Smooth Transition AR Models in Twinkle.
(Version 1.0)

Alexios Ghalanos

March 19, 2014

Contents

1 Introduction

The motivation for creating a package for the modelling of smooth transition ARMA models was the observation that many phenomena appear to go through different states which may be explained by some underlying and observable factors.[1] In financial markets, under different states of the business cycle, financial instruments have been observed to exhibit different characteristics with recessions (or the lead up to such) usually marked by increased volatility and lower or negative mean. Being able to model the evolution of the process driving such changes and hence the switch from one state to another must surely make for a better understanding of the underlying dynamics and perhaps lead to a better forecast model.

The precursor to smooth transition models appears to have been Carmichael (1928) who posited the use of the arctangent transformation and even considered the possibility of a double transition well before the plethora of papers which came more than 30 years later. While Quandt (1958) originally discussed a switching regression model, the pioneering contributions to the literature on more general threshold autoregression (TAR) have been Tong and Lim (1980) and Tong and Ghaddar (1981), with a more general class of nonlinear AR models introduced in a series of papers by Billings and Voon (1983), Billings and Voon (1986), and Zhu and Billings (1993). Many extensions to the basic TAR model have been considered in the literature , including smooth transition based on the Gaussian CDF in Chan and Tong (1986), the more widely adopted logistic (popularized by May (1976) and discussed in Tong (1990)) (LSTAR) and exponential (ESTAR) models discussed in Teräsvirta (1994), double transitions in . . . , nested transition in Astatkie et al. (1997), multiple states in van Dijk and Franses (1999), and the inclusion of GARCH dynamics in Chan and McAleer(**?**Chan2002), 2003). A review of recent extensions and the state of research in this area can be found in Dijk et al. (2002).

Common themes among the vast majority of research in the area of smooth transition AR models are the use of the self-exciting model, where the lagged value of the dependent variable (or some simple transformation of the same) is used and a representation for the state dynamics which appears overly restrictive. The twinkle package departs from the traditional representation which appears to have dominated research in the area of STAR models and reparameterizes the state dynamics to include a possible linear combination of multivariate variables and the use of autoregressive first order dynamics. The m-states model is also reparameterized to more closely resemble the representation of the multinomial logistic regression model in the way the probabilities are summed and weighted across states. Further extensions include a 2-state AR mixture model partially bridging the gap with the finite mixture models, the inclusion of GARCH dynamics, MA dynamics either inside or outside the states, and a large number of conditional distributions which follow from the rugarch package. Methods for model specification, estimation, filtering, forecasting and simulation are provided with similar interface and access methods as in related packages by the author. It is important for the interested user to be aware from the start that such models are difficult to estimate, may contain local minima and may be generally hard to solve with confidence. While the package has made efforts to provide for a number of solvers and strategies to estimate these models, confident estimation may prove challenging depending on properties of the dataset used and choice of model options. The model is naturally greedy in requiring a substantial amount of data to confidently identify the optimal classification of states given the conditional mean equation. From experience, it is this author's opinion that these types of models may not be as forgiving as linear models when it comes to forecasting, depending on whether the actual forecast state contains the type of nonlinearities under which the model was estimated. Thus, unlike linear models, the misclassification of the forecast state may be more costly.

[1]As opposed to unobservable factors which leads to a different class of models, such as the Markov switching models.

This paper is organized as follows: Section 2 discusses the representation of the model in the twinkle package and how the model can be specified. Section 3 discusses forecasting with a special emphasis on the different methods implemented for n-period ahead forecasts, followed by Section 4 on the simulation. Finally, Section 5 presents a number of examples using real and simulated data. While every possible effort has been made to test the model and its methods under different scenarios and squash any bugs, the package is still quite new and further testing is required. General questions on the package should be posted to the R-SIG-FINANCE mailing list, while bugs (with reproducible code and preferably a patch) and suggestions can be reported directly to me. Finally I would like to acknowledge the valuable help of Eduardo Rossi who collaborated on the new representation and research publication in this area.

