Quaderni di Statistica
Vol. 4, 2002

Nonlinear time series models with switching structure: a comparison of their forecast performances

Marcella Niglio
Dipartimento di Scienze Economiche e Statistiche, Università di Salerno
E-mail: mniglio@unisa.it

Summary: In the present paper the forecast performances of the Logistic Double Smooth Transition (LDST) model (Lee and Li, 1998) are investigated and compared to the Double Threshold ARCH (DTARCH) model (Li and Li, 1996). After the presentation of both models, the estimation procedure of the LDST model is illustrated and its point predictors are explicitly given. Further, in order to present their empirical distributions, the results of a Monte Carlo study are shown.

The LDST model is applied to study the series of the daily returns of the Dax 30 stock index and the generated forecasts are compared with those obtained from the DTARCH model.

Key words: Nonlinearity, Point predictor, Density forecasts, LDST model.

1. Introduction

In recent years nonlinear modelling has received increasing attention in financial time series analysis. These series are affected by a considerable number of sources of nonlinearity such as asymmetries in the returns, business cycles, asymmetric behaviour of the variability (usually called leverage effect), volatility clustering, and so on.

These well known features of financial data imply dynamic behaviours which can be well captured using regimes structures, as pointed out in

Priestley (1988). Some of them have been shown in Tong (1990), Granger and Teräsvirta (1993) and recently summarized in Potter (1999) and Franses and Van Dijk (2000).

Tong (1990), in order to overcome the traditional distinction between models for the conditional mean and models for the conditional variance, suggests to combine them to accommodate, for example in a financial context, structural changes and leverage effects of the data at the same time.

In this context, Li and Li (1996), Liu, Li and Li (1997) propose the Double Threshold Autoregressive ARCH (DTARCH) model, which combines the Self-Exciting Threshold AutoRegressive (SETAR) structure (Tong and Lim, 1980) with a threshold ARCH component. Lee and Li (1998) introduce the Double Smooth Transition model (DST), where the conditional mean and the conditional variance are modelled using regimes structures where the switch from one regime to another is smooth. Lundbergh and Tersvirta (1998) further generalize the DST structure with the STAR-STGARCH model and propose its application to high frequency time series.

The generation of forecasts from these models is often more involved than computing forecasts from linear models.

In Niglio (2000) multi-step ahead point forecasts for the DTARCH model are investigated.

In the present paper, in order to extend this investigation to a more general model, the generation of forecasts from the DST model are examined. In particular, the DTARCH and the DST models, which have a regimes structure both in conditional mean and conditional variance, are briefly presented in Section 2, giving more relevance to the second and less known model. The attention is prevalently confined to their forecast performances and those of the DST model are investigated in Section 3. An application to real data is shown in Section 4 where the forecast ability of the two models are compared and discussed. Some concluding remarks are given in the final section.

2. The examined models

The DTARCH(p_1, p_2, q_1, q_2) model proposed in Li and Li (1996) and Liu, Li and Li (1997) is characterized by the following regimes structure:

$$\begin{aligned} X_t &= (\phi_0^{(1)} + \sum_{i=1}^{p_1} \phi_i^{(1)} X_{t-i}) \cdot \mathbf{1}(X_{t-d} \in R_1) + \\ &+ (\phi_0^{(2)} + \sum_{i=1}^{p_2} \phi_i^{(2)} X_{t-i}) \cdot \mathbf{1}(X_{t-d} \in R_2) + \epsilon_t \\ h_t &= (\alpha_0^{(1)} + \sum_{i=1}^{q_1} \alpha_i^{(1)} \epsilon_{t-i}^2) \cdot \mathbf{1}(\epsilon_{t-s} \in S_1) + \\ &+ (\alpha_0^{(2)} + \sum_{i=1}^{q_2} \alpha_i^{(2)} \epsilon_{t-i}^2) \mathbf{1}(\epsilon_{t-s} \in S_2) \end{aligned} \tag{1}$$

where $\epsilon_t \sim N(0, h_t)$, p_i and q_i, $i = 1, 2$, are the order of the conditional mean and variance respectively, d and s are the threshold delaies, R_i and S_i ($i = 1, 2$) are partitions of the real line $\mathbf{R}$ with $R_i = (r_{i-1}, r_i]$, $S_i = (s_{i-1}, s_i]$, r_i is the threshold value for the conditional mean and s_i is the threshold value of the conditional variance. Further the parameters $\alpha_0^{(i)} > 0$ and $\alpha_j^{(i)} \geq 0$, for $j = 1, 2, ..., q_i$ and $i = 1, 2$.

