Multivariate Contemporaneous-Threshold Autoregressive Models[*]

Michael J. Dueker

Russell Investments, U.S.A.

Zacharias Psaradakis

Department of Economics, Mathematics & Statistics, Birkbeck, University of London, U.K.

Martin Sola

Department of Economics, Mathematics & Statistics, Birkbeck, University of London, U.K.

Department of Economics, Universidad Torcuato di Tella, Argentina

Fabio Spagnolo

Department of Economics and Finance, Brunel University, U.K.

This version: March 2010

First draft: April 2007

1

Abstract

This paper proposes a contemporaneous-threshold multivariate smooth transition autoregressive (C-MSTAR) model in which the regime weights depend on the ex ante probabilities that latent regime-specific variables exceed certain threshold values. A key feature of the model is that the transition function depends on all the parameters of the model as well as on the data. Since the mixing weights are also a function of the regime-specific innovation covariance matrix, the model can account for contemporaneous regime-specific co-movements of the variables. The stability and distributional properties of the proposed model are discussed, as well as issues of estimation, testing and forecasting. The practical usefulness of the C-MSTAR model is illustrated by examining the relationship between US stock prices and interest rates.

Keywords: Nonlinear autoregressive model; Smooth transition; Stability; Threshold.

JEL Classification: C32; G12.

1 Introduction

It has been long recognized that economic variables may behave very differently in different states of the economy such as, for example, high/low inflation, high/low growth, or high/low stock prices (relative to dividends). This behavior may be attributable not only to state-dependent response of economic variables to policy shocks but also to state-dependent response on the part of authorities responsible for fiscal and monetary policies. In an attempt to capture state-dependent or regime-switching behavior, a variety of nonlinear models have been proposed for describing the dynamics of economic time series subject to changes in regime [see, e.g., Tong (1983, 1990); Hamilton (1993); van Dijk et al. (2002); Dueker et al. (2007)].

Researchers are often interested in studying the interrelationships between several economic/financial variables. To this end, several multivariate models have been considered in the literature, including Markov-switching autoregressive (MSAR) models [e.g., Sola and Driffill (1994)], threshold autoregressive (TAR) models [Tsay (1998)], smooth transition autoregressive (STAR) models [van Dijk et al. (2002)], functional-coefficient autoregressive models [Harvill and Ray (2006)], and mixture autoregressive models [Fong et al. (2007); Bec et al. (2008)]. In spite of some obvious difficulties associated with the practical use of many of these models (e.g., choice of an appropriate threshold variable, number of regimes, transition function, functional forms), they are potentially very useful for analyzing state-dependent multivariate relationships. Well-known examples of such relationships, which have been the focus of recent research, are nonlinear money-output Granger causality patterns [e.g., Rothman et al. (2001); Psaradakis et al. (2005)], nonlinearities in the term structure of interest rates [e.g., Sola and Driffill (1994); Tsay (1998); De Gooijer and Vidiella-i-Anguera (2004)], and nonlinearities in business-cycle relationships [e.g., Altissimo and Violante (2001); Koop and Potter (2006)], inter alia.

One of the major challenges faced in a multivariate framework is how best to capture the state-dependent behavior that the components of a multiple time series may exhibit, as well as the potentially changing interrelationships between the variables, in a way which is both statistically sound and economically meaningful. In many instances, different states of the economy can be characterized in terms of high and low values of certain

economic/financial variables (e.g., high/low inflation or high/low growth). The economy typically behaves differently in these regimes and it is reasonable to expect that the contemporaneous and feedback relationships between variables will also be regime specific. An econometric model will be useful in such cases if it is capable of both identifying the periods associated with different states of nature and capturing the state-specific interrelationships among variables. An MSAR model, for example, which allows for shifts in the mean or the intercept can capture extreme events associated with the level of the series but cannot account for state-dependent interrelationships among the variables. The latter may be accounted for by allowing all the parameters of the MSAR to switch, but this usually results in identifying as separate regimes periods which do not necessarily correspond to economically meaningful states of nature (e.g., high/low growth rates). Multivariate TAR and STAR models typically associate different regimes with small and large values of the transition variables and are capable of characterizing state-dependent interactions among the variables.

This paper contributes to the literature on multivariate nonlinear models by proposing a contemporaneous-threshold multivariate STAR, or C-MSTAR, model. A key characteristic of this model is that the mixing (or regime) weights depend on the ex ante probabilities that latent regime-specific variables exceed certain (unknown) threshold values [cf. Dueker et al. (2007)]. What is more, the mixing (or transition) function of the C-MSTAR model depends on all the parameters of the model as well as on the data. This implies that, contrary to conventional STAR models, there is no need to choose an appropriate transition variable using a model selection criterion since, by construction, all the variables that enter the model's information set are also present in the transition function. Furthermore, the dependence of the mixing weights on the regime-specific innovation covariance matrices allows the model to capture contemporaneous regime-specific co-movements of the variables and to exploit the information in these covariance matrices in order to predict regimes. These important characteristics make the C-MSTAR model capable of describing successfully multiple time series with a wide variety of conditional distributions and of capturing state-dependent interrelationships among the variables of interest.

To convey the flavor of contemporaneous-threshold smooth transition autoregressive (C-STAR) models, the definition and main characteristics of the univariate model is re-

called in Section 2. The C-MSTAR model is introduced and discussed in Section 3. We examine the stability properties of the model and use artificial data to analyze the various types of conditional distributions that can be generated by a C-MSTAR model. Section 4 discusses estimation and testing, and reports the results of simulation experiments that assess the finite-sample performance of the maximum likelihood (ML) estimator and of the related statistics. In Section 5, we investigate the relationship between US stock prices and interest rates using a C-MSTAR model and evaluate its out-of-sample forecast performance. Our empirical results suggest that monetary policy has different effects on stock prices in different states of the economy and that Granger causality between stock prices and interest rates is regime dependent. A summary is given in Section 6.

2   Univariate Contemporaneous-Threshold Models

The C-STAR model of Dueker et al. (2007) is a member of the STAR family. A STAR process may be thought of as a function of two (or more) autoregressive processes which are averaged, at any given point in time, according to some continuous function $G(\cdot)$ taking values in $[0, 1]$. More specifically, a two-regime (conditionally heteroskedastic) STAR model for the univariate time series $\{x_t\}$ may be formulated as

$$x_t = G(\mathbf{z}_{t-1})x_{1t} + [1 - G(\mathbf{z}_{t-1})]x_{2t}, \quad t = 1, 2, \ldots, \tag{1}$$

where $\mathbf{z}_{t-1}$ is a vector of exogenous and/or pre-determined variables and

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

In (2), p is a positive integer, $\{u_t\}$ are independent and identically distributed (i.i.d.) random variables such that u_t is independent of $(x_{1-p}, \ldots, x_0)$ and $\mathbb{E}(u_t) = \mathbb{E}(u_t^2 - 1) = 0$, σ_1 and σ_2 are positive constants, and μ_i and $\alpha_j^{(i)}$ $(i = 1, 2; j = 1, \ldots, p)$ are real constants. The feature that differentiates alternative STAR models is the choice of the mixing function $G(\cdot)$ and transition variables $\mathbf{z}_{t-1}$ [cf. Teräsvirta (1998); van Dijk et al. (2002)].

Letting $\mathbf{z}_{t-1} = (x_{t-1}, ..., x_{t-p})'$ and $\boldsymbol{\alpha}_i = (\alpha_1^{(i)}, \ldots, \alpha_p^{(i)})'$ $(i = 1, 2)$, the (conditionally) Gaussian two-regime C-STAR model of order p is obtained by defining the mixing function

$G(\cdot)$ in (1) as

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

where $\Phi(\cdot)$ denotes the standard normal distribution function and x^* is a threshold parameter.[1] Notice that

$$G(\mathbf{z}_{t-1}) = \frac{\mathbb{P}(x_{1t} < x^*|\mathbf{z}_{t-1}; \boldsymbol{\vartheta}_1)}{\mathbb{P}(x_{1t} < x^*|\mathbf{z}_{t-1}; \boldsymbol{\vartheta}_1) + \mathbb{P}(x_{2t} \geqslant x^*|\mathbf{z}_{t-1}; \boldsymbol{\vartheta}_2)}$$

and

$$1 - G(\mathbf{z}_{t-1}) = \frac{\mathbb{P}(x_{2t} \geq x^*|\mathbf{z}_{t-1}; \boldsymbol{\vartheta}_1)}{\mathbb{P}(x_{1t} < x^*|\mathbf{z}_{t-1}; \boldsymbol{\vartheta}_1) + \mathbb{P}(x_{2t} \geq x^*|\mathbf{z}_{t-1}; \boldsymbol{\vartheta}_2)},$$

where $\boldsymbol{\vartheta}_i = (\mu_i, \alpha_1^{(i)}, \ldots, \alpha_p^{(i)}, \sigma_i^2)'$ is the vector of parameters associated with regime i. Hence, (1) may be rewritten as

$$x_t = \frac{\mathbb{P}(x_{1t} < x^*|\mathbf{z}_{t-1}; \boldsymbol{\vartheta}_1)x_{1t} + \mathbb{P}(x_{2t} \geq x^*|\mathbf{z}_{t-1}; \boldsymbol{\vartheta}_2)x_{2t}}{\mathbb{P}(x_{1t} < x^*|\mathbf{z}_{t-1}; \boldsymbol{\vartheta}_1) + \mathbb{P}(x_{2t} \geq x^*|\mathbf{z}_{t-1}; \boldsymbol{\vartheta}_2)}.$$

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

---

[1] Although conditional Gaussianity is used as a convenient assumption in much of what follows, $\Phi(\cdot)$ can in principle be replaced with another continuous distribution function.

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

One of the main purposes of the C-STAR model is to address two somewhat arbitrary features of conventional STAR models. First, STAR models specify a delay such that the mixing function for period t consists of a function of x_{t-j} for some $j \geq 1$. Second, STAR models specify which of and in what way the model parameters enter the mixing function. C-STAR models address these twin issues in an intuitive way: they use a forecasting function such that the mixing function depends on the ex ante regime-dependent probabilities that x_t will exceed the threshold value(s). Furthermore, the mixing function makes use of all of the model parameters in a coherent way.

3 Multivariate Contemporaneous-Threshold Models

In this section, we present a C-MSTAR model which is capable of both separating different regimes in terms of the probability of regime-specific latent variables being greater (or smaller) than relevant thresholds and allowing the interaction and feedback relationships between variables to differ between regimes. We begin by defining the model and then proceed to investigate some of its properties.

3.1 Definition

The C-MSTAR model belongs to the class of multivariate STAR models. An n-variate (conditionally heteroskedastic) STAR process $\{\mathbf{y}_t\}$ with m regimes may be defined as

$$\mathbf{y}_t = \sum_{i=1}^{m} G_i(\mathbf{z}_{t-1})\mathbf{y}_{it}, \quad t = 1, 2, \ldots, \tag{3}$$

where $G_i(\cdot)$ $(i = 1, \ldots, m)$ are continuous functions taking values in $[0, 1]$, $\mathbf{z}_{t-1}$ is a vector of exogenous and/or pre-determined variables, and

$$\mathbf{y}_{it} = \boldsymbol{\mu}_i + \sum_{j=1}^{p} \mathbf{A}_j^{(i)}\mathbf{y}_{t-j} + \boldsymbol{\Sigma}_i^{1/2}\mathbf{u}_t, \quad i = 1, \ldots, m. \tag{4}$$

In (4), p is a positive integer, $\{\mathbf{u}_t\}$ is a sequence of i.i.d. n-dimensional random vectors, with $\mathbb{E}(\mathbf{u}_t) = \mathbf{0}$ and $\mathbb{E}(\mathbf{u}_t\mathbf{u}_t') = \mathbf{I}_n$ ($\mathbf{I}_n$ being the $n \times n$ identity matrix), such that $\mathbf{u}_t$ is independent of $(\mathbf{y}_{1-p}, \ldots, \mathbf{y}_0)$, $\boldsymbol{\mu}_i$ $(i = 1, \ldots, m)$ are n-dimensional vectors of intercepts, $\mathbf{A}_j^{(i)}$ $(i = 1, \ldots, m; j = 1, \ldots, p)$ are $n \times n$ coefficient matrices, and $\boldsymbol{\Sigma}_i$ $(i = 1, \ldots, m)$ are symmetric, positive definite $n \times n$ matrices.[3]

---

[3]For a symmetric, positive definite matrix $\mathbf{C}$, $\mathbf{C}^{1/2}$ denotes its symmetric, positive definite square root.

For simplicity and clarity of exposition, we shall focus hereafter on the bivariate first-order C-MSTAR model, i.e., the case when $n = 2$, $m = 4$, and $p = 1$. To define this model, let

$$\mathbf{y}_t = (x_t, w_t)', \qquad \mathbf{y}_{it} = (x_{it}, w_{it})', \quad i = 1, \ldots, 4,$$

$$\mathbf{y}_1^* = (x^*, w^*)', \qquad \mathbf{y}_2^* = (x^*, -w^*)', \qquad \mathbf{y}_3^* = (-x^*, w^*)', \qquad \mathbf{y}_4^* = (-x^*, -w^*)',$$

where x^* and w^* are threshold parameters, and x_{it} and w_{it} $(i = 1, \ldots, 4)$ are latent regime-specific random variables. Then, $\{\mathbf{y}_t\}$ is said to follow a (conditionally) Gaussian first-order C-MSTAR model if it satisfies (3)–(4) with $\mathbf{u}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_2)$, $\mathbf{z}_{t-1} = \mathbf{y}_{t-1}$, and

$$G_i(\mathbf{z}_{t-1}) = (1/\delta_t)\Phi_2(\boldsymbol{\Sigma}_i^{-1/2}\{\mathbf{y}_i^* - \boldsymbol{\mu}_i - \mathbf{A}_1^{(i)}\mathbf{y}_{t-1}\}), \quad i = 1, \ldots, 4, \tag{5}$$

where $\Phi_2(\cdot)$ denotes the $\mathcal{N}(\mathbf{0}, \mathbf{I}_2)$ distribution function and

$$\delta_t = \sum_{i=1}^4 \Phi_2(\boldsymbol{\Sigma}_i^{-1/2}\{\mathbf{y}_i^* - \boldsymbol{\mu}_i - \mathbf{A}_1^{(i)}\mathbf{y}_{t-1}\}). \tag{6}$$

It can be readily seen that

$$G_1(\mathbf{z}_{t-1}) = (1/\delta_t)\mathbb{P}(x_{1t} < x^*, w_{1t} < w^*|\mathbf{y}_{t-1}; \boldsymbol{\theta}_1),$$

$$G_2(\mathbf{z}_{t-1}) = (1/\delta_t)\mathbb{P}(x_{2t} < x^*, w_{2t} \geq w^*|\mathbf{y}_{t-1}; \boldsymbol{\theta}_2),$$

$$G_3(\mathbf{z}_{t-1}) = (1/\delta_t)\mathbb{P}(x_{3t} \geq x^*, w_{3t} < w^*|\mathbf{y}_{t-1}; \boldsymbol{\theta}_3),$$

$$G_4(\mathbf{z}_{t-1}) = (1/\delta_t)\mathbb{P}(x_{4t} \geq x^*, w_{4t} \geq w^*|\mathbf{y}_{t-1}; \boldsymbol{\theta}_4),$$

where $\boldsymbol{\theta}_i = (\boldsymbol{\mu}_i', \text{vec}(\mathbf{A}_1^{(i)})', \text{vech}(\boldsymbol{\Sigma}_i)')'$ is the vector of parameters associated with regime i. Hence the mixing functions $G_i(\cdot)$ reflect the weighted probabilities that the regime-specific latent variables x_{it} and w_{it} are above or below the respective thresholds x^* and w^*.

