Central European Journal of Economic Modelling and Econometrics

State-dependent Autoregressive Models with p Lags: Properties, Estimation and Forecasting

Fabio Gobbi* and Sabrina Mulinacci†

Submitted: 4.06.2021, Accepted: 4.10.2021

Abstract

In this paper we consider a class of nonlinear autoregressive models in which a specific type of dependence structure between the error term and the lagged values of the state variable is assumed. We show that there exists an equivalent representation given by a p-th order state-dependent autoregressive (SDAR(p)) model where the error term is independent of the last p lagged values of the state variable $(y_{t-1}, \dots, y_{t-p})$ and the autoregressive coefficients are specific functions of them. We discuss a quasi-maximum likelihood estimator of the model parameters and we prove its consistency and asymptotic normality. To test the forecasting ability of the SDAR(p) model, we propose an empirical application to the quarterly Japan GDP growth rate which is a time series characterized by a *level-increment* dependence. A comparative analyses is conducted taking into consideration some alternative and competitive models for nonlinear time series such as SETAR and AR-GARCH models.

Keywords: convolution-based autoregressive models, level-increment dependence, nonlinear time series, maximum likelihood, forecasting accuracy

JEL Classification: C01, C5, C22

*Department of Economics and Statistics, University of Siena, Italy; e-mail: fabio.gobbi@unisi.it; ORCID: 0000-0003-3669-1619

†Department of Statistical Sciences, University of Bologna, Italy; e-mail: sabrina.mulinacci@unibo.it; ORCID: 0000-0001-6190-5127

1 Introduction

In this paper we study a generalized version of an autoregressive model where we assume a dependence structure which links the lagged state variable Y_{t-1} and the error term ξ_t, extending the standard autoregressive model, in line with the approach considered in Cherubini et al. (2016) (some temporal dependence properties of this model, such as stationarity and mixing, are derived in Gobbi and Mulinacci, 2020). The aim of this approach is to reconcile two needs: i) model nonlinearity in time series; ii) capture the level-increment dependence. In particular, the dependence between the state variable at time $t-1$, Y_{t-1} and its next increment $\Delta Y_t = Y_t - Y_{t-1}$ is of the utmost interest in our opinion. The issue has been investigated in Gobbi and Mulinacci (2021) where the assumption that the error term is a function of the current level plus some noise has been considered: the resulting model is applied to the realized volatility extracted from financial indexes on a weekly basis and its forecasting ability is analyzed in comparison with other nonlinear models. Aim of the present paper is to generalize the model there introduced and discussed by allowing the innovation to be a function of an arbitrary number of lags, in order to detect in a more efficient way the level-increment dependence. Here our interest is to show how the model can also be considered useful for low-frequency macroeconomic data where the dependence on the past can be more pronounced. In particular, the time series considered in this work is the Japan GDP growth rate which has two significant characteristics: nonlinearity combined with a marked and negative level-increment dependence. Two elements that make the model considered in this paper an ideal candidate for its modeling.

Following Cherubini et al. (2011, 2012) and Cherubini et al. (2016), we consider a stochastic process whose dynamics are governed by a relation of type

$$Y_t = \alpha + Y_{t-1} + \eta_t, \tag{1}$$

with $\alpha \in \mathbb{R}$ and Y_{t-1} dependent on η_t. More in general, we assume that η_t is dependent on the past history of the process and we model this dependence assuming that η_t is a specific function of the lagged variables $(Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p})$ plus a disturbance. In this paper we will focus on a representation of type

$$\eta_t = \sum_{k=1}^{p} \psi_k \left(Y_{t-k}; \boldsymbol{\gamma}_i\right) Y_{t-k} - Y_{t-1} + \xi_t,$$

where ψ_k are measurable functions depending on a set of parameters $\boldsymbol{\gamma}_k$ and ξ_t are i.i.d. and independent of $Y_{t-1}, Y_{t-2}, \ldots$. This way, we can rewrite (1) as

$$Y_t = \alpha + \sum_{k=1}^{p} \psi_k(Y_{t-k}; \boldsymbol{\gamma}_k) Y_{t-k} + \xi_t, \tag{2}$$

where the error term ξ_t is characterized by a distribution with zero mean and standard deviation σ, that is the model can be represented as a functional autoregressive model along the lines of those introduced and discussed in Hastie and Tibshirani (1990) and Chen and Tsay (1993), among others. We call the family of models of type (2) p-th order state-dependent autoregressive (SDAR(p)) models. Clearly, its whole specification strongly depends on the functional forms of the autoregressive coefficients ψ_k, with $k = 1, \dots, p$. The heart of the reasoning is that the dependence between η_t and $(Y_{t-1}, \dots, Y_{t-p})$ can be discharged onto the autoregression coefficients which will each be functions of the corresponding lagged variable.
When $p = 1$, (2) coincides with the model studied in Gobbi and Mulinacci (2021) and, as done in that paper for $p = 1$, we will assume suitable conditions that are not too restrictive on the functional coefficients so that the process is geometrically ergodic and stationary. Furthermore, the quasi-maximum likelihood (QML) estimator is proved to be consistent and asymptotically normal.
Our aim is to investigate the forecasting ability of the SDAR model with $p = 1$ and $p = 2$, after specifying the functional forms of the autoregressive coefficients. A comparison with two alternative classes of nonlinear models intensively used in the literature is proposed: the self-exciting threshold autoregressive (SETAR) models and the generalized autoregressive conditional heteroscedasticity (GARCH) models with autoregressive structure of the mean equation (AR-GARCH). As for the SETAR models, there is a rich literature since the 90's on this topic. For example, Tiao and Tsay (1994) and Potter (1995), among others, consider a self-exciting threshold autoregressive (SETAR) model to analyse nonlinearity in the US GDP growth rate. Krager and Kluger (1993) and Clements and Smith (1997) evaluate forecasts from SETAR models of exchange rates. As regards AR-GARCH models, in the last twenty years, many papers addressed the problem of the evolution of the Japan GDP growth rate from the point of view of nonlinear time series. Among others, Hamori (2000) provides an empirical analysis of the volatility of growth rates for different countries including Japan. On the same line, Ho and Tsuy (2003) and Fang and Miller (2009) study the structural decline of the variance of Japan GDP growth rate and they show how the output variability does not affect output growth. Our study is conducted on Japan GDP growth rate using quarterly data. The evaluation of the forecast accuracy of different models adopted is conducted according to two different measures, the average performance using the root mean square error (RMSE) and the mean absolute error (MAE) over different forecast horizons, from 1 to 8 quarters ahead. A simulation experiment is conducted using all the analyzed models and aims at identifying their ability to intercept the level-increment dependence. We will show how the SDAR(p) model with 2 lags is particularly effective for this purpose.
The paper is organized as follows. In Section 2 we introduce a process with dependent increments equivalent to a p-th order state-dependent autoregressive model. In Section 3 a QML estimator of the parameters of the SDAR(p) model is introduced

and discussed. Section 4 presents an empirical application to the Japan GDP growth rate. Section 5 concludes.

2 The SDAR(p) model

We consider stochastic processes $(Y_t)_{t\in\mathbb{N}}$ characterized by dynamics of type

$$Y_t = \alpha + Y_{t-1} + \eta_t, \tag{3}$$

with $\alpha \in \mathbb{R}$ and

$$\eta_t = h\left(Y_{t-1}, \ldots, Y_{t-p}\right) + \xi_t$$

where $h : \mathbb{R}^p \to \mathbb{R}$ is a measurable function and ξ_t is a sequence of i.i.d. centered random variables independent of $\{Y_{t-k}, k \geq 1\}$. If we restrict to the case in which $h(y_1, \ldots, y_p) = \sum_{k=1}^{p} \psi_k(y_k)y_k - y_1$, where, for $k = 1, \ldots, p$, $\psi_k : \mathbb{R} \to \mathbb{R}$ are measurable functions, then (3) can be rewritten as

$$Y_t = \alpha + \sum_{k=1}^{p} \psi_k\left(Y_{t-k}\right) Y_{t-k} + \xi_t. \tag{4}$$

Stochastic processes of this type are Markov processes of order p, that is for all $n \geq p+1$ and $t_1 < t_2 < \cdots < t_{n-p} < t_{n-p+1} < \cdots < t_n < t$ and for all $x \in \mathbb{R}$

$$\mathbb{P}\left(Y_t \leq x | Y_{t_1}, \ldots, Y_{t_{n-p}}, Y_{t_{n-p+1}}, \ldots, Y_{t_n}\right) = \mathbb{P}\left(Y_t \leq | Y_{t_{n-p+1}}, \ldots, Y_{t_n}\right).$$

