The European Journal of Finance
Vol. 12, No. 6–7, 455–472, September–October 2006

Extended Switching Regression Models with Time-varying Probabilities for Combining Forecasts

ARIE PREMINGER, URI BEN-ZION & DAVID WETTSTEIN*
Department of Economics, Ben-Gurion University of the Negev, Beer-Sheva, Israel

ABSTRACT *This paper introduces a new methodology, which extends the well-known switching regression model. The extension is via the introduction of several latent state variables, each one of which influencing a disjoint set of the model parameters. Furthermore, the probability distribution of the state variables is allowed to vary over time. This model is called the time varying extended switching regression (TV-ESR) model. The model is used to combine volatility forecasts of several currencies (JPY/USD, GBP/USD, and CHF/USD). A detailed comparison of the forecasts generated by the TV-ESR approach is made with those of traditional linear combining procedures and other methods for combining forecasts derived from the switching regression model. On the basis of out-of-sample forecast encompassing tests as well as other measures for forecasting accuracy, results indicate that the use of this new method yields overall better forecasts than those generated by competing models.*

KEY WORDS: Forecast combining, TV-ESR models, volatility modelling

1. Introduction

Many economic and financial time series are characterized by periods in which the behaviour of the series seems to change quite dramatically. Such apparent changes are often captured through the use of models with time-varying parameters. One notable class of models is given by the switching regression models where the whole set of parameters moves over a finite number of value sets. The switches between the sets of values are controlled by an unobserved state variable. Important methodological contributions include the work of Quandt (1958) and the more recent work of Le *et al.* (1996) and Li and Wong (2000), which extend the switching regression model to the case of dependent data or more specifically an autoregression. A related class of models are the hidden Markov model regressions (Hamilton, 1994), which differ from the switching regression models in that the unobserved state variable follows a latent Markov structure.

The switching regression models assume that there is only one state variable in each period which controls the switches in the model parameters. More specifically, these models are based on the assumption that the data generating process changes over time, and there is a latent model

Correspondence Address: Professor David Wettstein, Department of Economics, Ben-Gurion University of the Negev, Beer-Sheva, 84105, Israel. Email: wettstn@bgumail.bgu.ac.il

1351-847X Print/1466-4364 Online/06/06–70455–18 © 2006 Taylor & Francis
DOI: 10.1080/13518470500039360

selection procedure dependent on a discrete state variable which randomly picks a parametric model each time. This procedure is characterized by defining a set or a subset of the model parameters to be mutually dependent on the state variable.

Preminger *et al.* (2004a) introduced the extended switching regression (ESR) model, which is characterized by several state variables that independently influence the model selection procedure through the picking of a partial and disjoint group of the model parameters. The advantage of formulating state variables in such a way is that interesting qualitative information may result from the nature of the state variables. Furthermore, the assumption of independence among the state variables allows us to provide a parsimonious parameterization of the model, while expanding the possible number of states the model can assume.

In the ESR model, we assume that the state probabilities are constant, whereas economic as well as financial considerations suggest the desirability of allowing the state probabilities to vary over time. In this paper, we introduce the time-varying extended switching regression, TV-ESR, model in which the probability distribution changes over time. The ESR models are nested in the TV-ESR models; this implies that use of the TV-ESR models may allow for better description and prediction for the variables of interest. A related class of models is the switching regression model (*i.e.* models with one state variable) with time-varying probabilities, proposed by Li and Wong (2001) and which can be traced back to Goldfeld and Quandt (1972).

We apply the TV-ESR models to combine forecasts of exchange rate volatility and investigate the incremental value of going from the traditional linear forecast combining methods and other methods for combining forecasts, derived from switching regression models, to the class of TV-ESR models. The motivation for using such models is that when one performs a linear combination of several individual forecasts to obtain a single forecast, the weights which are given to each individual forecast may change over time. The changes in the weights may be associated with the realization of several independent latent state variables.

A detailed comparison of the forecast combination methods on the basis of the root mean squared error (RMSE), mean absolute error (MAE), Theil-U statistic, the correct direction change prediction and the forecast encompassing tests, is performed. We show that the TV-ESR modelling procedure performs at least, as well, and often better than, forecasts based on rival combining methods.

The layout of this paper is as follows: the next section describes the model with relation to switching regression models. The following section presents forecast combining models, one of which is the TV-ESR model. The fourth section describes the data set and discusses empirical findings, and the final section concludes.

2. The Model

Switching regression models were developed as a way of allowing data to arise from a combination of two or more distinct data generation processes. At each time t, the actual process generating the data is determined by the realization of an unobserved random discrete variable denoted by s_t which is called a state variable. A switching regression model is described by:

$$y_t = \mu_t(x_t, \psi(s_t)) + \varepsilon_t \tag{1}$$

where $\mu_t\colon \mathrm{X} \times \Psi \to \Re$ are known functions which satisfy standard measurability and continuity requirements[1] for all t, the error term ε_t is a zero-mean white noise and s_t is a state variable that can assume one of k integer values $\{1, 2, \ldots, k\}$ and $x_t \in \mathrm{X}$ is a vector of explanatory variables.

The function $\psi(s_t) \in \Psi$ associates with each realization of the state variable, a parameter vector which is chosen from the set $\{\psi_1, \psi_2, \ldots, \psi_k\}$.

On the other hand, in the Extended Switching Regression (ESR) model, which was proposed by Preminger *et al.* (2004a), the existence of p discrete switches in disjoint groups of the model parameters is assumed. The changes in the ith group depend only on the realization of s_t^i, unobserved i.i.d. state variables which can assume one of k integer values $\{1, 2, \ldots, k\}$. The ESR model can be described as follows:

$$y_t = \mu_t(x_t, \psi_1(s_t^1), \ldots, \psi_p(s_t^p)) + \varepsilon_t \tag{2}$$

where μ_t: $X \times \Psi_1 \cdots \times \Psi_p \to \Re$ and the parameters set Ψ is being partitioned into p distinct subsets such that $\Psi = \mathop{\times}\limits_{i=1}^{p} \Psi_i$ and the function $\psi_i(s_t^i) \in \Psi_i$ associates with each realization of the state variable s_t^i, a parameter vector which is chosen from the set $\{\psi_{i1}, \psi_{i2}, \ldots, \psi_{ik}\}$. In other words, the ESR model is characterized by p independent selections from the disjoint parameter sets. From each set we select one element among k possible ones. The selection is random and dependent on the realization of latent state variables. A concise comparison between the latent structures of the ESR and the switching regression (SR) models is given in Fig. 1.

From Figure 1 we can see that the ESR models have the advantage that they take into account situations where disjoint groups of the model parameters change independently over time. This assumption allows us to provide a parsimonious parameterization of the model while allowing us to consider a variety of k^p states which our model can assume, and hence describe more efficiently structural changes in the data. However, the probabilities governing these changes are constant over time. The time-varying ESR (TV-ESR) model relaxes this assumption.