The first-order model above can be generalized straightforwardly to the case of $p \geq 2$ lags. Furthermore, although we do not pursue this modelling strategy here, the number of lags in (4) may be allowed to differ over i and thus be regime-specific.[4] Regarding the number of regimes m, it should be remembered that m is always determined by the dimension n of the C-MSTAR model. When $n = 2$, we have $m = 4$ by construction since

---

[4]In either case, the number of lags may be selected adaptively by using complexity-penalized likelihood criteria [see Kapetanios (2001) and Psaradakis and Spagnolo (2006) for related results concerning univariate nonlinear autoregressive models].

there are four possible states of nature defined by the regime-specific latent variables and the thresholds, namely $\{x_{1t} < x^*, w_{1t} < w^*\}$, $\{w_{2t} < w^*, w_{2t} \geq w^*\}$, $\{w_{3t} \geq w^*, w_{3t} < w^*\}$, and $\{w_{4t} \geq w^*, w_{4t} \geq w^*\}$. For a model with $n = 3$, we have $m = 9$, and so on.[5] Finally, as in the univariate case, a (conditionally) non-Gaussian model can be obtained by replacing $\Phi_2(\cdot)$ in (5)–(6) by the distribution function $\Psi(\cdot)$, say, of another continuous distribution (having mean vector $\mathbf{0}$ and covariance matrix $\mathbf{I}_2$). The interpretation of the C-MSTAR model remains the same as long as $\mathbf{u}_t$ is assumed to be distributed according to $\Psi(\cdot)$.

3.2 Distributional Characteristics

To gain an understanding of the behavior of C-MSTAR time series, we illustrate some properties of the C-MSTAR model by using artificial data obtained from the data-generating processes (DGPs) given in Table 1. These DGPs have been chosen to highlight some relevant features of the model with respect to: (i) the response of the mixing function to changes in the parameters of the model; and (ii) the empirical distribution of C-MSTAR data. The errors $\mathbf{u}_t$ are contemporaneously uncorrelated under DGP-1, while DGP-2 and DGP-3 allow for positive and negative contemporaneous correlation, respectively.

Figure 1 shows the conditional density functions of the latent regime-specific random vectors $\mathbf{y}_{it}$ ($i = 1, \ldots, 4$) for DGP-1, given $\mathbf{y}_{t-1} = (0.4, 0.6)'$, along with the threshold $\mathbf{y}_1^* = (0.4, 0.6)'$ and the values of the mixing functions $G_i(\mathbf{y}_{t-1})$. Each plot shows the relevant area of the density (suitably rotated) for which each regime is defined. The regime-specific conditional means are $\mathbb{E}(\mathbf{y}_{1t}|\mathbf{y}_{t-1}) = (0.35, 0.57)'$, $\mathbb{E}(\mathbf{y}_{2t}|\mathbf{y}_{t-1}) = (0.29, 0.6)'$, $\mathbb{E}(\mathbf{y}_{3t}|\mathbf{y}_{t-1}) = (0.59, 0.39)'$, and $\mathbb{E}(\mathbf{y}_{4t}|\mathbf{y}_{t-1}) = (0.43, 0.66)'$. It can be seen that the values of the mixing weights $G_i(\mathbf{y}_{t-1})$ depend on the values of the regime-specific conditional means relative to

---

[5]Needless to say, the number of parameters in an C-MSTAR model increases considerably with the dimension of the model, and hence with the number of regimes (a problem which is, of course, common to many of the multiple-regime multivariate models mentioned in Section 1). One way of dealing with this difficulty may be to allow only some of the components of $\mathbf{y}_{it}$ in (4) to have regime-specific dynamics. To give an example, suppose that $\mathbf{y}_t = (x_t, w_t, r_t)'$, where x_t is output growth, w_t is inflation and r_t is the change in the exchange rate; since periods of high inflation are likely to coincide with periods of devaluation, one might allow the dynamics of output and of only one of the other two variables to be regime-specific. An alternative approach may be to consider a two-regime model in which the regimes are defined in terms of a linear combination of the latent variables being greater (or smaller) than a linear combination of the thresholds. The former approach has the advantage that the regimes have a clear economic interpretation.

the threshold. More specifically, the larger the area of the conditional distribution which lies above the threshold is, the larger $G_i(\mathbf{y}_{t-1})$ is. In our example, we have $G_1(\mathbf{y}_{t-1}) = 0.09$, $G_2(\mathbf{y}_{t-1}) = 0.48$, $G_3(\mathbf{y}_{t-1}) = 0.09$, and $G_4(\mathbf{y}_{t-1}) = 0.34$.

Conditioning on $\mathbf{y}_{t-1} = (-1.5, -2)'$ yields the density functions shown in Figure 2. The regime-specific conditional means now are $\mathbb{E}(\mathbf{y}_{1t}|\mathbf{y}_{t-1}) = (-1.44, -1.97)'$, $\mathbb{E}(\mathbf{y}_{2t}|\mathbf{y}_{t-1}) = (-1.26, -1.97)'$, $\mathbb{E}(\mathbf{y}_{3t}|\mathbf{y}_{t-1}) = (-1.37, -1.35)'$, and $\mathbb{E}(\mathbf{y}_{4t}|\mathbf{y}_{t-1}) = (-1.31, -1.59)'$. The mixing functions take the values $G_1(\mathbf{y}_{t-1}) = 0.88$, $G_2(\mathbf{y}_{t-1}) = 0.1$, $G_3(\mathbf{y}_{t-1}) = 0.02$, and $G_4(\mathbf{y}_{t-1}) = 0$. It is not surprising that the regime associated with $G_1(\cdot)$ is now the most prominent regime since the distance of $\mathbb{E}(\mathbf{y}_{1t}|\mathbf{y}_{t-1})$ from each of the thresholds is about one standard deviation.

The results for DGP-2 and DGP-3 can be summarized as follows. When we condition on $\mathbf{y}_{t-1} = (0.4, 0.6)'$, the values of the mixing functions do not change substantially as a result of the change in the shape of the conditional distributions (the relevant plots are omitted to save space). We have $G_1(\mathbf{y}_{t-1}) = 0$, $G_2(\mathbf{y}_{t-1}) = 0.52$, $G_3(\mathbf{y}_{t-1}) = 0.11$, and $G_4(\mathbf{y}_{t-1}) = 0.36$ under DGP-2 (positive contemporaneous correlation), while $G_1(\mathbf{y}_{t-1}) = 0$, $G_2(\mathbf{y}_{t-1}) = 0.54$, $G_3(\mathbf{y}_{t-1}) = 0.07$, and $G_4(\mathbf{y}_{t-1}) = 0.38$ under DGP-3 (negative contemporaneous correlation). Interestingly, the change in the sign of the correlation coefficient results in marginal changes in the values of the mixing functions; it is the location of the conditional means relative to the thresholds and the dispersion of the conditional densities that are of primary importance as far as the mixing weights are concerned. Similar results are obtained when we condition on $\mathbf{y}_{t-1} = (-1.5, -2)'$.

3.3 Stability

3.3.1 Probabilistic Properties

In this subsection we examine some probabilistic properties of the C-MSTAR model. Specifically, we give conditions under which the C-MSTAR model is stable in the sense of having a Markovian representation which is geometrically ergodic.[6] For simplicity and clarity of exposition, the discussion is once again focused on the Gaussian, bivariate, first-order C-MSTAR model.

---

[6] For a comprehensive account of the stability and convergence theory of Markov chains the interested reader is referred to Meyn and Tweedie (2009).

The stability concept considered here is that of Q-geometric ergodicity of a Markov chain introduced by Liebscher (2005). To recall the definition of this concept, suppose that $\{\boldsymbol{\xi}_t\}_{t\geq 0}$ is a Markov chain on a general state space $\mathcal{S}$ with k-step transition probability kernel $P^{(k)}(\cdot,\cdot)$ and an invariant distribution $\pi(\cdot)$, so that $P^{(k)}(\mathbf{v},B) = \mathbb{P}(\boldsymbol{\xi}_k \in B|\boldsymbol{\xi}_0 = \mathbf{v})$ and $\pi(B) = \int_{\mathcal{S}} P^{(1)}(\mathbf{v},B)\pi(\mathrm{d}\mathbf{v})$ for any Borel set B in $\mathcal{S}$ and $\mathbf{v} \in \mathcal{S}$. Then $\{\boldsymbol{\xi}_t\}$ is said to be Q-geometrically ergodic if there exists a non-negative function $Q(\cdot)$ on $\mathcal{S}$ satisfying $\int_{\mathcal{S}} Q(\mathbf{v})\pi(\mathrm{d}\mathbf{v}) < \infty$ and positive constants a_1, a_2 and $\gamma < 1$ such that, for all $\mathbf{v} \in \mathcal{S}$,

$$\left\| P^{(k)}(\mathbf{v},\cdot) - \pi(\cdot) \right\|_{\tau} \leq \{a_1 + a_2 Q(\mathbf{v})\}\gamma^k, \quad k = 1, 2, \ldots,$$

where $\|\cdot\|_{\tau}$ denotes the total variation norm.[7]

Geometric ergodicity entails that the total variation distance between the probability measures $P^{(k)}(\mathbf{v},\cdot)$ and $\pi(\cdot)$ converges to zero geometrically fast (as $k \to \infty$) for all $\mathbf{v} \in \mathcal{S}$. It is well known that, if the initial value $\boldsymbol{\xi}_0$ of the Markov chain has distribution $\pi(\cdot)$, then geometric ergodicity implies strict stationarity of $\{\boldsymbol{\xi}_t\}$. Furthermore, provided that $\boldsymbol{\xi}_0$ is such that $Q(\boldsymbol{\xi}_0)$ is integrable with respect to $\pi(\cdot)$, Q-geometric ergodicity implies that $\{\boldsymbol{\xi}_t\}$ is Harris ergodic (i.e., aperiodic, irreducible and positive Harris recurrent), as well as absolutely regular (or β-mixing) with a geometric mixing rate [see Liebscher (2005, Proposition 4)]. Such ergodicity and mixing properties are of much importance for the purposes of statistical inference in dynamic models since they ensure the validity of many conventional limit theorems [see, e.g., Doukhan (1994)].

To give sufficient conditions for Q-geometric ergodicity of a C-MSTAR process, the concept of the joint spectral radius of a set of matrices is needed. Suppose that $\mathcal{C}$ is a set of real square matrices and let $\mathcal{C}_h$ be the set of all products of length $h \geq 1$ of the elements of $\mathcal{C}$. Then the joint spectral radius of $\mathcal{C}$ is defined as

$$\rho(\mathcal{C}) = \limsup_{h\to\infty} \left(\sup_{\mathbf{C}\in\mathcal{C}_h} \|\mathbf{C}\| \right)^{1/h}, \tag{7}$$

where $\|\cdot\|$ is an arbitrary matrix norm. We note that the value of $\rho(\mathcal{C})$ is independent of the choice of matrix norm and that, if the set $\mathcal{C}$ trivially consists of a single matrix, then $\rho(\mathcal{C})$ coincides with the ordinary spectral radius (i.e., the maximal modulus of the

---

[7]Note that $\left\| P^{(k)}(\mathbf{v},\cdot) - \pi(\cdot) \right\|_{\tau} = 2\sup_B \left| P^{(k)}(\mathbf{v},B) - \pi(B) \right|$.

eigenvalues of the matrix).[8]

It is easy to see that the first-order C-MSTAR model defined by (3)–(6) belongs to the family of models studied by Liebscher (2005). By appealing to Theorem 2 and Proposition 5 in that paper, the following proposition is readily established.[9] Here and in the sequel, $\|\cdot\|$ is used to denote the Euclidean vector norm and its subordinate matrix norm (i.e., $\|\mathbf{v}\| = (\mathbf{v}'\mathbf{v})^{1/2}$ and $\|\mathbf{C}\| = \sup_{\|\mathbf{v}\|=1} \|\mathbf{C}\mathbf{v}\|$, for an n-dimensional vector $\mathbf{v}$ and an $n \times n$ matrix $\mathbf{C}$).

Proposition 1 *Suppose that, for every compact subset B of $\mathbb{R}^2$, there exist positive constants b_1 and b_2 such that $\left\|\mathbf{\Sigma}(\mathbf{v})^{-1}\right\| \leq b_1$ and $|\det\{\mathbf{\Sigma}(\mathbf{v})\}| \leq b_2$ for all $\mathbf{v} \in B$, where $\mathbf{\Sigma}(\mathbf{v}) = \sum_{i=1}^{4} G_i(\mathbf{v})\mathbf{\Sigma}_i^{1/2}$. If, in addition, the set $\mathcal{A} = \{\mathbf{A}_1^{(1)}, \mathbf{A}_1^{(2)}, \mathbf{A}_1^{(3)}, \mathbf{A}_1^{(4)}\}$ is such that $\rho(\mathcal{A}) < 1$, then the first-order C-MSTAR process $\{\mathbf{y}_t\}$ is a Q-geometrically ergodic Markov chain with $Q(\mathbf{v}) = \|\mathbf{v}\|$.*

It follows from our earlier discussion that $\rho(\mathcal{A}) < 1$ guarantees the existence of a unique invariant distribution for $\{\mathbf{y}_t\}$ with respect to which $\mathbb{E}(\|\mathbf{y}_t\|) < \infty$; furthermore, if $\{\mathbf{y}_t\}$ is initialized from this invariant distribution, then it is strictly stationary, as well as absolutely regular and hence ergodic (in the sense of ergodic theory). We also note that the conclusion of Proposition 1 remains true for a non-Gaussian C-MSTAR model in which the distribution of the noise $\mathbf{u}_t$ admits a positive Lebesgue density on $\mathbb{R}^2$.

Finally, it worth pointing out that Liebscher's (2005) approach, which we have followed here, delivers conditions for geometric ergodicity which can sometimes be weaker than alternative sufficient conditions [cf. Liebscher (2005, p. 682)]. A practical difficulty, however, is that exact or approximate computation of the joint spectral radius of a set of matrices is not an easy task, not even in the simplest non-trivial case of a two-element set [see, e.g., Tsitsiklis and Blondel (1997)].[10] One possibility is to use the algorithm presented in Gripenberg (1996) to obtain an arbitrarily small interval within which the

---

[8] Also note that the norm of $\mathbf{C}$ in the definition of $\rho(\mathcal{C})$ in (7) may be replaced by the spectral radius of $\mathbf{C}$ as long as $\mathcal{C}$ is a finite or bounded set.

[9] It can be easily seen that that, under the conditions of Proposition 1 below, the nonlinear functions that specify the conditional mean and conditional variance of $\mathbf{y}_t$, given $\mathbf{y}_{t-1}$, satisfy the assumptions in Section 4 of Liebscher (2005).

[10] The problem of determining whether $\rho(\mathcal{A}) < 1$ is, in fact, known to be NP-hard, that is it cannot be solved in a number of steps which is a polynomial function of the size of $\mathcal{A}$. It should also be remembered

12

joint spectral radius of $\mathcal{A}$ lies. An alternative approach, which may also provide useful information about the model in cases where the condition of Proposition 1 is not fulfilled, is to use simulation methods to investigate the properties of the skeleton of the C-MSTAR model. We turn our attention to this topic next.

3.3.2   Skeleton of the Model

As shown by Chan and Tong (1985), the stability properties of a nonlinear dynamic model may be analyzed by considering the noiseless part, or skeleton, of the model alone [see also Tong (1990)]. The skeleton of the bivariate first-order C-MSTAR model is the dynamic system

$$\overline{\mathbf{y}}_t = \mathbf{f}(\overline{\mathbf{y}}_{t-1}, \boldsymbol{\theta}), \tag{8}$$

where

$$\mathbf{f}(\overline{\mathbf{y}}_{t-1}, \boldsymbol{\theta}) = \sum_{i=1}^{4} G_i(\overline{\mathbf{y}}_{t-1})(\boldsymbol{\mu}_i + \mathbf{A}_1^{(i)}\overline{\mathbf{y}}_{t-1}) \tag{9}$$

and $\boldsymbol{\theta}$ denotes the vector of all the parameters of the model. A fixed point of the skeleton is any two-dimensional vector $\overline{\mathbf{y}}_e$ satisfying the equation

$$\mathbf{f}(\overline{\mathbf{y}}_e, \boldsymbol{\theta}) = \overline{\mathbf{y}}_e, \tag{10}$$

and $\overline{\mathbf{y}}_e$ is said to be an equilibrium point of the C-MSTAR model. Since the model is nonlinear, there may, of course, exist one, several or no equilibrium points satisfying (10). By a first-order Taylor expansion of $\mathbf{f}(\overline{\mathbf{y}}_{t-1}, \boldsymbol{\theta})$ about the point $\overline{\mathbf{y}}_e$, we have

$$\overline{\mathbf{y}}_t - \overline{\mathbf{y}}_e = \mathbf{f}(\overline{\mathbf{y}}_{t-1}, \boldsymbol{\theta}) - \mathbf{f}(\overline{\mathbf{y}}_e, \boldsymbol{\theta}) \approx \mathbf{D}(\overline{\mathbf{y}}_e)'(\overline{\mathbf{y}}_{t-1} - \overline{\mathbf{y}}_e), \tag{11}$$

where

$$\mathbf{D}(\overline{\mathbf{y}}_e) = \left. \frac{\partial \mathbf{f}(\overline{\mathbf{y}}_{t-1}, \boldsymbol{\theta})}{\partial \overline{\mathbf{y}}_{t-1}} \right|_{\overline{\mathbf{y}}_{t-1}=\overline{\mathbf{y}}_e}. \tag{12}$$

Thus, an examination of the local stability of each equilibrium point $\overline{\mathbf{y}}_e$ may be carried out by considering the spectrum of $\mathbf{D}(\overline{\mathbf{y}}_e)$. More specifically, if the spectral radius of $\mathbf{D}(\overline{\mathbf{y}}_e)$ is less than unity, then the equilibrium is locally stable and $\overline{\mathbf{y}}_t$ is a contraction in a neighborhood of $\overline{\mathbf{y}}_e$.

that the condition that each of the matrices in $\mathcal{A}$ has spectral radius less than unity is necessary but not sufficient for $\rho(\mathcal{A}) < 1$. A summary of some of the methods available for computing or approximating the joint spectral radius of a set of matrices can be found in Jungers (2009).