This model is based on the idea that the volatility of the series is not always a constant multiple of the squared past disturbances, as in the ARCH case (Engle, 1982), but positive and negative shocks affect differently the volatility and a switching structure can model this asymmetric behaviour.

A generalisation of model (1) is the Double Smooth Transition (DST) model, introduced in Lee and Li (1998). It combines a STAR structure (Teräsvirta, 1994) with a Smooth Transition ARCH model (Hagerud, 1996) in order to take into account, at the same time, the asymmetries in the returns and in the volatility of the series.

The main difference with the DTARCH model is that the switch among the regimes is not steep. Its gradual transition is empirically appreciated in a context characterised by market inefficiencies where the regimes switch between positive and negative returns and the decaying and the increase of the volatility is not abrupt.

The DST(z_1, z_2, w_1, w_2) model is defined as follows:

$$
\begin{aligned}
X_t &= \beta_0^{(1)} + \sum_{i=1}^{z_1} \beta_i^{(1)} X_{t-i} + \left\{1 + e^{-\gamma(x_{t-d}-c)}\right\}^{-1} \cdot \\
&\qquad \cdot \left\{\beta_0^{(2)} + \sum_{i=1}^{z_2} \beta_i^{(2)} X_{t-i}\right\} + \epsilon_t \\
h_t &= \xi_0^{(1)} + \sum_{i=1}^{w_1} \xi_i^{(1)} \epsilon_{t-i}^2 + \left\{1 + e^{-\kappa(x_{t-b}-r)}\right\}^{-1} \cdot \\
&\qquad \cdot \left\{\xi_0^{(2)} + \sum_{i=1}^{w_2} \xi_i^{(2)} \epsilon_{t-i}^2\right\}
\end{aligned} \tag{2}
$$

where $\epsilon_t \sim N(0, h_t)$, d and b are the *delay parameters*, c and r are called *transition parameters*, γ and κ are the *smoothness parameters*, x_{t-d} and x_{t-b} are the *transition variables* of the conditional mean and conditional variance respectively. Van Dijk, Teräsvirta and Franses (2002) argue that the transition variables cannot be restricted to x_t's, even though in the present context this choice appears appropriate. Further, the selection of x_{t-b}, as transition variable of the conditional variance, instead of ϵ_{t-b} (as in equation (1)) does not impacts on the model because, as highlighted in Liu, Li, Li (1997) and Lee and Li (1998), in real life situations, abrupt changes in the observations x_t are often accompained by those in the disturbances ϵ_t.

The logistic smoothing function in (2) give rise to the so called Logistic DST (LDST) model.

The assumptions on the LDST model, given in Lee and Li (1998), are the followings:

1. the time series X_t is second order stationary and ergodic;

2. all the parameters in the first regime of the conditional variance are such that $\xi_0^{(1)} > 0$, $\xi_j^{(1)} \geq 0$, for $j = 1, 2, ..., w_1$. Further given $\xi_j = \xi_j^{(1)} + \xi_j^{(2)}$, for $j = 1, 2, ..., w_i$ and $i = 1, 2$, all the ξ_j must be non-negative.

3. let $Z_{it} = (X_{t-1}, X_{t-2}, ..., X_{t-z_i})$ and $W_{it} = (\epsilon^2_{t-1}, \epsilon^2_{t-2}, ..., \epsilon^2_{t-w_i})$, for $i = 1, 2$, then $\{Z_{1t}, F(\cdot)Z_{2t}\}$ are assumed to be linearly independent and the same independence is assumed for $\{W_{1t}, G(\cdot)W_{2t}\}$, where $F(\cdot)$ and $G(\cdot)$ are the logistic functions.

Some of these conditions are necessary to guarantee the existence of the estimators in model (2) whereas the gaussian distribution of the errors ϵ_t, is assumed in order to have maximum likelihood estimators and to derive an LM linearity test (Lee and Li, 1998).

