial sample.

When $\delta = \delta_1^{(1)} = \delta_1^{(2)} = 0$ (so that the SDC-STARX model reduces to a C-STAR) and $\alpha_1^{(2)} = 0.9$, we find a unique locally stable equilibrium $Y_e = 0.62$ with $\lambda(Y_e, X_e) = 0.95$. When $\delta = 0.3$, $\delta_1^{(1)} = \delta_1^{(2)} = 0$ (so that the SDC-STARX model reduces to a SDC-STAR) and $\alpha_1^{(2)} = 0.9$, there is also a unique locally stable equilibrium point $Y_e = -2.89$ for which $\lambda(Y_e, X_e) = 0.94$. Figure 2 shows the plots of the skeleton and of the data generated under the latter scenario. The top left panel plots the simulated data and the skeleton, using $y_0 = 10$ and $x_0 = 0$ as initial values. It can be seen that the skeleton takes considerable time to converge to the fixed point $Y_e = -2.89$. The stability of this fixed point can be inferred from the the lower right panel, where ΔY_t is plotted against Y_{t-1}. The changes in the skeleton are positive to the left of $Y_e = -2.89$ (so that Y_t tends to increase towards the equilibrium value Y_e), while the changes are negative to the right of $Y_e = -2.89$ (so that Y_t tends to decrease towards Y_e). The values of the mixing weights $G(\cdot)$ for the SDC-STAR and C-STAR models, which are shown in the top right panel, differ substantially in several parts of the artificial sample.

In the case where $\delta = 0.3$, $\delta_1^{(1)} = 0$, $\delta_1^{(2)} = 0$ and $\alpha_1^{(2)} = 0.99$, we find three fixed points $Y_e = -4.33$, $Y_e = 1.165$, and $Y_e = 49.002$, for which $\lambda(Y_e, X_e) = 0.928$, $\lambda(Y_e, X_e) = 1.006$, and $\lambda(Y_e, X_e) = 0.99$, respectively. Notice that the points -4.33 and 49.002 are both stable attractors with respect to the skeleton of the model, and 1.165 is the boundary between the domains of attraction.[3] This means that, in the presence of stochastic shocks, it is reasonable to expect the observed time series to switch occasionally between attractors in a stationary fashion. Whenever y_t is near any of the two locally stable equilibria, then it will take a large shock to cause a transition from one equilibrium point to the other.

One may also study the effect of changes in y^* on the fixed point Y_e. This, however, is a rather complicated exercise because changes in y^* affect both the number and the values of equilibrium points. In general, it can be shown that for very low values of y^* such that $G(Y_e, X_e) \approx 0$ we have $Y_e \approx \mu_2/(1-\alpha_1^{(2)})$, while for high values of y^* such that $G(Y_e, X_e) \approx 1$ we have $Y_e \approx \mu_1/(1-\alpha_1^{(1)})$.

[3] By the domain of attraction for a point Y_e, with respect to the skeleton defined by (11)–(12), we mean the set of all real v for which the iterates $S^n(v, X_e)$ converge to Y_e as n tends to infinity.

3.3 Estimation

Once the probability distributions of u_{1t} and u_{2t} have been specified, the parameters of state-dependent threshold STAR models can be estimated by the ML method. Letting $\boldsymbol{\theta}$ denote the vector of all the unknown parameters specifying the model under consideration, the conditional log-likelihood function associated with a sample $(y_1, \ldots, y_T)$ from the SDC-STAR(p) model is

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_{t=p+1}^{T} \ln \ell_t(\boldsymbol{\theta}),$$

where

$$\ell_t(\boldsymbol{\theta}) = \frac{G(\mathbf{z}_{t-1})}{\sigma_1} f_1\left(\frac{y_t - \mu_1 - \boldsymbol{\alpha}_1' \mathbf{y}_{t-1}}{\sigma_1}\right) + \frac{1 - G(\mathbf{z}_{t-1})}{\sigma_2} f_2\left(\frac{y_t - \mu_2 - \boldsymbol{\alpha}_2' \mathbf{y}_{t-1}}{\sigma_2}\right), \qquad (18)$$

with $G(\mathbf{z}_{t-1})$ given by (7) and $f_i(\cdot)$ denoting, as before, the probability density function of u_{it} $(i = 1, 2)$. The conditional log-likelihood function under a SD-LSTAR(p) model is obtained by replacing $G(\mathbf{z}_{t-1})$ in (18) with the mixing function given in (5). Under a SDC-STARX(p) model, the contribution of the t-th observation to the conditional likelihood is

$$\ell_t(\boldsymbol{\theta}) = \frac{G(\mathbf{z}_{t-1})}{\sigma_1} f_1\left(\frac{y_t - \mu_1 - \boldsymbol{\beta}_1' \mathbf{z}_{t-1}}{\sigma_1}\right) + \frac{1 - G(\mathbf{z}_{t-1})}{\sigma_2} f_2\left(\frac{y_t - \mu_2 - \boldsymbol{\beta}_2' \mathbf{z}_{t-1}}{\sigma_2}\right),$$

with $G(\mathbf{z}_{t-1})$ given by (10).

In the remainder of the paper, rather than assume conditional Gaussianity, $f_i(\cdot)$ $(i = 1, 2)$ will be specified to be the probability density function of the standardized Student-t distribution with $\nu_i > 2$ degrees of freedom, i.e.,

$$f_i(z) = \frac{\Gamma\left(\{\nu_i + 1\}/2\right)}{\Gamma(\nu_i/2)\sqrt{(\nu_i - 2)\pi}} \left(1 + \frac{z^2}{\nu_i - 2}\right)^{-(\nu_i+1)/2}, \qquad -\infty < z < \infty,$$

where $\Gamma(\cdot)$ is the gamma function. This specification has greater flexibility to accommodate possible outliers and other heavy-tailed characteristics in the data.

The sampling properties of the ML estimator are investigated by simulation. The DGP in the experiments is the SDC-STAR(1) model with several combinations of parameter values. In order to save space, we will only present here the results obtained with the parameter configuration used in the previous sub-section (with $\delta_1^{(1)} = \delta_1^{(2)} = 0$ and $\delta = 0.3$).[4] Experiments proceed by first generating $50 + T$ data points for y_t, with $T = 100, 200, 400, 800$, and initial

[4] The full set of results is available upon request.

values set to zero; the first 50 data points are then discarded in order to eliminate start-up effects, while the remaining T points are used for estimation.

ML estimates are obtained by means of a quasi-Newton algorithm that approximates the Hessian according to the Broyden–Fletcher–Goldfarb–Shanno (BFGS) update computed from numerical derivatives. In each case, a grid of 5 initial values for each parameter (including the true parameter) are used as starting values for the BFGS iterations; the replication that achieves the highest likelihood value is then selected.[5] An estimate of the asymptotic covariance matrix of the ML estimator is obtained in a familiar manner from the inverse of the Hessian matrix of $\mathcal{L}(\boldsymbol{\theta})$ evaluated at the ML estimates. Finally, since the computation of ML estimates for multiple-regime models is particularly time consuming (given the large number of simulations and the grid for the initial values), the number of Monte Carlo replications per experiment is 2,000.

In Table 1, we report some of the characteristics of the finite-sample distribution of each estimated parameter. These include the bias of the ML estimator, a measure of the accuracy of estimated standard errors as approximations to the sampling standard deviation of the ML estimator, and a test for the normality of the sampling distribution of the ML estimator.

For most parameters the bias is significantly different from zero only for the smallest sample size under consideration. The bias is clearly a decreasing function of the sample size, becoming negligible in most cases when $T = 800$.

As a measure of the accuracy of estimated asymptotic standard errors as approximations to the sampling standard deviation of the ML estimator, the ratio of the exact standard deviation of the ML estimates to the estimated standard errors averaged across replications for each design point is shown (in parentheses) in Table 1. For most parameters, the estimated asymptotic standard errors are downward biased. These biases are not, however, substantial (even when $T = 100$) and should not have significant adverse effects on inference.

Finally, the Gaussianity of the finite-sample distributions of the ML estimators is assessed by means of a Kolmogorov–Smirnov goodness-of-fit test based on the difference between the empirical cumulative distribution function of the ML estimates (relocated and scaled so that the linearly transformed estimates have zero mean and unit variance) and the standard normal cumulative distribution function (cf. Lilliefors, 1967). As can be seen in Table 1, the

[5] It is worth noting that results were found to be robust with respect to the choice of starting values.

normality hypothesis is rejected at the 5% level for some of the parameters and sample sizes of less than 200 observations. However, we find that the values of the Kolmogorov–Smirnov statistic decrease as T increases, suggesting that the quality of the normal approximation is likely to improve with increasing sample sizes. In fact, while normality is rejected a few times when $T \leq 200$, it is never rejected when $T \geq 400$.

4 An Empirical Application

In this section we investigate whether the dynamics of U.S. short-term interests can be adequately described using STAR models with constant threshold or whether they are better represented by the models proposed in this paper, i.e., the SD-LSTAR, SDC-STAR and SDC-STARX models with a threshold which is determined by inflation and output growth. Our data set consists of quarterly observations from 1947:2 to 2008:4 on the secondary market rate on three-month U.S. Treasury bills (y_t), the growth rate of real gross domestic product (g_t), and the consumer price inflation rate (π_t).[6]

4.1 Estimation Results

We begin by comparing a constant-threshold LSTAR model with one which has a state-dependent threshold. Table 2 reports ML estimates of the parameters of LSTAR and SD-LSTAR models (assuming Student-t conditional distributions). After a specification search, both models are fitted with $p = 4$ lags and, in the case of the SD-LSTAR model, the state-dependent threshold y^*_{t-1} is specified as in (6) with $\mathbf{x}_{t-1} = (g_{t-1}, \pi_{t-1})'$. Table 2 also reports the values of: the Ljung–Box statistic based on standardized residuals (Q_m) with $m = 30$ lags; the Ljung–Box statistic based on squared standardized residuals (Q^2_m); the maximized log-likelihood ($\mathcal{L}_{\max}$); the Akaike information criterion (AIC); the Bayesian information criterion (BIC).

The LSTAR and SD-LSTAR models have standardized residuals which do not show any significant signs of linear or nonlinear dependence on the basis of the Ljung–Box portmanteau tests, suggesting that the fitted models are adequate.[7] The estimated parameters reveal

[6] The data are obtained from the FRED® database of the Federal Reserve Bank of St. Louis (`http://research.stlouisfed.org/fred2/`).