It can be readily verified that

$$\frac{\partial \mathbf{f}(\overline{\mathbf{y}}_{t-1}, \boldsymbol{\theta})}{\partial \overline{\mathbf{y}}_{t-1}} = \sum_{i=1}^{4} \left\{ \frac{\partial G_i(\overline{\mathbf{y}}_{t-1})}{\partial \overline{\mathbf{y}}_{t-1}} (\boldsymbol{\mu}_i + \mathbf{A}_1^{(i)}\overline{\mathbf{y}}_{t-1})' + G_i(\overline{\mathbf{y}}_{t-1})(\mathbf{A}_1^{(i)})' \right\} \tag{13}$$

and

$$\frac{\partial G_i(\overline{\mathbf{y}}_{t-1})}{\partial \overline{\mathbf{y}}_{t-1}} = \frac{1}{\delta_t^2} \left\{ -\delta_t (\boldsymbol{\Sigma}_i^{-1/2}\mathbf{A}_1^{(i)})'\nabla\Phi_2(\mathbf{v}_i) + \Phi_2(\mathbf{v}_i)\sum_{i=1}^{4}(\boldsymbol{\Sigma}_i^{-1/2}\mathbf{A}_1^{(i)})'\nabla\Phi_2(\mathbf{v}_i) \right\}, \tag{14}$$

where $\mathbf{v}_i = \boldsymbol{\Sigma}_i^{-1/2}(\mathbf{y}_i^* - \boldsymbol{\mu}_i - \mathbf{A}_1^{(i)}\overline{\mathbf{y}}_{t-1})$ and $\nabla\Phi_2(\mathbf{v}_i)$ is the gradient vector of $\Phi_2(\cdot)$ at $\mathbf{v}_i$.[11]

3.3.3 Numerical Examples

A wide variety of empirical distributions and time series can be generated by an C-MSTAR model. In Figure 3 we show, using the DGP-1 presented in Table 1, typical data generated according to a first-order C-MSTAR model, the corresponding mixing functions $G_i(\mathbf{y}_{t-1})$, and the skeleton $\overline{\mathbf{y}}_t$.[12]

When the covariance matrix of the noise is diagonal (DGP-1), the data appear to take values which correspond to all the regimes. When, on the other hand, there is positive contemporaneous correlation (DGP-2), the generated data assume values which are mostly associated with regimes 1 and 4 (corresponding to $G_1(\cdot)$ and $G_4(\cdot)$), while regimes 2 and 3 (associated with $G_2(\cdot)$ and $G_3(\cdot)$) appear to dominate in the presence of negative contemporaneous correlation (DGP-3).

In all three cases, the skeleton converges to its fixed point very quickly. Using numerical simulations, we found the fixed point $\overline{\mathbf{y}}_e$ to be unique for each DGP, taking the value $(0.0251, 0.2309)'$, $(0.0539, 0.3828)'$ and $(-0.1052, -0.0451)'$ for DGP-1, DGP-2 and DGP-3, respectively. To assess the stability of these fixed points, we compute the spectral radius of the matrix of partial derivatives given in (12) using the expansion in (13)–(14). The spectral radius of $\mathbf{D}(\overline{\mathbf{y}}_e)$ is 0.8357, 0.8320 and 0.8296 under DGP-1, DGP-2 and DGP-3, respectively, suggesting that the equilibrium points are locally stable. Furthermore, the

---

[11]Notice that, since $\Phi_2(\mathbf{v}_i) = \Phi(v_{1i})\Phi(v_{2i})$ for any $\mathbf{v}_i = (v_{1i}, v_{2i})' \in \mathbb{R}^2$, $\nabla\Phi_2(\mathbf{v}_i)$ may be computed as $\nabla\Phi_2(\mathbf{v}_i) = (\phi(v_{1i})\Phi(v_{2i}), \Phi(v_{1i})\phi(v_{2i}))'$, where $\phi(\cdot)$ is the standard normal density function.

[12]The corresponding plots for DGP-2 and DGP-3 (computed using the same realizations of shocks as for DGP-1) are omitted in order to save space.

Q-geometric ergodicity condition of Proposition 1 is also satisfied for these DGPs: an application of the algorithm in Gripenberg (1996) yields $0.9366025 < \rho(\mathcal{A}) < 0.9366125$.[13]

4 Estimation and Testing

4.1 ML Parameter Estimation

As in the univariate case, once the distribution of the noise $\mathbf{u}_t$ is specified, the parameters of the C-MSTAR model can be estimated by the ML method. Letting $\Psi(\cdot)$ denote the distribution function of $\mathbf{u}_t$, we assume that $\Psi(\cdot)$ admits a positive Lebesgue density $\psi(\cdot)$. Then, for a sample $(\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_T)$ of consecutive observations from the bivariate first-order model characterized by the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2', \boldsymbol{\theta}_3', \boldsymbol{\theta}_4', x^*, w^*)' \in \boldsymbol{\Theta} \subset \mathbb{R}^{\dim(\boldsymbol{\theta})}$, we define the log-likelihood function (conditional on $\mathbf{y}_0$) as

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

where

$$\ell_t(\boldsymbol{\theta}) = \sum_{i=1}^{4} G_i(\mathbf{y}_{t-1}) \det(\boldsymbol{\Sigma}_i^{-1/2}) \psi(\boldsymbol{\Sigma}_i^{-1/2}\{\mathbf{y}_t - \boldsymbol{\mu}_i - \mathbf{A}_1^{(i)}\mathbf{y}_{t-1}\}),$$

and the mixing weights $G_i(\mathbf{y}_{t-1})$ are given by (5)–(6) with $\Psi(\cdot)$ used in the place of $\Phi_2(\cdot)$.

If $\ell_t(\boldsymbol{\theta})$ is sufficiently smooth with respect to $\boldsymbol{\theta}$ and satisfies suitable stationarity, ergodicity and moment conditions, then standard asymptotic results hold for the ML estimator $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$, obtained as the maximizer of $(1/T)\mathcal{L}_T(\boldsymbol{\theta})$ over $\boldsymbol{\Theta}$. More specifically, $\widehat{\boldsymbol{\theta}}$ is strongly consistent for the (unknown) true value $\boldsymbol{\theta}_0$ of the parameter $\boldsymbol{\theta}$ and $\{-\nabla^2\mathcal{L}_T(\widehat{\boldsymbol{\theta}})\}^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ is asymptotically normal with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{I}_2$, where $\nabla^2\mathcal{L}_T(\widehat{\boldsymbol{\theta}})$ is the Hessian matrix of $\mathcal{L}_T(\boldsymbol{\theta})$ evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$. Sufficient conditions which ensure the validity of these asymptotic results are given in Appendix A.1, along with their proof.

4.2 Finite-Sample Properties of ML

To throw some light on the sampling properties of the ML estimator of the parameters of a C-MSTAR model, we now conduct an extensive simulation study. The DGP used

---

[13]The algorithm is implemented using Gustaf Gripenberg's `MATLAB` code, which is available at `http://math.tkk.fi/~ggripenb/ggsoftwa.htm`.

in the experiments is the bivariate first-order C-MSTAR model with Gaussian errors and several parameter configurations. In order to save space, we only report results for the three parameter configurations listed in Table 1.[14]

Experiments proceed by first generating $50 + T$ data points for $\mathbf{y}_t$, with $T = 100, 200$, 400, 800, 1000, and initial 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 to estimate the parameters of the model. The ML estimate $\widehat{\boldsymbol{\theta}}$ is obtained by means of a quasi-Newton algorithm that approximates the Hessian according to the Broyden–Fletcher–Goldfarb–Shanno update computed from numerical derivatives. Approximate standard errors for the elements of $\widehat{\boldsymbol{\theta}}$ are obtained from the inverted negative Hessian matrix of the log-likelihood function evaluated at the ML estimates. Since the computation of ML estimates is time consuming (given the large number of parameters), the number of Monte Carlo replications per experiment is 2,000.

In Tables 2–4, we report some of the characteristics of the finite-sample distributions of each of the elements of $\widehat{\boldsymbol{\theta}}$.[15] 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 when $T = 200$. The size of the bias depends somewhat on the DGP. For example, while large samples are needed to reduce the bias of $\widehat{\boldsymbol{\mu}}_3$ (DGP-1 and DGP-2) and $\widehat{\boldsymbol{\mu}}_4$ (DGP-3), we find that the bias of the elements of $\widehat{\mathbf{A}}_1^{(1)}$ (DGP-1) and $\widehat{\mathbf{A}}_1^{(2)}$ (DGP-2 and DGP-3) approaches zero even for relatively small sample sizes. Overall the results show that the ML estimator is slightly biased only for the smallest sample size under consideration, and the bias clearly decreases as the sample increases, becoming negligible in most cases when $T = 800$.

As a measure of the accuracy of estimated asymptotic standard errors, 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 Tables 2–4. For most parameters, the estimated asymptotic standard errors are downward biased. These biases

---

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

[15]In order to save space, only results for $T = 200$ and $T = 800$ are reported here.

are not, however, substantial (even when $T = 200$) and should not have significant adverse effects on inference.

Finally, the Gaussianity of the finite-sample distributions of the ML estimates is assessed by means of a Kolmogorov–Smirnov goodness-of-fit test based on the difference between the empirical 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 distribution function [see Lilliefors (1967)]. As can be seen in Tables 2–4, the normality hypothesis for estimators other than $\widehat{\boldsymbol{\mu}}_3$ and $\widehat{\boldsymbol{\mu}}_4$ (DGP-1and DGP-3) and $\widehat{x}^*$ (DGP-2) cannot be rejected (at the 5% level) for sample sizes larger than 200. Furthermore, 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 $T = 200$, it is never rejected when $T = 800$.

4.3 Testing for Nonlinearity

Although a linear specification is nested within the C-MSTAR model, testing the former against the latter by means of conventional Wald, likelihood ratio or score tests is not straightforward because the threshold parameters (x^* and w^* in the bivariate case) are not identified under linearity. It is well known that in problems of this type the asymptotic distributions of conventional test statistics typically depend on unknown parameters and are non-standard. As in Dueker et al. (2007), one may, in principle, adapt Hansen's (1992) procedure to obtain asymptotic P-values for a suitably modified likelihood ratio statistic. However, the computational demands of this procedure are rather prohibitive in our multivariate setting because ML parameter estimation for each point of a grid involving a large number of parameters is required ($\dim(\boldsymbol{\theta}) = 38$ when $n = 2$).

As an alternative, we will investigate here an approach based on a general portmanteau-type test that is designed to detect nonlinearity of an unspecified type in a multivariate time series. The test in question was proposed by Harvill and Ray (1999) and is a multivariate extension of Tsay's (1986) nonlinearity test. To describe the test procedure, let $\{\mathbf{e}_t\}$ be the least-squares residuals from a pth-order vector autoregressive (VAR) model for $\{\mathbf{y}_t\}$ and $\{\mathbf{e}_t^*\}$ be the least-squares residuals from the regression of

17

the $\{np(np + 1)/2\}$-dimensional vector $\mathbf{q}_t^* = \text{vech}(\mathbf{q}_t \otimes \mathbf{q}_t')$ on the (np)-dimensional vector $\mathbf{q}_t = (\mathbf{y}_{t-1}', \ldots, \mathbf{y}_{t-p}')'$, where $\otimes$ is the Kronecker product operator. Further, let $\mathbf{S}_1$ and $\mathbf{S}_2$ be the $n \times n$ matrices of residual sum of squares and regression sum of squares, respectively, in the least-squares regression of $\mathbf{e}_t$ on $\mathbf{e}_t^*$. Then, for a sample of size T, the Harvill–Ray test statistic is given by

$$\Re = \left(\frac{bd - nc + 1}{nc}\right)\left(\frac{1 - \omega^{1/2}}{\omega^{1/2}}\right),$$

where $c = np(np+1)/2$, $b = T - p - c - np - (n - c + 1)/2$, $d = \{(n^2c^2 - 4)/(n^2 + c^2 - 5)\}^{1/2}$, and $\omega = \det(\mathbf{S}_1)/\det(\mathbf{S}_1 + \mathbf{S}_2)$. Under the null hypothesis that $\{\mathbf{y}_t\}$ follows a (zero-mean) linear pth-order VAR model, $\Re$ has asymptotically a central F-distribution with nc and $bd - (nc/2) + 1$ degrees of freedom.

To examine whether a test based on $\Re$ has power to detect nonlinearity of the C-MSTAR type, we carry out some Monte Carlo experiments. Table 5 shows the empirical rejection frequencies of the tests for C-MSTAR time series generated according to the three DGPs in Table 1. It is clear that, even for time series of relatively short length, the test based on $\Re$ has significant power to reject a linear first-order VAR specification when the data come form a C-MSTAR model.

It should be emphasized, however, that the results of a test based on $\Re$ should be interpreted with caution in an empirical setting since the test is not designed against a C-MSTAR, or any other specific nonlinear alternative model, and can be expected to have non-trivial power against a wide range of nonlinear mechanisms. However, since the test appears to be powerful enough to detect nonlinearity of the C-MSTAR type, it should be useful as part of a modelling strategy which seeks to establish the usefulness of a C-MSTAR model by first checking a simpler linear VAR model for signs of misspecification. Of course, once the linear and C-MSTAR are estimated, they can be compared by using a complexity-penalized likelihood criterion such as the well-known Akaike information criterion (AIC) or one of its many variants. Such criteria have been shown to be useful when selecting among competing multiple-regime models [see Psaradakis et al. (2009)].

5 Empirical Application

As an illustration, we analyze the low-frequency relationship between stock prices and interest rates. The interactions between asset prices and monetary policy is a topic which has attracted considerable interest in the literature [see, e.g., Bernanke and Gertler (1999, 2001) and Cecchetti et al. (2000)]. Using a C-MSTAR model, we examine the possibly different effects that monetary policy may have on stock prices in different states of the economy. An interest rate shock may, for example, have very different effects on stock markets depending on whether the price-earnings ratio is (perceived to be) high or low. Our approach explicitly allows for four different regimes, which are associated with: (i) low price-earning ratio, low interest rates; (ii) low price-earning ratio, high interest rates; (iii) high price-earning ratio, low interest rates; and (iv) high price-earning ratio, high interest rates.