It is well known that, thanks to Sklar's theorem, the joint cumulative distribution function of a d-dimensional random vector $(X_1, \ldots, X_d)$, $F_{X_1,\ldots,X_d}(x_1, \ldots, x_d)$, is completely specified by the marginal distributions and the d-dimensional copula function $C(u_1, \ldots, u_d)$, that, linking them and modeling the dependence structure, allows to recover the joint distribution:

$$F_{X_1,\ldots,X_d}(x_1, \ldots, x_d) = C\left(F_{X_1}(x_1), \ldots, F_{X_d}(x_d)\right),$$

where F_{X_i} is the cumulative distribution function of X_i for $i = 1, \ldots, d$.
In the seminal paper of Darsow et al (1992), it is proved that the Markov property of a stochastic process can be expressed in terms of some suitable requirements on the copula functions representing the dependence structure of the finite dimensional distributions. This result has been extended to Markov processes of order $p \geq 1$ in Ibragimov (2009), in the following way.
Let $m, n \geq p \geq 1$ and A and B be m- and n-dimensional copulas, respectively, such that

$$A\left(1, \ldots, 1, v_1, \ldots, v_p\right) = B\left(v_1, \ldots, v_p, 1, \ldots, 1\right) = C\left(v_1, \ldots, v_p\right),$$

where C is a p dimensional copula. If

$$A_{1,\ldots,m|m-p+1,\ldots,m}(u_1,\ldots,u_{m-p},v_1,\ldots,v_p) = \\ = \frac{\partial^p A(u_1,\ldots,u_{m-p},v_1,\ldots,v_p)}{\partial w_{m-p+1}\ldots\partial w_m} \bigg/ \frac{\partial^p C(v_1,\ldots,v_p)}{\partial w_1\ldots\partial w_p}$$

and

$$B_{1,\ldots,n|1,\ldots,p}(v_1,\ldots,v_p,u_{m+1},\ldots,u_{m+n-p}) = \\ = \frac{\partial^p B(v_1,\ldots,v_p,u_{m+1},\ldots,u_{m+n-p})}{\partial w_1\ldots\partial w_p} \bigg/ \frac{\partial^p C(v_1,\ldots,v_p)}{\partial w_1\ldots\partial w_p},$$

the $\star^p$-product of the copulas A and B is defined as

$$A \star^p B(u_1,\ldots,u_{m+n-p}) = \\ = \int_0^{u_{m-p+1}} \cdots \int_0^{u_m} A_{1,\ldots,m|m-p+1,\ldots,m}(u_1,\ldots,u_{m-p},v_1,\ldots,v_p)\cdot \\ \cdot B_{1,\ldots,n|1,\ldots,p}(v_1,\ldots,v_p,u_{m+1},\ldots,u_{m+n-p})dC(v_1,\ldots,v_p).$$

Theorem 1 (Ibragimov, 2009). *A stochastic process $(Y_t)_t$ is a Markov process of order p, $p \geq 1$ if and only if for all $n \geq p+1$ and $t_1 < \ldots < t_n$*

$$C_{t_1,\ldots,t_n} = C_{t_1,\ldots,t_{p+1}} \star^p C_{t_2,\ldots,t_{p+2}} \star^p C_{t_{n-p},\ldots,t_n},$$

where $C_{s_1,\ldots,s_k}$ is the copula associated to the vector $(Y_{s_1},\ldots,Y_{s_k})$.

Hence the law of a discrete time Markov process of order p is uniquely defined by the family of the $p+1$ dimensional copulas $C_{t,t+1,\ldots,t+p}$ and the family of the marginal distributions of the variables Y_t. In next proposition we provide the corresponding expressions for the model dynamics defined in (4) extending the corresponding representations for $p = 1$ in Gobbi and Mulinacci (2021).

Proposition 2. *If G is the cumulative distribution function of $\frac{\xi_t}{\sigma}$, then the cumulative distribution function of Y_t, F_t, is*

$$F_t(y_t) = \int_0^1 \cdots \int_0^1 G\left(\frac{y_t - \alpha - \sum_{k=1}^p \psi_k\left(F_{t-k}^{-1}(v_{t-k})\right) F_{t-k}^{-1}(v_{t-k})}{\sigma}\right) \times \\ \times dC_{t-p,\ldots,t-1}(v_{t-p},\ldots,v_{t-1}) \quad (5)$$

and

$$C_{t-p,\ldots,t-1,t}(u_{t-p},\ldots,u_{t-1},u_t) = \\ = \int_0^{u_{t-p}} \cdots \int_0^{u_{t-1}} G\left(\frac{F_t^{-1}(u_t) - \alpha - \sum_{k=1}^p \psi_k\left(F_{t-k}^{-1}(v_{t-k})\right) F_{t-k}^{-1}(v_{t-k})}{\sigma}\right) \times \\ \times dC_{t-p,\ldots,t-1}(v_{t-p},\ldots,v_{t-1}). \quad (6)$$

Proof. The result follows immediately taking into account that

$$
\begin{aligned}
&\mathbb{P}\left(Y_t \leq y_t, Y_{t-1} \leq y_{t-1}, \ldots, Y_{t-p} \leq y_{t-p}\right) = \\
&= \mathbb{P}\left(\alpha + \sum_{k=1}^{p} \psi_k\left(Y_{t-k}\right) Y_{t-k} + \xi_t \leq y_t, Y_{t-1} \leq y_{t-1}, \ldots, Y_{t-p} \leq y_{t-p}\right) = \\
&= \int_{-\infty}^{y_{t-p}} \cdots \int_{-\infty}^{y_{t-1}} \mathbb{P}\left(\xi_t \leq y_t - \alpha - \sum_{k=1}^{p} \psi_k\left(z_{t-k}\right) z_{t-k}\right) \times \\
&\qquad \times dC_{t-p,\ldots,t-1}\left(F_{t-p}(z_{t-p}), \ldots, F_{t-1}(z_{t-1})\right) = \\
&= \int_{0}^{F_{t-p}(y_{t-p})} \cdots \int_{0}^{F_{t-1}(y_{t-1})} G\left(\frac{y_t - \alpha - \sum_{k=1}^{p} \psi_k\left(F_{t-k}^{-1}\left(v_{t-k}\right)\right) F_{t-k}^{-1}\left(v_{t-k}\right)}{\sigma}\right) \times \\
&\qquad \times dC_{t-p,\ldots,t-1}\left(v_{t-p}, \ldots, v_{t-1}\right).
\end{aligned}
$$

Letting $y_{t-k} \to +\infty$ for $k = 1, \ldots, p$, one gets (5), while setting $u_{t-k} = F_{t-k}(y_{t-k})$, for $k = 0, 1, \ldots, p$ one gets (6). □

Similarly, one can compute the cumulative distribution function of η_t, F_{η_t}, and its dependence relation with $Y_{t-1}, \ldots, Y_{t-p}$.

Proposition 3. *Under the model defined in* (4)*, the equivalent representation given in* (3) *is characterized by*

$$
F_{\eta_t}(x) = \int_0^1 \cdots \int_0^1 G\left(\frac{x - h\left(F_{t-1}^{-1}\left(v_{t-1}\right), \ldots, F_{t-p}^{-1}\left(v_{t-p}\right)\right)}{\sigma}\right) \times \\
\times dC_{t-p,\ldots,t-1}\left(v_{t-p}, \ldots, v_{t-1}\right) \quad (7)
$$

and

$$
\begin{aligned}
&C_{Y_{t-p},\ldots,Y_{t-1},\eta_t}(u_{t-p}, \ldots, u_{t-1}, u) = \\
&\qquad = \int_0^{u_{t-p}} \cdots \int_0^{u_{t-1}} G\left(\frac{F_{\eta_t}^{-1}(u) - h\left(F_{t-1}^{-1}\left(v_{t-1}\right), \ldots, F_{t-p}^{-1}\left(v_{t-p}\right)\right)}{\sigma}\right) \times \\
&\qquad\qquad \times dC_{t-p,\ldots,t-1}\left(v_{t-p}, \ldots, v_{t-1}\right). \quad (8)
\end{aligned}
$$

Proof. It follows immediately from

$$\begin{aligned}
&\mathbb{P}\left(\eta_t \leq x, Y_{t-1} \leq y_{t-1}, \ldots, Y_{t-p} \leq y_{t-p}\right) = \\
&= \mathbb{P}\left(h\left(Y_{t-1}, \ldots, Y_{t-p}\right) + \xi_t \leq x, Y_{t-1} \leq y_{t-1}, \ldots, Y_{t-p} \leq y_{t-p}\right) = \\
&= \int_{-\infty}^{y_{t-p}} \cdots \int_{-\infty}^{y_{t-1}} G\left(\frac{x - h\left(z_{t-1}, \ldots, z_{t-p}\right)}{\sigma}\right) \times \\
&\qquad \times dC_{t-p,\ldots,t-1}\left(F_{t-p}(z_{t-p}), \ldots, F_{t-1}(z_{t-1})\right) = \\
&= \int_{0}^{F_{t-p}(y_{t-p})} \cdots \int_{0}^{F_{t-1}(y_{t-1})} G\left(\frac{x - h\left(F_{t-1}^{-1}\left(v_{t-1}\right), \ldots, F_{t-p}^{-1}\left(v_{t-p}\right)\right)}{\sigma}\right) \times \\
&\qquad \times dC_{t-p,\ldots,t-1}\left(v_{t-p}, \ldots, v_{t-1}\right),
\end{aligned}$$

and conclusions trivially follow. □

In order to guarantee ergodicity and stationarity, we introduce the following assumptions relative to the functional autoregressive coefficients.

i) **Assumption A1**: there exist $\lambda_1, \ldots, \lambda_p$ such that $|\psi_k(x)| \leq \lambda_k$ and $\sum_{k=1}^{p} \lambda_k < 1$.

ii) **Assumption A2**: ξ_t are i.i.d. with zero mean, positive density and with finite fourth moment.

Assumptions **A1** and **A2** ensure that the Markov model of order p defined by (4) is geometrically ergodic (see Theorem 3.2 in An and Huang, 1996), which implies strict stationarity (see Theorem 2.2 in Fan and Yao, 2003). Moreover, they imply the finiteness of the fourth moment of Y_t which is a technical requirement for the validity of next Lemma 5.