In addition, we assume for simplicity, and without loss of generality, that $k = 2$. The distribution of the state variable is assumed to be time varying, and given by logistic functions:

$$\Pr(s_t^i = 1 \mid z_{it}; \gamma_i) = \frac{\exp(z_{it}'\gamma_i)}{1 + \exp(z_{it}'\gamma_i)} \quad i = 1, \ldots, p \tag{3}$$

where the $(l_i \times 1)$ conditioning vector z_{it} contains explanatory variables that affect the state probabilities and $\gamma_i \in \Gamma_i$ is an unknown set of logistic function parameters. It will be convenient to stack the parameters governing the probabilities of the state variables into one vector, $\gamma = [\gamma_1', \ldots, \gamma_p']' \in \mathop{\times}\limits_{i=1}^{p} \Gamma_i$, and let z_t be a vector containing all the information in the z_{it}. It is obvious, but worth noting, that when the last $(l_i - 1)$ terms of the $(l_i \times 1)$ state variable vector γ_i, are set to zero, the probability distribution of the state variables is time-invariant and our model reduces

- Switching regression (SR) model

$s_t \to \psi \in \{\psi_1, \ldots, \psi_k\} \subset \Psi$

- **Extended Switching regression (ESR) model**

$$\left.\begin{array}{l} s_t^1 \to \tilde{\psi}_1 \in \{\psi_{11}, \ldots, \psi_{1k}\} \subset \Psi_1 \\ \quad\vdots \\ s_t^p \to \tilde{\psi}_p \in \{\psi_{p1}, \ldots, \psi_{pk}\} \subset \Psi_p \end{array}\right\} \Rightarrow \psi = [\tilde{\psi}_1, \ldots, \tilde{\psi}_p]$$

Figure 1. Latent structure of ESR model and SR model

to the ESR model discussed above. The TV-ESR model we consider is given by

$$y_t = \sum_{i=1}^{d} \beta_{it} x_{it} + \varepsilon_t \tag{4}$$

where $\varepsilon_t \sim i.i.N(0, \sigma)$, $\beta_{it} \in \{\beta_{i1}, \beta_{i2}\}$ and $\{x_{it}\}$ are explanatory variables which could include lagged variables of y_t. The model's slopes are determined by d latent state variables which are distributed according to Equation 3. The conditional density of y_t may be obtained as follows:

$$f(y_t \mid x_t, z_t; \theta) = \sum_{j_1=1}^{2} \cdots \sum_{j_d}^{2} \left(\prod_{i=1}^{d} \Pr(s_t^i = j_i \mid z_{it}; \gamma_i) \right) \cdot f(y_t \mid x_t, \{s_t^i = j_i\}; \{\beta_{i1}, \beta_{i2}\}, \sigma) \tag{5}$$

$j_i \in \{1, 2\}$ and $\theta = [\beta_{11}, \beta_{12}, \ldots, \beta_{1d}, \beta_{d2}, \gamma, \sigma]$ is the vector of all model parameters and under the normality assumption, the density of y_t conditional upon x_t, $\{s_t^i\}_{i=1}^{p}$ is given by

$$f(y_t \mid x_t, \{s_t^i = j_i\}; \{\beta_{i1}, \beta_{i2}\}, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ \frac{-\left(y_t - \sum_{i=1}^{d} \beta_{ij_i} x_{it}\right)^2}{2\sigma^2} \right\} \tag{6}$$

where $\beta_{ij_i} \in \{\beta_{i1}, \beta_{i2}\}$. The log-likelihood function of the sample is

$$L_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} \log f(y_t \mid x_t, z_t; \theta) \tag{7}$$

The maximum likelihood estimates (MLE) are obtained by maximizing Equation 7 with respect to the unknown model parameters. We can obtain under mild regularity assumptions the consistency and the asymptotic normality of the MLE (see Preminger *et al.*, 2004b for more details).

From the conditional density (Equation 6) it is straightforward to predict y_t conditional on knowing $\{x_t, z_t, \{s_t^i\}_{i=1}^{d}\}$, such a forecast is given by

$$E(y_t \mid \{s_t^i = j_i\}_{i=1}^{d}, x_t, z_t; \theta) = \sum_{i=1}^{d} \beta_{j_i} x_{it} \quad j_i \in \{1, 2\}, \quad i = 1, \ldots, d \tag{8}$$

Since the state variables are not observable, there are 2^d conditional predictions associated with 2^d possible states. Therefore, the unconditional prediction based on the data, is calculated as follows:

$$\begin{aligned}
&E(y_t \mid x_t, z_t; \theta) \\
&\quad = \int y_t \cdot f(y_t \mid x_t, z_t; \theta) dy_t = \int y_t \cdot \left(\sum_{j_1=1}^{2} \cdots \sum_{j_d=1}^{2} \Pr(y_t, \{s_t^i = j_i\}_{i=1}^{d} \mid x_t, z_t; \theta) \right) dy_t \\
&\quad = \int y_t \cdot \left(\sum_{j_1=1}^{2} \cdots \sum_{j_d=1}^{2} f(y_t \mid \{s_t^i = j_i\}_{i=1}^{d}, x_t, z_t; \theta) \cdot \prod_{i=1}^{d} \Pr(s_t^i = j_i \mid x_t, z_t; \theta) \right) dy_t
\end{aligned}$$

$$= \sum_{j_1=1}^{2} \cdots \sum_{j_d=1}^{2} \left(\prod_{i=1}^{d} \Pr(s_t^i = j_i \mid x_t, z_t; \theta) \right) \cdot \int y_t \cdot (\cdot f(y_t \mid \{s_t^i = j_i\}_{i=1}^d, x_t, z_t; \theta)) dy_t$$

$$= \sum_{j_1=1}^{2} \cdots \sum_{j_d=1}^{2} \left\{ \left(\prod_{i=1}^{d} \Pr(s_t^i = j_i \mid x_t, z_t; \theta) \right) \cdot E(y_t \mid \{s_t^i = j_i\}_{i=1}^d, x_t, z_t; \theta) \right\} \quad (9)$$

Thus, the prediction for y_t is a weighted average of the predictions given in Equation 8. In addition it is a nonlinear function of the observations although our model is linear, since the state probabilities depend nonlinearly on the data. In the simple case where the state variables are perfectly correlated our model reduces to the time-varying switching regression model (Li and Wong, 2001).

In order to estimate the parameters of a linear TV-ESR model, we use the EM algorithm popularized by Dempster *et al.* (1977) – see also McLachlan and Krishnan (1997). This algorithm is particularly suitable to our model since the fact that the TV-ESR likelihood function is not concave hinders the performance of Newton–Raphson based optimization methods. In the EM algorithm, on the other hand, the likelihood values increase (weakly) in each iteration, thus ensuring the algorithm will converge to a local maximum in almost all cases. The development of the EM algorithm for the proposed model can be found in the Appendix.

3. TV-ESR Model for Combining Conditional Volatility Forecasts

Forecasting the exchange rate volatility has been a challenging area of research ever since the collapse of the Bretton Woods system of fixed parties. Volatility is important to policymakers and financial market participants because it can be used as a measure of risk. From the theoretical perspective, volatility plays a central role in pricing of derivative securities. Furthermore, for the purpose of forecasting, confidence intervals may be time varying so that more accurate intervals can be obtained by modelling the volatility of returns.

There is a vast literature on forecasting volatility, and many econometric models, most of which are likely to be misspecified, have been used. However, as noted by Hu and Tsoukalas (1999), no single model was found to be superior. This raises the question: how should volatility forecasts, if at all, be combined into a single forecast with better forecasting performance. The rationale behind combining forecasts is that in practice, forecasting models are intentional abstractions of a much more complex reality (hence the individual forecasts might be biased at times). By combining individual forecasts based on different specifications and/or information sets, we can improve the forecast accuracy. Granger (1989) summarizes the usefulness of combining forecasts: "The combination of forecasts is a simple, pragmatic and sensible way to possibly produce better forecasts".

A common approach for combining forecasts is the simple average of the individual forecasts, which according to Clemen (1989) tends to outperform more complicated combining methods. Another method of combining forecasts suggested by Granger and Ramanathan (1984) is a linear regression on a set of forecasts, where the dependent variable is the true value. Other combination methods such as the Bayesian time-varying weight methods see Min and Zellner (1993) have been proposed as well. However, for the forecasting horizon we investigate in this work, the same authors have demonstrated that there is no substantial benefit in using Bayesian techniques rather than linear regression.