5.1 A C-MSTAR Model for Stock Prices and Interest Rates

Let S_t and R_t denote the ratio of stock prices to earnings per share and the nominal interest rate, respectively. Further, let $s_t = S_t - \mu_s$ and $r_t = R_t - \mu_r$ denote the deviation of the two variables from their respective means. Putting $\mathbf{y}_t = (s_t, r_t)'$, our analysis is based on the C-MSTAR model

$$\mathbf{y}_t = \sum_{i=1}^{4} G_i(\mathbf{y}_{t-1})\mathbf{y}_{it}, \tag{15}$$

where $\mathbf{y}_{it} = (s_{it}, r_{it})'$ are latent regime-specific random vectors satisfying

$$\mathbf{y}_{it} = \boldsymbol{\mu}_i + \mathbf{A}_1^{(i)}\mathbf{y}_{t-1} + \boldsymbol{\Sigma}_i^{1/2}\mathbf{u}_t, \quad i = 1, \ldots, 4. \tag{16}$$

In (15)–(16),

$$G_1(\mathbf{y}_{t-1}) = (1/\delta_t)\mathbb{P}(s_{1t} < s^*, r_{1t} < r^*|\mathbf{y}_{t-1}; \boldsymbol{\theta}_1),$$

$$G_2(\mathbf{y}_{t-1}) = (1/\delta_t)\mathbb{P}(s_{2t} < s^*, r_{2t} \geq r^*|\mathbf{y}_{t-1}; \boldsymbol{\theta}_2), \tag{17}$$

$$G_3(\mathbf{y}_{t-1}) = (1/\delta_t)\mathbb{P}(s_{3t} \geq s^*, r_{3t} < r^*|\mathbf{y}_{t-1}; \boldsymbol{\theta}_3),$$

$$G_4(\mathbf{y}_{t-1}) = (1/\delta_t)\mathbb{P}(s_{4t} \geq s^*, r_{4t} \geq r^*|\mathbf{y}_{t-1}; \boldsymbol{\theta}_4),$$

$$\delta_t = \mathbb{P}(s_{1t} < s^*, r_{1t} < r^* | \mathbf{y}_{t-1}; \boldsymbol{\theta}_1) + \mathbb{P}(s_{2t} < s^*, r_{2t} \geq r^* | \mathbf{y}_{t-1}; \boldsymbol{\theta}_2)$$

$$+\mathbb{P}(s_{3t} \geq s^*, r_{3t} < r^* | \mathbf{y}_{t-1}; \boldsymbol{\theta}_3) + \mathbb{P}(s_{4t} \geq s^*, r_{4t} \geq r^* | \mathbf{y}_{t-1}; \boldsymbol{\theta}_4), \qquad (18)$$

$$\{\mathbf{u}_t\} \sim \text{i.i.d. } \mathcal{N}(\mathbf{0}, \mathbf{I}_2), \qquad (19)$$

and $\boldsymbol{\theta}_i = (\boldsymbol{\mu}_i', \text{vec}(\mathbf{A}_1^{(i)})', \text{vech}(\boldsymbol{\Sigma}_i)')'$.

We use Robert Shiller's well-known data set of annual observations, from 1900 to 2000, on the Standard and Poor's 500 composite stock price index to earnings per share (S_t) and the three-month Treasury Bill rate (R_t).[16] It is clear from Figure 4 that, for long periods of time, both series take values well above their sample means (which are $\widehat{\mu}_s = 13.731$ and $\widehat{\mu}_r = 4.809$). It is also clear that both time series tend to remain above or below the respective sample mean for relatively long periods.[17] It is reasonable to expect that the economy behaved differently in the 1970's and 1980's, when interest rates were relatively high and the price-earnings ratio was relatively low, and in periods such as the 1930's and late 1990's, when the price-earnings ratio was relatively high. When considering linear VAR models for (s_t, r_t), the AIC selects a first-order model. However, such a model is firmly rejected by the nonlinearity test discussed in Section 4.3: the value of $\Re$ is 7.44689, which has a zero asymptotic P-value.

Since we use annual data, we expect that the nonlinear dynamics of stock price and interest rate will be adequately captured by a first-order model such as the one in (15)–(19). ML estimates of the parameters of the C-MSTAR model and their asymptotic standard errors are reported in Table 6. The standardized residuals of the model exhibit no signs of serial correlation on the basis of conventional Ljung–Box portmanteau tests.

The estimated threshold parameters reported in the last row of Table 6 are $\widehat{s}^* = 3.40317$ and $\widehat{r}^* = -0.07214$. Adding to these values the corresponding sample means $\widehat{\mu}_s$ and $\widehat{\mu}_r$, we see that the estimated thresholds for the price-earnings ratio and interest rate are 17.1343 and 4.73695 , respectively.

The bottom four panels of Figure 4 plot the estimated mixing functions, for each point in sample, which specify the weight of regime 1 (associated with $G_1(\cdot)$), regime 2

---

[16] The date is available at `http://www.econ.yale.edu/~shiller/data/chapt26.xls`.

[17] The hypothesis that S_t and R_t are random walks (with drift) is rejected in favor of a stationary STAR alternative using Eklund's (2003) test statistic, which takes the value 6.38 and 2.68 for S_t and R_t, respectively.

(associated with $G_2(\cdot)$), regime 3 (associated with $G_3(\cdot)$), and regime 4 (associated with $G_4(\cdot)$). It is seen that the most prominent regime is the one characterized by a low price-earnings ratio and low interest rates (regime 1). This regime lasts from mid 1930's to the end of the 1960's. Much of the 1970's and 1980's appear to be associated with a regime with low price-earnings ratio and high interest rates (regime 2), a regime which also seems to characterize a few years in the beginning of the 1900's through 1930. The regime associated with high price-earnings ratio and low interest rates (regime 3) never lasts more than six years and is prevalent in only a few years during the 1930's, 1960's and 1990's. Finally, the regime associated with high price-earnings ratio and high interest rates (regime 4) seems to dominate for only short periods of time towards the end of the 1960's and the early 1990's.

Regarding the stability properties of the empirical model, we note that the ML estimates reported in Table 6 do not satisfy the condition of Proposition 1; specifically, we have $1.25346 < \rho(\widehat{\mathcal{A}}) < 1.27997$, where $\widehat{\mathcal{A}} = \{\widehat{\mathbf{A}}_1^{(1)}, \widehat{\mathbf{A}}_1^{(2)}, \widehat{\mathbf{A}}_1^{(3)}, \widehat{\mathbf{A}}_1^{(4)}\}$. It should be remembered, however, that a joint spectral radius less than unity is not necessary for Q-geometric ergodicity and is clearly a rather stringent condition.

To investigate further the stability characteristics of the empirical model, we examine the properties of its skeleton. Using numerical simulation and a grid of starting values, it is found that the skeleton of the empirical model in Table 6 has a unique fixed point $\overline{\mathbf{y}}_e = (0.478, -0.059)'$ and that the matrix of partial derivatives $\mathbf{D}(\overline{\mathbf{y}}_e)$ in (12) has spectral radius 0.801. This suggests that the model is locally stable. Furthermore, plots of the skeleton (not shown here) reveal that, for both the price-earning ratio and the interest rate, the skeleton converges very quickly to the respective long-run value, which provides further evidence of stability.

5.2   Regime-Specific Granger Causality

In the majority of applications, Granger causality has been analyzed in the context of linear VAR models for a set of variables of interest. A standard auxiliary assumption typically made is that the parameters of the VAR are constant over the sample period under consideration. This corresponds to an assumption that the causal links are stable over time, an assumption which is far from innocuous and may often not hold in practice

[see, e.g., Psaradakis et al. (2005)].

The potential non-constancy of causality patterns poses significant econometric problems. Although there may be many good reasons for time-dependence of causality patterns (changes in operating procedures and other aspects of monetary policy, large shocks to the economy, etc.), standard econometric techniques are not well-equipped to handle such situations, unless researchers have strong a priori indications of when and why the links between the variables of interest change. Furthermore, since there is no direct relationship between causality in an economic sense (one variable being responsible for changes in another variable) and Granger causality (one variable having predictive content for another), the information that is useful in terms of understanding changes in causality in an economic sense may not be useful, or sufficient, for guiding one's choice of sample points at which Granger causality changes. Finally, causality results obtained under the assumption of a constant-parameter linear model are likely to be biased in the presence Granger-casual links which are regime dependent.

To illustrate this point, we begin our analysis using a first-order VAR model, the estimated parameters of which are reported in Table 7. Clearly, none of the two variables appears to be Granger causal for the other. This result is very surprising since, not only do the two variables reflect alternative investing opportunities, but the interest rate is usually thought of as a policy variable that might be used to correct misalignments in stock prices. The lack of Granger causality in our system may well be a consequence of the issues described above.

Another potential difficulty is that causality tests based on VAR models may have low power in the presence of nonlinearities in the data. For this reason, we also carry out the nonparametric test for Granger non-causality proposed by Diks and Panchenko (2006). The test is implemented with one lag and bandwidth set equal to $\max\{8.62/T^{2/7}, 1.5\} = 2.3059$.[18] The test statistic for the null hypothesis that the price-earning ratio is Granger non-causal for the interest rate takes the value 0.195, which has asymptotic P-value 0.4226; the statistic for testing the null hypothesis that the interest rate is Granger non-causal for the price-earning ratio takes the value 1.1095, which has asymptotic P-value 0.0866.

---

[18] For details on the definition of the test statistic and the choice of bandwidth the reader is referred to Diks and Panchenko (2006).

Of course, neither the causality test based on the VAR nor the nonparametric test can provide information about the potential regime-specific nature of Granger causality in our bivariate system. To investigate this issue we adopt a slightly different approach to that of Psaradakis et al. (2005) and, instead of inquiring how causality patters change over time, we examine whether the two variables are useful for predicting each other in different economic regimes. Using the C-MSTAR model in Table 6, it can be seen that the off-diagonal elements of $\mathbf{A}_1^{(i)}$ vary significantly across regimes. Specifically, the interest rate Granger causes the price-earning ratio in regime 3. One may speculate that in regime 3 the stock price boom of the 1960's is associated with a long period of relatively low interest rates; the causality in regime 1 reflects the fact that stocks and bonds are substitute assets and that low interest rates may help to forecast high future stock prices. The price-earnings ratio Granger causes the interest rates in regimes 2, 3 and 4. This result may reflect the fact that the central bank reacts to the price-earning ratio by changing the interest rate when it is thought that a misalignment correction is needed. In regime 2, a low price-earnings ratio leads to a reduction in interest rates (from a high interest rate regime). In regime 3, a high price-earnings ratio leads to an increase in interest rates (from a low interest rate regime). Finally, in regime 4 a high price-earnings ratio leads to a reduction of the interest rate (from a high interest rate regime). Notice that regime 4 is followed by regime 2; for example, the period of high price-earnings ratio and interest rates of the 1920's is followed by a crash in the stock markets.[19]

5.3 An Alternative Nonlinear Model

In this sub-section, we compare the C-MSTAR with a closely related nonlinear model which may be capable of accounting for the regime-specific characteristics of the data, namely a logistic multivariate STAR model (LMSTAR), and discuss briefly some other types of nonlinear models that one might use.

The LMSTAR model is specified as in (15)–(16), with the mixing functions $G_i(\cdot)$ given

---

[19]Even though there is no reason, in general, for regime 4 to be short lived (as this is not an intrinsic property of the model), we expect this to be the case for our data set because a high enough interest rate tends to cool down the stock market.

by

$$G_i\left(\mathbf{y}_{t-1}\right) = \frac{\Lambda_i(\mathbf{y}_{t-1})}{\sum_{i=1}^{4} \Lambda_i(\mathbf{y}_{t-1})}, \quad i = 1, \ldots, 4, \tag{20}$$

where

$$
\begin{aligned}
\Lambda_1(\mathbf{y}_{t-1}) &= L(s_{t-1}; \lambda_s, s^*) L(r_{t-1}; \lambda_r, r^*), \\
\Lambda_2(\mathbf{y}_{t-1}) &= L(s_{t-1}; \lambda_s, s^*) \left[1 - L(r_{t-1}; \lambda_r, r^*)\right], \\
\Lambda_3(\mathbf{y}_{t-1}) &= \left[1 - L(s_{t-1}; \lambda_s, s^*)\right] L(r_{t-1}; \lambda_r, r^*), \\
\Lambda_4(\mathbf{y}_{t-1}) &= \left[1 - L(s_{t-1}; \lambda_s, s^*)\right] \left[1 - L(r_{t-1}; \lambda_r, r^*)\right],
\end{aligned}
\tag{21}
$$

and $L(\cdot; \cdot, \cdot)$ is the logistic function

$$L(x; \lambda, x^*) = \frac{1}{1 + \exp\{\lambda(x - x^*)\}}, \quad x \in \mathbb{R}, \quad \lambda > 0. \tag{22}$$

The parameter λ determines the speed of transition between the two regimes associated with the limiting values of $L(x; \lambda, x^*)$ (as $x \to \pm\infty$), while x^* can be interpreted as a threshold parameter.[20] It is easy to see that $G_i\left(\mathbf{y}_{t-1}\right) = \Lambda_i(\mathbf{y}_{t-1})$, $i = 1, \ldots, 4$.

ML estimates of the parameters of the LMSTAR model are reported in Table 8. The standardized residuals of the model exhibit no signs of serial correlation. The estimated threshold parameters are $\widehat{s}^* = 0.61643$ and $\widehat{r}^* = 2.58431$; adding these to the corresponding sample means $\widehat{\mu}_s$ and $\widehat{\mu}_r$, we see that the estimated thresholds for the price-earnings ratio and interest rates are 14.3475 and 7.39341, respectively. It is also clear that the off-diagonal elements of $\mathbf{A}_1^{(i)}$ vary significantly across regimes. In particular, the interest rate is Granger-causal for the price-earnings ratio only in regime 1 (when the probability of the latent variable r_{1t} being below the relevant threshold is high). The price-earnings ratio Granger causes interest rates only in regime 4 (when the probability of r_{4t} and s_{4t} being above their respective thresholds is high). We also note that the separation of regimes shown in Figure 5 is similar to that implied by the C-MSTAR model. Finally, we see that the C-MSTAR model is the preferred specification according to the

---

[20] The LMSTAR model is the closest model to the C-MSTAR in the sense that we can define the same four regimes in both models. The main differences between the two models concern the information that is included in the mixing function, and are extensively discussed in Dueker et al. (2007). To these differences we should add the presence of regime-specific covariance matrices in the mixing weights of the C-MSTAR model.

three complexity-penalized goodness-of-fit measures reported here, namely the AIC, the Bayesian information criterion (BIC) and the Hannan–Quinn (HQ) criterion.

Regarding the stability of the fitted LMSTAR model, the set of estimated coefficient matrices has a joint spectral radius that exceeds unity ($1.1968250 < \rho(\widehat{\mathcal{A}}) < 1.1970068$). However, numerical analysis of the skeleton of the LMSTAR model reveals that it has a single fixed point $\overline{\mathbf{y}}_e = (0.124, 0.001)'$ which is locally stable; the matrix of partial derivatives in (11) associated with the skeleton of the LMSTAR model (see Appendix A.2) has spectral radius 0.842.

Another nonlinear model that might be considered is the autoregressive conditional root (ACR) model of Bec et al. (2008). Despite some apparent similarities, the ACR and C-MSTAR models have different dynamic interpretations and are designed to capture different characteristics of the data, something which makes empirical comparisons rather problematic in our empirical setting. The ACR is primarily designed to capture dynamics which switch between seemingly stationary and nonstationary behavior. The ACR model is defined as a mixture of a finite number of vector autoregressions, only one of which is operative at each observation; the switching between these autoregressions is determined by a random variable the transition probabilities of which are a nonlinear function of the lagged endogenous variables. A common feature of the ACR and C-MSTAR models is that the probability of transitions between regimes is a function of lagged variables. The regimes are, however, defined in entirely different ways in the two models and regime separation will be quite different in practice (for example, extreme values of lagged variables of either sign will be associated with the same regime in the ACR, while this is not the case in the C-MSTAR).[21] The two classes of models are clearly best suited to different applications.

Finally, we note that we also considered a four-state bivariate MSAR model in which the unconditional means of the two series were allowed to be either high or low in each of the regimes. The MSAR model was found to be outperformed by the M-CSTAR, both in terms of in-sample goodness of fit and out-of-sample predictive ability.

---

[21]This issue is of relevance in our setting because the de-meaned variables s_t and r_t take both positive and negative values in the sample.

5.4 Forecast Accuracy

In this sub-section, we evaluate the accuracy of out-of-sample forecasts from the VAR, C-MSTAR and LMSTAR models. The comparisons are based on a series of recursive forecasts computed in the following way. Each of the three models is fitted to the bivariate time series $\{\mathbf{y}_t = (s_t, r_t)'\}_{t=1}^{T-N}$, where $T = 101$ is the number of observations in the full sample and $N = 25$ is the number of forecasts (the forecast period is 1976–2000). 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. Note that the one-step-ahead forecasts for the C-MSTAR and LMSTAR models are relatively straightforward to compute as both models involve a weighted average of the two linear relationships.