2 Smooth Transition ARMA Models Revisited

2.1 Dynamics and Extensions

Consider the standard representation of a STAR model (adapted from van Dijk and Franses (1999))

$$y_t = {\phi'}_1 y_t^{(p)} \left(1 - F\left(z_t; \gamma, \alpha, c\right)\right) + {\phi'}_2 y_t^{(p)} \left(1 - F\left(z_t; \gamma, \alpha, c\right)\right) + \varepsilon_t \tag{1}$$

where $y_t^{(p)} = \left(1, \tilde{y}_t^{(p)}\right)'$, $\tilde{y}_t^{(p)} = (y_{t-1}, \ldots, y_{t-p})'$,$\phi_i = (\phi_{i0}, \phi_{i1}, \ldots, \phi_{ip})'$, $i = 1, 2$ and ε_t is a white noise zero mean error process with standard deviation σ. The state transition function $F\left(z_t; \gamma, \alpha, c\right)$ is a continuous function bounded on the unit interval and usually taken to be the logistic CDF[2] such that:

$$F\left(z_t; \gamma, \alpha, c\right) = \left(1 + \exp\left\{-\gamma\left(\alpha' z_t - c\right)\right\}\right)^{-1}, \gamma > 0 \tag{2}$$

where $z_t = (z_{1t}, \ldots, z_{jt})'$, $j = 1, \ldots, k$ is a vector of k observed variables which are assumed to explain the state transition. These can be a set of explanatory variables or the lagged values of y_t in which case the model is called 'self-exciting'. It is also possible that the variable is time in which case the model can be used to identify breaks in the mean as in Lin and Teräsvirta (1994), or a combination of time and other variables giving rise to the time varying STAR (TVSTAR) model discussed in Lundbergh et al. (2003). As correctly noted by van Dijk and Franses (1999), the vector of parameters α needs to be nornalized in some way in order to achieve identification (i.e by setting α_1=1). The parameter γ is then a type of scaling factor which determines the smoothness (or speed) of the transition, with values at the limits, $[0, \infty]$, representing linear and SETAR type transitions respectively. By far the most popular test of STAR nonlinearity is described in Luukkonen et al. (1988) using a Taylor series expansion around equation 2, effectively testing whether $\gamma = 0$ (via a auxilliary regression), which would in turn imply that the α vector is also zero and hence in favour of linearity. However, there is really little reason for estimating γ seperately in the STAR model since we can allow it to be subsumed by the vector of state parameters (α, c). In doing so we also gain the additional advantage of extending the type of dynamics to include autoregression as follows:

$$\begin{aligned} &F\left(z_t; \alpha, c, \beta\right) = \left(1 + \exp\left\{-\pi_t\right\}\right)^{-1} \\ &\pi_t = \hat{c} + \hat{\alpha}' z_t + \beta' \pi_t^{(q)} \\ &\pi_t^{(q)} = \left(\pi_{t-1}, \ldots, \pi_{t-q}\right)' \end{aligned} \tag{3}$$

where the unconstrained state dynamics π_t can be initialized by setting $\pi_0 = \frac{\hat{c} + \hat{\alpha}' E[z]}{1 - \beta' \mathbf{1}}$ which requires that $|\beta' \mathbf{1}| < 1$. It should be clear from this representation that $\hat{c} = \gamma c, \hat{\alpha}' = \gamma(1, \alpha_2, \ldots, \alpha_j)', j =$

[2] At present only the Logistic STAR model is entertained and it is not likely that the exponential STAR model will be considered at all.

$1, \ldots, k$ recovers the original representation in equation 1. The use of autoregressive dynamics in the state equation follows related work in the area of dynamic binary response models of Kauppi and Saikkonen (2008) and Nyberg (2010). Generally, estimation becomes quite difficult for more than one autoregressive parameter in the state dynamics which is why at present only a lag-1 autoregressive model is allowed in the package.