In this work, we illustrate the use of the TV-ESR models as a forecast combining tool for exchange rates volatility forecasts and compare their performances to other common combining methods. We employ five alternative models of combining forecasts given by: the simple average of the individual forecasts (Average), the linear regression where the coefficients are estimated by ordinary least squares (OLS), the switching regression (SR) model with constant and time-varying probabilities, the extended switching regression (ESR) model and the time-varying extended switching regression (TV-ESR) model.

We expect the SR\ESR models to perform better relative to the traditional linear combining methods because these models would account for situations where the "best" model switches over time, by allowing one to change the weighting scheme for each of the individual forecasts over time. The ESR model allows the individual weights to shift at different unknown change-points (according to the realization of several latent state variables). However, if we assume that all the change-points are the same, the ESR model reduces to the traditional switching regression model in which all the combining weights shift at the same point in time. Thus, the switching regression incorporates a potentially binding constraint; all the individual combining weights are mutually dependent on one state variable. Practical considerations suggest the desirability of allowing the individual weights to switch independently over time and to describe more efficiently structural changes which can be related to shifts in the relative performances of the individual forecasts. Furthermore, in the TV-ESR we induce more flexibility by allowing the probability distributions of the latent state variables to change over time.

We consider two models for forecasting the conditional volatility. The GARCH(1, 1) and the moving average variance (MAV) model. Our choice is motivated by the fact that these models are widely employed in the financial econometric literature. Hansen and Lunde (2004), compare the out-of-sample performances of 330 different volatility models to the GARCH(1, 1) model, concluding: "Interestingly, the best models do not provide a significantly better forecast than the GARCH(1, 1) model". See also Donaldson and Kamstra (1997) and Pagan and Schwert (1990) for similar results regarding the simple GARCH and MAV models, respectively.

In the GARCH(1, 1) model, the current conditional volatility of the currency's return depends on the lagged squared error term of the return, and the conditional volatility in the previous period. That is $\sigma_t^2 = \omega + \lambda\sigma_{t-1}^2 + \delta\varepsilon_{t-1}^2$; where σ_t^2 and ε_t^2 are the conditional variance and the squared error term of the return at time t respectively. The model (GARCH) parameters are estimated jointly by maximum likelihood methods assuming conditional normality. Previous studies (Bollerslev *et al.*, 1992) have shown that lower order GARCH models in general and the GARCH (1,1) in particular, provide a parsimonious representation of the temporal dependence in conditional volatility. The second model we use is the MAV model (Pagan and Schwert, 1990) where the volatility is modelled as a simple average of the lagged squared error terms $\sigma_t^2 = 1/H \sum_{h=1}^{H} \varepsilon_{t-h}^2$, where H, the number of lags, is chosen to minimize the Schwarz criterion (1978). In order to combine our individual volatility forecasts we estimate the following TV-ESR model:

$$\sigma_t^2 = \alpha + \beta_{1t}\hat{\sigma}_{1t}^2 + \beta_{2t}\hat{\sigma}_{2t}^2 + \varepsilon_t \tag{10}$$

where ε_t is a Gaussian white noise, $\beta_{it} \in \{\beta_{i1}, \beta_{i2}\}$ for $i = 1, 2$ and σ_t^2 is the actual volatility for time t, and $\hat{\sigma}_{1t}^2, \hat{\sigma}_{2t}^2$ are the individual forecasts for the GARCH and MAV models respectively. The weights which are given to each individual forecast are determined by the realization of its latent state variable, s_t^i, where the probabilities of the state variables are evolving according to Equation 3. Using the maximum likelihood estimates of the TV-ESR model, described in Equation 10, the two individual forecasts are combined through Equations 8 and 9 to produce a one-step-ahead volatility forecast. The EM algorithm, mentioned above, is used for estimating the weighting scheme of the individual forecasts. The application of TV-ESR models for combining forecasts requires modelling the dynamics which characterize the probabilities of the state variables. In the absence of a specific theory for modelling, we consider two simple alternative models for the dynamics.

In the first model, we assume that the probability distribution of the state variables is a function of the sign of the lagged return of the exchange rate, thus the probabilities which correspond to each of the state variables respond asymmetrically to changes in past returns.[2] This variable was shown to be important in modeling conditional mean and the conditional variance in exchange rate markets (Laopodis, 2001) and in stock markets (Koutmos, 1998, 1999). Therefore, it seems reasonable to assume that this variable influences unobserved state variables of the model. We denote this model as TV-ESR1 and the state probabilities are given as follows:

$$\Pr(s_t^i = 1 \mid r_{t-1,i}; \eta_{0i}, \eta_{1i}) = \frac{\exp(\eta_{0i} + \eta_{1i} \cdot sign(r_{t-1}))}{1 + \exp(\eta_{0i} + \eta_{1i} \cdot \mathrm{sign}(r_{t-1}))} \quad i = 1, 2 \tag{11}$$

In the second model, we assume that the probability distribution of the state variables depends on the performance of our individual forecasts relative to each other. The performance is measured by the absolute forecast error of the forecasting models. Such a model could be useful since information about the forecast errors could aid considerably in determining the states. We denote this model as TV-ESR2. Let $e_{t,1}$, $e_{t,2}$ denote the forecast errors of the first (GARCH) and second (MAV) forecasting models respectively, the state variable probabilities are then given by

$$\Pr(s_t^i = 1 \mid I_{t-1}; \gamma_{0i}, \gamma_{1i}) = \frac{\exp(\gamma_{0i} + \gamma_{1i} \cdot I_{t-1})}{1 + \exp(\gamma_{0i} + \gamma_{1i} \cdot I_{t-1})} \quad i = 1, 2 \tag{12}$$

$I_t = 1$ when $\left|\frac{e_{t,1}}{e_{t,2}}\right| > 1$ $I_t = 0$, otherwise.

Now, when the combination of forecasts is derived from the usage of switching regression models, with time-varying probabilities, we will estimate the parameters of Equations 11 and 12 under the restriction that $s_t^1 = s_t^2$. Thus, there exists only one state variable which influences simultaneously the weights given to each of the individual forecasts. The model resulting from Equation 11 is denoted by TV-SR1. The model, where the probability distribution of the state variable changes according to Equation 12, is denoted by TV-SR2 and when the probability of the state variable is constant we denote the model as SR.

4. Results

4.1 Data and the Estimation Procedure

The exchange rate data, used in the empirical analysis, consist of noon bid rates on the Japanese yen (JPY), the British pound (GBP) and the Swiss franc (CHF). All rates are against the US dollar (USD). Let E_t be the exchange rate at time t where the returns are measured as $r_t = \ln(E_t/E_{t-1})$ and are calculated for each exchange rate on a weekly frequency. Our exchange rate data is from the first week of January 1980 to the last week of December 1996. These rates have been taken from the BIS database, giving us 889 observations per exchange rate. Each r_t series is plotted in Fig. 2. In order to calculate the volatility we follow the standard approach suggested by Pagan and Schwert (1990) and Day and Lewis (1992). A proxy for the true volatility is given by $\hat{\sigma}_t = \mid r_t - \bar{r} \mid$, where $\bar{r}$ is the average return over the sample period.