3 Quasi-maximum likelihood estimation

In the sequel, we will suppose that the functions $\psi_k(y; \boldsymbol{\gamma}_k)$ with $k = 1, \ldots, p$ are measurable functions that depend on a d_k-dimensional vector of parameters $\boldsymbol{\gamma}_k = (\gamma_{1,k}, \ldots, \gamma_{d_k,k})$ and we rewrite (4) as

$$\begin{cases} Y_t = \alpha + \sum_{k=1}^{p} \psi_k(Y_{t-k}; \boldsymbol{\gamma}_k) Y_{t-k} + \xi_t, \\ \xi_t/\sigma \sim g, \end{cases} \tag{9}$$

where g is the density of the standardized random variable ξ_t/σ.
As a consequence, if $D = \sum_{k=1}^{p} d_k$, model (9) depends on the $D+2$ parameters $\boldsymbol{\theta} = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_{d_k}, \alpha, \sigma)$.
In order to ease the notation, we set $\psi_k^{\gamma_{i,k}}(y; \boldsymbol{\gamma}_k) = (\partial/\partial\gamma_{i,k})\psi_k(y; \boldsymbol{\gamma}_k)$ and $\psi_k^{\gamma_{i,k},\gamma_{j,k}}(y; \boldsymbol{\gamma}_k) = (\partial^2/\partial\gamma_{i,k}\partial\gamma_{j,k})\psi_k(y; \boldsymbol{\gamma}_k)$. Moreover, we introduce the additional assumptions:

i) **Assumption A3**. The parameter space Θ is a compact subset of $\mathbb{R}^{D+2}$.

ii) **Assumption A4**. First and second-order partial derivatives of the persistence functions ψ_k with respect to the parameters are continuous for $k = 1, \dots, p$.

iii) **Assumption A5**. $|\psi_k^{\gamma_{i,k}}(y; \boldsymbol{\gamma}_k)| \leq C_{i,k}$ uniformly on $\mathbb{R} \times \Theta$, for $i = 1, \dots, d_k$ and $k = 1, \dots, p$ and $|\psi^{\gamma_{i,k}\gamma_{j,k}}(y; \boldsymbol{\gamma}_k)| \leq D_{i,j,k}$ uniformly on $\mathbb{R} \times \Theta$, for $i, j = 1, \dots, d_k$ and $k = 1, \dots, p$.

iv) **Assumption A6**. The density g of the standardized random variable ξ_t/σ is twice differentiable.

Let us now assume a time series $\mathbf{y}^n = (y_1, \dots, y_n)$ generated by (9) be given. If $p_t^{\boldsymbol{\theta}}$ is the conditional distribution of Y_t given $Y_{t-1}, \dots, Y_{t-p}$, since the process given in (9) is a Markov process of order p, the quasi log-likelihood is given by

$$L_n(\mathbf{y}^n; \boldsymbol{\theta}) = \frac{1}{n} \sum_{t=p+1}^{n} \ell_t(y^t; \boldsymbol{\theta}),$$

where $\ell_t(y^t; \boldsymbol{\theta}) = \ln p_t^{\boldsymbol{\theta}}(y^t)$ and $y^t = y_t | (y_{t-1}, \dots, y_{t-p})$. Since

$$p_t^{\boldsymbol{\theta}}(y^t) = \frac{1}{\sigma} g\left(\frac{y_t - \alpha - \sum_{k=1}^{p} \psi_k(y_{t-k}; \boldsymbol{\gamma}_k) y_{t-k}}{\sigma}\right) = \frac{1}{\sigma} g(z_t)$$

with $z_t = \xi_t/\sigma$, we have that $\ell_t(y^t; \boldsymbol{\theta}) = -\ln \sigma + \ln g(z_t)$.
The QML estimator is the solution of the maximization problem

$$\hat{\boldsymbol{\theta}}_n = \underset{\boldsymbol{\theta} \in \Theta}{\operatorname{argmax}} \frac{1}{n} \sum_{t=p+1}^{n} \ell_t(y^t; \boldsymbol{\theta}). \tag{10}$$

If we denote by $\nabla_{\boldsymbol{\theta}} \ell_t(y^t; \boldsymbol{\theta})$ and by $\nabla^2_{\boldsymbol{\theta}} \ell_t(y^t; \boldsymbol{\theta})$ the $D+2$-dimensional gradient and the $(D+2) \times (D+2)$-dimensional hessian matrix of $\ell_t(y^t; \boldsymbol{\theta})$ respectively, we are in the position to introduce the following theorem.

Theorem 4. *Under Assumptions* ***A1**–**A6*** *the QML estimator* $\hat{\boldsymbol{\theta}}_n$ *is strongly consistent for* $\boldsymbol{\theta}^0$ *and moreover it satisfies*

$$\bar{H}_n^0 \left(G_n^0\right)^{-1/2} \sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0) \xrightarrow{d} N(0, I),$$

where

$$\bar{H}_n^0 = \mathbb{E}\left[\frac{1}{n} \sum_t \nabla^2_{\boldsymbol{\theta}} \ell_t(Y^t; \boldsymbol{\theta}^0)\right],$$

$$G_n^0 = \mathbb{E}\left[\frac{1}{n} \sum_t \left(\nabla_{\boldsymbol{\theta}} \ell_t(Y^t; \boldsymbol{\theta}^0)\right) \left(\nabla_{\boldsymbol{\theta}} \ell_t(Y^t; \boldsymbol{\theta}^0)\right)^T\right]$$

and I *is the identity matrix of order* $D+2$.

Proof. Even if (9) is more general than the case of only one lag studied in Gobbi and Mulinacci (2021), the proof is a straightforward adaptation of that of the analogous result provided by Theorem 2.1 in Gobbi and Mulinacci (2021), where the asymptotic normality of the estimator is proved in the SDAR(1) case. The proof there given is based on the application of Theorems 3.13 and 6.4 in White (1994) whose assumptions are proved to be satisfied by the model. Since the derivatives of the likelihood are exactly of the same type as those considered there for the SDAR(1) model, apart from the partial second order derivatives of type

$$\frac{\partial^2}{\partial\gamma_{i,k}\partial\gamma_{j,h}}\ell_t\left(y;\boldsymbol{\theta}\right)=\frac{1}{\sigma^2}\psi_k^{\gamma_{i,k}}\left(y_{t-k};\boldsymbol{\gamma}_k\right)\psi_h^{\gamma_{j,h}}\left(y_{t-h};\boldsymbol{\gamma}_h\right)y_{t-k}y_{t-h}v(z_t)$$

where $v(x)=\left(g''(x)g(x)-\left(g'(x)\right)^2\right)/g^2(x)$, in the present framework we need to additionally consider a more general formulation of the uniform strong law of large numbers provided by Lemma 5.1 introduced in Gobbi and Mulinacci (2021). For the sake of completeness we state here the required result, without presenting the proof, that, under our assumptions follows exactly the same steps as that of Lemma 7.1 in Gobbi and Mulinacci (2021).

Lemma 5. *We assume that the stochastic process $(Y_t)_{t\in\mathbb{N}}$ follows the SDAR dynamics in* (9) *and that Assumptions* **A1**, **A2** *and* **A3** *are fulfilled. Let $q:\mathbb{R}^3\times\Theta\to\mathbb{R}$ given by $q(y_1,y_2,z,\boldsymbol{\theta})=f(y_1,y_2,\boldsymbol{\theta})h(z)$ so that $f:\mathbb{R}^2\times\Theta\to\mathbb{R}$ is continuous and such that $|f(y_1,y_2,\boldsymbol{\theta})|\leq H|y_1y_2|$ for some $H>0$ and all $(y_1,y_2,\boldsymbol{\theta})\in\mathbb{R}^2\times\Theta$ and $h:\mathbb{R}\to\mathbb{R}$ is a measurable function such that $\mathbb{E}\left[|h(z_t)|^\beta\right]<+\infty$, for some $\beta\geq 2$. Then*

$$\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{1}{n}\sum_{t=1}^{n}q(Y_{t-i},Y_{t-j},z_t,\boldsymbol{\theta})-\mathbb{E}\left[q(Y_{t-i},Y_{t-j},z_t,\boldsymbol{\theta})\right]\right|\to 0. \tag{11}$$

□

3.1 Two particular specifications of the model

The choice of the functional form of the autoregressive coefficients ψ_k in line with the Assumptions **A1**, **A4** and **A5** is not a trivial matter. In Gobbi and Mulinacci (2021) the authors discuss two families of continuous functions which are particularly suitable for shape and functional properties: both have, in fact, the characteristic of reducing the effect of persistence as the value of the state variable increases. In particular, we will focus on the following two particular type of persistence functions ψ_k:

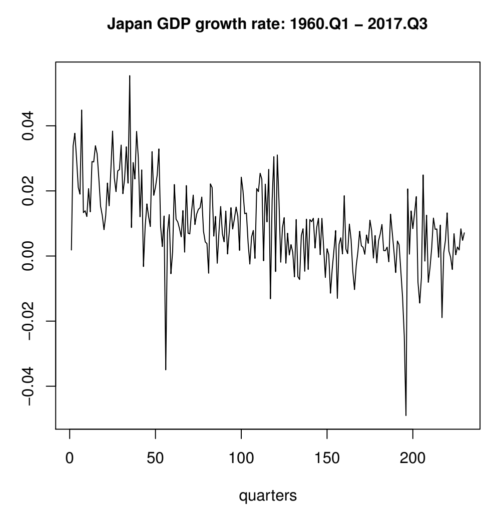
1. $\phi_e(x)=\pm\exp\left\{-\left(\gamma_1+\gamma_2x^{2\gamma_3}\right)\right\}$, with $\gamma_1>0$, $\gamma_2\geq 0$ and $\gamma_3>0$,
2. $\phi_p(x)=1/(\gamma_1+\gamma_2x^{2\gamma_3})$ with $\gamma_1>1$, $\gamma_2\geq 0$ and $\gamma_3>0$,

that fulfil Assumptions **A4** and **A5**. Since Assumption **A1** additionally requires that $\sum_{k=1}^{p} \sup_x |\psi_k(x)| < 1$ and being $\sup_x |\phi_e(x)| = \exp\{-\gamma_1\}$ and $\sup_x |\phi_p(x)| = 1/\gamma_1$, additional constraints on the parameters must be considered case by case, depending on the particular specification of the model and of the persistence functions involved. Some properties relating to the temporal dependence structure induced by these families of persistence functions are investigate in Gobbi (2021) through a Monte Carlo simulation experiment.