The conditional mean dynamics are not limited to AR terms but may include external regressors (ARX) and moving average (MA) terms in the states giving rise to a full STARMAX model specification:

$$y_t = \left(\phi'_1 y_t^{(p)} + \xi'_1 x_t + \psi'_1 e_t^{(q)}\right)(1 - F(z_t;\alpha, c, \beta)) + \left(\phi'_2 y_t^{(p)} + \xi'_2 x_t + \psi'_2 e_t^{(q)}\right)(1 - F(z_t;\alpha, c, \beta)) + \varepsilon_t \tag{4}$$

where $\varepsilon_t^{(q)} = (\varepsilon_{t-1}, \ldots, \varepsilon_{t-q})', \psi'_i = (\psi_{i1}, \ldots, \psi_{iq})'$ represent the q moving average terms and parameters per state i ($i = 1, 2$), and $x_t = (x_1, \ldots, x_l)', \xi'_1 = (\xi_{i1}, \ldots, \xi_{il})'$ the l external regressors and their parameters per state i. It is also possible that the MA term enters outside of the states instead of inside giving rise to a STARX with Linear MA terms (STARXLMA):

$$y_t = \left(\phi'_1 y_t^{(p)} + \xi'_1 x_t\right)(1 - F(z_t;\alpha, c, \beta)) + \left(\phi'_2 y_t^{(p)} + \xi'_2 x_t\right)(1 - F(z_t;\alpha, c, \beta)) + \psi' e_t^{(q)} + \varepsilon_t \tag{5}$$

The STARMAX model therefore encompasses a very wide range of sub-models based on the type of restrictions placed in the conditional mean and state dynamics, and choice of switching variables in the latter.

A natural question which arises from the representation is whether it is reasonable to assume that the conditional variance is the same in both states. Re-write the 2-state STARX equation as follows:

$$\varepsilon_t = y_t - p_t(\mu_{1t}) - (1 - p_t)(\mu_{2t}) \tag{6}$$

where μ_{i1} and $\mu_{2,t}$ represent the conditional mean dynamics per state at time t and p_t the conditional probability. Add and subtract $p_t y_t$, and re-arrange:

$$\begin{aligned}
\varepsilon_t &= +p_t y_t - p_t(\mu_{1t}) + y_t - p_t y_t - (1 - p_t)(\mu_{2t}) \\
\varepsilon_t &= p_t y_t - p_t(\mu_{1t}) + y_t(1 - p_t) - (1 - p_t)(\mu_{2t}) \\
\varepsilon_t &= p_t(y_t - \mu_{1t}) + (1 - p_t)(y_t - \mu_{2t}) \\
\varepsilon_t &= p_t(\varepsilon_{1,t}) + (1 - p_t)(\varepsilon_{2,t}) \\
\varepsilon_{1,t} &\sim N\left(0, \sigma_1^2\right) \qquad , \varepsilon_{2,t} \sim N\left(0, \sigma_2^2\right) \\
\varepsilon_t &\sim N\left(0, p_t\sigma_1^2 + (1 - p_t)\sigma_2^2\right)
\end{aligned} \tag{7}$$

Thus, the model can naturally be re-formulated as a mixture of Normals with mixing probabilities derived from the state dynamics. This provides for a more parsimonious and clear extension than using GARCH dynamics on the mixed state residuals. Alternatively, it can be thought of as the time-invariant version of a STARX-STGARCH model where the STARX and STGARCH models have common transition dynamics. This also provides a partial bridge between finite mixture and time-series autoregressive models.

2.2 Multiple States

van Dijk and Franses (1999) consider extending the 2-state STAR model in equation 1 to a 4-state models as follows:

$$\begin{aligned}
y_t = &\left[\phi'_1 y_t^{(p)}(1 - F(z_t;\gamma_1, \alpha, c)) + \phi'_2 y_t^{(p)}(1 - F(z_t;\gamma_1, \alpha, c))\right](1 - F(z_t;\gamma_2, b, d)) \\
+ &\left[\phi'_3 y_t^{(p)}(1 - F(z_t;\gamma_1, \alpha, c)) + \phi'_4 y_t^{(p)}(1 - F(z_t;\gamma_1, \alpha, c))\right] F(z_t;\gamma_2, b, d) + \varepsilon_t
\end{aligned} \tag{8}$$