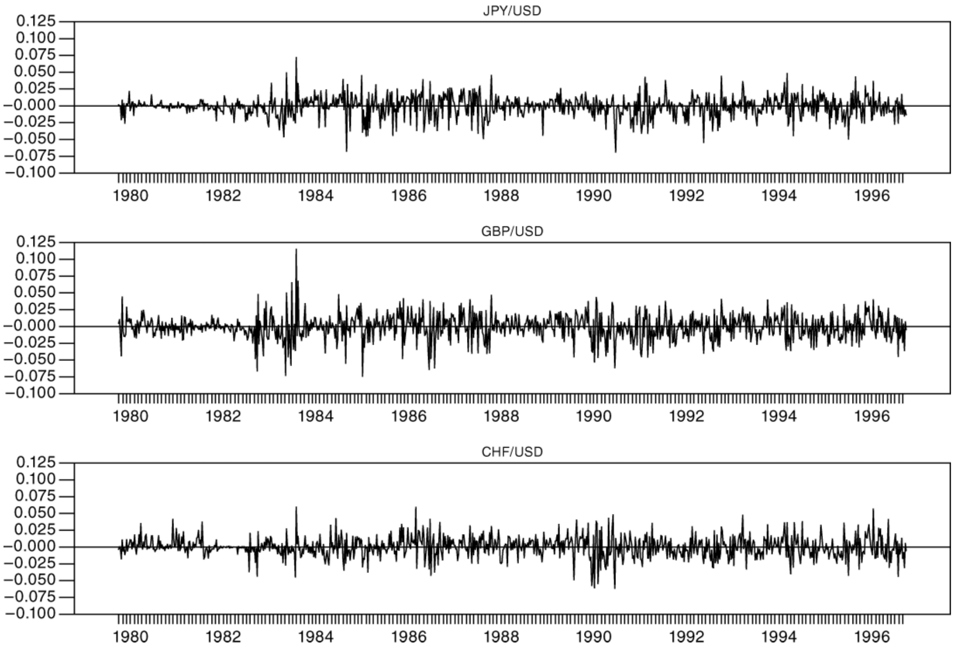

Figure 2. Rate of return of currencies against the dollar

The exchange rates are modeled as a random walk, in line with the analysis of the data and many extensive empirical studies (Meese and Rogoff, 1983; Diebold and Nerlove, 1989; West and Cho, 1995). Visual inspection of the series, which are presented in Figure 2, reveals no evidence of serial correlation, although the conditional variances are characterized by typical "volatility clustering", that is, periods of high volatility followed by periods of low volatility.

The descriptive statistics, presented in Table 1, clearly indicate that all the series have excessive kurtosis and asymmetry and the Bera and Jarque (1982) normality test strongly rejects the normality hypothesis for all series. The first three autocorrelations (r1, r2, r3) along with their standard errors (in parentheses), calculated for each exchange rate, indicate white noise for each series. For the joint test of autocorrelation, we compute the Ljung–Box statistic (LB1) up to the 24th-order serial correlation. Under the null hypothesis of no autocorrelation, the statistic is distributed chi-square asymptotically with 24 degrees of freedom. The test does not reject the white noise hypothesis for all currencies at the 10% significance level. Since our currencies have a mean very close to zero, we can use the squared returns as a measure of their variance and the absolute return as a measure for a standard deviation. The squared returns are clearly not uncorrelated over time, as reflected by the highly significant Ljung–Box (LB2) test for 12th serial correlation, implying conditional heteroscedasticity.

The 17-year study period is split up into three subsamples; the first subsample contains 574 observations from the years 1980 to 1990, which we use to estimate the MAV and the GARCH parameters and then produce a one-step-ahead volatility forecast. Next, the data is updated by adding the first week of 1991 and dropping the last week of 1980, and the model parameters are re-estimated again in order to produce a one-step-ahead forecast from our models for the second week of 1991. This procedure is repeated until we get volatility forecasts for each week for the period 1991–1996. The correlation coefficient between the MAV and GARCH models is low

Table 1. Summary statistics for the weekly return data in the period 1980–1996

Statistic	JPY/USD	GBP/USD	CHF/USD
Mean	−0.0010	−0.0007	0.0003
Std. Dev.	0.0155	0.0187	0.0155
Skewness	−0.2748	−0.0237	0.0161
Kurtosis	4.8740	5.5668	4.5129
BJ test	141.2800	244.1200	84.8200
Maximum	0.0719	0.1148	0.0594
Q3	0.0075	0.0108	0.0092
Median	−0.0001	−0.0007	−0.0006
Q1	−0.0085	−0.0110	−0.0082
Minimum	−0.0687	−0.0741	−0.0614
r1	−0.0083	−0.0620	−0.0012
	(0.335)	(0.335)	(0.340)
r2	0.1112	0.0628	0.0024
	(0.335)	(0.0337)	(0.338)
r3	0.0088	0.0460	0.0616
	(0.335)	(0.335)	(0.335)
LB1(24)	31.3429	24.1347	13.8146
LB2(12)	39.5171	101.8860	62.9393

The data is from the first week of January 1980 to the last week of December 1996 (889 observations). Q1 and Q3 are the first and third quartile respectively and BJ test is the Bera and Jarque (1982) joint test of normality, based on skewness and kurtosis and follows chi-square distribution with two degrees of freedom. r1, r2 and r3 are the first three autocorrelations along with their standard errors in parentheses. LB1(24) is the Ljung and Box (1978) test for the 24th serial correlation. LB2(12) is the same test estimated for the 12th serial correlation for the squared returns of our data.

(0.46), since we use different information sets for the two models. The MAV model uses a longer lag structure than the GARCH to avoid any overlapping effects. Thus, there may be an advantage to using a combined forecast as opposed to either of the individual volatility forecasts.[3]

The second subsample is from the first week of 1991 to the last week of 1994 and is used to estimate the parameters of our combining models. The estimation results from this subsample show that the weighing scheme given to each of the individual forecasts shifts between a high-weight state with low probability (5–10% for both forecasting models) and a low-weight state with high probability. The weights in the high-weight state are substantially higher than the ones in the low-weight state. In addition, some of the weights assigned to the individual forecasts are not significantly different from zero, which indicates that in some periods the information from one volatility forecasting model would add no information beyond that provided by the other model. For the SR\ESR models, the combined forecasts are generated by multiplying our GARCH and MAV volatility forecasts by their expected weights. In the time-varying SR\ESR models, the expected weights vary over time, and in the forecast combining method derived from Equation 12, the better individual forecast, that is the one who performs better relative to the other model in the previous period, receives a higher expected weight and vise versa. Due to space constraints the in-sample estimation results for all our combining methods are available upon request from the authors. The third subsample, from the period 1995–1996, which contains

106 observations, serves as the out-of-sample forecasting period. Given our estimation results, we calculated the one-step-ahead combined volatility forecasts for this period.

4.2 Forecasting Accuracy

We use four measures to compare the one-step-ahead volatility forecasts obtained from our models. The statistics are the root mean square error (RMSE), the mean absolute error (MAE), the Theil-U statistic (Theil-U),[4] and the "correct directional change" (CDC) statistic which measures the ability of our models to correctly predict the actual change which has subsequently occurred in the volatility. These statistics are given as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\sigma_i^2 - \hat{\sigma}_i^2)^2} \tag{13}$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|\sigma_i^2 - \hat{\sigma}_i^2| \tag{14}$$

$$\text{Theil} - U = \frac{\sum_{i=1}^{N}(\sigma_i^2 - \hat{\sigma}_i^2)^2}{\sum_{i=1}^{N}(\sigma_i^2 - \sigma_{i-1}^2)^2} \tag{15}$$

$$\text{CDC} = \frac{100}{T}\sum_{i=1}^{N} D_i \quad \text{where} \quad D_i = \begin{cases} 1 & \text{if } (\sigma_i^2 - \sigma_{i-1}^2)\cdot(\hat{\sigma}_i^2 - \sigma_{i-1}^2) > 0 \\ 0 & \text{if } (\sigma_i^2 - \sigma_{i-1}^2)\cdot(\hat{\sigma}_i^2 - \sigma_{i-1}^2) \leq 0 \end{cases} \tag{16}$$

The RMSE and the MAE are two popular measures to test the forecasting power of a model. However, these measures are not invariant to scale transformations. In the Theil-U statistic, the forecasting error is standardized by the error from a random walk. For the random walk, the Theil-U statistic equals one. Of course, the random walk is not a naïve rival, particularly in many financial and economic series; therefore, a value of the Theil-U statistic close to one is not necessarily an indication of bad forecasting performance. The advantage of the Theil-U statistic is that it is independent of the scale of the variables.