[7] We should note that the results of these tests should be viewed with caution since the asymptotic properties

significant evidence of nonlinear behavior in the interest rate. For both models the estimated standard deviation of the errors is larger in regime 1.

In the case of the SD-LSTAR model, the coefficients on output growth (δ_1) and inflation (δ_2) in the equation which determines the state-dependent threshold suggest that the interest rate threshold values are such that regime 1 is favoured for lower output growth and higher inflation, while regime 2 is favoured for high values of output growth and low values of inflation. This is because high values of inflation (and low values of output growth) result in small values for the threshold, which makes it more likely that the interest rate would exceed the threshold value. Note also that the AIC and BIC favour the SD-LSTAR model over the LSTAR. Finally, the likelihood ratio (LR) statistic for the null hypothesis that $\delta_1 = \delta_2 = 0$ has a value of 36.25, implying that the hypothesis of a constant threshold can be rejected at the 1% significance level.

Comparing the transition functions of the LSTAR and SD-LSTAR models, shown in Figures 3 and 4, we see that the LSTAR gives weights close to unity to regime 2 in periods when the interest rate is greater than the threshold (around 5.6), while the SD-LSTAR gives weights close to unity in several periods which are mostly associated with the NBER dating of economic recessions. In the SD-LSTAR case, this happens in many periods when the interest rate was below the fixed threshold but above the state-dependent threshold. In other words, an interest rate of around 3% in 1957 was high enough given the relatively high inflation and low output growth.

Table 3 reports the ML estimation results for C-STAR(4), SDC-STAR(4), RC-STARX(4), C-STARX(4) and SDC-STARX(4) models (under the assumption of Student-t conditional distributions). We find that the models have standardized residuals which do not show any significant signs of linear or nonlinear dependence on the basis of the portmanteau tests. The estimated parameters reveal significant evidence of nonlinear behavior in the interest rate. The estimated standard deviation of the errors is larger in regime 2 for all models.

For the SDC-STAR and the SDC-STARX models, the coefficients on output growth (δ_1) and inflation (δ_2) in the equation which determines the state-dependent threshold are both significantly different from zero. The results suggest that, for both models, the interest rate

of residual autocorrelations from nonlinear models such as those considered here are not necessarily the same as those obtained under linearity (see, e.g., Li, 1992; Li and Mak, 1994).

threshold values are such that regime 1 is favoured the lower output growth is and the higher inflation is, while regime 2 is favoured for high values of output growth and low values of inflation. This is because high values of inflation (and low values of output growth) result in small values for the threshold, which makes it more likely that the interest rate would exceed the threshold value. Comparing the mixing functions of the C-STAR and SDC-STAR models, shown in Figure 5 and Figure 6, it can be seen that the C-STAR gives weights close to unity to regime 2 only in the periods associated with the highest interest rates, while the SDC-STAR gives weights close to unity in several periods which can be mostly associated with economic recessions. We find that the C-STAR is a not a valid restriction of the SDC-STAR since the LR statistic for the null hypothesis that $\delta_1 = \delta_2 = 0$ has a value of 26.932, implying that the hypothesis of a constant threshold is rejected at the 1% significance level. The AIC and BIC also favour the SDC-STAR model over the C-STAR.

Within the class of SD-CSTARX models, there are several testable hypotheses of interest. We find that the C-STARX specification is not a valid reduction of the SD-CSTARX model since the LR statistic for the null hypothesis that $\delta_1 = \delta_2 = 0$ has a value of 20.818, leading to a rejection of the hypothesis of a constant threshold at the 1% significance level. The RSDC-STARX specification is not a valid simplification either since the LR statistic for the null hypothesis that $\delta_1^{(1)} - \delta_1^{(2)} = \delta_2^{(1)} - \delta_2^{(2)} = 0$ is 20.814. When the SD-CSTAR is tested against the SD-CSTARX, the LR statistic for the null hypothesis that $\delta_1^{(1)} = \delta_1^{(2)} = \delta_2^{(1)} = \delta_2^{(2)} = 0$ has a value of 24.368, implying that the model in which output growth and inflation are allowed to influence the regime-specific variables directly enjoys more support by the data. The SDC-STARX model is favoured by the AIC, while the SDC-STAR is favoured by the BIC. When comparing LSTAR and the C-STAR models, the SDC-STAR model is favoured over the LSTAR and SD-LSTAR models by the BIC, while the AIC favours the SD-CSTARX.[8] Overall, there is significant evidence in favour of specifications with a state-dependent threshold.

The mixing functions for various C-STARX models are shown in Figures 8, 9 and 10. It is clear from Figures 3, 5, 8 and 9 that all models with a constant threshold mainly separate, as expected, periods of high and low interest rates regardless of the economic conditions. The separation of the regimes implied by the models where the threshold is state dependent,

[8] Psaradakis et al. (2009) examined the properties of information criteria as a means of selecting among alternative nonlinear autoregressive models.

which can be seen in Figures 4, 6 and 10, reveals that the two regimes are primarily associated with periods of economic booms and recessions.

The stability of the empirical SDC-STAR(4) is assessed by using numerical simulation. The skeleton of the model is found to have a single stable fixed point $Y_e = 2.2129$. The values of the skeleton are plotted in the top left panel of Figure 7, along with artificial data obtained by using the fitted model as the DGP. We note that the simulated data appear to replicate the qualitative features of the observed data – there is a fairly long period during which the series diverges from its long-run value, which takes the series above the threshold. The top right panel in Figure 7 shows the values of the mixing function $G(\cdot)$ evaluated under the skeleton of the SDC-STAR(4) and C-STAR(4) models, while the bottom left panel shows the evolution of the state-dependent threshold and its convergence to its long-run value 7.591. Finally, the bottom right panel shows the evolution of the mixing functions for the simulated data, which reveals substantial differences between the absolute and relative (to the exogenous variable) mixing functions. The stability of the empirical SDC-STARX(4) is assessed in a similar way, with the corresponding plots shown in Figure 11. We find a single stable fixed point $Y_e = 5.50$ with a long-run value of 5.57 for the threshold.

In summary, the results presented in this section suggest that models with state-dependent thresholds, i.e., the SD-LSTAR, SDC-STAR and SDC-STARX models, are capable of characterizing the data successfully. In our application at least, what seems to matter is not so much the absolute value of the threshold but rather its level relative to other variables that dictate the evolution of the economy as a whole. We believe that, for economic variables such as those considered in this paper, a relative threshold with respect to other variables that dictate the evolution of the economy is of more importance for the dynamics of the variables under consideration than a fixed threshold. Furthermore, since constant-threshold specifications are special cases of state-dependent specifications, it seems reasonable to use state-dependent specifications as a first step in a modelling cycle.

4.2 Out-of-Sample Forecasting

In this subsection we evaluate the accuracy of out-of-sample forecasts from the empirical models discussed earlier. The comparisons are based on a series of recursive forecasts computed in the following way. Each of the five competing models is fitted to the time series

$\{y_t, 1 \leq t \leq T-N\}$, where $T = 248$ is the number of observations in the full sample and $N = 100$ is the number of forecasts (this corresponds to a forecast period of 25 years, from 1984:1 to 2008.:4). Using $T-N$ as the forecast origin, a sequence of one-step-ahead forecasts are generated from each of the fitted models. The forecast origin is then rolled forward one period to $T-N+1$, the parameters of the forecast models are re-estimated, and another sequence of one-step-ahead forecasts is generated. The procedure is repeated until N forecasts are obtained, which are then used to compute measures of forecast accuracy. These measures include the mean squared prediction error (MSPE), the mean absolute prediction error (MAPE), the mean squared percentage prediction error (MSPPE), and the mean absolute percentage prediction error (MAPPE).

Out-of-sample forecasts for all the models under consideration involve a weighted average of the two linear relationships. The one-step-ahead forecast from a SDC-STARX(4) model at forecast origin T is obtained as

$$\widehat{y}_{T+1} = G(\widehat{\mathbf{z}}_T)(\widehat{\mu}_1 + \widehat{\boldsymbol{\alpha}}_1'\mathbf{y}_T + \widehat{\boldsymbol{\delta}}_1'\mathbf{x}_T) + \{1 - G(\widehat{\mathbf{z}}_T)\}(\widehat{\mu}_2 + \widehat{\boldsymbol{\alpha}}_2'\mathbf{y}_T + \widehat{\boldsymbol{\delta}}_2'\mathbf{x}_T), \tag{19}$$

where

$$G(\widehat{\mathbf{z}}_T) = \frac{F_1(\{\widehat{y}^* + \widehat{\boldsymbol{\delta}}'\mathbf{x}_T - \widehat{\mu}_1 - \widehat{\boldsymbol{\beta}}_1'\mathbf{z}_T\}/\widehat{\sigma}_1)}{F_1(\{\widehat{y}^* + \widehat{\boldsymbol{\delta}}'\mathbf{x}_T - \widehat{\mu}_1 - \widehat{\boldsymbol{\beta}}_1'\mathbf{z}_T\}/\widehat{\sigma}_1) + 1 - F_2(\{\widehat{y}^* + \widehat{\boldsymbol{\delta}}'\mathbf{x}_T - \widehat{\mu}_2 - \widehat{\boldsymbol{\beta}}_2'\mathbf{z}_T\}/\widehat{\sigma}_2)}. \tag{20}$$

Here and in the sequel, a hat on a parameter denotes the ML estimate, $\mathbf{x}_T = (g_T, \pi_T)'$, $\mathbf{y}_T = (y_T, y_{T-1}, y_{T-2}, y_{T-3})'$ and $\mathbf{z}_T = (\mathbf{y}_T', \mathbf{x}_T')'$. Forecasts for the other contemporaneous-threshold models are obtained by imposing the following restrictions in (19)–(20): $\widehat{\boldsymbol{\delta}}_1 - \widehat{\boldsymbol{\delta}}_2 = \widehat{\boldsymbol{\delta}} = \mathbf{0}$ for RSDC-STARX(4); $\widehat{\boldsymbol{\delta}}_1 = \widehat{\boldsymbol{\delta}}_2 = \mathbf{0}$ for SDC-STAR(4); $\widehat{\boldsymbol{\delta}} = \mathbf{0}$ for C-STARX(4); $\widehat{\boldsymbol{\delta}}_1 = \widehat{\boldsymbol{\delta}}_2 = \widehat{\boldsymbol{\delta}} = \mathbf{0}$ for C-STAR(4). For LSTAR-type models, the transition equation (20) is replaced by

$$G(\widehat{\mathbf{z}}_T) = \frac{\exp\{-\widehat{\gamma}(y_T - \widehat{y}^* - \widehat{\boldsymbol{\delta}}'\mathbf{x}_T)\}}{1 + \exp\{-\widehat{\gamma}(y_T - \widehat{y}^* - \widehat{\boldsymbol{\delta}}'\mathbf{x}_T)\}}. \tag{21}$$

One-step-ahead forecasts are then obtained using the following restrictions on (19) and (21): $\widehat{\boldsymbol{\delta}}_1 = \widehat{\boldsymbol{\delta}}_2 = \widehat{\boldsymbol{\delta}} = \mathbf{0}$ for LSTAR(4); $\widehat{\boldsymbol{\delta}}_1 = \widehat{\boldsymbol{\delta}}_2 = \mathbf{0}$ for SD-LSTAR(4). We note that, unlike LSTAR and SD-LSTAR models, the regimes in the C-STAR, SDC-STAR, RSDC-STARX, C-STARX and SDC-STARX specifications are not predetermined, so forecasts are obtained by computing the tree of possible future values and evaluating the probability that the regimes would follow different paths in the future.