Alternatively, the implementation followed in this package takes a page out of multinomial regression and models multiple states using the following representation:

$$y_t = \sum_{i=1}^{s} \left[\left({\phi'}_i y_t^{(p)} + {\xi'}_i x_t + {\psi'}_i e_t^{(q)} \right) F_i \left(z_t; \alpha_i, c_i \right) \right] + \varepsilon_t \tag{9}$$

with

$$\begin{aligned} F_i \left(z_t; \alpha_i, c_i \right) &= \frac{e^{\pi_{i.t}}}{1 + \sum\limits_{i=1}^{s-1} e^{\pi_{i,t}}} \\ F_s \left(z_t; \alpha_i, c_i \right) &= \frac{1}{1 + \sum\limits_{i=1}^{s-1} e^{\pi_{i,t}}} \end{aligned} \tag{10}$$

where the s states are weighted to sum to unity. This appears, at least to this author, to be a more natural representation for a multi-state setup.

2.3 Estimation

Estimation of the STARMAX models is done by maximizing the likelihood without imposing any particular inequality restrictions on the state dynamic intercepts or any parameter bound restrictions (except for positivity bounds on the variance). Since unconstrained optimizers appear to do quite well for hard nonlinear/non-smooth problems, the main solver in the twinkle package is BFGS from the optim function. It is possible to include parameter bounds in which case a logistic-transformation is used with the unconstrained solvers. Additional solvers included are 'nlminb', 'solnp', 'cmaes' and 'deoptim'. However, it is suggested that either a multi-start strategy is followed (by choosing 'msoptim') or an iterative search strategy ('strategy') which cycles between fixing the state parameters to and estimate the conditional mean parameters (linear), fixing the conditional mean parameters to estimate the state parameters (nonlinear) and a random start estimation. As with general nonlinear optimization problems, scaling of the variables prior to estimation may help, an ica or pca transformation if they are highly correlated, or hinge basis transformation of the dataset via the earth package is another interesting option in the case of relevant feature extraction.

3 Forecasting

Consider a general nonlinear first order autoregressive model:

$$y_t = F \left(y_{t-1}; \theta \right) + \varepsilon_t \tag{11}$$

where $F \left(y_{t-1}; \theta \right)$ is some nonlinear function mapping y_{t-1} to y_t given the parameter set θ. The optimal h-step ahead point forecast, using a least squares criterion, of y_{t+h} at time t is given by:

$$\hat{y}_{t+h|t} = E \left[y_{t+h} \left| \Im_t \right. \right] \tag{12}$$

where $\Im_t$ is the information set upto time t. Given that $E \left[\varepsilon_{t+1} \left| \Im_t \right. \right] = 0$, then the 1-step-ahead optimal forecast is:

$$\hat{y}_{t+1|t} = E \left[y_{t+1} \left| \Im_t \right. \right] = F \left(y_t; \theta \right) \tag{13}$$

which is the same as when $F(.)$ is linear. However, for horizons greater than 1, this is not the case since $E \left[F \left(. \right) \right] \neq F \left(E \left[. \right] \right)$, which means that simple recursive relationship found in the linear case do not exist in the nonlinear case. Instead, consider the h-step-ahead point forecast using the following closed form representation:[3]

$$E\left[y_{t+h}\left|\Im_t\right.\right] = \int_{-\infty}^{\infty} E\left[y_{t+h}\left|y_{t+h-1}\right.\right] g\left(y_{t+h-1}\left|\Im_t\right.\right) dy_{t+h-1} \tag{15}$$

where $g\left(y_{t+h}\left|\Im_t\right.\right) = f\left(y_{t+h} - F\left(y_{t+h-1};\theta\right)\right)$, is the distribution of the shock ε_{t+h} with mean $F\left(y_{t+h-1}\right)$, though the distribution ε_t is never known with certainty. A number of approaches have been used in the literature to estimate this integral. It is simple to see that the conditional distribution of $g\left(y_{t+h-1}\left|\Im_t\right.\right)$ can be obtained recursively starting at h=2 and noting that $g\left(y_{t+1}\left|\Im_t\right.\right) = f\left(y_{t+1} - F\left(y_t;\theta\right)\right)$. To obtain the forecasts, numerical integration can be used (applied recursively) or monte carlo methods. In the former case, the form of the conditional distribution $f\left(.\left|\Im_t\right.\right)$ can be replaced by a kernel estimator, whereas in the latter case one has an option of using an empirical bootstrap, simulating from the conditional distribution $f\left(.\left|\Im_t\right.\right)$ or a kernel estimator. For instance, the 2-step ahead monte carlo forecast is given by:

$$\hat{y}_{t+2|t} = \frac{1}{T}\sum_{i=1}^{T} F\left(\hat{y}_{t+1|t} + \varepsilon_i;\theta\right) \tag{16}$$