When comparing different models, it can also be useful to measure the number of times a given model correctly predicts the direction of change of the actual values being forecast. Furthermore, as was noted by Dunis and Huang (2002), the RMSE, MAE and Theil-U statistics are important measures for forecasting accuracy of the model concerned, but ignore the profitability point of view. There are, however, cases when the forecaster is less interested in the forecasting accuracy of his volatility models and cares more about the ability of these models to correctly forecast the direction of the change. This is an important issue in trading strategy that relies on the direction of a forecast rather than its level. The CDC statistic, as a measure of the direction, addresses this issue.

Table 2 reports the RMSE, MAE, Theil-U and the CDC statistics for each of the individual forecasts and our combining methods. In terms of the RMSE the ESR and TV-ESR1/TV-ESR2 models as a group show superior out-of-sample forecasting performance compared to other models, for the JPY/USD and the GBP/USD volatility. Conversely, on the basis of the MAE criterion MAV, GARCH and the Average perform as well as, or better than the other models for the GBP/USD volatility. The TV-SR1 and TV-SR2 combination methods provide the most accurate forecasts for the JPY/USD volatility, while the SR model is ranked second. For the CHF/USD volatility,

Table 2. Forecasting performance of competing forecast combining methods for the period 1995–1996

		GARCH	MAV	Average	OLS	ESR	TV-ESR1	TV-ESR2	SR	TV-SR1	TV-SR2
JPY/USD	RMSE	4.43E-04	4.58E-04	4.45E-04	4.50E-04	4.42E-04	4.44E-04	4.43E-04	4.56E-04	4.56E-04	4.56E-04
	MAE	2.76E-04	2.82E-04	2.76E-04	2.62E-04	2.69E-04	2.72E-04	2.72E-04	2.59E-04	2.56E-04	2.56E-04
	Theil-U	0.545	0.580	0.550	0.561	0.541	0.547	0.543	0.576	0.575	0.576
	CDC	65.71%	76.19%	74.29%	67.62%	70.48%	72.38%	71.43%	69.52%	68.57%	69.52%
GBP/USD	RMSE	6.83E-04	7.08E-04	6.89E-04	6.79E-04	6.85E-04	6.73E-04	6.84E-04	6.89E-04	6.85E-04	6.89E-04
	MAE	3.57E-04	3.53E-04	3.50E-04	3.80E-04	3.76E-04	3.76E-04	3.84E-04	3.88E-04	3.84E-04	3.96E-04
	Theil-U	0.481	0.517	0.490	0.477	0.484	0.467	0.483	0.490	0.485	0.491
	CDC	83.95%	83.95%	85.08%	81.06%	83.38%	82.23%	82.23%	80.47%	81.06%	81.65%
CHF/USD	RMSE	1.72E-04	1.70E-04	1.67E-04	2.37E-04	1.87E-04	1.90E-04	1.90E-04	1.96E-04	1.97E-04	1.97E-04
	MAE	1.37E-04	1.12E-04	1.23E-04	2.22E-04	1.61E-04	1.64E-04	1.64E-04	1.74E-04	1.74E-04	1.72E-04
	Theil-U	0.597	0.581	0.565	1.128	0.702	0.729	0.724	0.778	0.783	0.780
	CDC	79.88%	81.06%	82.23%	74.96%	78.07%	78.07%	78.68%	77.46%	77.46%	78.68%

RMSE is the squared root of the mean squared deviation between the volatility forecasted by our combining methods and the actual volatility. MAE is the mean of absolute deviation of the volatility forecasted by our combining method and the actual volatility. Theil-U is the ratio between the squared forecasting error of our models and the squared error from the random walk forecast. CDC is a measure of the ability of our models to forecast changes in the direction of the actual volatility. The results are calculated on weekly data in the period 1995–1996.

the GARCH, MAV and the Average combining methods outperform the other models under all forecast accuracy measures and they also provide the best combining models of the directional change for all the exchange rates.

Examining the Theil-U statistic shows only one model performs worse than the random walk model. This is the OLS combined forecast for the CHF/USD volatility for which the Theil-U is greater than one. The ESR and the TV-ESR models as a group provide the best forecasting performance for the JPY/USD and GBP/USD volatility.

It is hard to identify the best combining model overall since the simple average, the GARCH and the MAV models, perform quite well under any statistic for the CHF/USD volatility. There is also no clear preference between the ESR model and the TV-ESR models, where comparisons across different measures and currencies yield different conclusions. However, on the basis of RMSE, Theil-U and the CDC statistics, the ESR and TV-ESR emerge as the best forecast combining methods in comparison to the OLS and SR models.

4.3 The Encompassing Test

Although useful, the forecasting evaluation measures, discussed in the previous section, cannot determine whether a given forecasting model is in fact "significantly" better than another. In order to evaluate the statistical significance of rival models, we conduct a Chong and Hendry (1986) forecast encompassing test. Applications of this test for the out-of-sample comparison of forecasts in financial markets, can be found in Donaldson and Kamstra (1997) and Darrat and Zhong (2000). To clarify the notion of forecast encompassing, note that the forecast error from a correctly specified model should be orthogonal to any additional information available to the forecaster. Thus, a model claiming to congruently represent the data generating process must be able to account for the salient features of rival models. In more specific terms, model k encompasses model j if model k can explain what model j cannot explain, without model j being able to explain what model k cannot explain. The encompassing tests are therefore based on a set of linear regressions of the forecast error from one model on the forecast from the other model. Thus, with $(\hat{\sigma}_t^2 - \hat{\sigma}_{jt}^2)$ and $(\hat{\sigma}_t^2 - \hat{\sigma}_{kt}^2)$ being the forecast errors from model j and model k respectively and $\hat{\sigma}_{jt}^2$, $\hat{\sigma}_{kt}^2$ being the forecasts of the two models, we test the significance of the δ_{jk} and π_{kj} coefficients in the following regressions:

$$(\hat{\sigma}_t^2 - \hat{\sigma}_{jt}^2) = \lambda_1 + \delta_{jk}\hat{\sigma}_{kt}^2 + \eta_t \quad (17)$$

$$(\hat{\sigma}_t^2 - \hat{\sigma}_{kt}^2) = \lambda_2 + \pi_{kj}\hat{\sigma}_{jt}^2 + \upsilon_t \quad (18)$$

in which η_t and υ_t are random errors.

The null hypothesis is that neither model encompasses the other. If δ_{jk} is not significant at some predetermined level, but π_{kj} is significant, we reject the null hypothesis in favour of the alternative hypothesis that model j encompasses model k. Conversely, if δ_{jk} is significant but π_{kj} is not significant, we say that model k encompasses model j. If both δ_{jk} and π_{kj} are not significant, or if δ_{jk} and π_{kj} are significant, we accept the null hypothesis that neither model encompasses the other.