The forecast performance of the different models is evaluated using traditional accuracy measures such as mean square percentage error (MSPE), mean absolute percentage error (MAPE), and root mean square percentage error (RMSPE). In addition, the ability of the models to correctly identify turning points (i.e., the direction of change in the variable of interest regardless of the accuracy with which the magnitude of the change is predicted) is evaluated using the so-called confusion rate (CR), which is computed as the percentage of times the direction of change is wrongly predicted.

From the results reported in Table 9, it is clear that the C-MSTAR model yields the smallest MSPE, MAPE and RMSPE for the price-earnings ratio, while the VAR outperforms the competing models in forecasting the interest rate. Turning to the outcomes for the bivariate system (sum of the individual results), the C-MSTAR evidently outperforms the competing models. In particular, the gain of using the C-MSTAR over the VAR model is 2% in terms of both MSPE and MAPE, and 1% in terms of the RMSPE. The marginal gain of the C-MSTAR over the LMSTAR is 31% when using the RMSPE, 71% with MAPE and over 300% with MSPE. Interestingly, but not entirely surprisingly, a comparison between the models on the basis of confusion rates shows that the C-MSTAR produces better results for both series. In fact, the C-MSTAR wrongly predicts the direction of the change in the price-earnings ratio only 25% of time, with the corresponding figure for the

26

alternative nonlinear (linear) model being 58% (29%). Turning to the interest rate, it is interesting to note that while the C-MSTAR wrongly predicts the direction of change one third of the time, the LMSTAR does much worse predicting the wrong direction of change 75% of the time.

To assess which model is more successful over time (that is, which model outperforms the alternatives most of the time as opposed to being more successful on average), we compute the number of times each model achieves the smallest MAPE over the 25 forecast points. On the basis of the individual series, we find that C-MSTAR outperforms the alternative specifications 76% of the time when forecasting the price-earnings ratio and 60% of the time when forecasting the interest rates.

To summarize, the results presented in this section illustrate the importance of capturing the regime-specific properties of the data in order to understand the complex interrelationships between economic variables. Not accounting for such regime-specific characteristics may lead to results which, like those obtained from a linear VAR, may appear to be counterintuitive. The C-MSTAR model characterizes adequately the dynamics of interest rates and stock prices, yields economically meaningful results, and has good out-of-sample forecast performance.[22]

6 Summary

In this paper we have introduced a new class of contemporaneous-threshold multivariate STAR models in which the mixing weights are determined by the probability that contemporaneous latent variables exceed certain threshold values. We have discussed issues related to the stability of the model, estimation and testing. We have also illustrated the practical use of the proposed model by analyzing the bivariate relationship between US stock prices and interest rates. We have found that the proposed model is capable of outperforming some competing linear and nonlinear models, especially in terms of out-of-sample forecast performance, and that the regime-specific Granger causality patterns between the two variables typically differ from those obtained from a linear model in a

---

[22]The forecast results are particularly noteworthy because one of the major weaknesses of many nonlinear models is their relatively poor out-of-sample performance [see also Dueker et al. (2007)].

27

way which is economically meaningful.

A Appendix

A.1 Asymptotic Properties of the ML Estimator

Sufficient conditions which ensure the strong consistency and asymptotic normality of the ML estimator of $\boldsymbol{\theta}$ mentioned in Section 4.1 are given below. Here, we write $\nabla f(\bar{\boldsymbol{\theta}})$ and $\nabla^2 f(\bar{\boldsymbol{\theta}})$ for the gradient vector and Hessian matrix, respectively, of a scalar-valued function $\boldsymbol{\theta} \mapsto f(\boldsymbol{\theta})$ at $\bar{\boldsymbol{\theta}}$, and use $\|\cdot\|_2$ to denote the Frobenius matrix norm (i.e., $\|\mathbf{C}\|_2 = \{\mathrm{tr}(\mathbf{C}'\mathbf{C})\}^{1/2}$).

(C.1) For each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $\{\mathbf{y}_t\}$ is strictly stationary and ergodic.

(C.2) $\Psi(\cdot)$ and $\psi(\cdot)$ are twice continuously differentiable.

(C.3) $\boldsymbol{\theta}_0$ is an interior point of the compact and convex parameter space $\boldsymbol{\Theta}$.

(C.4) $\mathbb{P}[\ell_t(\boldsymbol{\theta}) - \ell_t(\boldsymbol{\theta}_0) \neq 0] > 0$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta} \backslash \{\boldsymbol{\theta}_0\}$.

(C.5) $\mathbb{E}\left(\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} |\ln \ell_t(\boldsymbol{\theta})|\right) < \infty$.

(C.6) $\mathbb{E}\left(\sup_{\boldsymbol{\theta} \in \mathcal{B}(\boldsymbol{\theta}_0)} \left\|\nabla^2 \ln \ell_t(\boldsymbol{\theta})\right\|_2\right) < \infty$ for some open neighborhood $\mathcal{B}(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$.

(C.7) $\mathbb{E}\left(\sup_{\boldsymbol{\theta} \in \mathcal{B}(\boldsymbol{\theta}_0)} \left\|\nabla^2 \ell_t(\boldsymbol{\theta})\right\|_2\right) < \infty$.

(C.8) $\mathbb{E}\left(\|\nabla \ln \ell_t(\boldsymbol{\theta}_0)\|^2\right) < \infty$.

(C.9) $\boldsymbol{\Omega}(\boldsymbol{\theta}_0) = -\mathbb{E}[\nabla^2 \ln \ell_t(\boldsymbol{\theta}_0)]$ is nonsingular.

These are fairly standard regularity conditions for ML estimation. We note that for (C.1) to hold it is sufficient that the conditions of Proposition 1 are satisfied and $\{\mathbf{y}_t\}$ is initialized from its invariant distribution.

We have the following result for the ML estimator $\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} (1/T)\mathcal{L}_T(\boldsymbol{\theta})$.

Proposition 2 *If conditions* (C.1)–(C.5) *are satisfied, then* $\widehat{\boldsymbol{\theta}}$ *is strongly consistent for* $\boldsymbol{\theta}_0$. *If, in addition, conditions* (C.6)–(C.9) *are satisfied, then* $\sqrt{T}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ *is asymptotically normal with mean vector* $\mathbf{0}$ *and covariance matrix* $\boldsymbol{\Omega}(\boldsymbol{\theta}_0)^{-1}$.

Proof. It is easy to see that $\mathcal{L}_T(\boldsymbol{\theta})$ is a measurable function of the data for each fixed $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and almost surely continuous in $\boldsymbol{\theta}$. Moreover, since the sequence $\{\ln \ell_t(\boldsymbol{\theta})\}$ is strictly stationary and ergodic under (C.1)–(C.2), it follows from (C.5) and the uniform strong law of large numbers in Theorem 2.7 of Straumann and Mikosch (2006) that

$$\lim_{T \to \infty} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left| \frac{1}{T} \sum_{t=1}^{T} \ln \ell_t(\boldsymbol{\theta}) - \mathbb{E}[\ln \ell_t(\boldsymbol{\theta})] \right| = 0 \quad \text{almost surely.}$$

Thus, using the compactness of $\boldsymbol{\Theta}$, along with the fact that $\mathbb{E}[\ln \ell_t(\boldsymbol{\theta})]$ attains a unique maximum at $\boldsymbol{\theta} = \boldsymbol{\theta_0}$ under conditions (C.3)–(C.5), we conclude by a standard argument [cf. Amemiya (1973, Lemma 3)] that $\lim_{T \to \infty} \widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}_0$ almost surely.

Turning to the root-T asymptotic normality of $\widehat{\boldsymbol{\theta}}$, we note that $\mathcal{L}_T(\boldsymbol{\theta})$ is almost surely twice continuously differentiable in $\boldsymbol{\theta}$ and $\sum_{t=1}^{T} \nabla \ln \ell_t(\widehat{\boldsymbol{\theta}}) = \mathbf{0}$ for all T sufficiently large because $\widehat{\boldsymbol{\theta}}$ is strongly consistent for $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_0$ is interior to $\boldsymbol{\Theta}$. Hence, by a mean-value expansion of $\sum_{t=1}^{T} \nabla \ln \ell_t(\widehat{\boldsymbol{\theta}})$ about $\boldsymbol{\theta}_0$, we have

$$\mathbf{0} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \nabla \ln \ell_t(\boldsymbol{\theta}_0) + \left(\frac{1}{T} \sum_{t=1}^{T} \nabla^2 \ln \ell_t(\boldsymbol{\theta}^*) \right) \left\{ \sqrt{T}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \right\}, \tag{23}$$

for all T sufficiently large and some $\boldsymbol{\theta}^* \in \boldsymbol{\Theta}$ satisfying $\|\boldsymbol{\theta}^* - \boldsymbol{\theta}_0\| \leq \left\| \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \right\|$. Since $\{\nabla^2 \ln \ell_t(\boldsymbol{\theta})\}$ is a strictly stationary and ergodic sequence, and $\lim_{T \to \infty} \boldsymbol{\theta}^* = \boldsymbol{\theta}_0$ almost surely by virtue of the strong consistency of $\widehat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}_0$, it follows from (C.6), Theorem 2.7 of Straumann and Mikosch (2006), and Lemma 4 of Amemiya (1973) that

$$\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \nabla^2 \ln \ell_t(\boldsymbol{\theta}^*) = -\boldsymbol{\Omega}(\boldsymbol{\theta}_0) \quad \text{almost surely.} \tag{24}$$

Furthermore, since the model is correctly specified, $\{\nabla \ln \ell_t(\boldsymbol{\theta}_0)\}$ forms a strictly stationary and ergodic vector-valued martingale-difference sequence relative to the σ-field generated by $\{\mathbf{y}_t, \mathbf{y}_{t-1}, \ldots, \mathbf{y}_0\}$, and $\mathbb{E}[\{\nabla \ln \ell_t(\boldsymbol{\theta}_0)\}\{\nabla \ln \ell_t(\boldsymbol{\theta}_0)\}']$ is finite and equal to $\boldsymbol{\Omega}(\boldsymbol{\theta}_0)$ under (C.4)–(C.8). Thus, we may use the Billingsley–Ibragimov martingale central limit theorem [e.g., Taniguchi and Kakizawa (2000, Theorem A.2.14)] and the Cramér–Wold device to conclude that $(1/\sqrt{T}) \sum_{t=1}^{T} \nabla \ln \ell_t(\boldsymbol{\theta}_0)$ is asymptotically normal with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Omega}(\boldsymbol{\theta}_0)$. This result, together with (23), (24) and (C.9), delivers the claimed asymptotic distribution of $\sqrt{T}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$. $\square$

The asymptotic normality of $\{-\nabla^2 \mathcal{L}_T(\widehat{\boldsymbol{\theta}})\}^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ mentioned in Section 4.1 is an immediate consequence of Proposition 2 and of the fact that $\lim_{T\to\infty}(1/T)\sum_{t=1}^{T}\nabla^2 \ln \ell_t(\widehat{\boldsymbol{\theta}}) = -\boldsymbol{\Omega}(\boldsymbol{\theta}_0)$ almost surely (which also guarantees that there exists T large enough such that $\nabla^2 \mathcal{L}_T(\widehat{\boldsymbol{\theta}})$ is negative definite almost surely).

A.2 Skeleton of the LMSTAR Model

The skeleton of the LMSTAR model discussed in Section 5.3 is given by equations (8) and (9), with the mixing functions $G_i(\cdot)$, $i = 1, ..., 4$, defined as in (20)–(22). Consequently, the local stability of a fixed point of the skeleton can be assessed by considering the spectrum of the matrix of partial derivatives given in (12)–(13), with

$$\frac{\partial G_1(\overline{\mathbf{y}}_{t-1})}{\partial \overline{\mathbf{y}}_{t-1}} = \begin{bmatrix} l(\overline{x}_{t-1})L(\overline{w}_{t-1}) \\ L(\overline{x}_{t-1})l(\overline{w}_{t-1}) \end{bmatrix},$$

$$\frac{\partial G_2(\overline{\mathbf{y}}_{t-1})}{\partial \overline{\mathbf{y}}_{t-1}} = \begin{bmatrix} l(\overline{x}_{t-1})\left[1 - L(\overline{w}_{t-1})\right] \\ -L(\overline{x}_{t-1})l(\overline{w}_{t-1}) \end{bmatrix},$$

$$\frac{\partial G_3(\overline{\mathbf{y}}_{t-1})}{\partial \overline{\mathbf{y}}_{t-1}} = \begin{bmatrix} -l(\overline{x}_{t-1})L(\overline{w}_{t-1}) \\ \left[1 - L(\overline{x}_{t-1})\right]l(\overline{w}_{t-1}) \end{bmatrix},$$

$$\frac{\partial G_4(\overline{\mathbf{y}}_{t-1})}{\partial \overline{\mathbf{y}}_{t-1}} = \begin{bmatrix} -l(\overline{x}_{t-1})\left[1 - L(\overline{w}_{t-1})\right] \\ -\left[1 - L(\overline{x}_{t-1})\right]l(\overline{w}_{t-1}) \end{bmatrix},$$

and

$$l(x) = \frac{\partial L(x)}{\partial x} = \frac{-\lambda \exp\{\lambda(x - x^*)\}}{[1 + \exp\{\lambda(x - x^*)\}]^2}.$$

References

[1] Altissimo, F. and Violante, G. (2001), The non-linear dynamics of output and unemployment in the U.S., *Journal of Applied Econometrics* 16, 461–486.

[2] Amemiya, T. (1973), Regression analysis when the dependent variable is truncated normal, *Econometrica* 41, 997–1016.

[3] Bec, F., Rahbeck, A. and Shephard, N. (2008), The ACR model: a multivariate dynamic mixture autoregression, *Oxford Bulletin of Economics and Statistics* 70, 583–618.

[4] Bernanke, B. and Gertler, M. (1999), Monetary policy and asset price volatility, in *New Challenges for Monetary Policy*, Kansas City: Federal Reserve Bank of Kansas City, pp. 77–128.

[5] Bernanke, B. and Gertler, M. (2001), Should central banks respond to movements in asset prices?, *American Economic Review* 91, 253–257.

[6] Cecchetti S.G., Genberg, H., Lipsky, J. and Wadhwani, S.B. (2000), *Asset Prices and Central Bank Policy*, Geneva Reports on the World Economy, No. 2, International Center for Monetary and Banking Studies and Centre for Economic Policy Research.

[7] Chan, K.S. and Tong, H. (1985), On the use of the deterministic Lyapunov function for the ergodicity of stochastic difference equations, *Advances in Applied Probability* 17, 666–678.

[8] De Gooijer, J.G. and Vidiella-i-Anguera, A. (2004), Forecasting threshold cointegrated systems, *International Journal of Forecasting* 20, 237–253.

[9] Diks, C. and Panchenko, V. (2006), A new statistic and practical guidelines for nonparametric Granger causality testing, *Journal of Economic Dynamics and Control* 30, 1647–1669.

[10] Doukhan, P. (1994), *Mixing: Properties and Examples*, Lecture Notes in Statistics 85, Berlin: Springer-Verlag.

[11] Dueker, M. J., Sola, M. and Spagnolo, F. (2007), Contemporaneous threshold autoregressive models: estimation, testing and forecasting, *Journal of Econometrics* 141, 517–547.

[12] Eklund, B. (2003), A nonlinear alternative to the unit root hypothesis, SSE/EFI Working Paper No. 547, Stockholm School of Economics.

[13] Fong, P.W., Li, W.K., Yau, C.W. and Wong, C.S. (2007), On a mixture vector autoregressive model, *Canadian Journal of Statistics* 35, 135–150.

[14] Gripenberg, G. (1996), Computing the joint spectral radius, *Linear Algebra and its Applications* 234, 43–60.

[15] Hamilton, J.D. (1993), Estimation, inference and forecasting of time series subject to changes in regime, in Maddala, G.S., Rao, C.R. and H.D. Vinod (eds.), *Handbook of Statistics*, Vol. 11, Amsterdam: Elsevier Science Publishers, pp. 231–260.

[16] Hansen, B.E. (1992), The likelihood ratio test under nonstandard conditions: testing the Markov switching model of GNP, *Journal of Applied Econometrics* 7, S61–S82 (Erratum: 11, 195–198).

[17] Harvill, J.L. and Ray, B.K. (1999), A note on tests for nonlinearity in a vector time series, *Biometrika* 86, 728–734.

[18] Harvill, J.L. and Ray, B.K. (2006), Functional coefficient autoregressive models for vector time series, *Computational Statistics and Data Analysis* 50, 3547–3566.