Table 4 reports the results of the forecast exercise. It is clear that, while the C-STARX model yields the smallest MSPE and MAPE, the SDC-STAR outperforms the competing models on the basis of the MSPPE and the MAPPE. Also, models with state-dependent thresholds (SDC-STAR, SD-LSTAR) outperform their constant-threshold counterparts (C-STAR, LSTAR). In particular, the gain of using the SDC-STAR over the C-STAR model varies from around 5% in terms of the MAPE to over 100% in terms of the MSPPE. Similarly, the marginal gain of the SD-LSTAR over the LSTAR varies from 4% when using the MAPE to over 90% with the MSPPE. Interestingly, but not entirely surprisingly, a direct comparison between the contemporaneous and logistic threshold models (i.e., C-STAR vs LSTAR and SDC-STAR vs SD-LSTAR models) favours the contemporaneous-threshold specifications. A possible explanation for this result is that the mixing function of the C-STAR and SDC-STAR gives a probability forecast of the latent regime-specific variables at period $t+1$, while LSTAR and SD-LSTAR specifications evaluate the mixing function at period t (cf. Dueker et al., 2007).

As pointed out in Section 3, the SDC-STARX model nests several more parsimonious specifications, while the SD-LSTAR and SDC-STAR models nest their corresponding fixed-threshold alternatives. This suggests a direct comparison of forecast errors from parsimonious models with those from the more general models within which they are nested. A recent study of the problem of comparing the MSPE for nested models is Clark and West (2007), who suggest the use of MSPE-based statistics which are adjusted for the noise that is introduced into the forecasting process by the more general model when its additional parameters are not helpful for prediction.

Table 5 contains the MSPE of the parsimonious model (M1) and the corresponding larger model (M2) that nests M1, denoted by τ_1 and τ_2, respectively. In each case, the Clark–West adjustment term is computed as $d = (1/N)\sum_{j=1}^{N}(\widehat{y}_{1,T-N+j} - \widehat{y}_{2,T-N+j})^2$, where $\widehat{y}_{1,T-N+j}$ and $\widehat{y}_{2,T-N+j}$ are the one-step-ahead forecasts from M1 and M2, respectively. We also report the adjusted MSPE of the larger model ($\tau_2 - d$), the difference between the MSPE of M1 and the adjusted MSPE of M2 ($\tau_1 - \tau_2 + d$), and a t-statistic for testing the null hypothesis of equal predictive accuracy based on $\tau_1 - \tau_2 + d$ (the alternative hypothesis is that M2 has a smaller MSPE than M1). A comparison of models with state-dependent thresholds against their constant-threshold alternatives show an adjustment of 0.0152 (0.0336) for the

SD-LSTAR (SDC-STAR), corresponding to over 5% (10%) of the alternative model's MSPE. The adjustments seem to be large enough to give the larger models an advantage over the restricted models. In fact, on the basis of the MSPE-adjusted differences, 0.0391 and 0.0627 for the LSTAR and C-STAR specifications, respectively, the restricted models are rejected at the 5% significance level. Thus, after accounting for the estimation noise associated with the additional parameters, there is significant evidence that inflation and output growth have additional predictive content for U.S. short-term interests.

Turning to a comparison of the SDC-STARX specification against the parsimonious models that it nests, the MSPE-adjusted tests reject the hypothesis of equal predictive accuracy at the 1% significance level for the RSDC-STARX and C-STAR models but not for the SDC-STAR and C-STARX. This corroborates the findings that the information content of inflation and output growth is helpful in forecasting interest rates. However, it also shows that parsimonious models, such as the C-STARX and the SDC-STAR, are capable of producing as accurate forecasts as the general SDC-STARX specification even though the restrictions they entail are rejected by the data. This suggests that models with a superior in-sample fit do not necessarily produce the most accurate out-of-sample forecasts.[9]

5 Summary

In this paper we have considered STAR models with a threshold which evolves as a function of exogenous and/or predetermined variables. We have discussed two general classes of models with state-dependent threshold, namely logistic STAR models and contemporaneous-threshold STAR models. We have also introduced a state-dependent contemporaneous-threshold STAR model with exogenous variables which nests many other contemporaneous-threshold specifications as special cases. Several issues related to the stability properties of the models and the estimation of their parameters have been discussed. The proposed models have been used to analyze the dynamics of short-term U.S. interest rates. We have shown that it is possible to distinguish between situations where the threshold is modeled in absolute terms or relative to state variables which affect the evolution of the interest rates. Models with a state-dependent threshold associate different regimes with periods of economic booms

[9] Driffill et al. (2009) highlight this issue in the context of Markov-switching models for interest rates.

and recessions, while models with constant thresholds, by construction, associate different regimes with periods of high and low interest rates. Models with constant thresholds have been found to be outperformed by models with a state-dependent threshold both in terms of in-sample fit and out-of-sample forecast accuracy.

References

[1] Clark, T.E. and West, K.D. (2007), Approximately normal tests for equal predictive accuracy in nested models, *Journal of Econometrics* 138, 291–311.

[2] Driffill, J., Kenc, T., Sola, M. and Spagnolo, F. (2009), The effects of different parameterizations of Markov-switching in a CIR model of bond pricing, *Studies in Nonlinear Dynamics & Econometrics* 13, 1–22.

[3] Dueker, M.J., Sola, M. and Spagnolo, F. (2007), Contemporaneous threshold autoregressive models: estimation, testing and forecasting, *Journal of Econometrics* 138, 291–311.

[4] Li, W.K. (1992), On the asymptotic standard errors of residual autocorrelations in nonlinear time series modelling, *Biometrika* 79, 435–437.

[5] Li, W.K. and Mak, T.K. (1994), On the squared residual autocorrelations in non-linear time series with conditional heteroskedasticity, *Journal of Time Series Analysis* 15, 627–636.

[6] Lilliefors, W.H. (1967), On the Kolmogorov–Smirnov test for normality with mean and variance unknown, *Journal of the American Statistical Association* 62, 399–402.

[7] Psaradakis, Z., Sola, M., Spagnolo, F. and Spagnolo, N. (2009), Selecting nonlinear time series models using information criteria, *Journal of Time Series Analysis* 30, 369–394.

[8] Teräsvirta, T. (1994), Specification, estimation, and evaluation of smooth transition autoregressive models, *Journal of the American Statistical Association* 89, 208–218.

[9] Teräsvirta, T. (1998), Modelling economic relationships with smooth transition regressions, in Ullah, A. and D.E.A. Giles (eds.), *Handbook of Applied Economic Statistics*, New York: Marcel Dekker, pp. 507–552.

[10] Tong, H. (1990), *Non-linear Time Series: A Dynamical System Approach*, Oxford: Oxford University Press.

[11] van Dijk, D., Teräsvirta, T. and Franses, P.H. (2002), Smooth transition autoregressive models – a survey of recent developments, *Econometric Reviews* 21, 1–47.

Table 1: Sampling Properties of the ML Estimator

$\alpha_1^{(2)}$	T	$\widehat{\mu}_0$	$\widehat{\mu}_1$	$\widehat{\alpha}_1^{(1)}$	$\widehat{\alpha}_1^{(2)}$	$\widehat{\sigma}_0$	$\widehat{\sigma}_1$	$\widehat{y}^*$	$\widehat{\gamma}$	$\widehat{\nu}_1$	$\widehat{\nu}_2$
		Bias									
0.90	100	0.072*	0.026	0.046	0.017	-0.392*	-0.193*	0.007*	0.043*	0.210*	0.377*
	200	-0.042	0.021	0.032	0.016	-0.126	-0.180	-0.005	0.033*	0.203*	0.329*
	400	0.023	0.011	-0.004	0.003	0.107	-0.119	0.002	0.021	0.109	0.162
	800	0.008	-0.004	0.002	0.003	-0.059	0.098	0.000	0.012	0.087	0.101
0.99	100	0.085*	0.044*	0.065	0.016	-0.355*	-0.266*	0.006*	0.052*	0.309*	0.420*
	200	0.046*	0.027	0.045	0.012	-0.203	0.194	0.004	0.034	0.222*	0.291*
	400	0.028	0.017	0.022	0.033	-0.121	0.125	0.002	0.023	0.198	0.235
	800	0.012	0.010	0.012	0.021	-0.067	-0.107	0.000	0.011	0.096	0.112
		Sampling Standard Deviation/Estimated Standard Error									
0.90	100	1.202	1.021	0.955	0.963	1.130	1.152	1.077	1.123	1.239	1.300
	200	1.063	0.983	0.971	0.975	1.029	0.954	1.045	1.089	1.109	1.125
	400	1.023	1.028	0.985	1.024	0.987	1.036	1.022	1.034	1.057	1.060
	800	1.008	1.010	0.990	1.007	0.993	1.011	1.010	1.009	1.041	1.053
0.99	100	1.190	1.035	0.952	1.042	1.236	1.140	1.110	1.109	1.451	1.531
	200	1.113	0.980	1.018	1.023	0.959	1.061	1.057	1.034	1.142	1.209
	400	1.034	1.019	1.012	1.021	1.029	1.011	1.043	0.986	1.019	1.056
	800	1.021	1.012	0.992	1.012	1.015	1.006	1.025	0.996	0.997	1.023

An asterisk indicates that the Kolmogorov–Smirnov statistic for normality is significant at the 5% level.