The encompassing tests results are presented in Table 3 for the JPY/USD, GBP/USD and CHF/USD exchange rates, respectively. The name of the dependent variable, the model forecasting error, is listed down the left side of the table, while the independent variable which is the model forecast, is listed at the top of the table. The entries in the table contain p-values associated with the heteroscedasticity robust t-statistics (White, 1980) of δ_{jk} and $\pi_{kj} \cdot p$-values less than

Table 3. Encompassing tests results – marginal significance level

Exchange rate	Forecast error $\hat{\sigma}_t^2 - \hat{\sigma}_{jt}^2$ from ↓	Forecast $\hat{\sigma}_{kt}^2$ from ↓ GARCH	MAV	Average	OLS	ESR	TV-ESR1	TV-ESR2	SR	TV-SR1	TV-SR2
JPY/USD	GARCH	NA	0.8172	0.8989	0.5681	0.8902	0.9456	0.9413	0.3764	0.3884	0.4438
	MAV	0.0102	NA	0.0065	0.0832	0.0065	0.0067	0.0074	0.7471	0.6651	0.7109
	Average	0.1738	0.2054	NA	0.2415	0.1901	0.1804	0.1864	0.5448	0.5148	0.5677
	OLS	0.3781	0.1529	0.2011	NA	0.1953	0.2329	0.2351	0.2679	0.3036	0.3577
	ESR	0.6859	0.7667	0.7378	0.6443	NA	0.6899	0.7188	0.7046	0.7004	0.7559
	TV-ESR1	0.656	0.7109	0.6895	0.6552	0.6916	NA	0.6639	0.7575	0.7083	0.807
	TV-ESR2	0.7627	0.8812	0.8420	0.6554	0.8426	0.7801	NA	0.6442	0.6444	0.8048
	SR	0.0822	0.0244	0.0337	0.5369	0.0325	0.0459	0.0434	NA	0.5142	0.578
	TV-SR1	0.0891	0.0269	0.0371	0.5547	0.0358	0.0524	0.0476	0.4423	NA	0.5782
	TV-SR2	0.0903	0.0273	0.0376	0.5592	0.0363	0.0506	0.0427	0.4392	0.5049	NA
GBP/USD	GARCH	NA	0.6333	0.5153	0.9264	0.6263	0.3327	0.0479	0.3249	0.6735	0.8357
	MAV	0.0021	NA	0.0017	0.0049	0.0021	0.5421	0.0548	0.3190	0.4054	0.7421
	Average	0.0394	0.0715	NA	0.1411	0.0704	0.8680	0.2675	0.9804	0.5248	0.9471
	OLS	0.9458	0.8323	0.8644	NA	0.8341	0.0295	0.8279	0.7065	0.5070	0.9241
	ESR	0.1766	0.2319	0.2054	0.3236	NA	0.5559	0.4792	0.9634	0.4888	0.8953
	TV-ESR1	0.8102	0.5593	0.6261	0.4504	0.5630	NA	0.6139	0.3549	0.8932	0.9917
	TV-ESR2	0.6341	0.4238	0.4756	0.3480	0.4266	0.0584	NA	0.3423	0.5781	0.5417
	SR	0.3152	0.1339	0.1731	0.0918	0.1356	0.0233	0.298	NA	0.7002	0.8087
	TV-SR1	0.3431	0.1587	0.1966	0.1134	0.1605	0.0790	0.3081	0.1516	NA	0.8775
	TV-SR2	0.3462	0.1609	0.1992	0.1152	0.1627	0.0241	0.5852	0.1530	0.6484	NA
CHF/USD	GARCH	NA	0.4838	0.3071	0.8383	0.4153	0.0298	0.6564	0.5641	0.7371	0.7917
	MAV	0.0067	NA	0.0019	0.0035	0.0018	0.0019	0.0552	0.0121	0.0028	0.0077
	Average	0.0337	0.0538	NA	0.1146	0.0466	0.0366	0.2304	0.0639	0.0933	0.1384
	OLS	0.9223	0.9757	0.9872	NA	0.9883	0.7527	0.8182	0.9626	0.9531	0.9945
	ESR	0.1483	0.1471	0.1317	0.2057	NA	0.1032	0.4143	0.1574	0.1888	0.2501
	TV-ESR1	0.0144	0.1444	0.1287	0.2037	0.1369	NA	0.3402	0.1549	0.1936	0.2147
	TV-ESR2	0.3328	0.4396	0.3805	0.567	0.4164	0.2483	NA	0.4672	0.5391	0.1459
	SR	0.1493	0.0813	0.0888	0.0941	0.0823	0.0629	0.3348	NA	0.0905	0.1357
	TV-SR1	0.1038	0.0194	0.0303	0.0162	0.0202	0.0202	0.1951	0.0176	NA	0.0343
	TV-SR2	0.1586	0.0431	0.0607	0.0375	0.0475	0.0297	0.0255	0.0424	0.0394	NA

The table reports robust p-values on δ_{jk} from the linear regression: $(\hat{\sigma}_t^2 - \hat{\sigma}_{jt}^2) = \lambda_1 + \delta_{jk}\hat{\sigma}_{kt}^2 + \eta_t$ where $\hat{\sigma}_{kt}^2$ is model k's one-step-ahead forecast of the variance and $(\hat{\sigma}_t^2 - \hat{\sigma}_{jt}^2)$ is the out-of-sample forecasting error of model j for the exchange rates on weekly data in the period 1995–1996 for our forecast combination methods, mentioned above.

0.10 indicate that the forecast from the model listed along the top of each table explains, with 10% significance, the forecast error from the model listed down the left side of the table and thus the model listed on the left side cannot encompass the model listed on the top.

Consider first the results for the JPY/USD data as reported in Table 3. The absence of any *p*-values less than 0.1 in the GARCH, Average, OLS, ESR and TV-ESR1/TV-ESR2 rows reveals that none of these models' forecast errors can be explained by other models' forecasts and therefore these models are not encompassed by other models. Conversely, these models (except the OLS) can explain at the 10% significance level, the forecast errors for the MAV, and the switching regression models. Therefore, we conclude that the GARCH, Average, ESR and the TV-ESR1/TV-ESR2 encompass the MAV, SR, TV-SR1, and TV-SR2 models.

We now consider the results for the GBP/USD data. Pair-wise comparisons show that the ESR, TV-ESR1 and the TV-ESR2 forecast errors are not explained by any other model at 10% significance level, therefore they are not encompassed, while these model forecasts explain significantly the forecast errors from other models. Finally, the last section of Table 3 reports the results for the CHF/USD data. The GARCH, OLS, ESR and the TV-ESR models are superior to other models without being encompassed.

The results from the encompassing tests imply that the ESR and the TV-ESR models often encompass rival models in terms of out-of-sample forecasting ability. But these models are not encompassed by rival models while every other model is encompassed at least once. Therefore, we conclude that the TV-ESR, which nests the ESR, tends to be better than other forecast combining techniques on the basis of the encompassing tests.

5. Summary and Conclusions

This paper presents a new class of models which generalize the concept of switching regression models, and develops an EM algorithm in order to estimate their parameters. These models, denoted by ESR and TV-ESR, can capture occasional but recurrent independent switches in disjoint subsets of the model parameters, which are determined by latent state variables. This approach increases the number of states of the model with parsimonious parameterization. In the ESR model, the probability distribution underlying the parameter switches is constant over time, while in the TV-ESR model, it is allowed to vary over time as a function of relevant explanatory variables.

The new methodology is used to combine forecasts of exchange rate volatility. For simplicity, the individual forecasts used are those given by GARCH and MAV forecasting models. Alternative methods of combining forecasts that are considered are the linear regression (OLS), the simple average (Average) and the switching regression (SR) models. The forecasts obtained are compared on the basis of forecast accuracy measures and encompassing tests. The results presented suggest the ESR and TV-ESR models are overall preferred to a variety of competing models.

We should note that although the TV-ESR model nests the ESR model and the traditional linear combining methods as special cases, there is no guarantee that the TV-ESR model would dominate out-of-sample, especially if it over-fits the in-sample data. The empirical findings which suggest that the TV-ESR model performs overall better than the rival models indicates that this model should be preferred to more restrictive forecasts combining methods such as the OLS or the SR models.

Further empirical work should apply these models to the study of other financial time series. This approach can also be used to estimate the relationships between financial variables, *e.g.* the transmission of volatility across financial markets. In addition, the approach adopted in this paper can be extended to allow for more than two latent state variables as well as other types of probability distributions for the state variables (*e.g.*, the probit function). Furthermore, our model can also be used to generate volatility forecasts. These extensions leave several interesting and challenging areas for future research.