4 Empirical application and Monte Carlo experiment

The empirical data analysis has been carried out on the quarterly Japan GDP growth rate. The observation period goes from 1960.Q1 until 2017.Q3 (230 observations) and is depicted in Figure 1. The series appears mean-stationary but with a slightly decreasing dynamics which highlights the trend of low economic growth in Japan since the 90s. In fact, the average quarterly growth in the first half of the sample is around 1.5% while in the second half (precisely from the 1990s onwards) is 0.3%. The variance of the growth rate features the volatility clustering phenomenon with periods with high volatility followed by periods of low volatility. Furthermore, volatility is higher in the last part of the time series (indicatively after the great recession).

Figure 1: Quarterly Japan GDP growth rate



4.1 Descriptive statistics of the Japan GDP growth rate

Table 1 reports the summary of the descriptive statistics of the Japan GDP growth rate. The series is characterized by excess kurtosis and negative asymmetry. The asymmetry characterized by negative skewness means that in the sample period a greater probability exists of large decreases in GDP growth than larger increases while the kurtosis exhibits leptokurticity with fat tails highlighting that extreme changes can occur more frequently. The Jarque-Bera test (Jarque and Bera, 1980 and 1987) strongly rejects the normality hypothesis. Furthermore, the Ljung-Box test (Ljung and Box, 1978) indicates autocorrelation (up to 20 lags) in the time series. The McLeod-Li test (McLeod and Li, 1983) suggests a time-varying variance structure leading to the rejection of the null of no ARCH components up to 20 lags.

Table 1: Descriptive statistics for the Japan GDP growth rate

Japan GDP growth rate	
N. of obs.	230
Mean	0.00938
Median	0.00836
Maximum	0.05541
Minimum	-0.04921
Std. dev.	0.01288
Skewness	-0.11356
Kurtosis	2.24206
Jarque-Bera (p-value)	0.00000
Ljung-Box (p-value)	0.00000
McLeod-Li (p-value)	0.00000

4.2 Linearity tests

Table 2 reports the p-values of four different linearity tests performed on the full sample and on the last ten years of observations. For each test we consider different lag structures (lag=1,2,3). We employ four different linearity tests intensively used in the literature: the TNN test, the WNN test, the Tlrt test and the Tsay test. In the Teraesvirta Neural Network test (TNN test), introduced in Teraesvirta, Lin and Granger, (1993), and in the White Neural Network test (WNN test), discussed in Lee, White and Granger (1993), the null is the hypotheses of linearity in mean. The Tlrt test carry out the likelihood ratio test for threshold nonlinearity and was implemented by Chan (1990). The null hypothesis is that the fitted model to the time series is an AR model with a specified lag structure and the alternative is that the fitted model is a threshold autoregressive model with the same lag structure for each regime. Finally, the Tsay test, which was introduced and implemented in Tsay (1986) is a test for quadratic nonlinearity in a time series in which the null hypothesis is a normal AR process. The results show that there is a strong evidence of nonlinearity in the full series, since in a number of cases tests lead to the rejection of linearity. In particular, all tests highlight low p-values (less than 10%) when the lag structure considered is 1. On the other hand, whether the lag structure increases to 2 or 3 the nonlinearity is less strong. Tlrt test and the Tsay test do not reject the null of linearity for all lag structures, reflecting a weakness of the hypothesis of quadratic and threshold autoregressive nonlinearities.

Table 2: Linearity tests. p-values for different lag structures and different portions of the observed time series

Japan GDP: full sample 1960.Q1–2017.Q3			
	lag=1	lag=2	lag=3
TNN test	**0.00374**	**0.00298**	**0.00553**
WNN test	**0.00375**	**0.02571**	**0.02628**
Tlrt test	**0.06754**	0.17411	0.19391
Tsay test	**0.09310**	**0.04539**	0.13110
Japan GDP: last 10 years 2008.Q1–2017.Q3			
	lag=1	lag=2	lag=3
TNN test	**0.00053**	**0.00456**	**0.00000**
WNN test	**0.00056**	0.82980	0.31672
Tlrt test	0.28371	0.27641	0.32904
Tsay test	0.42711	0.18231	0.58571

Note: The cases highlighted in bold lead to the rejection of the null.

In order to realize whether the nonlinearity structure strengthens or not in more recent period, we conduct the same linearity tests in a portion of the sample corresponding to the last 10 years of observations. Table 2 shows a weakening of the nonlinearity to the point that only in one case (TNN test) the hypothesis is rejected for all lag structures. It is possible that this result can have consequences from the point of view of the forecasting evaluation as argued in Granger and Terasvirta (1993) and in Terasvirta and Anderson (1992). Indeed, in that papers the authors suggest that the superior in-sample performance of nonlinear models will only be matched out-of-sample if the nonlinear features also characterize the latest period of observation. Furthermore, even Ljung-Box and McLeod-Li tests provide p-values significantly high (0.8227 and 0.9986) indicating that this last portion of the time series is free from autocorrelation and heteroskedasticity.

4.3 Estimation

This section discusses the estimation results of the selected nonlinear models. The benchmark model is the standard linear autoregressive model of order 3 (AR(3)). To evaluate the forecasting accuracy of SDAR models with two lags we estimate five alternative nonlinear models within three different classes: SDAR models with one lag, SETAR models with two or three regimes and AR-GARCH models. We consider Gaussian error terms for all models. The estimation results are collected in Appendix A where we also report the functional form of the alternative models.

Results for each model are contained in Tables A1–A6 in Appendix A. The selection of the best combination of parameters within each class of nonlinear models is based on the AIC criterion. The AR lag order p is selected by fixing a maximum lag length. In Table A1 are reported the estimates for the AR(3) model which is the optimal linear model and will be used as benchmark regarding the forecasting accuracy. We can observe that all parameter estimates are positive and in particular the second coefficient is higher than the first and the third ones, i.e., for the second lag the persistence appears stronger. Regarding the self-exciting threshold autoregressive model with two regimes the AIC criterion leads to a SETAR(2,3,5) model whereas in the case of three regimes leads to select a SETAR(3,3,3,5) model as documented in Tables A2 and Table A3. A few words on the estimates are necessary. As for the SETAR(2,3,5) model all estimates relative to the low regime are positive. This can be interpreted as the poor ability of Japan GDP growth rate to react to the recessionary phases and could be consistent with the long-term trend which, as seen in Figure 1, is slightly negative. On the contrary, the second coefficient relating to the second lagged variable within the high regime is negative. On the other hand, in the SETAR(3,3,3,5) model the coefficient of y_{t-2} is negative in all three regimes, to emphasize that the effect of the GDP growth of two quarters earlier produced a contribution in the opposite direction to its value. As for AR-GARCH models, in line with the number of lags estimated in the linear model, we select an AR(3)-GARCH(1,1) model after checking that a higher order in the AR component or in the conditional variance structure produced a lower AIC. Table A4 shows the results. We note that the GARCH component, measured by ω_2, is strong, indicating a temporal structure in the variance dynamics.

Finally Table A5 and Table A6 report the results of estimating SDAR models with Gaussian error term for both specifications adopted, SDAR(1) and SDAR(2), depending on the chosen persistence functions ϕ_e and ϕ_p as introduced in Subsection 3.1. We consider a maximum of $p = 2$ lags to avoid over-parameterization problems. As the tables clearly indicate, all parameters are significant. However, as the estimates relating to previous models have shown, the coefficient of y_{t-2} appears to be particularly important. For this reason we expect that the SDAR model with one lag could not be adequate for the explanation and forecast of the dynamics of the Japan GDP growth rate. On the other hand, in the SDAR models with two lags the functional coefficients take into account the differences in temporal dynamics induced by y_{t-1} and y_{t-2}. In particular, if we observe the parameters estimated in the four SDAR models with two lags, $\hat{\gamma}_1$ and $\hat{\gamma}_2$, we can make some interesting observations. Let's take the first model SDAR(2;ϕ_e,ϕ_e) as an example. For each couple (y_{t-1}, y_{t-2}) we have $\phi_e(y_{t-2}; \hat{\gamma}_2) > \phi_e(y_{t-1}; \hat{\gamma}_1)$: this means that the second persistence function which refers to the second lag of the Japan growth rate, y_{t-2}, is constantly higher than the first persistence function which refers to the first lag of the Japan growth rate, y_{t-1}. A similar consideration holds for the remaining employed SDAR models. In other words, our SDAR models with two lags attribute a greater weight to the second lag than to the first and this is consistently true throughout the period under consideration.