Table 2: LSTAR and SD-LSTAR Models

	LSTAR(4)	SD-LSTAR(4)
μ_1	0.0472 (0.0527)	0.0085 (0.0476)
μ_2	0.3205 (0.2477)	−0.5855 (0.3492)
$\alpha_1^{(1)}$	1.5214 (0.0698)	1.2401 (0.0665)
$\alpha_1^{(2)}$	0.8054 (0.1359)	1.2253 (0.1147)
$\alpha_2^{(1)}$	−0.7088 (0.0800)	−0.2167 (0.0854)
$\alpha_2^{(2)}$	0.3871 (0.2049)	−0.1222 (0.1492)
$\alpha_3^{(1)}$	0.3587 (0.1040)	0.0889 (0.0823)
$\alpha_3^{(2)}$	−0.0868 (0.1646)	0.3355 (0.1290)
$\alpha_4^{(1)}$	−0.1610 (0.0823)	−0.0749 (0.0554)
$\alpha_4^{(2)}$	−0.1781 (0.0941)	−0.4178 (0.0752)
σ_1	0.3712 (0.1362)	0.3780 (0.0755)
σ_2	1.1429 (0.4463)	1.8215 (0.6407)
y^*	5.4980 (0.9288)	4.8337 (0.9330)
γ	0.5583 (0.1707)	0.9845 (0.3047)
δ_1	—	0.62762 (0.1639)
δ_2	—	−0.6959 (0.2451)
ν_1	3.0993 (1.2831)	3.2163 (0.9565)
ν_2	2.7796 (1.1175)	2.200 (0.1565)
Q_{30}	27.8388	32.4640
Q_{30}^2	69.3870	24.2222
$\mathcal{L}_{\max}$	−193.511	−175.386
AIC	419.022	386.772
BIC	474.911	449.648

Figures in parentheses are asymptotic standard errors.

Table 3: C-STAR, SDC-STAR and C-STARX Models

	C-STAR(4)	SDC-STAR(4)	RSDC-STARX(4)	C-STARX(4)	SDC-STARX(4)
μ_1	0.0491 (0.0503)	0.1440 (0.0479)	−0.0465 (0.0603)	0.0063 (0.0558)	−0.0607 (0.0513)
μ_2	−0.1046 (0.3341)	−0.0381 (0.2771)	−0.2268. (0.3734)	−0.5744 (0.2242)	−0.8349 (0.2893)
$\alpha_1^{(1)}$	1.3072 (0.0812)	1.2230 (0.0634)	1.3080 (0.1000)	1.3003 (0.0718)	1.2706 (0.0644)
$\alpha_2^{(1)}$	−0.3042 (0.1313)	−0.2114 (0.0857)	−0.3358 (0.1615)	−0.2928 (0.1163)	−0.2508 (0.0792)
$\alpha_3^{(1)}$	0.1142 (0.1035)	0.1102 (0.0846)	0.1152 (0.1108)	0.0980 (0.0956)	0.0537 (0.0783)
$\alpha_4^{(1)}$	−0.1023 (0.0635)	−0.0859 (0.0565)	−0.0708 (0.0650)	−0.0889 (0.0590)	−0.0397 (0.0531)
$\alpha_1^{(2)}$	1.0726 (0.1248)	1.2128 (0.1288)	1.0613. (0.1309)	0.9639 (0.09483)	0.9949 (0.1055)
$\alpha_2^{(2)}$	−0.0771 (0.1885)	−0.0989 (0.1719)	−0.0657 (0.1932)	−0.0773 (0.1458)	−0.0319 (0.1501)
$\alpha_3^{(2)}$	0.2700 (0.1604)	0.2859 (0.1445)	0.2969 (0.1693)	0.4162 (0.1320)	0.4729 (0.1376)
$\alpha_4^{(2)}$	−0.2693 (0.1053)	−0.4070. (0.0818)	−0.2940 (0.1062)	−0.3273 (0.0838)	−0.4510 (0.0728)
σ_1	0.3078 (0.0924)	0.3998 (0.1215)	0.3100 (0.0874)	0.2303 (0.0481)	0.4458 (0.1269)
σ_2	1.2893 (0.2901)	1.2189 (0.2445)	1.2648 (0.2877)	1.5630 (0.4293)	0.8953 (0.1317)
y^*	5.4196 (0.3286)	4.7219 (1.2844)	5.5435 (0.3303)	5.3874 (0.2720)	5.0555 (0.9534)
$\delta_1^{(1)}$	—	—	0.0156 (0.0062)	0.0054 (0.0051)	0.0102 (0.0049)
$\delta_2^{(1)}$	—	—	−0.0026 (0.0087)	0.0033 (0.0082)	0.0027 (0.0089)
$\delta_1^{(2)}$	—	—	0.0156 (0.0062)	0.1025 (0.0182)	0.0988 (0.0226)
$\delta_2^{(2)}$	—	—	−0.0026 (0.0087)	−0.0122 (0.0242)	0.0174 (0.0311)
δ_1	—	0.9385 (0.1403)	—	—	0.4337 (0.0834)
δ_2	—	−0.9461 (0.2105)	—	—	−0.6798 (0.2732)
ν_1	3.4019 (1.4695)	3.0337 (1.1981)	3.4342 (1.8676)	5.7408 (5.5767)	2.4909 (0.3935)
ν_2	2.7620 (0.5728)	2.5497 (0.3183)	2.5194 (0.5290)	2.2577 (0.1821)	3.1796 (0.6434)
Q_{30}	27.0962	39.8058	30.2040	33.1596	41.2709
Q_{30}^2	39.6623	35.6250	42.3384	39.9753	39.1482
$\mathcal{L}_{\max}$	−191.184	−177.718	−186.350	−175.943	−165.534
AIC	412.369	389.437	406.701	389.885	373.069
BIC	464.765	413.355	466.083	456.253	446.423

Figures in parentheses are asymptotic standard errors.

Table 4: Out-of-Sample Predictive Accuracy

	MSPE	MAPE	MSPPE	MAPPE
SDC-STAR	0.2551	0.3661	0.0251	0.0984
SD-LSTAR	0.2636	0.3779	0.0254	0.1015
C-STAR	0.2842	0.3832	0.0502	0.1165
LSTAR	0.2875	0.3930	0.0521	0.1174
RSDC-STARX	0.2819	0.3875	0.0422	0.1127
C-STARX	0.2482	0.3630	0.0408	0.1089
SDC-STARX	0.2608	0.3754	0.0417	0.1124

MSPE is the mean squared prediction error, MAPE is the mean absolute prediction error, MSPPE is the mean squared percentage prediction error, and MAPPE is the mean absolute percentage prediction error.

.

Table 5: Tests for Equal Predictive Accuracy

		τ_1	τ_2	d	$\tau_2 - d$	$\tau_1 - \tau_2 + d$	t-stat
M2:	SD-LSTAR	0.2875	0.2636	0.0152	0.2484	0.0391	2.0690
M1:	LSTAR						
M2:	SDC-STAR	0.2842	0.2551	0.0336	0.2215	0.0627	2.2891
M1:	C-STAR						
M2:	SDC-STARX	0.2819	0.2608	0.0232	0.2377	0.0442	2.3417
M1:	RSDC-STARX						
M2:	SDC-STARX	0.2551	0.2608	0.0328	0.2280	0.0271	1.4366
M1:	SDC-STAR						
M2:	SDC-STARX	0.2482	0.2608	0.0111	0.2497	–0.0015	–0.1310
M1:	C-STARX						
M2:	SDC-STARX	0.2842	0.2608	0.0348	0.2261	0.0581	2.4286
M1:	C-STAR						

τ_1 is the MSPE of M1, τ_2 is the MSPE of M2, d is the Clark–West adjustment term, and t-stat is the t-ratio associated with $\tau_1 - \tau_2 + d$.