Acknowledgement

We are grateful for several comments from seminar participants at the FFM 2003 conference, Tel Aviv University and the Technion – Israel institute of Technology. We thank Ezra Einy for several insightful comments and suggestions. We also thank two anonymous referees and an editor for several important questions and suggestions. We are grateful to the Federal Reserve Bank of St. Louis and BIS for providing access to their databases. Preminger gratefully acknowledges research support from the Kreitman Foundation.

Notes

[1] The standard measurability and continuity requirements can be found in Wooldridge (p. 2726, 1994).

[2] The asymmetry assumption implies in effect that one model performs better (worse) for a rising exchange rate, inverting the exchange rate will mean that this model is now better (worse) when the exchange rate declines.

[3] Note that, combining forecasts is a practical alternative if we do not have access to the forecasters' information sets (Diebold and Lopez, 1996). However, we do have the information sets used to produce both our individual volatility forecasts. Hence, the application to conditional volatility forecasts in our paper should be viewed primarily as an exercise to compare the combining methods. Use of the TV-ESR model to forecast exchange rate volatility using all available information is a subject for future research.

[4] One should note that, at different times, Theil proposed several forecast accuracy measures, all under the same symbol U. The Theil-U statistic used in this paper is based on Theil (1966). This statistic is simple and widely used (see Bliemel, 1973 and Diebold, 2004). Moreover in contrast to the Theil (1958) version it is not bounded between zero and one.

References

Bera, A. K. and Jarque, C. M. (1982) Model specification tests: a simultaneous approach, *Journal of Econometrics*, 20, pp. 59–82.

Bliemel, F. W. (1973) Theil's forecast accuracy coefficient: a clarification, *Journal of Marketing Research*, 10, pp. 444–446.

Bollerslev, T. R., Chou, Y. and Kroner, K. F. (1992) ARCH modeling in finance: a selective review of the theory and empirical evidence, *Journal of Econometrics*, 52, pp. 5–59.

Chong, Y. Y. and Hendry, D. F. (1986) Econometric evaluation of linear macroeconomics models, *Review of Economics Studies*, 53, pp. 671–690.

Clemen, R. T. (1989) Combining forecasts: a review and annotated bibliography, *International Journal of Forecasting*, 5, pp. 559–583.

Darrat, A. F. and Zhong, M. (2000) On testing the random-walk hypothesis: a model comparison approach, *The Financial Review*, 35, pp. 105–124.

Day, T. E. and Lewis, C. M. (1992) Stock market volatility and information content of stock index options, *Journal of Econometrics*, 52, pp. 267–287.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society* B, 39, pp. 1–38.

Diebold, F. X. (2004) *Elements of Forecasting*, 3rd edn (Thomson, South-Western).

Diebold, F. X. and Lopez, J. A. (1996) Forecast evaluation and combination, in: G. S. Maddala and C. R. Rao (Eds) *Handbook of Statistics* (Amsterdam: North-Holland).

Diebold, F. X. and Nerlove, M. (1989) The dynamics of exchange rate volatility: a multivariate latent factor ARCH model, *Journal of Applied Econometrics*, 4, pp. 1–21.

Donaldson, R. G. and Kamstra, M. (1997) An artificial neural network-GARCH model for international stock return volatility, *Journal of Empirical Finance*, 4, pp. 17–46.

Dunis, C. L. and Huang, X. (2002) Forecasting and trading currency volatility: an application of recurrent neural regression and model combination, *Journal of Forecasting*, 21, pp. 317–354.

Goldfeld, S. M. and Quandt, R. E. (1972) *Nonlinear Methods in Econometrics* (Amsterdam: North Holland).

Granger, C. W. J. (1989) Invited review combining forecasts – twenty years later, *Journal of Forecasting*, 8, pp. 167–173.

Granger, C. W. J. and Ramanathan, R. (1984) Improved methods of forecasting, *Journal of Forecasting*, 3, pp. 197–204.

Hamilton, J. D. (1994) *Time Series Analysis* (Princeton: Princeton University Press).

Hansen, P. R. and Lunde, A. (2004) A comparison of volatility models: does anything beat a GARCH (1,1), *Journal of Applied Econometrics* (forthcoming).

Hu, M. Y. and Tsoukalas, C. (1999) Combining conditional forecasts using neural networks: an application to the EMS exchange rates, *Journal of International Financial Markets, Institutions and Money*, 9, pp. 407–422.

Koutmos, G. (1998) Asymmetries in the conditional mean and the conditional variance: evidence from nine stock markets, *Journal of Economics and Business*, 50, pp. 277–290.

Koutmos, G. (1999) Asymmetric price and volatility adjustments in emerging Asian stock markets, *Journal of Business, Finance and Accounting*, 26, pp. 83–101.

Laopodis, N. T. (2001) Time varying behaviour and asymmetry in EMS exchange rates, *International Economic Journal*, 15, pp. 81–94.

Le, N. D., Martin, D. and Raftery, A. E. (1996) Modeling flat stretches, bursts and outliers in time series using mixture transition distribution models, *Journal of the American Statistical Association*, 91, pp. 1504–1515.

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

Li, W. K. and Wong, C. S. (2001) On a logistic mixture autoregressive model, *Biometrika*, 88, pp. 833–846.

Ljung, G. and Box, G. (1978) On a measure of lack of fit in time-series models, *Biometrika*, 65, pp. 297–303.

McLachlan, G. J. and Krishnan, T. (1997) *The EM Algorithm and Extensions* (New York: Wiley).

Meese, R. A. and Rogoff, K. (1983) Empirical exchange rates models of the seventies: do they fit out-of-sample, *Journal of International Economics*, 14, pp. 2–24.

Min, C. and Zellner, A. (1993) Bayesian and non-Bayesian methods for combining models and forecasts with applications to forecasting international growth rates, *Journal of Econometrics*, 56, pp. 89–118.

Pagan, A. R. and Schwert, G. W. (1990) Alternative models for conditional stock volatility, *Journal of Econometrics*, 45, pp. 267–290.

Preminger, A., Ben-Zion, U. and Wettstein, D. (2004a) Extended switching regression models: allowing for multiple latent state variables. Working Paper, Monaster Center for Economic Research, Ben-Gurion University of the Negev.

Preminger, A., Ben-Zion, U. and Wettstein, D. (2004b) Extended switching regression models with time varying probabilities for combining forecasts. Working Paper, Monaster Center for Economic Research, Ben-Gurion University of the Negev.

Quandt, R. E. (1958) The estimation of parameters of linear regression systems obeying two separate regimes, *Journal of the American Statistical Association*, 53, pp. 873–880.

Schwarz, G. (1978) Estimating the dimension of a model, *Annals of Statistics*, 6, pp. 461–464.

Theil, H. (1958) *Economic Forecasts and Policy* (Amsterdam: North Holland).

Theil, H. (1966) *Applied Economic Forecasting* (Amsterdam: North Holland).

West, K. D. and Cho, D. (1995) The predictive ability of several models of exchange rate volatility, *Journal of Econometrics*, 69, pp. 367–391.

White, H. (1980) A heteroskedasticity consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica*, 48, pp. 817–838.

Wooldridge, J. M. (1994) Estimation and inference for dependent processes, in: R. F. Engle and D. L. McFadden (Eds) *Handbook of Econometrics* Vol. 4 (Amsterdam: Elsevier Science/North-Holland).