To evaluate if the proposed models are well specified, we consider residuals variance and p-values associated to the McLeod-Li test up to 20 lags. Results reflect a good specification for all models under considerations since the hypothesis of absence of serial autocorrelation of squared residuals can be accepted.

4.4 A simulation experiment to detect the "level-increment" dependence

In this subsection we perform a Monte Carlo simulation experiment based on the models estimated above. In particular we simulate trajectories from the models specified in Appendix A: Equation (14) for the SETAR(2,3,5) model, Equation (15) for the SETAR(3,3,3,5) model, Equation (16) for the AR(3)-GARCH(1,1) model and Equations (17), (18), (19), (20), (21) and (22) for the SDAR models with one and two lags.

The aim is to investigate the ability of the considered models to capture the "level-increment" dependence in the time series of Japan GDP growth rate. As measure of dependence we use the Kendall's tau parameter that is a measure of concordance between pairs of random variables that only depends on the copula function linking them, that is it only depends on the dependence structure (see, among the others, Nelsen, 2006, for more details). We denote by τ_{GDP} the empirical Kendall's τ coefficient between y_{t-1} and $\Delta y_t = y_t - y_{t-1}$. It is likely that this ability or disability may have effects on the forecast skills of the models themselves. The simulation design is arranged in such a way that a number of trajectories will be generated from each selected model exactly with estimated parameters (Tables A2–A6 in Appendix A). For each trajectory we compute the Kendall's τ coefficient associated to the simulated pairs $(\tilde{y}_{t-1}^{(i)}, \Delta\tilde{y}_t^{(i)})$ getting a vector of realizations of τ, i.e., $(\tau^{(1)}, \dots, \tau^{(M)})$, where M is the number of simulations. The measure of the "level-increment" dependence is the average value of τ's, denoted by $\tilde{\tau}$. For each simulated model we compute the absolute relative distance (ARD) between the empirical Kendall's τ_{GDP} and $\tilde{\tau}$ as a measure of accuracy that each model has in intercepting the level-increment dependence

$$ARD = \frac{|\tilde{\tau} - \tau_{GDP}|}{\tau_{GDP}}.$$

We expect that the absolute relative distance between the dependence implied by each model and the actual data dependence increases, the model is less appropriate to provide a good forecasts.

Table 3 provides a summary of the simulation results. It is noted that the level-increment dependence is not intercepted with sufficient accuracy by at least 4 models: SDAR(2; ϕ_e, ϕ_p), SETAR(2,3,5) and SETAR(3,3,3,5). On the contrary, the SDAR(2; ϕ_e, ϕ_e), SDAR(2; ϕ_p, ϕ_e), SDAR(2; ϕ_p, ϕ_p) and AR(3)-GARCH(1,1) models are capable of reproducing this dependence effectively. In particular, the models report a very low ARD value compared to the simulated average values.

Table 3: Data dependence vs. model dependence

	Kendall's τ	ARD
Data dependence	**-0.3544**	
SDAR(1; ϕ_e)	-0.3174	0.1044
SDAR(1; ϕ_p)	-0.3227	0.0894
SDAR(2; ϕ_e, ϕ_e)	-0.3717	0.0488
SDAR(2; ϕ_e, ϕ_p)	-0.2249	0.3654
SDAR(2; ϕ_p, ϕ_e)	-0.3763	0.0618
SDAR(2; ϕ_p, ϕ_p)	-0.3465	0.0223
SETAR(2,3,5)	-0.4635	0.3078
SETAR(3,3,3,5)	-0.4693	0.3242
AR(3)-GARCH(1,1)	-0.3564	0.0056

Note: Kendall's τ coefficient: average values from the set of simulated trajectories and absolute relative distance (ARD) from data dependence.

4.5 Forecasting accuracy

This section presents results about the forecasting ability of the estimated models. In particular, our approach is that to generate multi-period forecasts using a Monte Carlo simulation. This technique is particularly efficient when nonlinear models for time series are used (see, for example, Granger and Terasvirta, 1993). As for SDAR(2) models, we follow the same simulation setting proposed in Gobbi and Mulinacci (2021) for the case of one lag. Denote with H the forecast horizon. Let $\tilde{y}^{(m)}_{n+h}$ be the $(n+h)$-th forecast with $h = 1, \dots, H$ obtained in the m-th simulation where n is the time of the last observation in the sample. The forecast simulation scheme for a SDAR(2) model is

$$\begin{cases} \tilde{y}^{(m)}_{n+1} = \hat{\alpha} + \psi_1(y_n; \hat{\boldsymbol{\gamma}}_1)y_n + \psi_2(y_{n-1}; \hat{\boldsymbol{\gamma}}_2)y_{n-1} + \tilde{\xi}^{(m)}_{n+1}, \\ \tilde{y}^{(m)}_{n+2} = \hat{\alpha} + \psi_1(\tilde{y}^{(m)}_{n+1}; \hat{\boldsymbol{\gamma}}_1)\tilde{y}_{n+1} + \psi_2(y_n; \hat{\boldsymbol{\gamma}}_2)y_n + \tilde{\xi}^{(m)}_{n+2}, \\ \tilde{y}^{(m)}_{n+h} = \hat{\alpha} + \psi_1(\tilde{y}^{(m)}_{n+h-1}; \hat{\boldsymbol{\gamma}}_1)\tilde{y}^{(m)}_{n+h-1} + \psi_2(\tilde{y}^{(m)}_{n+h-2}; \hat{\boldsymbol{\gamma}}_1)\tilde{y}^{(m)}_{n+h-2} + \tilde{\xi}^{(m)}_{n+h}, \; h = 3, ..., H \end{cases} \tag{12}$$

where ψ_1 and ψ_2 are equal to ϕ_e or ϕ_p, $\tilde{\xi}^{(m)}_{n+1}$ and $\tilde{\xi}^{(m)}_{n+h}$ are normally distributed with zero mean and standard deviation $\hat{\sigma}$ for each simulation m. Now, averaging these forecasts across the $m = 1, .., M$ iterations of the MC yields

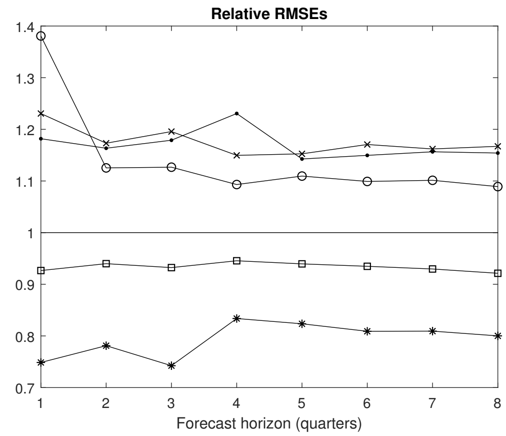
$$\tilde{y}_{n+h} = \sum_{m=1}^{M} \tilde{y}_{n+h}^{(m)}, \quad h = 1, .., H.$$

We use 8 values out-of-sample of the Japan GDP growth rate from 2017.Q4 to 2019.Q3. Therefore, the forecast errors are given by $e_{n+h} = y_{n+h} - \tilde{y}_{n+h}$, with $h = 1, \ldots, H$, with $H = 8$. To compare the average accuracy of the forecasts we use two measures: the root mean square error (RMSE) and the mean absolute error (MAE). The RMSE is defined as $\sqrt{\frac{1}{H}\sum_{h=1}^{H} e_{n+h}^2}$ whereas the MAE is given by $\frac{1}{H}\sum_{h=1}^{H} |e_{n+h}|$. Table 4, Figures 2–5 summarize the results in terms of relative efficiencies for a forecast horizon from 1 to 8 quarters ahead. The relative efficiency (RE) is obtained as the ratio of the RMSE (or MAE) of the model under consideration and the RMSE (MAE) of the model used as benchmark, i.e., the linear AR(3). A value of RE greater or equal than one indicates that the benchmark model provides more accurate forecast than the alternative nonlinear model. Results are very significant. As we see there are two models which provide more accuracy than the benchmark for all forecast horizons, i.e., the SETAR(2,3,5) model and the AR(3)-GARCH(1,1) model, even if the first one offers a superior performance reaching a gain of over 30% for $H = 3$. The SETAR(3,3,3,5) and both SDAR(1;ϕ_e) and SDAR(1;ϕ_p) models do worse than the benchmark.

On the other hand, SDAR models with two lags guarantee a performance that depends on their specification. In fact, as highlighted in Table 4 and Figures 3 and 5, except for the model SDAR(2;ϕ_e,ϕ_p), the other specifications provide a forecasting accuracy which is higher than the benchmark from the third quarter onwards. But even more significant is the fact that the performance in the last four quarters is the best among all tested models. What seems to be concluded is the ability of the SDAR(2) models, in at least three out of four specifications, to guarantee a gain in forecast accuracy in the medium-long term. For example, for $H = 8$ (8 quarters ahead) the SDAR(2;ϕ_e,ϕ_e) has a forecasting capacity 40% higher than that of the linear AR(3) model.