However, in the case when a GARCH model is used for the modelling of the conditional variance, then the monte carlo forecast needs to be adjusted as follows:

$$\hat{y}_{t+2|t} = \frac{1}{T}\sum_{i=1}^{T} F\left(\hat{y}_{t+1|t} + z_i\hat{\sigma}_{t+2|t};\theta\right) \tag{17}$$

where z_i represent draws from either the parametric standardized distribution of the model or the standardized in-sample innovations (or draws from a kernel estimated density of the standardized in-sample innovations), which are then multiplied by the forecast GARCH volatility $\hat{\sigma}_{t+2|t}$ to obtain the forecast residuals ε_i. One benefit of using a monte-carlo or bootstrap approach is that they immediately give rise to the density of each point forecast thus allowing for the creation of interval forecasts.

4 Simulation

5 Generalized Impulse Response

[3] This is based on the Chapman-Kolmogorov relation:

$$g\left(y_{t+h}\left|\Im_t\right.\right) = \int_{-\infty}^{\infty} g\left(y_{t+h}\left|y_{t+h-1}\right.\right) g\left(y_{t+h-1}\left|\Im_t\right.\right) dy_{t+h-1} \tag{14}$$

which leads to Equation 15 after taking conditional expectations from both sides.

6 Software Implementation

The entry point to defining and estimating a STARMAX model in the twinkle package is the starspec function:

```
>starspec
function(
mean.model = list(states=2, include.intercept=c(1,1), arOrder=c(1,1),
   maOrder=c(0, 0), matype="linear", statevar=c("y","x"), x=NULL,
   statear=FALSE, statelags=1, external.regressors=NULL, yfun=NULL,
   transform="log"),
variance.model=list(dynamic=FALSE, model="sGARCH", garchOrder=c(1,1),
   submodel=NULL, external.regressors=NULL,
   variance.targeting=FALSE),
distribution.model="norm", start.pars=list(), fixed.pars=list(),
fixed.prob=NULL, ...)
}
```

The **mean.model** defines the equation for the conditional mean dynamics including the state dynamics. Upto 4 **states** are allowed, with the 1-state option having a special implementation in that the **fixed.probs** list is an xts matrix (alined to the dataset which will be passed to the estimation routine) of weights. By default this is set to a vector of ones in this case but may be any other 'time-weighting' scheme the user wishes. The options for **intercept**, **arOrder** and **maOrder** should be integer vectors of length equal to the number of **states**.The **matype** denotes whether the moving average terms enters inside the states ('states') or outside ('linear'). The **statevar** indicates whether the model will switch based on its own value ('y') or an exogenouse set of regressors ('x'), in which case an xts matrix (aligned to the index of the dataset) is passed to **x**. This matrix must *NOT* be lagged, instead use the **statelags** option to pass an integer vector of lags (of length equal to the number of columns of x).[4] In the case that **statevar** is 'y', then **statelags** can be an integer vector of the unique lags to use as a linear combination. The **yfun** option allows the user to pass a function to transform the value of y[5] prior to being used in the state dynamics equation. While it may appear at first that the same can be achieved by passing a pre-transformed value and using 'x' as the **statevar**, consider that simulation and n-ahead forecasts (which depend on simulation methods) on transformed values of 'y' can then be used directly, where it would have been impossible to do so otherwise. Unlike the **x** option, the **external.regressors** needs to be passed pre-lagged (again, as an xts matrix aligned to the index of the dataset). The **statear** indicates whether to include lag-1 autoregression in the state dynamics as discussed previously in equation 3. Finally, the **transform** is currently fixed to use only the logistic transformation, and there are no plans to extend to the exponential at present. The variance model can be **dynamic**, in which case a choice of 'mixture','sGARCH','gjrGARCH' and 'eGARCH' are implemented. For the GARCH flavors, the rest of the options follow from the rugarch package, whilst the 'mixture' model is based on equation 7. All distributions implemented in rugarch are included as options in **distribution.model**, while fixed and starting parameters can be passed directly via **fixed.pars** and **start.pars** respectively, else later on via the **setfixed¡-** and **setpars¡-** methods on the star specification (note that there is also a **setbounds¡-** methods for setting and enforcing parameter bounds). Finally, the **fixed.probs** list allows the user to pass an xts matrix of fixed probabilities for each state (aligned to the index of the dataset and with columns equal to **states**). This could for instance be the forecast probabilities from another model (e.g. logistic regression) representing market