[19] Jungers, R. (2009), *The Joint Spectral Radius: Theory and Applications*, Lecture Notes in Control and Information Sciences 385, Berlin: Springer-Verlag.

[20] Kapetanios, G. (2001), Model selection in threshold models, *Journal of Time Series Analysis* 22, 733–754.

[21] Koop, G., and Potter, S. (2006), The vector floor and ceiling model, in Milas, C., Rothman, P. and D. van Dijk (eds.), *Nonlinear Time Series Analysis of Business Cycles*, Contributions to Economic Analysis 276, Amsterdam: Elsevier, pp. 97–131.

[22] Liebscher, E. (2005), Towards a unified approach for proving geometric ergodicity and mixing properties of nonlinear autoregressive processes, *Journal of Time Series Analysis* 26, 669–689.

[23] 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.

[24] Meyn, S.P. and Tweedie, R.L. (2009), *Markov Chains and Stochastic Stability*, 2nd ed., Cambridge: Cambridge University Press.

[25] Psaradakis, Z., Ravn, M.O. and Sola, M. (2005), Markov switching causality and the money–output relationship, *Journal of Applied Econometrics* 20, 665–683.

[26] 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.

[27] Psaradakis, Z. and Spagnolo, N. (2006), Joint determination of the state dimension and autoregressive order for models with Markov regime switching, *Journal of Time Series Analysis* 27, 753–766.

[28] Rothman, P., van Dijk, D. and Franses, P.H. (2001), A multivariate STAR analysis of the relationship between money and output, *Macroeconomic Dynamics* 5, 506–532.

[29] Sola, M. and Driffill, J. (1994), Testing the term structure of interest rates using a stationary vector autoregression with regime switching, *Journal of Economic Dynamics and Control* 18, 601–628.

[30] Straumann, D. and Mikosch, T. (2006), Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equations approach, *Annals of Statistics* 34, 2449–2495.

[31] Taniguchi, M. and Y. Kakizawa (2000), *Asymptotic Theory of Statistical Inference for Time Series*, New York: Springer-Verlag.

[32] 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.

[33] Tong, H. (1983), *Threshold Models in Non-Linear Time Series Analysis*, New York: Springer-Verlag.

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

[35] Tsay, R.S. (1986), Nonlinearity tests for time series, *Biometrika* 73, 461–466.

[36] Tsay, R.S. (1998), Testing and modeling multivariate threshold models, *Journal of the American Statistical Association* 93, 1188–1202.

[37] Tsitsiklis, J.N. and Blondel, V.D. (1997), The Lyapunov exponent and joint spectral radius of pairs of matrices are hard –when not impossible– to compute and to approximate, *Mathematics of Control, Signals, and Systems* 10, 31–40.

[38] 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. Data-Generating Processes

DGP-1

$$\boldsymbol{\mu}_1 = \begin{bmatrix} -0.05 \\ -0.05 \end{bmatrix}, \qquad \mathbf{A}_1^{(1)} = \begin{bmatrix} 0.80 & 0.05 \\ 0.10 & 0.90 \end{bmatrix}, \qquad \boldsymbol{\Sigma}_1 = \mathbf{I}_2$$

$$\boldsymbol{\mu}_2 = \begin{bmatrix} -0.05 \\ 0.05 \end{bmatrix}, \qquad \mathbf{A}_1^{(2)} = \begin{bmatrix} 0.75 & -0.05 \\ 0.05 & 0.85 \end{bmatrix}, \qquad \boldsymbol{\Sigma}_2 = \mathbf{I}_2$$

$$\boldsymbol{\mu}_3 = \begin{bmatrix} 0.15 \\ -0.05 \end{bmatrix}, \qquad \mathbf{A}_1^{(3)} = \begin{bmatrix} 0.75 & -0.30 \\ 0.20 & 0.85 \end{bmatrix}, \qquad \boldsymbol{\Sigma}_3 = \mathbf{I}_2$$

$$\boldsymbol{\mu}_4 = \begin{bmatrix} 0.05 \\ 0.10 \end{bmatrix}, \qquad \mathbf{A}_1^{(4)} = \begin{bmatrix} 0.90 & -0.10 \\ 0.01 & 0.90 \end{bmatrix}, \qquad \boldsymbol{\Sigma}_4 = \mathbf{I}_2$$

$$(x^*, w^*) = (0.6, -0.4)$$

DGP-2

Intercepts, autoregressive coefficients and threshold parameters are the same as for DGP-1.

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}, \quad \boldsymbol{\Sigma}_2 = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}, \quad \boldsymbol{\Sigma}_3 = \begin{bmatrix} 1 & 0.3 \\ 0.3 & 1 \end{bmatrix}$$

$$\boldsymbol{\Sigma}_4 = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}$$

DGP-3

Intercepts, autoregressive coefficients and threshold parameters are the same as for DGP-1.

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 1 & -0.9 \\ -0.9 & 1 \end{bmatrix}, \quad \boldsymbol{\Sigma}_2 = \begin{bmatrix} 1 & -0.8 \\ -0.8 & 1 \end{bmatrix}, \quad \boldsymbol{\Sigma}_3 = \begin{bmatrix} 1 & -0.3 \\ -0.3 & 1 \end{bmatrix}$$

$$\boldsymbol{\Sigma}_4 = \begin{bmatrix} 1 & -0.8 \\ -0.8 & 1 \end{bmatrix}$$

Table 2. Monte Carlo Results: DGP-1

$$T = 200$$

$$
\widehat{\boldsymbol{\mu}}_1 = \begin{bmatrix} 0.085 \\ (1.054) \\ 0.071 \\ (1.032) \end{bmatrix}, \quad
\widehat{\mathbf{A}}_1^{(1)} = \begin{bmatrix} 0.006 & 0.008 \\ (1.019) & (1.013) \\ 0.005 & 0.009 \\ (0.995) & (1.015) \end{bmatrix}, \quad
\widehat{\boldsymbol{\Sigma}}_1 = \begin{bmatrix} -0.057 & 0.024 \\ (1.055) & (0.989) \\ & -0.066 \\ & (1.042) \end{bmatrix}
$$

$$
\widehat{\boldsymbol{\mu}}_2 = \begin{bmatrix} 0.077 \\ (1.061) \\ 0.053 \\ (1.026) \end{bmatrix}, \quad
\widehat{\mathbf{A}}_1^{(2)} = \begin{bmatrix} -0.004 & -0.023 \\ (1.022) & (0.994) \\ 0.013 & -0.006 \\ (0.996) & (1.028) \end{bmatrix}, \quad
\widehat{\boldsymbol{\Sigma}}_2 = \begin{bmatrix} -0.046 & 0.033 \\ (1.049) & (0.992) \\ & -0.041 \\ & (1.076) \end{bmatrix}
$$

$$
\widehat{\boldsymbol{\mu}}_3 = \begin{bmatrix} 0.180 \\ (1.130)^\dagger \\ 0.156 \\ (1.106)^\dagger \end{bmatrix}, \quad
\widehat{\mathbf{A}}_1^{(3)} = \begin{bmatrix} 0.030 & 0.093 \\ (1.043) & (1.075) \\ 0.084 & 0.012 \\ (0.982) & (1.068) \end{bmatrix}, \quad
\widehat{\boldsymbol{\Sigma}}_3 = \begin{bmatrix} -0.072 & 0.083 \\ (0.969) & (0.955) \\ & -0.093 \\ & (1.022) \end{bmatrix}
$$

$$
\widehat{\boldsymbol{\mu}}_4 = \begin{bmatrix} 0.171 \\ (1.149)^\dagger \\ 0.109 \\ (1.165)^\dagger \end{bmatrix}, \quad
\widehat{\mathbf{A}}_1^{(4)} = \begin{bmatrix} 0.052 & 0.078 \\ (1.029) & (0.948) \\ 0.084 & 0.075 \\ (0.982) & (0.930) \end{bmatrix}, \quad
\widehat{\boldsymbol{\Sigma}}_4 = \begin{bmatrix} -0.066 & 0.065 \\ (0.950) & (1.033) \\ & -0.069 \\ & (1.041) \end{bmatrix}
$$

$$\widehat{x}^* = \frac{-0.042}{(1.061)}, \qquad \widehat{w}^* = \frac{-0.054}{(1.093)}$$

$$T = 800$$

$$
\widehat{\boldsymbol{\mu}}_1 = \begin{bmatrix} -0.007 \\ (1.003) \\ 0.010 \\ (1.009) \end{bmatrix}, \quad
\widehat{\mathbf{A}}_1^{(1)} = \begin{bmatrix} -0.001 & -0.005 \\ (1.011) & (0.992) \\ 0.006 & 0.002 \\ (0.998) & (0.996) \end{bmatrix}, \quad
\widehat{\boldsymbol{\Sigma}}_1 = \begin{bmatrix} -0.021 & 0.011 \\ (1.019) & (0.997) \\ & -0.020 \\ & (1.010) \end{bmatrix}
$$

$$
\widehat{\boldsymbol{\mu}}_2 = \begin{bmatrix} 0.011 \\ (1.005) \\ 0.004 \\ (1.006) \end{bmatrix}, \quad
\widehat{\mathbf{A}}_1^{(2)} = \begin{bmatrix} -0.002 & 0.016 \\ (1.003) & (1.005) \\ 0.009 & -0.002 \\ (0.999) & (1.007) \end{bmatrix}, \quad
\widehat{\boldsymbol{\Sigma}}_2 = \begin{bmatrix} 0.003 & 0.010 \\ (1.008) & (1.002) \\ & -0.005 \\ & (1.015) \end{bmatrix}
$$

$$
\widehat{\boldsymbol{\mu}}_3 = \begin{bmatrix} 0.077 \\ (1.053) \\ 0.111 \\ (1.041) \end{bmatrix}, \quad
\widehat{\mathbf{A}}_1^{(3)} = \begin{bmatrix} 0.009 & 0.044 \\ (1.018) & (0.951) \\ 0.022 & 0.010 \\ (1.012) & (1.040) \end{bmatrix}, \quad
\widehat{\boldsymbol{\Sigma}}_3 = \begin{bmatrix} -0.020 & 0.060 \\ (1.006) & (0.994) \\ & -0.052 \\ & (1.011) \end{bmatrix}
$$

$$
\widehat{\boldsymbol{\mu}}_4 = \begin{bmatrix} -0.088 \\ (1.058) \\ 0.061 \\ (1.044) \end{bmatrix}, \quad
\widehat{\mathbf{A}}_1^{(4)} = \begin{bmatrix} 0.019 & -0.038 \\ (1.006) & (1.056) \\ 0.058 & -0.049 \\ (0.992) & (1.031) \end{bmatrix}, \quad
\widehat{\boldsymbol{\Sigma}}_4 = \begin{bmatrix} -0.006 & 0.008 \\ (1.008) & (0.992) \\ & 0.029 \\ & (0.910) \end{bmatrix}
$$

$$\widehat{x}^* = \frac{-0.012}{(1.009)}, \qquad \widehat{w}^* = \frac{0.026}{(0.988)}$$

Figures are the finite-sample bias of the ML estimator and the ratios of sampling standard deviations to estimated standard errors (in parentheses).

† indicates that the Kolmogorov–Smirnov statistic is significant at the 5% level.

Table 3. Monte Carlo Results: DGP-2

$$T = 200$$

$$\widehat{\boldsymbol{\mu}}_1 = \begin{bmatrix} -0.072 \\ (1.047) \\ 0.099 \\ (1.018) \end{bmatrix}, \quad \widehat{\mathbf{A}}_1^{(1)} = \begin{bmatrix} -0.005 & 0.050 \\ (1.020) & (1.008) \\ 0.022 & 0.010 \\ (1.007) & (1.033) \end{bmatrix}, \quad \widehat{\boldsymbol{\Sigma}}_1 = \begin{bmatrix} -0.038 & -0.010 \\ (1.053) & (1.012) \\ & -0.033 \\ & (1.040) \end{bmatrix}$$

$$\widehat{\boldsymbol{\mu}}_2 = \begin{bmatrix} -0.051 \\ (1.062) \\ 0.072 \\ (1.009) \end{bmatrix}, \quad \widehat{\mathbf{A}}_1^{(2)} = \begin{bmatrix} 0.010 & -0.009 \\ (0.990) & (0.998) \\ 0.004 & -0.002 \\ (1.000) & (1.004) \end{bmatrix}, \quad \widehat{\boldsymbol{\Sigma}}_2 = \begin{bmatrix} -0.031 & 0.005 \\ (1.022) & (0.999) \\ & -0.032 \\ & (1.040) \end{bmatrix}$$

$$\widehat{\boldsymbol{\mu}}_3 = \begin{bmatrix} 0.201 \\ (1.301)^\dagger \\ 0.210 \\ (1.227)^\dagger \end{bmatrix}, \quad \widehat{\mathbf{A}}_1^{(3)} = \begin{bmatrix} 0.041 & 0.106 \\ (1.078) & (0.929) \\ 0.079 & 0.062 \\ (1.012) & (1.078) \end{bmatrix}, \quad \widehat{\boldsymbol{\Sigma}}_3 = \begin{bmatrix} -0.090 & 0.079 \\ (0.920) & (0.943) \\ & -0.101 \\ & (1.061) \end{bmatrix}$$

$$\widehat{\boldsymbol{\mu}}_4 = \begin{bmatrix} 0.032 \\ (1.031) \\ 0.080 \\ (1.029) \end{bmatrix}, \quad \widehat{\mathbf{A}}_1^{(4)} = \begin{bmatrix} 0.002 & 0.016 \\ (0.991) & (0.997) \\ 0.008 & 0.002 \\ (0.998) & (1.007) \end{bmatrix}, \quad \widehat{\boldsymbol{\Sigma}}_4 = \begin{bmatrix} 0.011 & 0.010 \\ (1.022) & (1.000) \\ & -0.015 \\ & (1.051) \end{bmatrix}$$

$$\widehat{x}^* = \frac{-0.055}{(1.091)^\dagger}, \quad \widehat{w}^* = \frac{-0.061}{(1.072)}$$

$$T = 800$$

$$\widehat{\boldsymbol{\mu}}_1 = \begin{bmatrix} 0.020 \\ (1.008) \\ 0.007 \\ (1.001) \end{bmatrix}, \quad \widehat{\mathbf{A}}_1^{(1)} = \begin{bmatrix} -0.001 & 0.007 \\ (1.000) & (1.002) \\ 0.007 & -0.005 \\ (1.002) & (1.014) \end{bmatrix}, \quad \widehat{\boldsymbol{\Sigma}}_1 = \begin{bmatrix} -0.022 & 0.005 \\ (1.010) & (1.006) \\ & -0.011 \\ & (1.004) \end{bmatrix}$$

$$\widehat{\boldsymbol{\mu}}_2 = \begin{bmatrix} 0.029 \\ (0.996) \\ 0.030 \\ (1.002) \end{bmatrix}, \quad \widehat{\mathbf{A}}_1^{(2)} = \begin{bmatrix} -0.003 & 0.005 \\ (0.996) & (0.999) \\ 0.002 & 0.002 \\ (0.999) & (1.002) \end{bmatrix}, \quad \widehat{\boldsymbol{\Sigma}}_2 = \begin{bmatrix} -0.009 & -0.002 \\ (1.005) & (0.998) \\ & -0.005 \\ & (1.009) \end{bmatrix}$$

$$\widehat{\boldsymbol{\mu}}_3 = \begin{bmatrix} 0.098 \\ (1.033) \\ 0.056 \\ (1.058) \end{bmatrix}, \quad \widehat{\mathbf{A}}_1^{(3)} = \begin{bmatrix} 0.026 & 0.031 \\ (1.009) & (1.082) \\ 0.044 & 0.049 \\ (1.006) & (1.051) \end{bmatrix}, \quad \widehat{\boldsymbol{\Sigma}}_3 = \begin{bmatrix} -0.007 & 0.044 \\ (1.012) & (1.013) \\ & -0.008 \\ & (1.010) \end{bmatrix}$$

$$\widehat{\boldsymbol{\mu}}_4 = \begin{bmatrix} 0.007 \\ (1.009) \\ 0.023 \\ (1.007) \end{bmatrix}, \quad \widehat{\mathbf{A}}_1^{(4)} = \begin{bmatrix} -0.001 & 0.008 \\ (1.003) & (1.002) \\ -0.003 & -0.002 \\ (1.001) & (0.997) \end{bmatrix}, \quad \widehat{\boldsymbol{\Sigma}}_4 = \begin{bmatrix} -0.001 & -0.004 \\ (1.004) & (0.999) \\ & -0.004 \\ & (0.989) \end{bmatrix}$$

$$\widehat{x}^* = \frac{0.012}{(0.993)}, \quad \widehat{w}^* = \frac{-0.010}{(1.004)}$$

Figures are the finite-sample bias of the ML estimator and the ratios of sampling standard deviations to estimated standard errors (in parentheses).