The estimation of the parameters follows the modelling cycle of Teräsvirta (1994):

1. select an upper bound for $z_i, i = 1, 2$, observing the Autocorrelation function (ACF) and the Partial ACF plots;

2. estimate the parameters d and b such that the p-values of the LM test for the conditional mean and conditional variance are minimized over the grid values D_m and B_m, with $1 \leq d \leq D_m$ and $1 \leq b \leq B_m$;

3. estimate the parameters $\beta_i^{(j)}$ (for $i = 1, ..., z_j$ and $j = 1, 2$), $\xi_i^{(j)}$ (with $i = 1, ..., w_j$ and $j = 1, 2$), γ, κ, c and r using the method of maximum likelihood, based on the Newton-Raphson algorithm, through the following steps:

   a. given the vectors $\boldsymbol{\beta}^j = (\beta_0^{(j)}, \beta_1^{(j)}, ..., \beta_{z_j}^{(j)})$ and $\boldsymbol{\xi}^j = (\xi_0^{(j)}, \xi_1^{(j)}, ..., \xi_{w_j}^{(j)})$, estimate the parameters $\left\{\boldsymbol{\beta}^j, \boldsymbol{\xi}^j\right\}$, $j = 1, 2$, after fixing γ, κ, c and r temporarily;

   b. using $\left\{\hat{\boldsymbol{\beta}}^j, \hat{\boldsymbol{\xi}}^j\right\}$ in step a., estimate c and r;

   c. based on $\left\{\hat{\boldsymbol{\beta}}^j, \hat{\boldsymbol{\xi}}^j\right\}$, $\hat{c}$ and $\hat{r}$, estimate the smoothness parameters γ and κ;

   d. repeat a. - c. until the convergence is reached.

The estimates obtained through a. - d. are evaluated in Lee and Li (1998) using a simulation study. The results show that the specification of the smoothness parameters (γ and κ) have minor effects on the estimation of the other parameters even if the larger are γ and κ more accurate are the estimation results because easier is the identification of the transition point and the intercepts of the regimes. Obviously the estimates of γ and κ may affect the forecast ability of the LDST model. This performance has been evaluated and some results are described in the next section.

3. The generation of point forecasts from the LDST model

The complex structure of the LDST model makes the generation of forecasts not an easy task.

In the present section point forecasts are generated from model (2) and the distribution of the predictor used to generate them has been explored at different lead times through a Monte Carlo study.

It is known that the best predictor of $X_{t+\ell}$, in term of minimum mean square errors, is the expected value of $X_{t+\ell}$ conditioned to the informations given up to time t, (I_t):

$$\hat{X}_t(\ell) = E[X_{t+\ell}|I_t]$$

with ℓ the *lead time* and $\hat{X}_t(\ell)$ the predicted value.

In this context, the ℓ steps ahead point forecasts (with $\ell \geq 1$) of the LDST model, are given as:

$$\hat{X}_t(\ell) = \beta_0^{(1)} + \sum_{i=1}^{\ell-1} \beta_i^{(1)} \hat{X}_t(\ell-i) + \sum_{i=\ell}^{z_1} \beta_i^{(1)} X_{t+\ell-i} + \left\{1 + e^{-\gamma(\hat{X}_t(\ell-d)-c)}\right\}^{-1} \cdot$$

$$\left(\beta_0^{(2)} + \sum_{i=1}^{\ell-1} \beta_i^{(2)} \hat{X}_t(\ell-i) + \sum_{i=\ell}^{z_2} \beta_i^{(2)} X_{t+\ell-i}\right) \quad (3)$$

$$\hat{h}_t(\ell) = \xi_0^{(1)} + \sum_{i=1}^{\ell} \xi_i^{(1)} \hat{h}_t(\ell - i) + \sum_{i=\ell+1}^{w_1} \xi_i^{(1)} \epsilon_{t+\ell-i}^2 + \left\{1 + e^{-\kappa(\hat{X}_t(\ell-d)-c)}\right\}^{-1} \cdot$$

$$\left(\xi_0^{(2)} + \sum_{i=1}^{\ell} \xi_i^{(2)} \hat{h}_t(\ell - i) + \sum_{i=\ell+1}^{w_2} \xi_i^{(2)} \epsilon_{t+\ell-i}^2\right) \tag{4}$$

where $E[\epsilon_{t+\ell}|I_t] = 0$, for $\ell \geq 1$, $\hat{X}_t(\ell - d) = X_{t+\ell-d}$ and $\hat{X}_t(\ell - b) = X_{t+\ell-b}$ if $\ell \leq d$ and $\ell \leq b$, respectively.