Generated data, histogram and skeleton for SDC−STARX

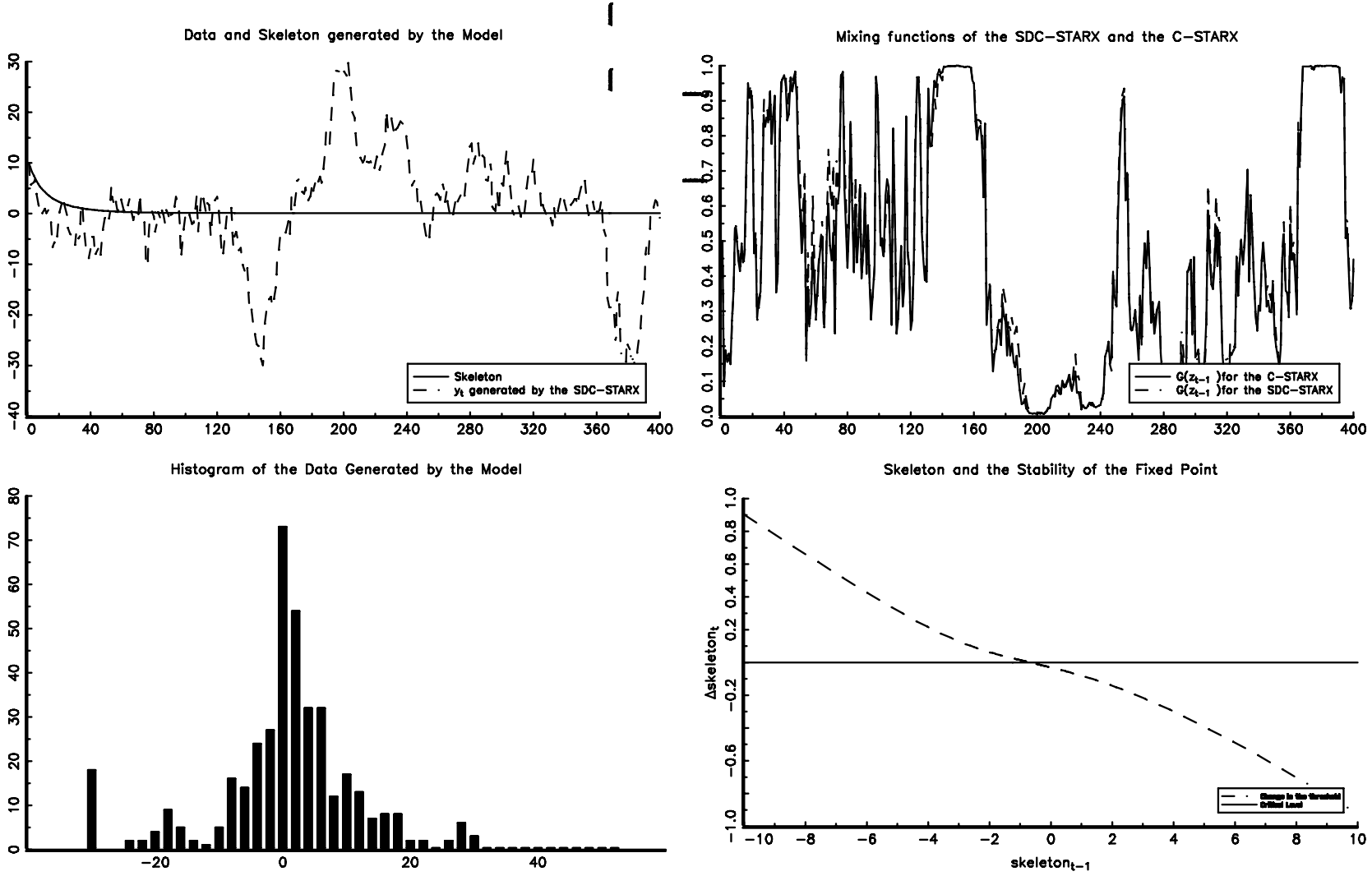

Figure 1

Generated data, histogram and skeleton for SDC−STAR

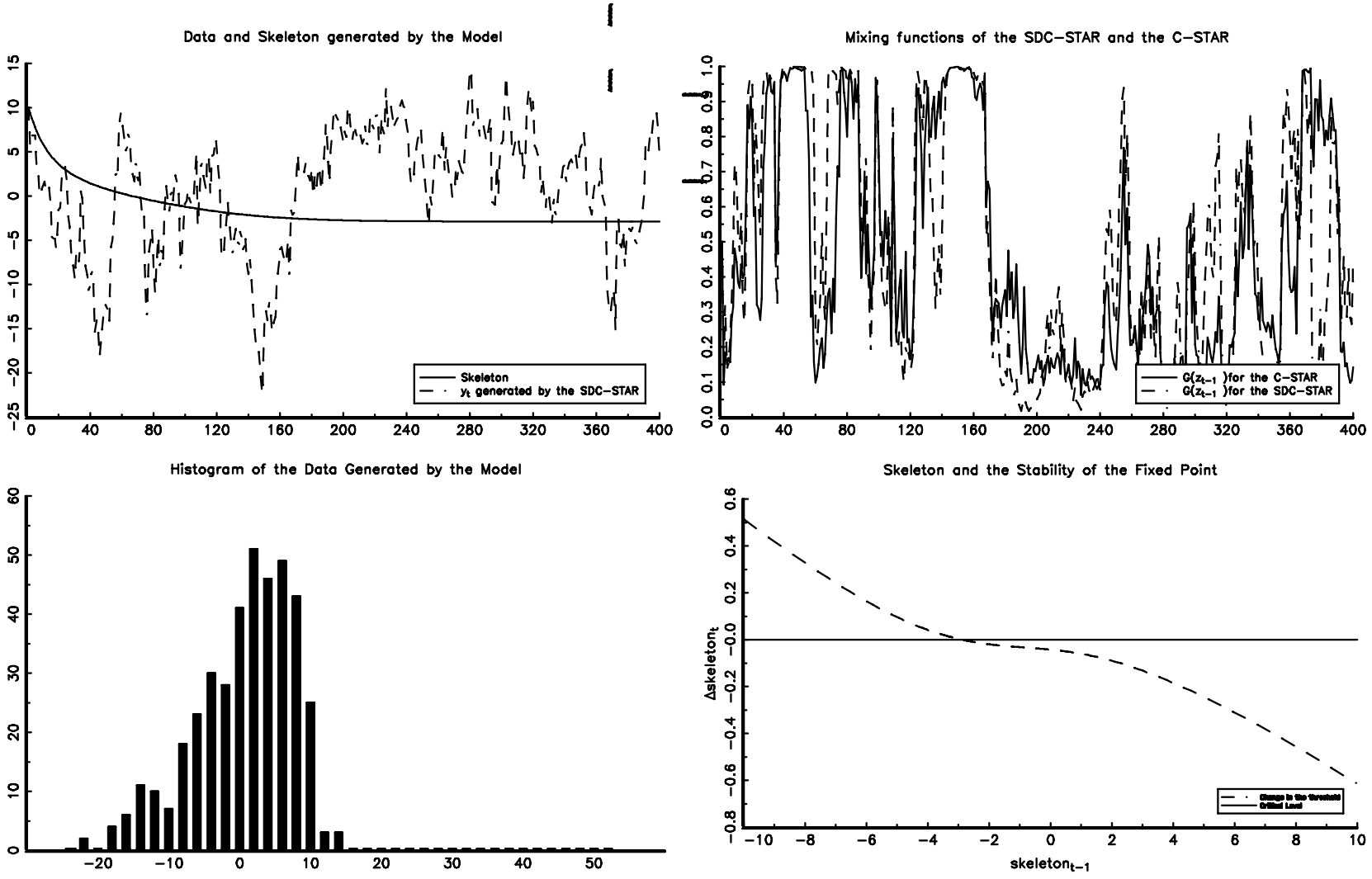

Figure 2

Threshold and Mixing Function for the LSTAR

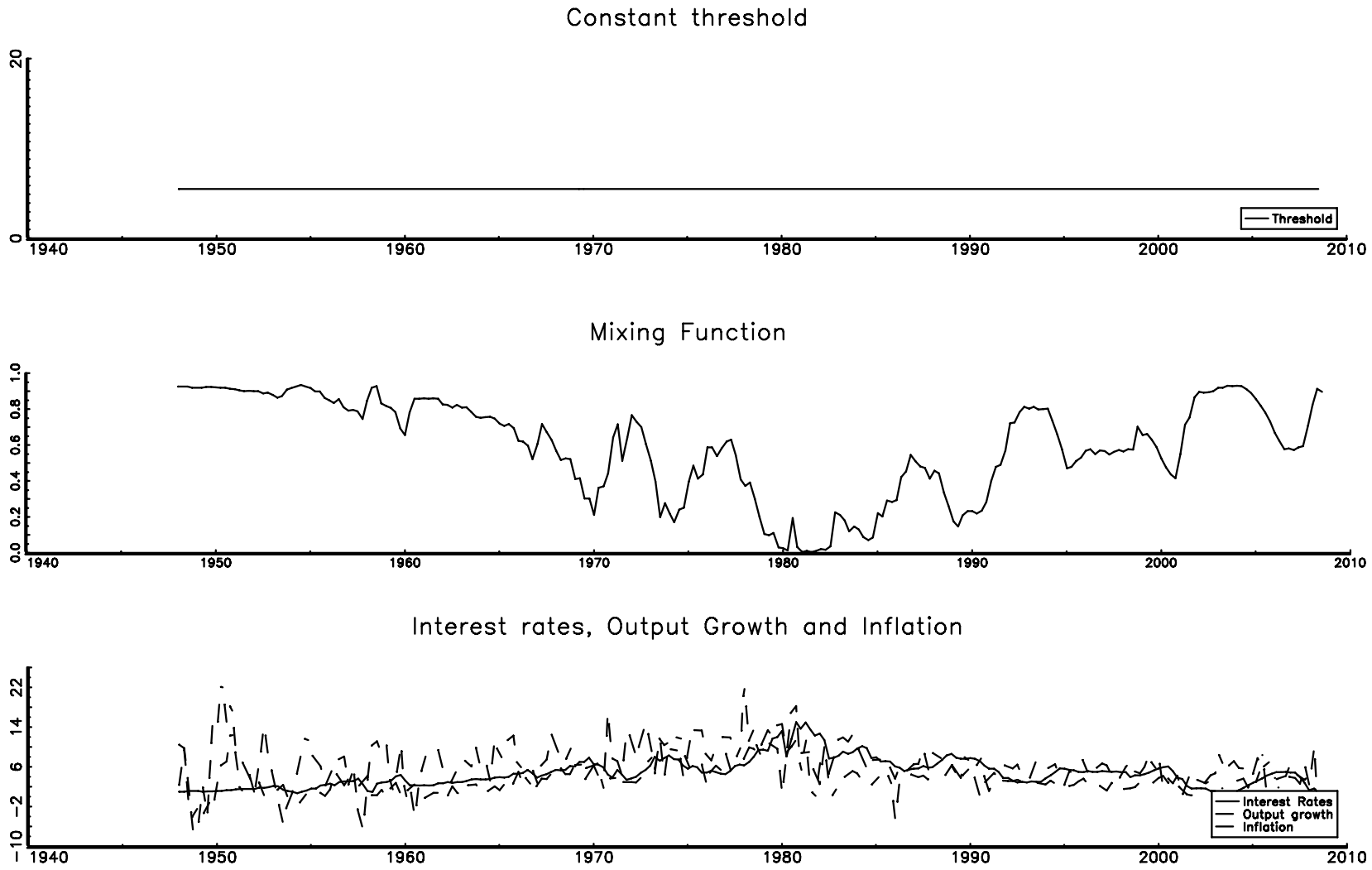