$\dagger$ indicates that the Kolmogorov–Smirnov statistic is significant at the 5% level.

Table 4. Monte Carlo Results: DGP-3

$$T = 200$$

$$\widehat{\boldsymbol{\mu}}_1 = \begin{bmatrix} 0.071 \\ (1.033) \\ 0.060 \\ (1.011) \end{bmatrix}, \quad \widehat{\mathbf{A}}_1^{(1)} = \begin{bmatrix} 0.004 & 0.007 \\ (0.972) & (1.015) \\ 0.002 & 0.001 \\ (0.998) & (1.007) \end{bmatrix}, \quad \widehat{\boldsymbol{\Sigma}}_1 = \begin{bmatrix} 0.013 & 0.020 \\ (0.967) & (1.005) \\ & -0.004 \\ & (1.006) \end{bmatrix}$$

$$\widehat{\boldsymbol{\mu}}_2 = \begin{bmatrix} 0.031 \\ (1.015) \\ 0.032 \\ (1.008) \end{bmatrix}, \quad \widehat{\mathbf{A}}_1^{(2)} = \begin{bmatrix} 0.002 & 0.003 \\ (1.005) & (1.001) \\ 0.003 & 0.002 \\ (0.999) & (0.999) \end{bmatrix}, \quad \widehat{\boldsymbol{\Sigma}}_2 = \begin{bmatrix} -0.048 & 0.037 \\ (1.048) & (0.990) \\ & -0.047 \\ & (1.033) \end{bmatrix}$$

$$\widehat{\boldsymbol{\mu}}_3 = \begin{bmatrix} 0.099 \\ (1.152)^\dagger \\ 0.163 \\ (1.076) \end{bmatrix}, \quad \widehat{\mathbf{A}}_1^{(3)} = \begin{bmatrix} 0.043 & 0.074 \\ (1.019) & (1.050) \\ 0.073 & 0.046 \\ (1.012) & (1.069) \end{bmatrix}, \quad \widehat{\boldsymbol{\Sigma}}_3 = \begin{bmatrix} -0.081 & 0.045 \\ (1.042) & (1.029) \\ & -0.054 \\ & (1.033) \end{bmatrix}$$

$$\widehat{\boldsymbol{\mu}}_4 = \begin{bmatrix} 0.165 \\ (1.262)^\dagger \\ 0.111 \\ (1.186)^\dagger \end{bmatrix}, \quad \widehat{\mathbf{A}}_1^{(4)} = \begin{bmatrix} 0.056 & 0.081 \\ (1.082) & (1.079) \\ 0.085 & 0.050 \\ (1.076) & (1.062) \end{bmatrix}, \quad \widehat{\boldsymbol{\Sigma}}_4 = \begin{bmatrix} -0.088 & 0.067 \\ (1.076) & (1.029) \\ & -0.078 \\ & (1.050) \end{bmatrix}$$

$$\widehat{x}^* = \underset{(1.031)}{0.022}, \qquad \widehat{w}^* = \underset{(0.955)}{-0.040}$$

$$T = 800$$

$$\widehat{\boldsymbol{\mu}}_1 = \begin{bmatrix} 0.030 \\ (1.002) \\ 0.043 \\ (1.007) \end{bmatrix}, \quad \widehat{\mathbf{A}}_1^{(1)} = \begin{bmatrix} 0.001 & 0.004 \\ (1.002) & (1.009) \\ 0.000 & -0.006 \\ (1.001) & (1.019) \end{bmatrix}, \quad \widehat{\boldsymbol{\Sigma}}_1 = \begin{bmatrix} -0.010 & 0.002 \\ (1.002) & (1.002) \\ & -0.005 \\ & (1.006) \end{bmatrix}$$

$$\widehat{\boldsymbol{\mu}}_2 = \begin{bmatrix} 0.019 \\ (1.002) \\ 0.001 \\ (1.003) \end{bmatrix}, \quad \widehat{\mathbf{A}}_1^{(2)} = \begin{bmatrix} 0.001 & 0.001 \\ (1.000) & (0.999) \\ 0.000 & 0.001 \\ (1.001) & (0.999) \end{bmatrix}, \quad \widehat{\boldsymbol{\Sigma}}_2 = \begin{bmatrix} -0.009 & 0.001 \\ (1.001) & (1.000) \\ & -0.004 \\ & (0.997) \end{bmatrix}$$

$$\widehat{\boldsymbol{\mu}}_3 = \begin{bmatrix} 0.067 \\ (1.047) \\ 0.101 \\ (1.032) \end{bmatrix}, \quad \widehat{\mathbf{A}}_1^{(3)} = \begin{bmatrix} 0.015 & 0.021 \\ (1.006) & (1.036) \\ 0.022 & 0.017 \\ (0.998) & (0.989) \end{bmatrix}, \quad \widehat{\boldsymbol{\Sigma}}_3 = \begin{bmatrix} -0.031 & 0.010 \\ (1.004) & (1.007) \\ & -0.018 \\ & (1.009) \end{bmatrix}$$

$$\widehat{\boldsymbol{\mu}}_4 = \begin{bmatrix} 0.087 \\ (1.052) \\ 0.020 \\ (1.019) \end{bmatrix}, \quad \widehat{\mathbf{A}}_1^{(4)} = \begin{bmatrix} 0.012 & 0.030 \\ (1.020) & (1.022) \\ -0.009 & 0.020 \\ (0.967) & (1.012) \end{bmatrix}, \quad \widehat{\boldsymbol{\Sigma}}_4 = \begin{bmatrix} 0.035 & 0.045 \\ (1.009) & (0.995) \\ & -0.021 \\ & (0.988) \end{bmatrix}$$

$$\widehat{x}^* = \underset{(1.009)}{0.006}, \qquad \widehat{w}^* = \underset{(1.004)}{-0.011}$$

Figures are the finite-sample bias of the ML estimator and the ratios of sampling standard deviations to estimated standard errors (in parentheses).

$\dagger$ indicates that the Kolmogorov–Smirnov statistic is significant at the 5% level.

Table 5. Power of Nonlinearity Test

T	1%	5%	10%
		DGP-1	
100	82.20	88.84	91.28
200	86.36	91.68	93.48
400	93.80	96.20	97.60
800	99.04	99.48	99.60
		DGP-2	
100	78.08	86.76	90.20
200	85.16	92.36	95.04
400	95.16	97.84	98.68
800	99.64	99.88	99.96
		DGP-3	
100	93.28	96.08	97.24
200	98.96	99.56	99.68
400	99.96	100	100
800	100	100	100

Entries are percentage rejection frequencies.

Table 6. ML Estimates for a C-MSTAR Model

Regime 1: Low Price-Earnings Ratio, Low Interest Rate

$$\widehat{\boldsymbol{\mu}}_1 = \begin{bmatrix} -0.54000 \\ (0.72558) \\ 0.52756 \\ (0.07339) \end{bmatrix}, \quad \widehat{\mathbf{A}}_1^{(1)} = \begin{bmatrix} 1.03181 & -0.32729 \\ (0.11118) & (0.25067) \\ 0.00836 & 1.11187 \\ (0.01090) & (0.02438) \end{bmatrix}, \quad \widehat{\boldsymbol{\Sigma}}_1 = \begin{bmatrix} 2.72805 & 0.07130 \\ (0.78069) & (0.00967) \\ 0.07130 & 0.02642 \\ (0.00967) & (0.04115) \end{bmatrix}$$

Regime 2: Low Price-Earnings Ratio, High Interest Rate

$$\widehat{\boldsymbol{\mu}}_2 = \begin{bmatrix} 0.52803 \\ (0.34762) \\ 0.59252 \\ (0.27681) \end{bmatrix}, \quad \widehat{\mathbf{A}}_1^{(2)} = \begin{bmatrix} 0.90321 & -0.03038 \\ (0.08047) & (0.10508) \\ -0.15747 & 0.79461 \\ (0.07017) & (0.08365) \end{bmatrix}, \quad \widehat{\boldsymbol{\Sigma}}_2 = \begin{bmatrix} 3.61557 & -0.29982 \\ (0.78726) & (0.74264) \\ -0.29982 & 3.00545 \\ (0.74264) & (0.33928) \end{bmatrix}$$

Regime 3: High Price-Earnings Ratio, Low Interest Rate

$$\widehat{\boldsymbol{\mu}}_3 = \begin{bmatrix} 0.19662 \\ (1.3002) \\ -1.08184 \\ (0.34393) \end{bmatrix}, \quad \widehat{\mathbf{A}}_1^{(3)} = \begin{bmatrix} 0.95186 & 1.04940 \\ (0.20444) & (0.44486) \\ 0.10149 & 0.90033 \\ (0.04864) & (0.09666) \end{bmatrix}, \quad \widehat{\boldsymbol{\Sigma}}_3 = \begin{bmatrix} 15.2194 & -0.43588 \\ (4.90575) & (0.00489) \\ -0.43588 & 0.63892 \\ (0.00489) & (0.55733) \end{bmatrix}$$

Regime 4: High Price-Earnings Ratio, High Interest Rate

$$\widehat{\boldsymbol{\mu}}_4 = \begin{bmatrix} -3.73793 \\ (1.82851) \\ -0.82893 \\ (0.24954) \end{bmatrix}, \quad \widehat{\mathbf{A}}_1^{(4)} = \begin{bmatrix} -0.46101 & 0.18350 \\ (0.25623) & (0.29623) \\ -0.13939 & 0.53549 \\ (0.03497) & (0.04044) \end{bmatrix}, \quad \widehat{\boldsymbol{\Sigma}}_4 = \begin{bmatrix} 22.9615 & 3.13327 \\ (10.4492) & (0.00489) \\ 3.13327 & 0.42766 \\ (0.00489) & (0.71372) \end{bmatrix}$$

$$\widehat{s}^* = \underset{(0.71359)}{3.40317}, \qquad \widehat{r}^* = \underset{(0.10159)}{-0.07214},$$

$$\max \mathcal{L} = -351.160, \quad \text{AIC} = 778.320, \text{ BIC} = 877.317, \text{ HQ} = 818.386$$

Figures in parentheses are asymptotic standard errors and $\max \mathcal{L}$ is the maximized log-likelihood.
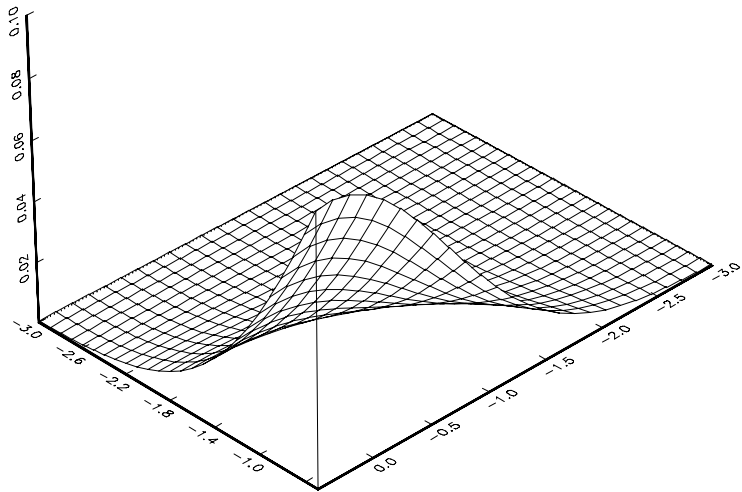
Table 7. ML Estimates for a VAR Model

$$\mathbf{y_t} = \boldsymbol{\mu} + \mathbf{A}\mathbf{y_{t-1}} + \boldsymbol{\Sigma}^{1/2}\mathbf{u}_t$$

$$\widehat{\boldsymbol{\mu}} = \begin{bmatrix} 0.1301 \\ (0.3200) \\ 0.0111 \\ (0.1503) \end{bmatrix}, \quad \widehat{\mathbf{A}} = \begin{bmatrix} 0.7938 & -0.0590 \\ (0.0706) & (0.0332) \\ 0.0988 & 0.8661 \\ (0.1047) & (0.0492) \end{bmatrix}, \quad \widehat{\boldsymbol{\Sigma}} = \begin{bmatrix} 10.2291 & 0.0577 \\ 0.0577 & 2.2561 \end{bmatrix}$$

$$\max \mathcal{L} = -437.679, \quad \text{AIC} = 893.358, \text{ BIC} = 916.805, \text{ HQ} = 902.847$$

Figures in parentheses are asymptotic standard errors and $\max \mathcal{L}$ is the maximized log-likelihood.

Table 8. ML Estimates for a LMSTAR Model

Regime 1: Low Price-Earnings Ratio, Low Interest Rate

$$
\widehat{\boldsymbol{\mu}}_1 = \begin{bmatrix} 1.02473 \\ (0.28692) \\ 0.92660 \\ (0.10919) \end{bmatrix}, \quad
\widehat{\mathbf{A}}_1^{(1)} = \begin{bmatrix} 1.02161 & -0.12392 \\ (0.05848) & (0.10759) \\ -0.00319 & 1.19457 \\ (0.02149) & (0.03978) \end{bmatrix}, \quad
\widehat{\boldsymbol{\Sigma}}_1 = \begin{bmatrix} 1.44019 & 0.32393 \\ (0.43671) & (0.07138) \\ 0.32393 & 0.20652 \\ (0.07138) & (0.08464) \end{bmatrix}
$$

Regime 2: Low Price-Earnings Ratio, High Interest Rate

$$
\widehat{\boldsymbol{\mu}}_2 = \begin{bmatrix} -1.79071 \\ (0.52315) \\ 1.04670 \\ (1.21128) \end{bmatrix}, \quad
\widehat{\mathbf{A}}_1^{(2)} = \begin{bmatrix} 0.76229 & 0.13731 \\ (0.09344) & (0.08808) \\ -0.29126 & 0.53078 \\ (0.21669) & (0.21931) \end{bmatrix}, \quad
\widehat{\boldsymbol{\Sigma}}_2 = \begin{bmatrix} 1.22248 & -0.65899 \\ (0.50228) & (3.61650) \\ -0.65899 & 6.94336 \\ (3.61650) & (0.54162) \end{bmatrix}
$$

Regime 3: High Price-Earnings Ratio, Low Interest Rate

$$
\widehat{\boldsymbol{\mu}}_3 = \begin{bmatrix} 0.23872 \\ (0.57517) \\ -0.66698 \\ (0.12075) \end{bmatrix}, \quad
\widehat{\mathbf{A}}_1^{(3)} = \begin{bmatrix} 0.88645 & 0.86029 \\ (0.11388) & (0.25165) \\ -0.00141 & 0.82499 \\ (0.02452) & (0.04672) \end{bmatrix}, \quad
\widehat{\boldsymbol{\Sigma}}_3 = \begin{bmatrix} 9.12349 & -0.32945 \\ (2.22135) & (0.11884) \\ -0.32945 & 0.35104 \\ (0.11884) & (0.26687) \end{bmatrix}
$$

Regime 4: High Price-Earnings Ratio, High Interest Rate

$$
\widehat{\boldsymbol{\mu}}_4 = \begin{bmatrix} 3.58485 \\ (2.62699) \\ 1.87520 \\ (0.41219) \end{bmatrix}, \quad
\widehat{\mathbf{A}}_1^{(4)} = \begin{bmatrix} -0.38696 & 0.15326 \\ (0.39798) & (0.94790) \\ -0.30991 & 0.24298 \\ (0.06200) & (0.14860) \end{bmatrix}, \quad
\widehat{\boldsymbol{\Sigma}}_4 = \begin{bmatrix} 19.4180 & -3.00842 \\ (10.0201) & (0.06049) \\ -3.00842 & 0.48008 \\ (0.06049) & (0.35632) \end{bmatrix}
$$

$$
\widehat{s}^* = \underset{(1.50367)}{0.616434}, \qquad
\widehat{r}^* = \underset{(0.72959)}{2.58431}, \qquad
\widehat{\lambda}_s = \underset{(0.06978)}{0.19228}, \
\widehat{\lambda}_r = \underset{(0.17905)}{0.63536}
$$

$$
\max \mathcal{L} = -359.705, \quad \text{AIC} = 799.411, \text{BIC} = 903.617, \text{HQ} = 841.585
$$

Figures in parentheses are asymptotic standard errors and $\max\mathcal{L}$ is the maximized log-likelihood.