The forecasts generated from the LDST model have relevant features, some of which are in common with the DTARCH model. When the lead time ℓ is greater than the threshold delaies, the threshold values are estimated and the prediction performance depends strongly on the accuracy of these estimates. Further, the LDST forecasts distributions show asymmetric behaviour and, in some cases, multimodality which is more evident as the smoothing parameter γ grows (in this case the rise of γ implies the convergence of model (2) to model (1)) and as the difference $|\beta_0^{(1)}| - |\beta_0^{(2)}|$ in (3) increases, highlighting a dependence of the forecast distribution shape upon the regimes intercepts.

To show this dependence, one thousand series of length 500 (obtained after dropping the first 100 observations in order to limit the dependence of the results on the initial values) have been generated from the following model:

$$\begin{aligned} X_t &= -0.05 + 0.10X_{t-1} + 0.15X_{t-2} + \left\{1 + e^{-\gamma \cdot X_{t-1}}\right\}^{-1} \cdot \\ &\qquad (0.20 - 0.24X_{t-1} + 0.10X_{t-2}) + \epsilon_t \\ h_t &= 0.10 + 0.15\epsilon_{t-1}^2 + 0.05\epsilon_{t-2}^2 + \left\{1 + e^{-1 \cdot X_{t-1}}\right\}^{-1} \cdot \\ &\qquad \left(0.27 + 0.32\epsilon_{t-1}^2 + 0.10\epsilon_{t-2}^2\right) \end{aligned} \tag{5}$$

with $\epsilon_t \sim N(0, h_t)$, $r = 0$ and $c = 0$ and where the simulation has been repeated for two different values of the smoothing parameter $\gamma = 1$ and

$\gamma = 100$. The parameters in (5) have been chosen in order to guarentee an equal partition of the X_ts between the two regimes which rise when $\gamma = 100$, as consequence of the approximation of the logistic function to an indicator function.

Assuming the parameters known, the forecasts $\hat{X}_t(\ell)$ and $\hat{h}_t(\ell)$ have been generated for each series at four different step lengths, $\ell \in [1, 4]$ in both cases and with forecast horizon $H = t + 4$.

The forecast distributions of model (2) with $\gamma = 1$ and $\gamma = 100$, estimated following Silverman (1986), are shown in Figure 1 and Figure 2 respectively.

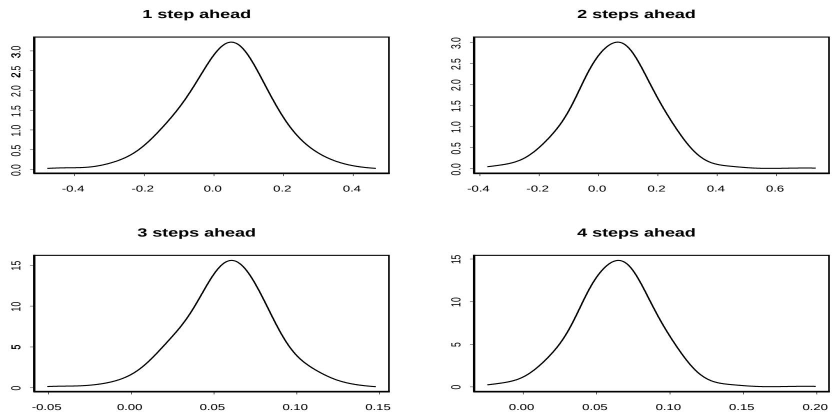

Figure 1: *Predictor densities of the LDST model with* $\gamma = 1$ *at four different lead times* $\ell = 1, 2, 3, 4$

As expected, the volatility and the asymmetry of the forecasts increase as the lead time grows. It is interesting to note that the main results obtained in the simulation study is that, for small values of γ, the number of modes of the $\hat{X}_t(\ell)$ distribution does not depend from the intercept values. In particular, the marked multimodal distribution of the DTARCH predictors as the difference beetwen $|\phi_0^1|$ and $|\phi_0^2|$ increases in (1) (Amendola and Niglio, 2001), is not recognised in the LDST model when γ is small and so in this model the intercepts do not impact the distribution shape. The main consequence of this result is on the forecast accuracy