[4] The decision to enforce this choice was motivated by the need to control the way the state equation is initialized so that it uses all available information at any given time without waiting for all lags to be in play before being used (i.e. the typical use of starting the recursion at max(lags) is completely wasteful.)

[5] The function must return the same length as the value it receives without any NAs or NaNs.

up and down periods, recessions etc. In this case the state equation is not used and the model is effectively linear and extremely fast to estimate for the conditional mean dynamics.

7 Examples

References

T Astatkie, DG Watts, and WE Watt. Nested threshold autoregressive (netar) models. *International Journal of Forecasting*, 13(1):105–116, 1997.

SA Billings and WSF Voon. Structure detection and model validity tests in the identification of nonlinear systems. In *IEE Proceedings D (Control Theory and Applications)*, volume 130, pages 193–199. IET, 1983.

SA Billings and WSF Voon. Correlation based model validity tests for non-linear models. *International journal of Control*, 44(1):235–244, 1986.

FL Carmichael. The arc tangent in trend determination. *Journal of the American Statistical Association*, 23(163):253–262, 1928.

Felix Chan and Michael McAleer. Maximum likelihood estimation of star and star-garch models: theory and monte carlo evidence. *Journal of Applied Econometrics*, 17(5):509–534, 2002.

Felix Chan and Michael McAleer. Estimating smooth transition autoregressive models with garch errors in the presence of extreme observations and outliers. *Applied Financial Economics*, 13 (8):581–592, 2003.

K. S. Chan and H. Tong. On estimating thresholds in autoregressive models. *Journal of Time Series Analysis*, 7:178–190., 1986.

Dick van Dijk, Timo Teräsvirta, and Philip Hans Franses. Smooth transition autoregressive modelsa survey of recent developments. *Econometric Reviews*, 21(1):1–47, 2002.

Heikki Kauppi and Pentti Saikkonen. Predicting us recessions with dynamic binary response models. *The Review of Economics and Statistics*, 90(4):777–791, 2008.

Chien-Fu Jeff Lin and Timo Teräsvirta. Testing the constancy of regression parameters against continuous structural change. *Journal of Econometrics*, 62(2):211–228, 1994.

Stefan Lundbergh, Timo Teräsvirta, and Dick van Dijk. Time-varying smooth transition autoregressive models. *Journal of Business & Economic Statistics*, 21(1):104–121, 2003.

Ritva Luukkonen, Pentti Saikkonen, and Timo Teräsvirta. Testing linearity against smooth transition autoregressive models. *Biometrika*, 75(3):491–499, 1988.

Robert M. May. Simple mathematical models with very complicated dynamics. *Nature*, 261 (5560):459–467, 1976.

Henri Nyberg. Dynamic probit models and financial variables in recession forecasting. *Journal of Forecasting*, 29(1-2):215–230, 2010.

Richard E Quandt. The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the american statistical association*, 53(284):873–880, 1958.

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

Howell Tong. *Non-linear time series: a dynamical system approach.* Oxford University Press, 1990.

Howell Tong and DK Ghaddar. Data transformation and self-exciting threshold autoregression. *Journal of the Royal Statistical Society, series C*, 30:238–248, 1981.

Howell Tong and K.S Lim. Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society, Series B*, 42(3):245–292, 1980.

Dick van Dijk and Philip Hans Franses. Modeling multiple regimes in the business cycle. *Macroeconomic Dynamics*, 3:311–340, 1999.

Q.M. Zhu and S. Billings. Parameter estimation for stochastic nonlinear rational models. *International Journal of Control*, 57(2):309–333, 1993.