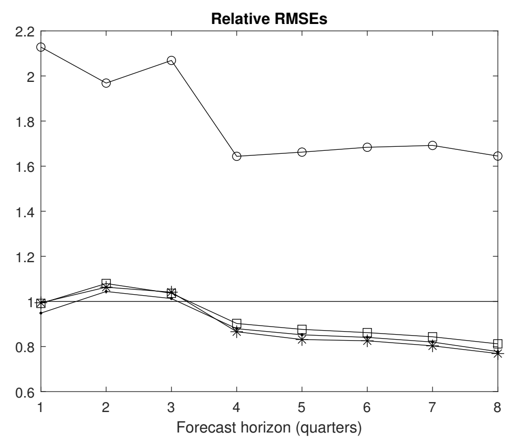
We can make some considerations by comparing the results contained in Table 3 and Table 4. As we can observe, the AR(3)-GARCH(1,1) model is the one most able to intercept the level-increment dependence followed by the SDAR(2; ϕ_p, ϕ_p) and SDAR (2; ϕ_e, ϕ_e). All three of these models have a good predictive ability. The AR(3)-GARCH (1,1) model is constantly better than the benchmark without being influenced by the forecast horizon, whereas SDAR(2; ϕ_p, ϕ_p) and SDAR (2; ϕ_e, ϕ_e) models offer excellent performance, especially in the medium-long term. On the other hand, the SDAR models with a single lag provide a worse forecasting ability and at the same time worse capacity to reproduce the level-increment dependence as we deduce from Table 3. Differently, for the SETAR models does not seem to be any correlation between the two behaviors. Both do not intercept the level-increment dependence

Figure 2: Relative RMSEs for each nonlinear model expressed in terms of that for AR(3)



Legend: "*" for SETAR(2,3,5), "o" for SETAR(3,3,3,5), "□" for AR(3)-GARCH(1,1), "." for SDAR(1;ϕ_e) and "x" for SDAR(1;ϕ_p).

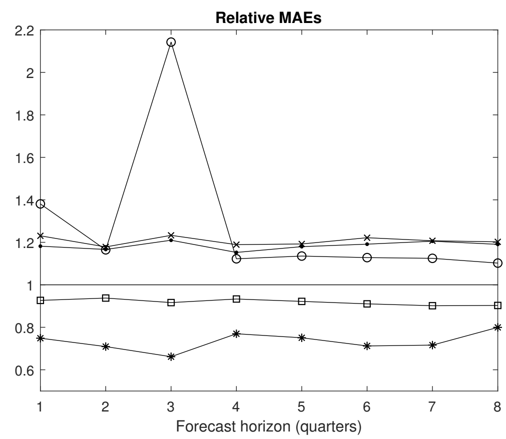
Figure 3: Relative RMSEs for each nonlinear model expressed in terms of that for AR(3)



Legend: "*" for SDAR(2;ϕ_e, ϕ_e), "o" for SDAR(2;ϕ_e, ϕ_p), "□" for SDAR(2;ϕ_p, ϕ_e), "." for SDAR(2;ϕ_p, ϕ_p).

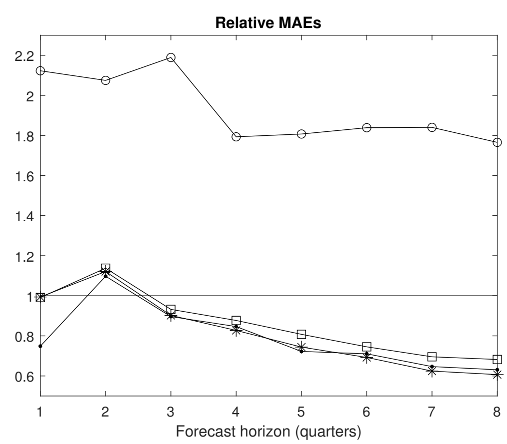
implied in the time series of the Japan GDP growth rate but the model with two thresholds offers excellent forecasts for all the time horizons analyzed.

Figure 4: Relative MAEs for each nonlinear model expressed in terms of that for AR(3)



Legend: "*" for SETAR(2,3,5), "o" for SETAR(3,3,3,5), "□" for AR(3)-GARCH(1,1), "." for SDAR(1;ϕ_e) and "x" for SDAR(1;ϕ_p).

Figure 5: Relative MAEs for each nonlinear model expressed in terms of that for AR(3)



Legend: "*" for SDAR(2;ϕ_e, ϕ_e), "o" for SDAR(2;ϕ_e, ϕ_p), "□" for SDAR(2;ϕ_p, ϕ_e), "." for SDAR(2;ϕ_p, ϕ_p).

Table 4: Relative efficiency of the forecasting accuracy measures RMSE (bottom) and MAE (down) with the AR(3) model as benchmark

	Number of quarters ahead							
H	1	2	3	4	5	6	7	8
SETAR(2,3,5)	0.7486	0.7811	0.7424	0.8337	0.8235	0.8090	0.8093	0.8001
	0.7486	0.7092	0.6615	0.7695	0.7506	0.7122	0.7161	0.7995
SETAR(3,3,3,5)	1.3808	1.1253	1.1267	1.0931	1.1095	1.0991	1.1013	1.0891
	1.3808	1.1646	2.1424	1.1224	1.1353	1.1280	1.1246	1.1022
AR(3)-GARCH(1,1)	0.9266	0.9399	0.9323	0.9455	0.9395	0.9348	0.9296	0.9213
	0.9266	0.9376	0.9163	0.9331	0.9219	0.9103	0.9015	0.9028
SDAR(1;ϕ_e)	1.1818	1.1634	1.1789	1.2306	1.1426	1.1496	1.1565	1.1541
	1.1818	1.1664	1.2098	1.1524	1.1796	1.1913	1.2052	1.1904
SDAR(1;ϕ_p)	1.2306	1.1731	1.1957	1.1496	1.1526	1.1706	1.1620	1.1671
	1.2306	1.1783	1.2329	1.1890	1.1922	1.2216	1.2073	1.2022
SDAR(2;ϕ_e,ϕ_e)	0.9941	1.0631	1.0410	0.8656	0.8310	0.8253	0.8030	0.7688
	0.9941	1.1211	0.9021	0.8272	0.7440	0.6924	0.6242	0.6067
SDAR(2;ϕ_e,ϕ_p)	2.1281	1.9684	2.0689	1.6436	1.6622	1.6837	1.6919	1.6448
	2.1228	2.0747	2.1890	1.7932	1.8072	1.8387	1.8402	1.7655
SDAR(2;ϕ_p,ϕ_e)	0.9913	1.0794	1.0361	0.9022	0.8760	0.8616	0.8431	0.8119
	0.9913	1.1377	0.9322	0.8769	0.8077	0.7453	0.6955	0.6822
SDAR(2;ϕ_p,ϕ_p)	0.9482	1.0441	1.0127	0.8791	0.8520	0.8405	0.8198	0.7775
	0.7486	1.0976	0.8963	0.8464	0.7224	0.7105	0.6467	0.6309

Note: A value of the ratio lesser than 1 indicates that the nonlinear model ensures more accuracy than the AR(3) model.

5 Concluding remarks

We introduce a class of nonlinear time series where the error term is dependent on a number of lagged values of the state variable. To let the model be tractable in practice, we focus on an autoregressive model in which the error term is independent of the lagged values of the state variables but the autoregressive coefficients are specified functions of the state variables themselves. We call this model p-order state-dependent autoregressive (SDAR(p)) model. We consider the case with p lags and a generic distribution of the error term. This model is strictly stationary and geometrically ergodic under very mild conditions satisfied by the functional autoregressive coefficients $\psi_k(y_{t-k})$, $k = 1, \ldots, p$. Furthermore, we propose a maximum likelihood estimator of parameters and we prove its consistency and asymptotic normality. A simulation experiment shows the ability of the model to capture the level-increment dependence. Finally, an empirical analysis of the Japan GDP growth rate is conducted, to evaluate the forecasting performance of a normal-SDAR(2) model compared with three alternative nonlinear models such as SETAR, AR-GARCH and normal-SDAR(1). The results show that if the autoregressive coefficients are appropriately chosen, the normal-SDAR(2) model provides a better forecast performance especially for medium-long horizons.

References

[1] An H. Z., Huang F. C., (1996), The geometrical ergodicity of nonlinear autoregressive models, *Statistica Sinica* 6, 943–956.

[2] Bollerslev T., (1986), Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics* 31(3), 307–327.

[3] Bollerlsev T., Chou R. Y., Kroner K. F., (1992), ARCH modelling in finance: a review of the theory and empirical evidence, *Journal of Econometrics* 52, 5–59.

[4] Brock W. A., Dechert W. D., Sheinkman J. A., (1987), A Test of Independence Based on the Correlation Dimension, SSRI no. 8702, Department of Economics, University of Wisconsin, Madison.

[5] Brock W. A., Dechert W. D., Sheinkman J. A., LeBaron B., (1996), A Test of Independence Based on the Correlation Dimension, *Econometric Reviews* 15, 197–235.

[6] Cai Z., Fan J., Yao Q., (2000), Functional-Coefficient Regressive Models for Nonlinear Time Series, *Journal of the American Statistical Association* 95(451), 941–956.

[7] Chan K. S., (1991), Percentage points of likelihood ratio tests for threshold autoregression, *Journal of Royal Statistical Society B* 53(3), 691–696.