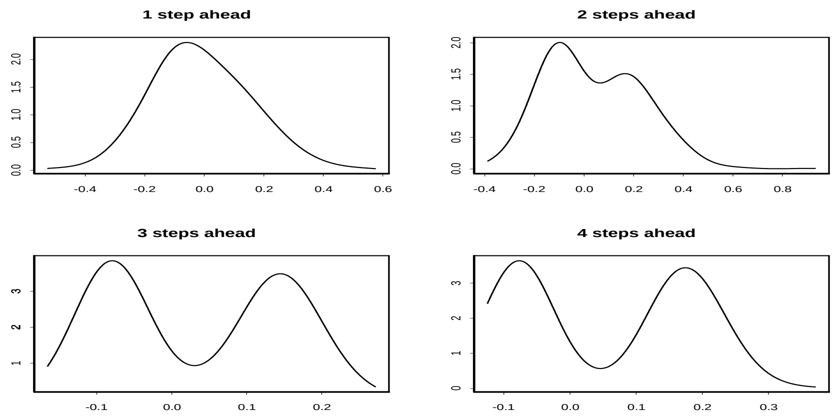

Figure 2: *Predictor densities of the LDST model with* $\gamma = 100$ *at four different lead times* $\ell = 1, 2, 3, 4$

and on the construction of the forecast regions which are easier to obtain in the unimodal case (Hyndman, 1995).

The multimodality of the LDST multi-step forecasts becomes more noticeable as γ grows (Figure 2) due to the convergence of the LDST model to the DTARCH model.

This is true when $\beta_0^{(1)} \neq \beta_0^{(2)}$ but when $\beta_0^{(1)} = \beta_0^{(2)}$ an unimodal distribution is given for the LDST model with no dependence from the values assigned to the γ parameter.

To highlight this result the distributions of the forecasts generated from three different models are shown in Appendix with $\ell = 4$ and $\gamma = 1$ (frame (a)) and $\gamma = 100$ (frame (b)).

In all three cases the unimodality of $\hat{X}_t(\ell)$ is preserved. In particular, *model 1*, obtained fixing the mean intercepts in (5) to zero, shows a symmetric unimodal distribution whereas in *model 2* and *model 3* a more pronounced asymmetry is highlighted.

4. An application to the German Dax 30 stock index

The LDST model described in Section 2 has been applied to the German stock market index, Dax 30, using daily observations from the 2nd of January 1996 to the 31 of December 2000, for a total of 1304 observations.

The model has been estimated on the returns R_t, given as the first difference of the logarithm of the daily closing index, $R_t = lnX_t - lnX_{t-1}$, whereas the ex-post forecasts have been generated for the month of January 2001 which counts 23 working days.

In order to evaluate the prediction performance of the LDST model, its forecasts are compared with those obtained from the DTARCH model estimated on the same data set.

The estimation of the DTARCH parameters are obtained maximizing the likelihood function explicitly given in Liu *et al.* (1997) whereas the specification and estimation of the LDST model follow the steps in Section 2.

The upper bound of z_i and w_i is fixed from the autocorrelation functions of the returns and of the squared returns respectively. The linearity LM test with LDST alternative has been used and, following Lee and Li (1998), it has been separated for the conditional mean and variance

Given a significance level $\alpha = 0.05$, linearity is widely rejected for the mean and the variance when the delaies are $d = 5$ and $b = 1$ and the corresponding values of the LM statistics are $LM_M = 43.14$ and $LM_V = 42.18$, respectively.

The remaining parameters, whose starting values are obtained separately for the conditional mean parameters first (Teräsvirta, 1994) followed by the conditional variance parameters (Hagerud, 1996), are estimated through the iterative procedure shown in Section 2 where steps 3.a - 3.d are repeated until the convergence is reached.

Using these estimates 23 ex-post forecasts are generated for the month of January 2001. The predictions have been obtained using 5 different lead times $\ell = 1, 2, 3, 4, 5$ in order to have multi-period forecasts generated through different step-lengths.