Figure 3

State Dependent Threshold and Mixing Funtion for the SD−LSTAR

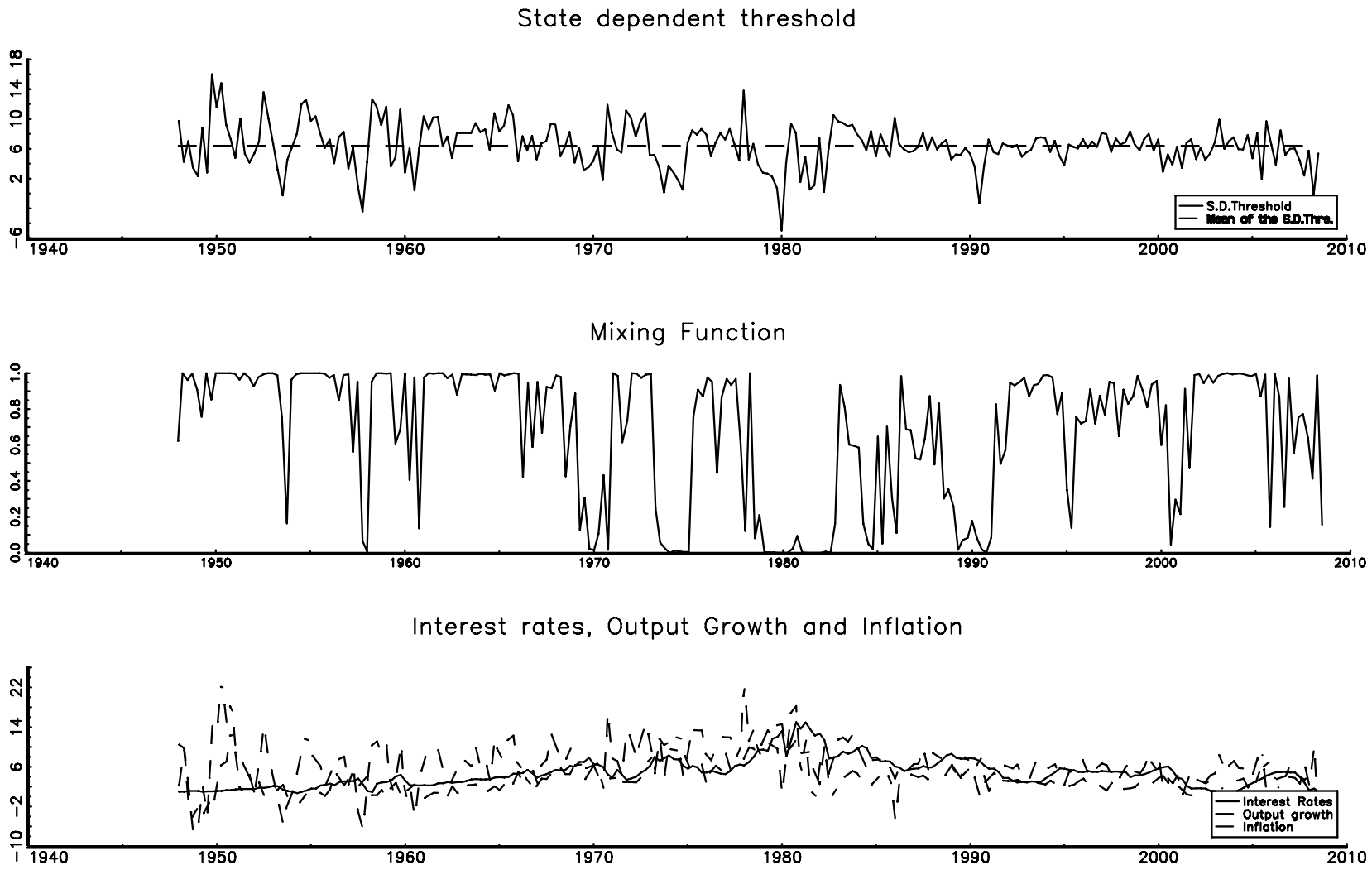

Figure 4

Threshold and Mixing Funtion for the C–STAR

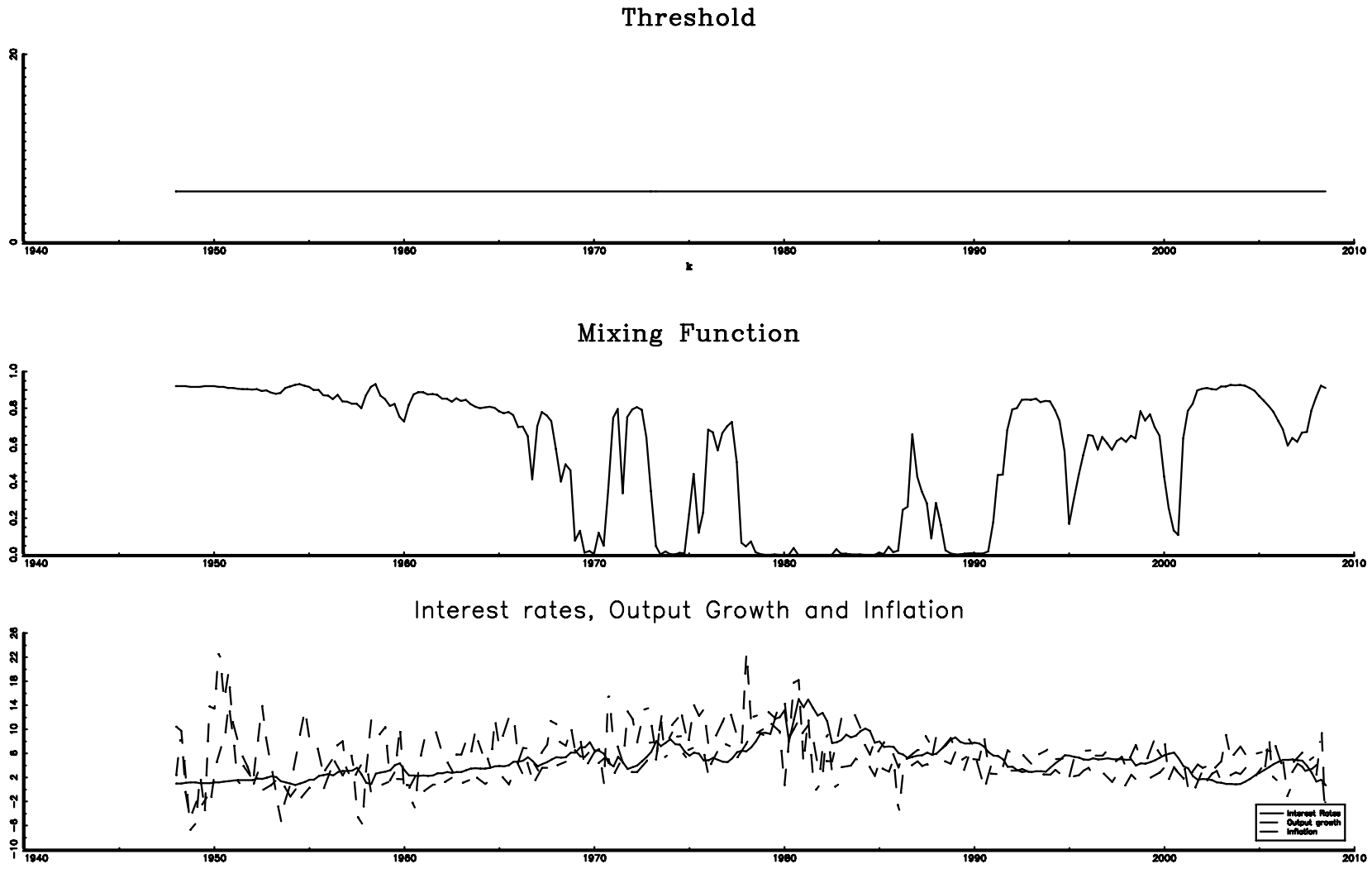

Figure 5

State Dependent Threshold, Mixing Funtion for the SDC−STAR