Table 9. Out-of-Sample Performance

	MSPE	MAPE	RMSPE	CR
VAR (PE)	0.0562	0.1946	0.4412	0.2917
VAR (IR)	0.0658	0.2132	0.4617	0.4167
Overall	0.1220	0.4078	0.9029	–
C-MSTAR (PE)	0.0487	0.1734	0.4164	0.2500
C-MSTAR (IR)	0.0710	0.2266	0.4760	0.3333
Overall	0.1197	0.4000	0.8924	–
LMSTAR (PE)	0.2413	0.3671	0.6059	0.5833
LMSTAR (IR)	0.1445	0.3175	0.5635	0.7500
Overall	0.3858	0.6846	1.1694	–

PE refers to the price-earnings ratio and IR to the interest rate. MSPE is mean square

percentage error, MAPE is mean absolute percentage error, and RMSPE is root

mean square percentage error of the difference between the forecast data and

and the actual data. CR are confusion rates.

DGP1: Distributions Conditional on $x_{t-1}=.4$ and $w_{t-1}=.6$

$G_1(y_{t-1}) = 0.09$ $G_2(y_{t-1}) = 0.48$ $G_3(y_{t-1}) = 0.09$ $G_4(y_{t-1}) = 0.34$ $x^* = 0.6$ $w^* = -0.4$

Regime 1: $E(x_{1t}|I_{t-1})=0.35$, $E(w_{1t}|I_{t-1})=0.57$.

Regime 2: $E(x_{2t}|I_{t-1})=0.29$, $E(w_{2t}|I_{t-1})=0.6$.
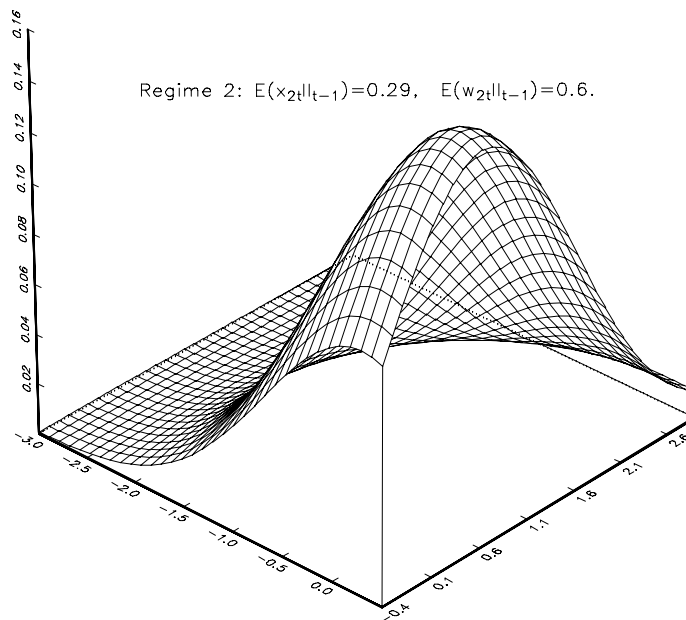
Regime 3: $E(x_{3t}|I_{t-1})=0.59$, $E(w_{3t}|I_{t-1})=0.39$.
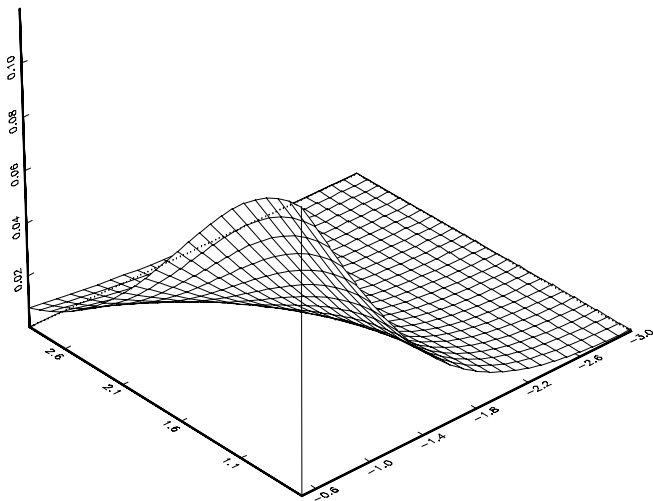
Regime 4: $E(x_{4t}|I_{t-1})=0.43$, $E(w_{4t}|I_{t-1})=0.66$.
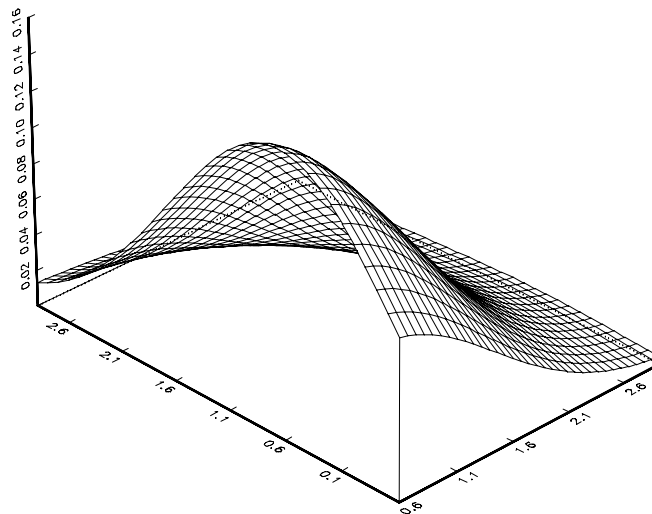
Figure 1

DGP1: Distributions Conditional on $x_{t-1}=-1.5$ and $w_{t-1}=-2$

$G_1(y_{t-1}) = 0.88 \quad G_2(y_{t-1}) = 0.1 \quad G_3(y_{t-1}) = 0.02 \quad G_4(y_{t-1}) = 0.0 \quad x^* = 0.6 \quad w^* = -0.4$

Regime 1: $E(x_{1t}|I_{t-1})=-1.44, \quad E(w_{1t}|I_{t-1})=-1.97.$
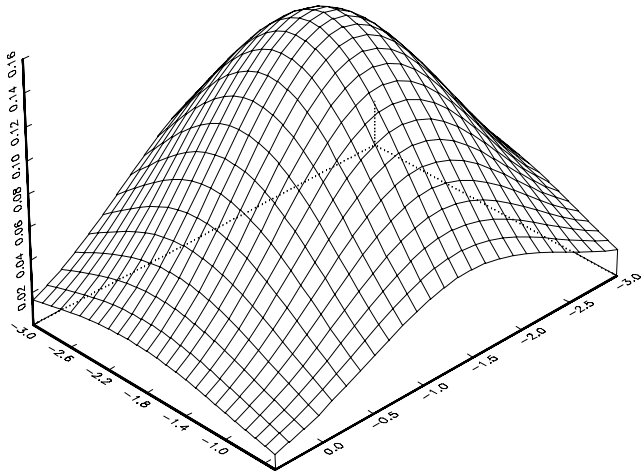
Regime 2: $E(x_{2t}|I_{t-1})=-1.26, \quad E(w_{2t}|I_{t-1})=-1.61.$
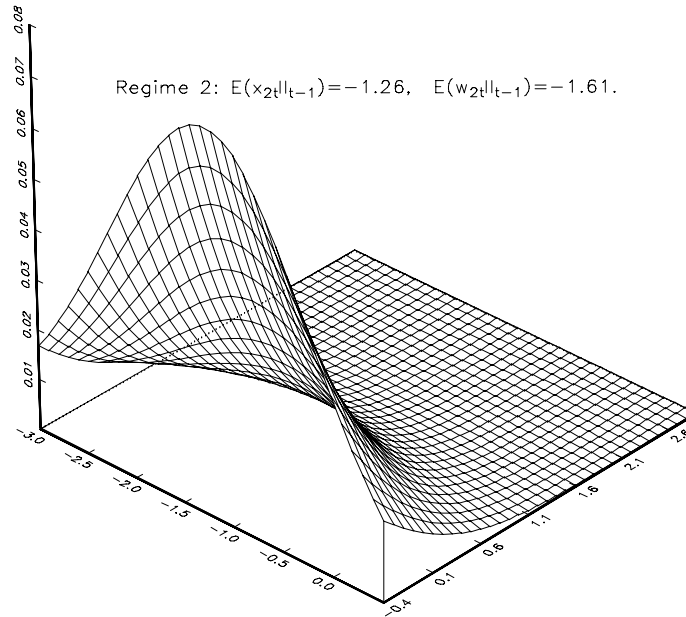
Regime 3: $E(x_{3t}|I_{t-1})=-1.37, \quad E(w_{3t}|I_{t-1})=-1.35.$
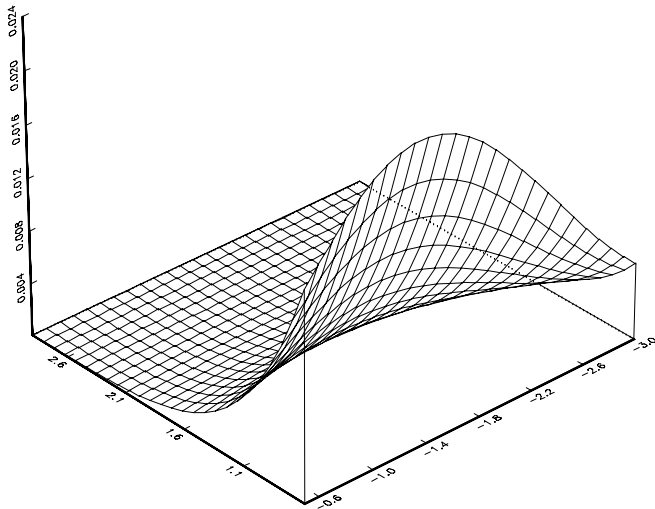
Regime 4: $E(x_{4t}|I_{t-1})=-1.31, \quad E(w_{4t}|I_{t-1})=-1.59.$
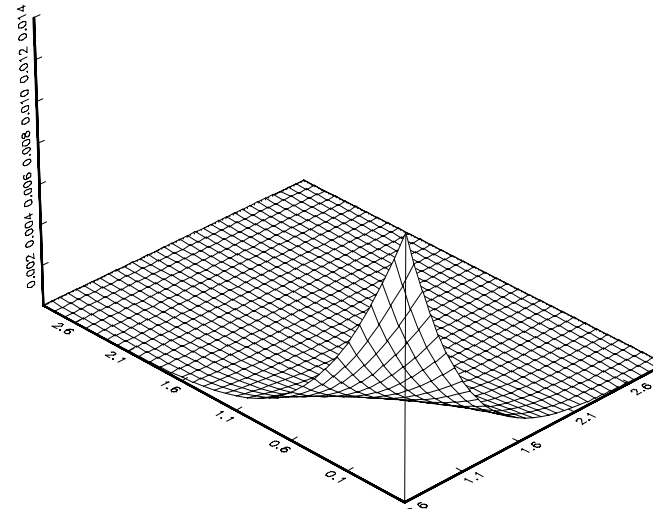
Figure 2

Figure 3

Data and Mixing Functions Using the C−MSTAR(1) Model
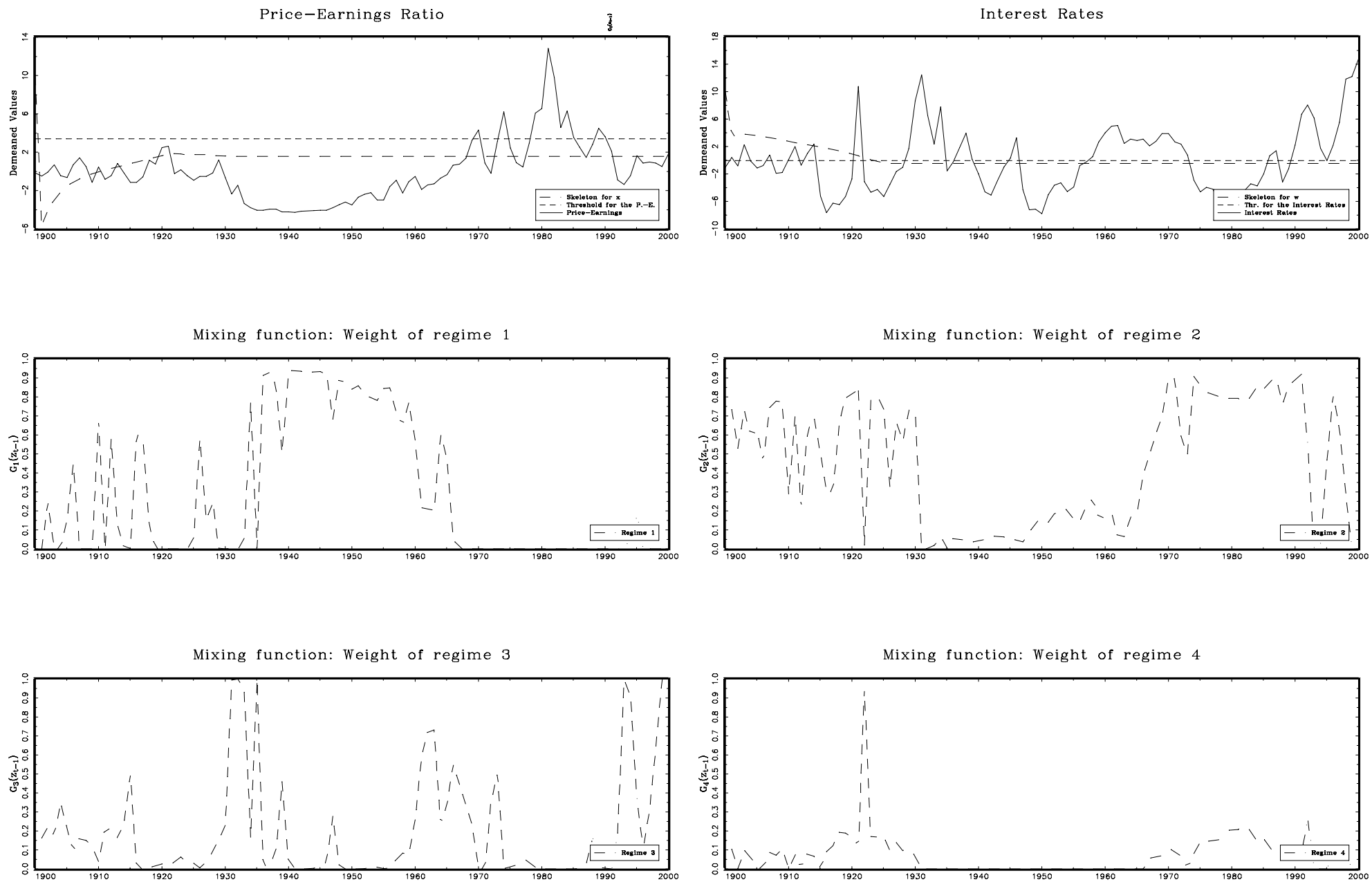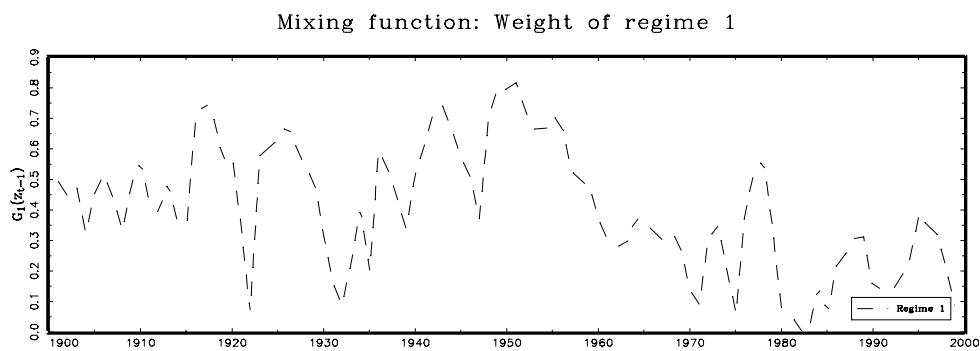

Figure 4

Data and Mixing Functions Using the LMSTAR(1) Model

Figure 5