The generated forecasts, for the conditional mean and the conditional variance, are compared with those obtained from the DTARCH model. The comparison of the conditional mean predictions has been performed using the Mean Absolute Error (MAE) whereas, following Bollerslev, Engle and Nelson (1994), the forecasts of the conditional variance have been evaluated through the HMSE (Heteroskedasticity adjusted mean square error) asymmetric index (Bollerslev, Ghysels, 1996):

$$HMSE = \frac{1}{T}\sum_{t=1}^{T}\left[\frac{\epsilon_{t+\ell}^{2}}{\hat{h}_{t+\ell}} - 1\right]^2$$

in order to avoid the use of symmetric indexes (such us the RMSE) which are often meaningless in the evaluation of the forecast volatility.

Further, it is conceivable that many investors will not attribute equal importance to both over and under prediction of volatility of similar magnitude. In this context, Brailsford and Faff (1996), propose an index which penalizes under prediction (U) more heavily that over prediction. It is called Mean Mixed Error, MME(U):

$$MME(U) = \frac{1}{T}\left[\sum_{t=1}^{O}|\hat{h}_t - h_t| + \sum_{t=1}^{U}\sqrt{|\hat{h}_t - h_t|}\right]$$

where O is the number of over-predictions and U is the number of under-predictions.

The results summarized in Table 1, show that the conditional mean DTARCH forecasts outperform those obtained from the LDST in terms of MAE even if the performance of the latter model is superior than the former in the conditional variance prediction.

Wong e Li (2000) give an explanation to similar results saying that if the volatility of the series is high, bimodal predictors, such us those obtained from the DTARCH model, can give more accurate forecasts for the returns because the chance of sharp increase or decrease in the level of the series is higher than the chance of a moderate increase or decrease.

Table 1 shows that the conditional mean forecast performance of the DTARCH model improves, with respect to the LDST model, as ℓ grows but opposite results are obtained for the prediction of the volatility where the LDST outperforms the DTARCH model at all lead times.

From a financial point of view, where operators are often interested in market volatility, the results obtained from the LDST model are remark-

Table 1: *Indexes to evaluate the forecast performance of the LDST and DTARCH models*

LDST forecasts			
lead time	MAE	HMSE	MME(U)$\times 10^{-2}$
1 step	0.01422	0.71796	0.05748
2 steps	0.01087	0.70971	0.05771
3 steps	0.01275	0.74097	0.06083
4 steps	0.01008	0.77678	0.06355
5 steps	0.01450	0.78520	0.06559
DTARCH forecasts			
lead time	MAE	HMSE	MME(U)$\times 10^{-2}$
1 step	0.00733	1.97304	0.37459
2 steps	0.00759	1.76677	0.38957
3 steps	0.00694	1.31422	0.28229
4 steps	0.00626	0.83649	0.18401
5 steps	0.00585	1.29545	0.20008

able. Further the accuracy of $\hat{X}_t(\ell)$ can be improved using other forecasting procedures such as Monte Carlo forecasts (Granger and Teräsvirta, 1993) which are widely employed in nonlinear context.

5. Concluding remarks

The point forecasts of the LDST model have been explored. After the presentation of the model, which can be considered a generalization of the DTARCH model, the LDST point predictor has been given and its distribution has been shown through a simulation study. It has highlighted relevant features of the LDST predictors which can impact on the forecast accuracy.

The LDST forecasts have been compared to the DTARCH forecasts through an application to the German stock market index, Dax 30, at different lead-times.

The forecast performance of the conditional mean and variance, has been evaluated differently.

The LDST model has shown a significant forecast gain for the conditional variance which is not recognised in the conditional mean case.

Acknowledgments: This paper is partially supported by MURST 2000 "Modelli stocastici e metodi di simulazione per l'analisi di dati dipendenti".

References

Bollerslev T., Engle R. F., Nelson D.F. (1994) ARCH models. In R.F. Engle and D. McFadden, eds. *Handbook of Econometrics* , vol.4, North Holland, Amsterdam.

Bollerslev T., Ghysels E. (1996) Periodic autoregressive conditional heteroskedasticity. *Journal of Business & Economic Statistics*, 14, 139-151.

Brailsford T.J., Faff R.W. (1996) An evaluation of volatility forecasting techniques, *Journal of Banking & Finance* , 20, 419-438.