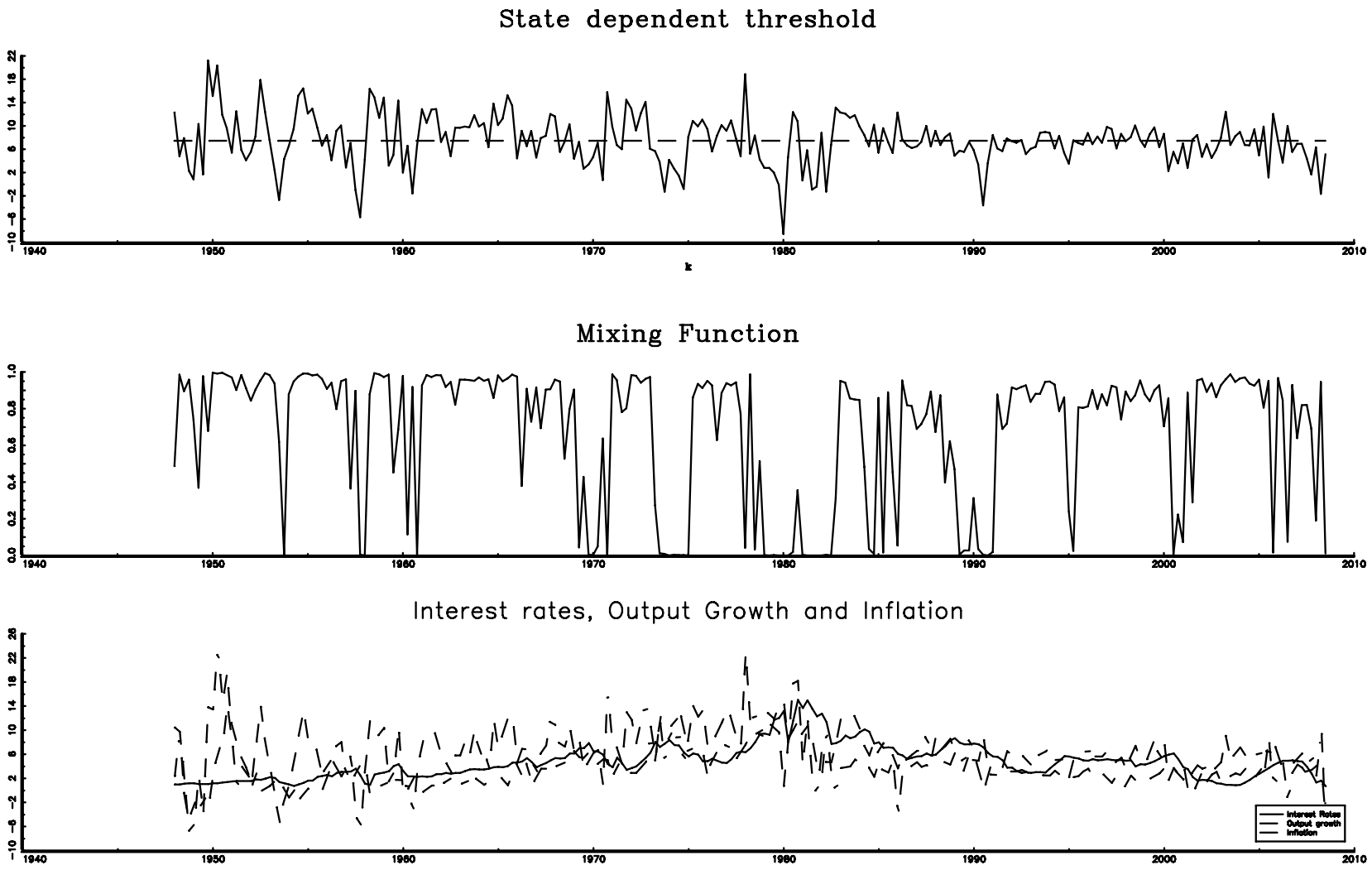

Figure 6

Generated Data and Skeleton for the Empirical DGP

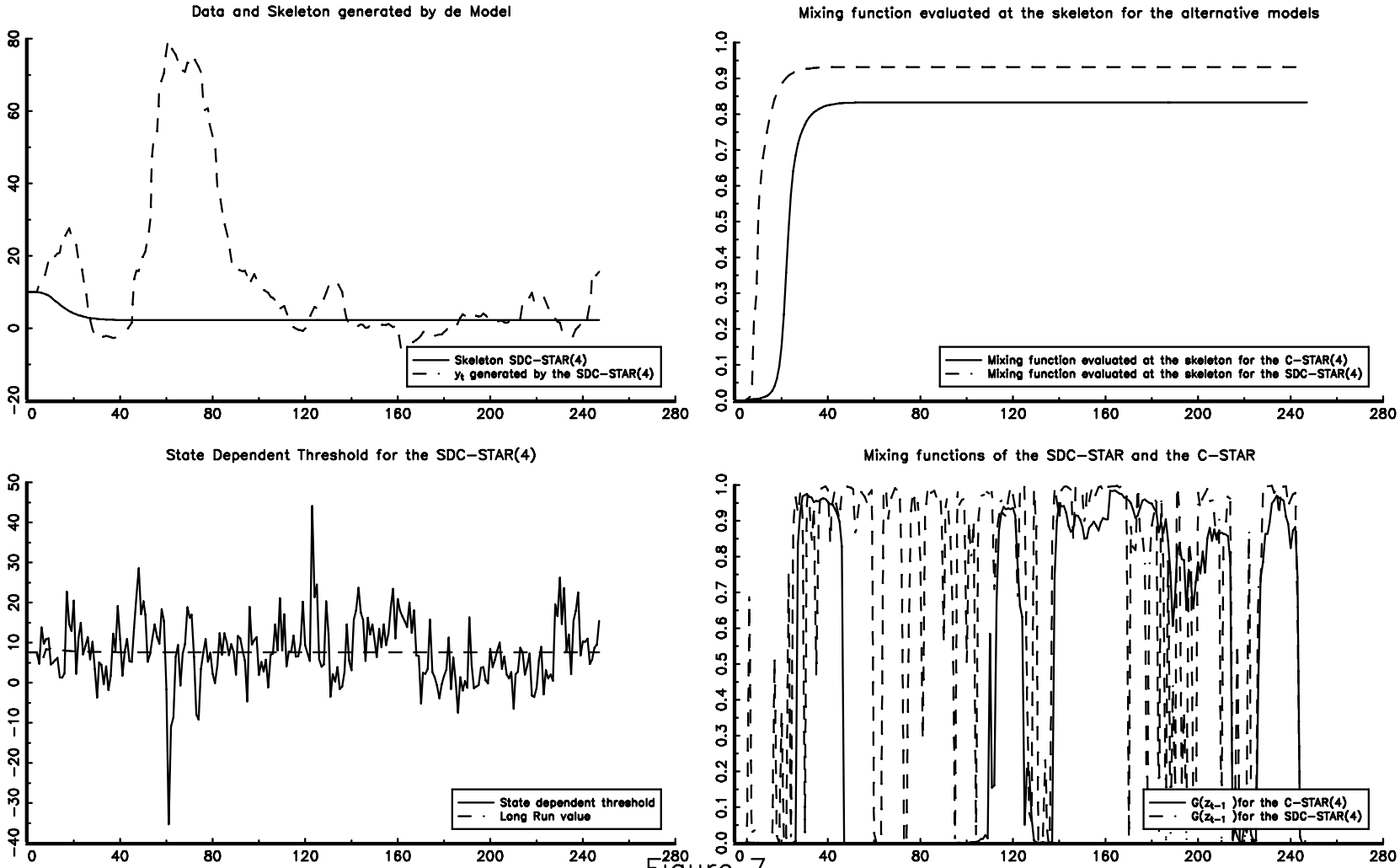

Figure 7

Threshold and Mixing Function for the RC–STARX

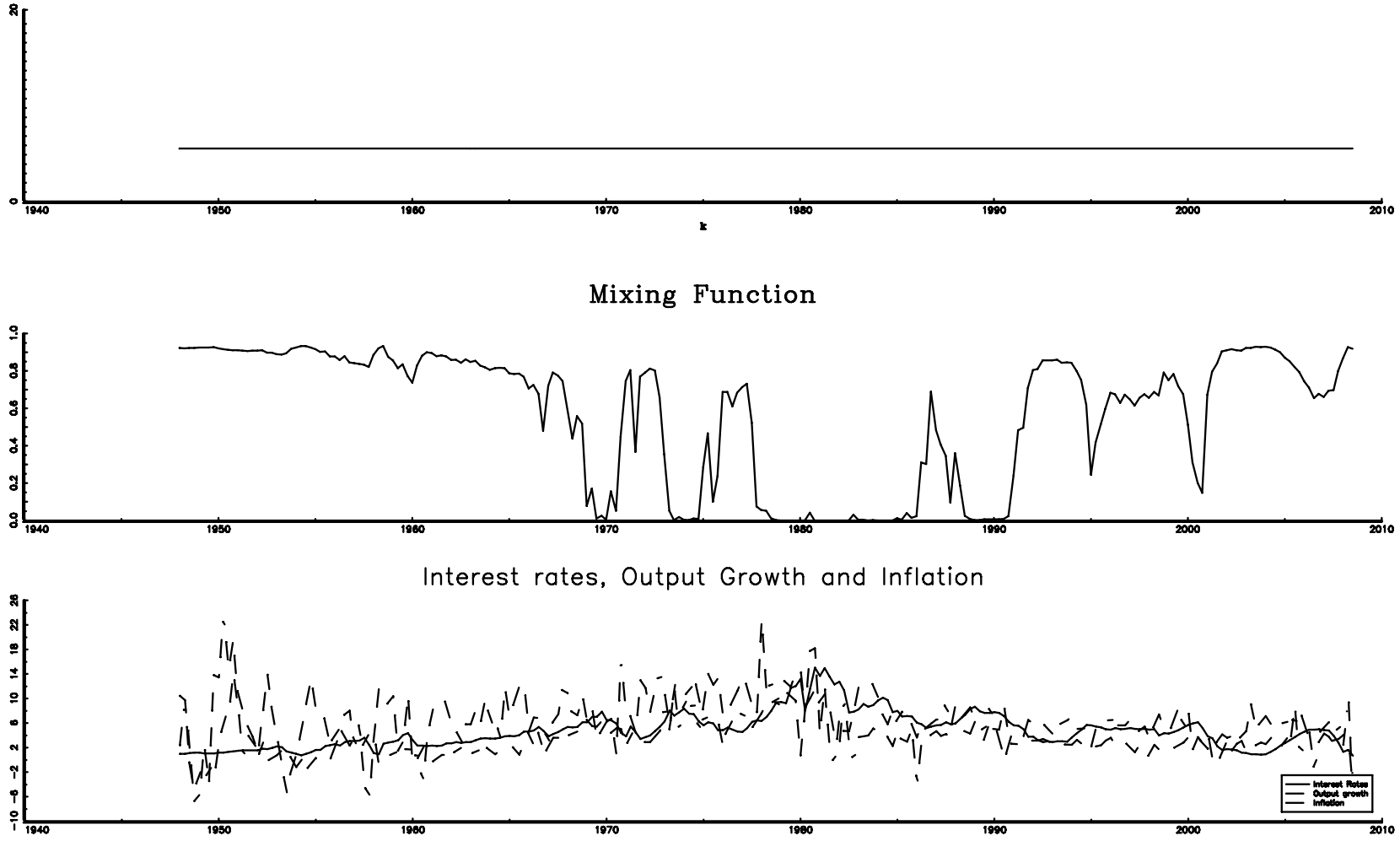

Figure 8

Threshold and Mixing Function for the C–STARX