Appendix – The EM algorithm

In order to estimate the TV-ESR model let $S_t = \{S_t^1, \ldots, S_t^d\}'$ where S_t^i is a two-dimensional unobserved vector $(S_t^{i1}, 1 - S_t^{i1})$, with S_t^{i1} equal one if $s_t^i = 1$ and zero if $s_t^i = 2$ and let $B = (\beta_{11}, \beta_{12}, \ldots \beta_{d1}, \beta_{d2})'$, $X_t = (x_{1t}, x_{1t}, x_{2t}, x_{2t}, \ldots, x_{dt}, x_{dt})'$ and $\theta^\ell = [B^\ell, \sigma^\ell, \gamma_1^\ell, \ldots, \gamma_d^\ell]$ denote the parameters estimated in the $\ell - th$ iteration. We can write

Equation 4 more succinctly as

$$y_t = (B * X_t)' \cdot S_t + \varepsilon_t \tag{A.1}$$

The symbol "$*$" denotes the Hadamard product, which means element-by-element multiplication. The log-likelihood function of the complete data, assuming that the values of the state variables are known, is given by

$$L_T^C(\theta) = \sum_{t=1}^{T}\left(\sum_{i=1}^{d}[S_t^{i1}\log p_{i1}(\cdot) + (1 - S_t^{i1})\log(1 - p_{i1}(\cdot)]\right.$$
$$\left. - \sum_{i=1}^{d}\sum_{j=1}^{2}\frac{(y_t - (B * X_t)'S_t))^2}{2\sigma^2} - 0.5 \cdot \log(2\pi\sigma^2)\right) \tag{A.2}$$

where $p_{i1}(\cdot) = \Pr(s_t^i = 1 \mid z_{it}, \gamma_i)$, the probability of $s_t^i = 1$, conditional on some information set, $z_{it} \in \Im_{t-1}$. Formally, $\Im_{t-1}$ is the σ – algebra induced by all variables that are observed at time $t - 1$. Thus, $\Im_{t-1}$ contains the lagged values of y_t and other predetermined variables. The iterative EM procedure for estimating the model parameters consists of an E-step in which the expectation of Equation A.2 is taken with respect to the distribution of the state variables given the data and the parameters estimated in the previous iteration, and an M-step where a new set of parameters is generated through the maximization of the expectation. In the E-step we get:

$$E(L_T^C(\theta \mid \Im_T; \theta^{\ell-1}))$$
$$= \frac{-1}{2\sigma^2}\sum_{t=1}^{T}\{y_t^2 - 2y_t \cdot (B * X_t)' \cdot \hat{S}_t(\theta^{\ell-1}) + (B * X_t)' \cdot \Lambda_t(\theta^{\ell-1}) \cdot (B * X_t)\}$$
$$+ \sum_{t=1}^{T}\sum_{i=1}^{p}[\hat{S}_t^{i1}\log p_{i1}(\cdot) + (1 - \hat{S}_t^{i1}) \cdot \log(1 - p_{i1}(\cdot))] - 0.5 \cdot T\log(2\pi\sigma^2) \tag{A.3}$$

where $\Lambda_t(\theta^{\ell-1}) = E(S_t \cdot S_t' | y_t, \Im_{t-1}; \theta^{\ell-1})$, $\hat{S}_t(\theta^{\ell-1}) = E(S_t \mid y_t, \Im_{t-1}; \theta^{\ell-1})$ and $\hat{S}_t^{i1}$ is the conditional expectation of S_t^{i1}. The elements of $\Lambda_t(\theta^{\ell-1})$, $\hat{S}_t(\theta^{\ell-1})$, can be deduced from the following calculations:

$$\hat{S}_t^{i1} = \Pr(S_t^{i1} = 1 \mid y_t, \Im_{t-1}; \theta^{\ell-1}) = \frac{\Pr(y_t, S_t^{i1} = 1 \mid \Im_{t-1}; \theta^{\ell-1})}{\Pr(y_t \mid \Im_{t-1}; \theta^{\ell-1})}$$
$$= \frac{p_{i1}(\cdot)\Pr(y_t \mid S_t^{i1} = 1, \Im_{t-1}; \theta^{\ell-1})}{\Pr(y_t \mid \Im_{t-1}; \theta^{\ell-1})}$$
$$= \frac{p_{i1}(\cdot)\sum_{\{S_t^{rj}\}_{r\neq i}}\Pr(y_t, \{S_t^{rj}\}_{r\neq i} \mid S_t^{i1} = 1, \Im_{t-1}; \theta^{\ell-1})}{\Pr(y_t \mid \Im_{t-1}; \theta^{\ell-1})} \tag{A.4}$$

$$\Lambda_{\text{mnt}}(\theta^{\ell-1}) = E(S_t^{mj}S_t^{nj} \mid y_t, \Im_{t-1})_{m\neq n} = \Pr(S_t^{mj} = 1, S_t^{nj} = 1 \mid y_t, \Im_{t-1}; \theta^{\ell-1})_{m\neq n}$$
$$= \frac{Pr(S_t^{mj} = 1, S_t^{nj} = 1, y_t \mid \Im_{t-1}; \theta^{\ell-1})}{Pr(y_t \mid \Im_{t-1}; \theta^{\ell-1})}$$

$$= \frac{p_{mj}(\cdot) \cdot p_{nj}(\cdot) \cdot \Pr(y_t \mid \Im_{t-1}, S_t^{mj} = 1, S_t^{nj} = 1; \theta^{\ell-1})}{\Pr(y_t \mid \Im_{t-1}; \theta^{\ell-1})}$$

$$= \frac{p_{mj}(\cdot) \cdot p_{nj}(\cdot) \sum_{\{S_t^{rj}\}_{r \neq m,n}} \Pr(y_t, \{S_t^{rj}\}_{r \neq m,n} \mid \Im_{t-1}, S_t^{mj} = 1, S_t^{nj} = 1; \theta^{\ell-1})}{\Pr(y_t \mid \Im_{t-1}; \theta^{\ell-1})} \tag{A.5}$$

Next, we perform the M-step, in which the estimates of the parameters are obtained by maximizing Equation A.3 above. Solution of the first-order conditions yields parameter estimates for B and σ given by

$$B^{\ell} = \left(\sum_{t=1}^{T} X_t \cdot X_t' * \Lambda_t(\theta^{\ell-1}) \right)^{-1} \sum_{t=1}^{T} y_t X_t * \hat{S}_t(\theta^{\ell-1}) \tag{A.6}$$

$$\sigma^{\ell} = \sqrt{\frac{1}{T} \sum_{t=1}^{T} [y_t^2 - 2y_t(B^{\ell} * X_t)' \hat{S}_t(\theta^{\ell-1}) + (B^{\ell} * X_t)' \Lambda_t(\theta^{\ell-1})(B^{\ell} * X_t)]} \tag{A.7}$$

By differentiating Equation A.3 with respect to γ_i we get

$$\sum_{t=1}^{T} (\hat{S}_t^{i1} - \Pr(s_t^i = 1 \mid z_{it}; \gamma_i) z_{it} = \underline{0} \tag{A.8}$$

In order to solve the nonlinear Equation A.8 for γ_i, we use the Newton method. That is, given initial guess γ_i^0, the values of γ_i in the subsequent iterations are given by

$$\gamma_i^{h+1} = \gamma_i^h + \left(\sum_{t=1}^{T} \Pr(s_t^i = 1 \mid z_{it}, \gamma_i^h)(1 - \Pr(s_t^i = 1 \mid z_{it}, \gamma_i^h)) \cdot z_{it}{}' z_{it} \right)^{-1} \times \sum_{t=1}^{T} (\hat{S}_t^{i1} - \Pr(s_t^i = 1 \mid z_{it}, \gamma_i^h)) z_{it} \tag{A.9}$$

For $h = 1$, we set $\gamma_i^h = \gamma_i^{\ell-1}$ (our initial guess equals the estimated parameters in the previous iteration) and update the values of γ_i according to Equation A.9 until convergence and for this h, we set $\gamma_i^h = \gamma_i^{\ell}$. Starting from an initial guess of the model parameters, the parameters are then obtained by iterating Equations A.4 to A.9 until convergence.