Franses P.H., Van Dijk D. (2000) *Non-linear time series models in empirical finance* , Cambridge University Press, Cambridge.

Granger C.W.J. and Teräsvirta T. (1993) *Modelling nonlinear relationships*, Oxford University Press, Oxford.

Hagerud G.H. (1996) A smooth transition ARCH model for asset returns, *Working Paper Series in Economics & Finance n. 162*, Stockholm School of Economics.

Hyndman R.J. (1995) Highest-density forecast regions for non- linear and non-normal time series models, *Journal of Forecasting*, 14, 431-441

Lee Y.N., Li W.K. (1998) On smooth transition double threshold models, *Working Paper n. 198, Department of Statistics*, University of Hong Kong.

Li C.W., Li W.K.(1996) On a double threshold autoregressive heteroskedastic time series model, *Journal of Applied Econometrics*, 11, 253-274

Liu J., Li W.K., Li C.W. (1997) On a threshold autoregression with conditional heteroskedasticity, *Journal of Statistical Planning & Inference*, 62, 279-300.

Lundbergh S., Teräsvirta T. (1998) Modelling economic high-frequency time series with STAR-STGARCH models, *Working Paper Series in Economics & Finance n. 291*, Stockholm School of Economics.

Niglio M. (2000) *Previsioni per modelli non lineari di serie storiche*, Tesi di Dottorato, Universitá ”‘G. D’Annunzio”’, Chieti.

Potter S.M. (1999), Nonlinear time series modelling: an introduction *Journal of Economic Surveys*, 13, 505-524.

Silverman B.W. (1986), *Density estimation*, Chapman & Hall: London.

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

Priestley M. B. (1988) *Non-Linear and Non-Stationary Time Series Analysis*, Academic Press, San Diego.

Van Dijk D., Teräsvirta T., Franses P.H. (2002) Smooth transition autoregressive models: A survey of recent developments, *Econometric review*, 21, 1-47.

Tong H. (1990) *Non-linear time series: A dynamical system approach*. Clarendon Press, Oxford.

Tong H., Lim K.S. (1980) Threshold autoregression, limit cycles and cyclical data, *Journal of Royal Statistical Society* (B), 42, 245-292

Wong C.S., Li W.K. (2000) On a mixture autoregressive model, *Journal of the Royal Statistical Society (B)*, 62, 95-115.

Appendix

Table A.1: *Conditional mean coefficients of 3 LDST simulated models*

	First Regime			Second Regime		
	$\beta_0^{(1)}$	$\beta_1^{(1)}$	$\beta_2^{(1)}$	$\beta_0^{(2)}$	$\beta_1^{(2)}$	$\beta_2^{(2)}$
model 1	0	0.1	0.15	0	-0.24	0.1
model 2	0.25	0.1	-0.45	0.2	-0.57	0.1
model 3	0	0.1	-0.45	0	-0.57	0.1

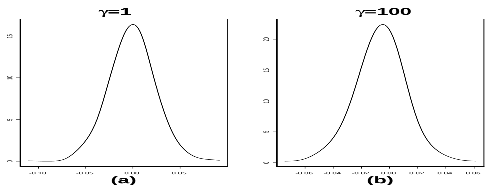

Figure A.1: $\hat{X}_t(4)$ *distribution from Model 1 at* $\gamma = 1$ *(a),* $\gamma = 100$ *(b)*

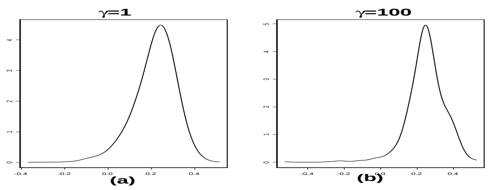

Figure A.2: $\hat{X}_t(4)$ *distribution from Model 2 at* $\gamma = 1$ *(a),* $\gamma = 100$ *(b)*

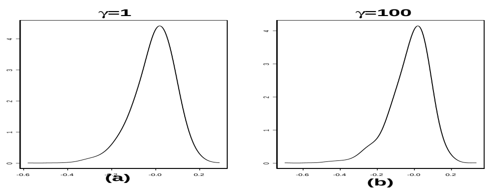

Figure A.3: $\hat{X}_t(4)$ *distribution from Model 3 at* $\gamma = 1$ *(a),* $\gamma = 100$ *(b)*