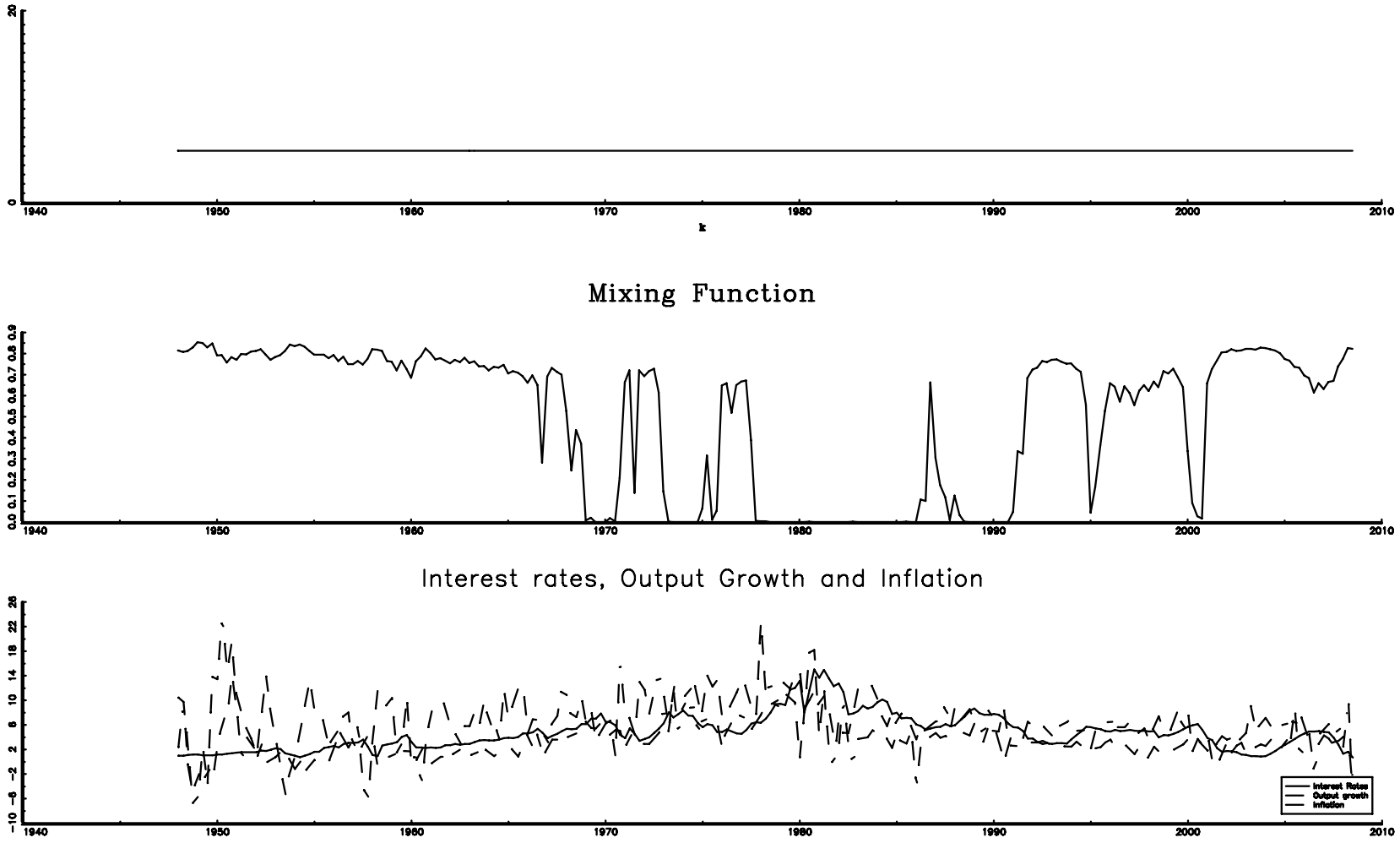

Figure 9

State Dependent Threshold and Mixing Function for the SDC–STARX

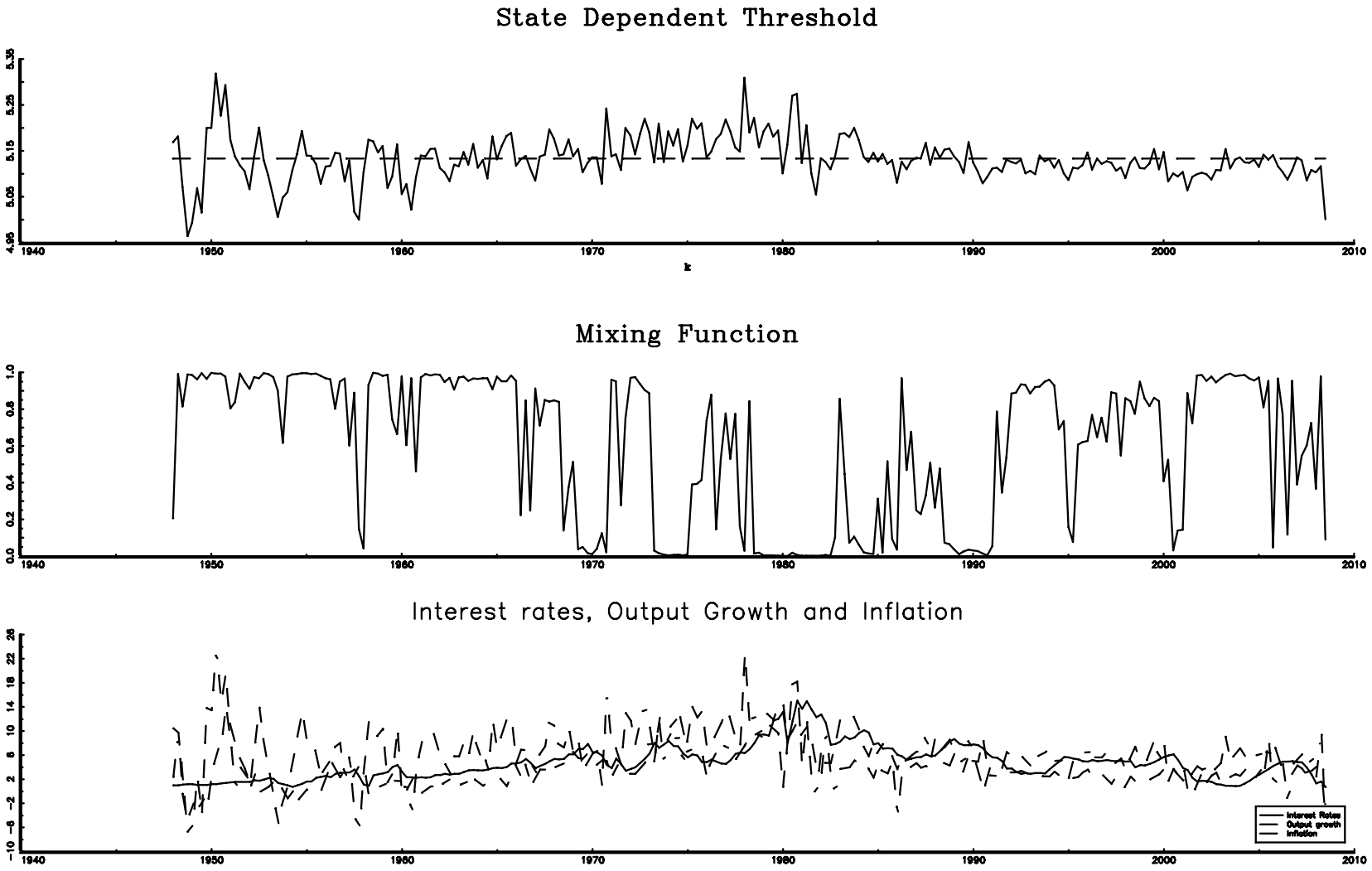

Figure 10

Generated Data and Skeleton for the Empirical DGP

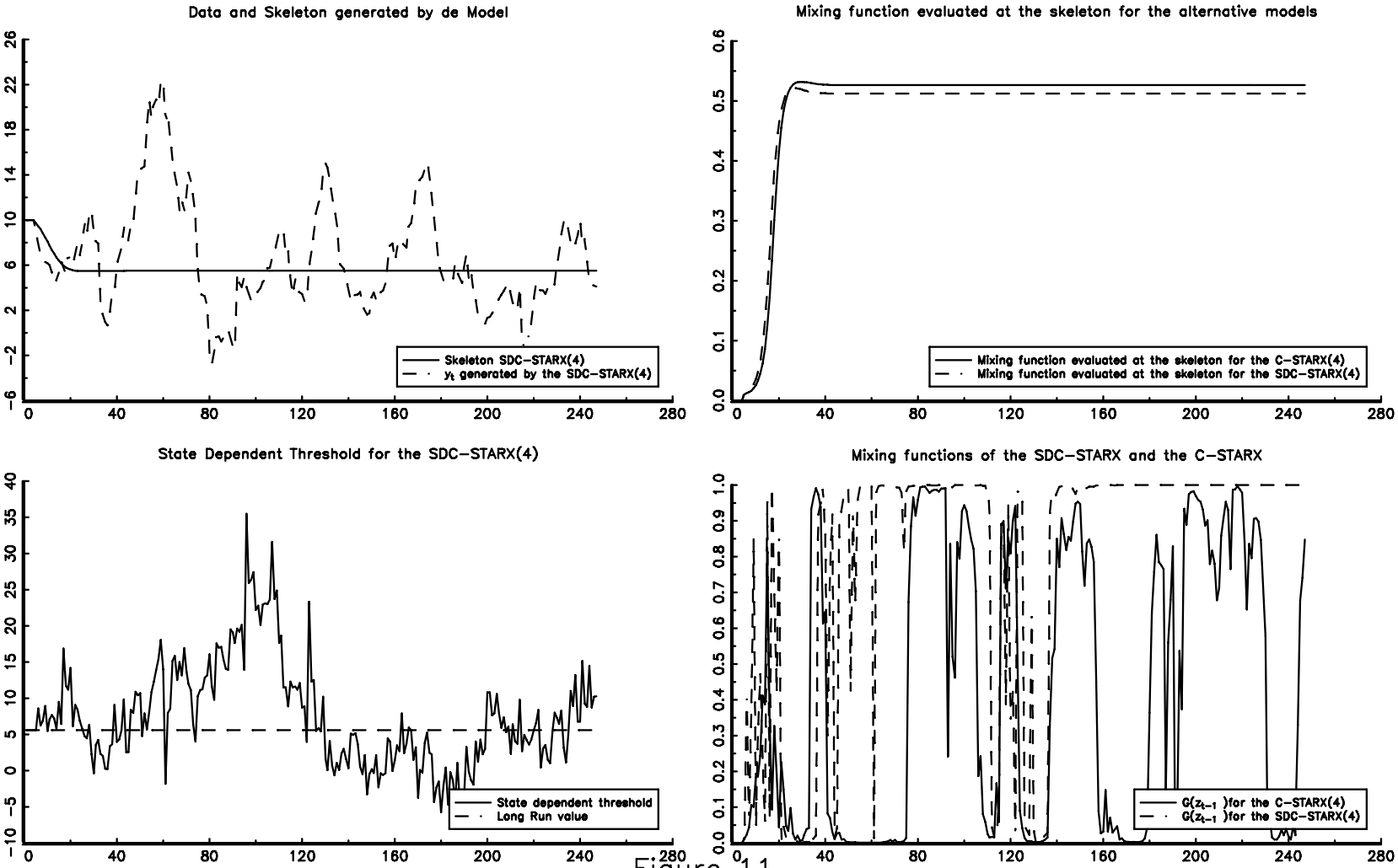

Figure 11