[8] Chen R., Liu M. L., (2001), Functional coefficient autoregressive models: estimation and tests of hypotheses, *Journal of Time Series Analysis* 22(2), 151–173.

[9] Chen R., Tsay R., (1993), Functional-coefficient autoregressive models, *Journal of the American Statistical Association* 88, 298–308.

[10] Cherubini U., Mulinacci S., Romagnoli S., (2011), A Copula-based Model of Speculative Price Dynamics in Discrete Time, *Journal of Multivariate Analysis* 102, 1047–1063.

[11] Cherubini U., Gobbi F., Mulinacci S., Romagnoli S., (2012), *Dynamic copula methods in Finance*, John Wiley & Sons.

[12] Cherubini U., Gobbi F., (2013), A Convolution-based Autoregressive Process, [in:] Workshop on Copula in Mathematics and Quantitative Finance. Lecture Notes in Statistics-Proceedings, [eds.:] F. Durante, W. Haerdle, P. Jaworski, Springer, Berlin/Heidelberg.

[13] Cherubini U., Gobbi F., Mulinacci S., (2016), *Convolution Copula Econometrics*, Springer Briefs in Statistics.

[14] Clements M. P., Smith J., (1997), The performance of alternative forecasting methods for SETAR models, *International Journal of Forecasting* 13, 463–475.

[15] Crowder M. J., (1976), Maximum likelihood estimation for dependent observations, *Journal of Royal Statistical Society B* 38, 45–53.

[16] Darsow W. F., Nguyen B., Olsen E. T., (1992), Copulas and Markov processes, *Illinois Journal of Mathematics* 36(4), 600–642.

[17] Douc R., Moulines E., Stoffer D. S., (2014), *Nonlinear Time Series. Theory, Methods and Applications with R Examples*, CRC Press.

[18] Engle R., (1982), Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica* 50(4), 987–1007.

[19] Fan J., Yao Q., (2003), *Nonlinear Time Series: Nonparametric and Patrametric Methods*, Springer-Verlag, New York.

[20] Fang W., Miller S. M., (2009), Modeling the volatility of real GDP growth: The case of Japan revisited, *Japan and the World Economy* 21(3), 312–324.

[21] Fountas S., Karanasos M., Mendoza A., (2004), Output variability and economic growth: the Japanese case, *Bullettin of Economic Research* 56, 353–363.

[22] Gobbi F., Mulinacci S., (2020), Mixing and moments properties of a non-stationary copula-based Markov process, *Communications in Statistics – Theory and Methods* 49(18), 4559–4570.

[23] Gobbi F., Mulinacci S., (2020), State-Dependent Autoregressive Models: Properties, Estimation and Forecasting, available at: arXiv:2002.03134v1.

[24] Gobbi F., (2021), The problem of detecting nonlinearity in time series generated by a state-dependent autoregressive model. A simulation study, to appear in: *International Journal of Operational Research.*

[25] Granger C. W. J., Terasvirta T., (1993), *Modelling Nonlinear Economics Relationships*, Oxford University Press, Oxford.

[26] Haggan V., Ozaki T., (1981), Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model, *Biometrica* 68(1), 189–196.

[27] Hamori S., (2000), Volatility of real GDP: Some evidence from the United States, the United Kingdom and Japan, *Japan and the World Economy* 12, 143–152.

[28] Hastie T., Tibshirani R., (1990), *Generalized additive models*, Chapman & Hall, New York.

[29] Ho K. Y., Tsui A. K. C., (2003), Asymmetric volatility of real GDP: some evidence from Canada, Japan, United Kingdom, and the United States, *Japan and the World Economy* 15, 437–445.

[30] Ibragimov R., (2009), Copula-based Characterizations for Higher Order Markov Processes, *Econometric Theory* 25, 819–846.

[31] Jarque C. M., Bera A. K., (1980), Efficient tests for normality homoscedasticity and serial independence of regression residuals, *Economics Letters* 6, 255–259.

[32] Jarque C. M., Bera A. K., (1987), A test for normality of observations and regression residuals, *International Statistical Review* 55, 163–172.

[33] Krager H., Kluger P., (1993), Nonlinearities in foreign exchange markets: a different perspective, *Journal of International Money and Finance* 12, 195–208.

[34] Lee T. H., White H., Granger C. W. J., (1993), Testing for neglected nonlinearity in time series models: A comparison of neural network methods and alternative tests, *Journal of Econometrics* 56, 269–290.

[35] Ljung G. M., Box G. E. P., (1978), On a measure of lack of fit in time series models, *Biometrika* 65, 297–303.

[36] McLeod A. I., Li W. K., (1983), Diagnostic checking ARMA time series models using squared residual autocorrelations, *Journal of Time Series Analysis* 4, 269–273.

[37] Nelsen R. B., (2006), *An Introduction to Copulas*, Springer.

[38] Potter S., (1995), A nonlinear approach to U.S. GNP, *Journal of Applied Econometrics* 10, 109–125.

[39] Priestley M. B., (1980), State-Dependent Models: A General Approach to Nonlinear Time Series Analysis, *Journal of Time Series Analysis* 1, 47–71.

[40] Terasvirta T., Anderson H., (1992), Characterizing Nonlinearities in Business Cycle Using Smooth Transition Autoregressive Models, *Journal of Applied Econometrics* 7, 119–136.

[41] Terasvirta T., Lin C. F., Granger C. W. J., (1993), Power of the Neural Network Linearity Test, *Journal of Time Series Analysis* 14, 209–220.

[42] Tjostheim D., (1986), Estimation in nonlinear time series models, *Stochastic Processes and their Applications* 21, 251–273.

[43] Tiao G. C., Tsay R. S., (1994), Some advances in nonlinear and adaptive modelling time series, *Journal of Forecasting* 13, 109–131.

[44] Tong H., (1978), On a threshold model, [in:] Pattern Recognition and Signal Processing, [eds.:] C. H. Chen, 101-141, Amsterdam: Sijhoff and Noordoff.

[45] Tong H., Lim K. S., (1980), Threshold autoregression, limit cycles and cyclical data, *Journal of the Royal Statistical Society* B 42, 245–268.

[46] Tong H., (1983), *Threshold Models in Nonlinear Time Series Analysis*, Springer-Verlag, New York.

[47] Tong H., (1986), On estimating thresholds in autoregressive models, *Journal of Time Series Analysis* 7, 178–190.

[48] Tong H., (1995), *Nonlinear Time Series: A Dynamical System Approach*, Oxford University Press.

[49] Tsay R.S., (1986), Nonlinearity Tests For Time Series, *Biometrika* 73, 461–466.

[50] White H., (1984), *Asymptotic theory for econometricians*, Academic Press, New York.

[51] White H., (1994), *Estimation, Inference and Specification Analysis*, Econometric Society Monographs n. 22, Cambridge University Press, Cambridge.

Appendix A Estimation results

In this appendix we report the parameter estimates and relative standard errors for each model we have considered.

i) The benchmark is the linear AR(3) model.

$$\begin{cases} y_t = \alpha + \sum_{i=1}^{3} \phi_i y_{t-i} + \xi_t, \quad t \geq 2, \\ \xi_t \sim IID\ N(0, \sigma), \end{cases} \tag{13}$$

where ξ_t is independent of the lagged variables $y_{t-1}, \dots, y_{t-p}$. The vector of parameters is $\boldsymbol{\theta} = (\alpha, \phi_1, \phi_2, \phi_3, \sigma)$.

ii) Self-exciting threshold autoregressive (SETAR) models were first proposed in Tong (1978, 1983), Tong and Lim (1980) and discussed in detail in Tong (1995). SETAR models considered in this paper assume that a variable y_t is a linear autoregression within a regime, but may move between regimes depending on the value assumed by the first lag y_{t-1}. We estimate two SETAR models, the

Table A1: AR(3) model

	AR(3)	
Parameters	Estimate	SE
α	-0.0001744	0.000722
ϕ_1	0.2108	0.064422***
ϕ_2	0.2481	0.063667***
ϕ_3	0.1922	0.063629***
Residuals variance	0.0001189	
ML test (p-value)	0.1161	

first with two regimes and the second with three regimes. We denote SETAR(2, p_1, p_2) the model with two regimes whose specification is

$$\begin{cases} y_t = \alpha_1 + \sum_{i=1}^{p_1} \phi_{1,i} y_{t-i} + \xi_{1,t}, & y_{t-1} \leq v, \\ y_t = \alpha_2 + \sum_{i=1}^{p_2} \phi_{2,i} y_{t-i} + \xi_{2,t}, & y_{t-1} > v, \end{cases} \tag{14}$$

where v is the threshold variable, p_1 and p_2 are the orders of the linear AR within each regime, $\xi_{j,t} \sim IID\ N(0, \sigma_j)$, $j = 1, 2$. Furthermore $\xi_{1,t}$ and $\xi_{2,t}$ are independent for all t. The vector of parameters is $\boldsymbol{\theta} = (\alpha_1, \alpha_2, \phi_{1,1}, \ldots, \phi_{1,p_1}, \phi_{2,1}, \ldots, \phi_{2,p_2}, \sigma_1, \sigma_2)$. SETAR model with three regimes, denoted by SETAR(3, p_1, p_2, p_3) is defined as

$$\begin{cases} y_t = \alpha_1 + \sum_{i=1}^{p_1} \phi_{1,i} y_{t-i} + \xi_{1,t}, & y_{t-1} \leq v_1, \\ y_t = \alpha_2 + \sum_{i=1}^{p_2} \phi_{2,i} y_{t-i} + \xi_{2,t}, & v_1 < y_{t-1} \leq v_2, \\ y_t = \alpha_3 + \sum_{i=1}^{p_3} \phi_{3,i} y_{t-i} + \xi_{3,t}, & y_{t-1} > v_2, \end{cases} \tag{15}$$

where v_1 and v_2, with $v_1 < v_2$ are two threshold variables, p_1, p_2 and p_3 are the orders of the linear AR within each regime $\xi_{j,t} \sim IID\ N(0, \sigma_j)$, $j = 1, 2, 3$. The vector of parameters is $\boldsymbol{\theta} = (\alpha_1, \alpha_2, \alpha_3, \phi_{1,1}, \ldots, \phi_{1,p_1}, \phi_{2,1}, \ldots, \phi_{2,p_2}, \phi_{3,1}, \ldots, \phi_{3,p_3} \sigma_1, \sigma_2, \sigma_3)$.

iii) GARCH models were proposed in Bollerslev (1986) as a generalization of ARCH model introduced in Engle (1982). In this paper we consider an AR(p) component in place of a constant mean for the Equation of the variable y_t in light of the preliminary analysis carried out in the previous section on the time series of Japan GDP growth rate. Therefore, our specification of the model is the following

$$\begin{cases} y_t = \alpha + \sum_{i=1}^{p} \phi_i y_{t-i} + \xi_t, & t \geq 1, \\ \xi_t | \mathcal{F}_{t-1} \sim IID\ N(0, h_t), \\ h_t^2 = \omega_0 + \omega_1 y_{t-1}^2 + \omega_2 h_{t-1}^2, \end{cases} \tag{16}$$

where $\mathcal{F}_{t-1}$ is the information set which includes the lagged values of the variable

Table A2: SETAR(2,3,5) model

	SETAR(2,3,5)	
Parameters	Estimate	SE
α_1	0.0030231	0.0010076**
$\phi_{1,1}$	0.1875983	0.0733370*
$\phi_{1,2}$	0.0097874	0.0883190
$\phi_{1,3}$	0.2226274	0.0753082**
α_2	0.0163617	0.0066186*
$\phi_{2,1}$	0.0065853	0.1338019
$\phi_{2,2}$	-0.1024677	0.2301083
$\phi_{2,3}$	0.0536636	0.1274458
$\phi_{2,4}$	0.0481957	0.1252912
$\phi_{2,5}$	0.2998508	0.1418985*
Residuals variance	0.0001046	
ML test (p-value)	0.3197	

Table A3: SETAR(3,3,3,5) model

	SETAR(3,3,3,5)	
Parameters	Estimate	SE
α_1	0.0014215	0.0013503
$\phi_{1,1}$	0.2190270	0.1024155*
$\phi_{1,2}$	-0.2143976	0.1346354
$\phi_{1,3}$	0.1410807	0.1139681
α_2	0.0065093	0.0039263*
$\phi_{2,1}$	0.1548477	0.1040731
$\phi_{2,2}$	-0.1990346	0.3437589
$\phi_{2,3}$	0.2796797	0.1000679**
α_3	0.0163617	0.0065656*
$\phi_{3,1}$	0.0065853	0.1327299
$\phi_{3,2}$	-0.1024677	0.2282648
$\phi_{3,3}$	0.0536636	0.1264248
$\phi_{3,4}$	0.0481957	0.1242874
$\phi_{3,5}$	0.2998508	0.1407616*
Residuals variance	0.0001011	
ML test (p-value)	0.4315	

$y_{t-1}, y_{t-2}, \ldots$ and the conditional variance has a GARCH(1,1) specification. The vector of parameters is $\boldsymbol{\theta} = (\alpha, \phi_1, \ldots, \phi_p, \omega_0, \omega_1, \omega_2)$.

Table A4: AR(3)-GARCH(1,1) model

	AR(3)-GARCH(1,1)	
Parameters	Estimate	SE
α	0.00215	0.00085*
ϕ_1	0.20200	0.07170**
ϕ_2	0.28701	0.06854***
ϕ_3	0.24822	0.06853***
ω_0	0.00001	0.00000
ω_1	0.20791	0.08197*
ω_2	0.69401	0.10071***
Residuals variance	0.0001183	
ML test (p-value)	0.7963	

iv) Let us consider SDAR models with one or two lags with persistence function of type specified in Subsection 3.1. We list explicitly the models in order to facilitate the reading of the estimate results.

$$SDAR(1;\phi_e):$$
$$y_t = \alpha + \exp\left(-\left(\gamma_{1,1} + \gamma_{2,1} y_{t-1}^{2\gamma_{3,1}}\right)\right) y_{t-1} + +\xi_t, \quad t \geq 1, \tag{17}$$

$$SDAR(1;\phi_p):$$
$$y_t = \alpha + \frac{1}{\gamma_{1,1} + \gamma_{2,1} y_{t-1}^{2\gamma_{3,1}}} y_{t-1} + \xi_t, \quad t \geq 1, \tag{18}$$

$$SDAR(2;\phi_e,\phi_e):$$
$$y_t = \alpha + \sum_{k=1}^{2} \exp\left(-\left(\gamma_{1,k} + \gamma_{2,k} y_{t-k}^{2\gamma_{3,k}}\right)\right) y_{t-k} + \xi_t, \quad t \geq 2, \tag{19}$$

$$SDAR(2;\phi_e,\phi_p):$$
$$y_t = \alpha + \exp\left(-\left(\gamma_{1,1} + \gamma_{2,1} y_{t-1}^{2\gamma_{3,1}}\right)\right) y_{t-1} + \frac{1}{\gamma_{1,2} + \gamma_{2,2} y_{t-2}^{2\gamma_{3,2}}} y_{t-2} + \xi_t, \quad t \geq 2, \tag{20}$$

$$SDAR(2;\phi_p,\phi_e):$$
$$y_t = \alpha + \frac{1}{\gamma_{1,1} + \gamma_{2,1} y_{t-1}^{2\gamma_{3,1}}} y_{t-1} + \exp\left(-\left(\gamma_{1,2} + \gamma_{2,2} y_{t-2}^{2\gamma_{3,2}}\right)\right) y_{t-2} + \xi_t, \quad t \geq 2, \tag{21}$$

$$SDAR(2; \phi_p, \phi_p):$$

$$y_t = \alpha + \sum_{k=1}^{2} \frac{1}{\gamma_{1,k} + \gamma_{2,k} y_{t-k}^{2\gamma_{3,k}}} y_{t-k} + \xi_t, \quad t \geq 2. \tag{22}$$

In all four specified models we have $\xi_t \sim \ IID\ N(0, \sigma^2)$ and independent of y_{t-j} for $j \geq 1$.

Table A5: SDAR(1) models

	SDAR(1,ϕ_e)		SDAR(1, ϕ_p)	
Parameters	Estimate	SE	Estimate	SE
α	0.00421	0.00054***	0.00438	0.00055***
$\gamma_{1,1}$	0.56975	0.12981***	1.78480	0.18149***
$\gamma_{2,1}$	0.42785	0.16243***	3.77341	0.51811***
$\gamma_{3,1}$	0.37173	0.07017***	0.50019	0.00003***
σ	0.01189	0.00287***	0.011715	0.00289***
Residuals variance	0.0001484		0.0001422	
ML test (p-value)	0.02251		0.04134	

Table A6: SDAR(2) models

	SDAR(2,ϕ_e,ϕ_e)		SDAR(2, ϕ_e,ϕ_p)	
Parameters	Estimate	SE	Estimate	SE
α	0.00312	0.00045***	0.00665	0.00081***
$\gamma_{1,1}$	1.00962	0.12941***	0.84511	0.09783***
$\gamma_{2,1}$	1.49491	0.21675***	1.72401	0.41976***
$\gamma_{3,1}$	0.18518	0.01828***	0.72136	0.09058***
$\gamma_{1,2}$	0.52757	0.12439**	1.94412	0.16744***
$\gamma_{2,2}$	0.88476	0.23307***	0.97394	0.24906***
$\gamma_{3,2}$	0.11285	0.02439***	0.98946	0.05156***
σ	0.01138	0.00129***	0.01350	0.00124***
Residuals variance	0.0001291		0.0001441	
ML test (p-value)	0.1480		0.1425	
	SDAR(2, ϕ_p,ϕ_e)		SDAR(2,ϕ_p ,ϕ_p)	
Parameters	Estimate	SE	Estimate	SE
α	0.00386	0.00064***	0.00324	0.00074***
$\gamma_{1,1}$	3.65489	0.60401***	2.96076	0.60564***
$\gamma_{2,1}$	3.42116	0.70877***	3.31371	0.55915***
$\gamma_{3,1}$	0.58821	0.04357***	0.50018	0.00088***
$\gamma_{1,2}$	1.17677	0.16892***	3.18532	0.47609***
$\gamma_{2,2}$	0.07385	0.03075***	1.70891	0.72061***
$\gamma_{3,2}$	0.54835	0.10944***	0.86593	0.08491***
σ	0.01137	0.01291***	0.01138	0.00149***
Residuals variance	0.0001533		0.0001274	
ML test (p-value)	0.1645		0.1325	