

Structural Breaks and Non-Linearities for Predicting the Probability of US Recessions using the Spread

Ana Beatriz C. Galvão*
Department of Economics
European University Institute
agalvao@iue.it

January 2002

Abstract

This paper proposes a structural break threshold model (SBT) to the dynamic relationship between US output growth and the spread between long- and short-term interest rates. This model is able to account for non-linearities, parameter changes and the reduction of the variability of output growth. The SBT model gives better in-sample predictions of the probability of US recessions during 1955-1999 than models with only non-linearity or structural breaks. The presence of a structural break affects the timing and the size of the predictions of the probability of recession for 2001.

Key words: predictions of event probabilities, threshold models, structural breaks, recessions;

1 Introduction

Economic forecasters do not enjoy a good reputation when the prediction of US recessions is concerned: “the dismal scientists have a dismal record in predicting recessions” (*Don't Mention the R-word*, 2001). The problem is that recessions are relatively rare events with large potential consequences for individuals and companies. Better predictions of recessions can help monetary authorities and the private sector to take economic decisions not only in the US but also in countries that are likely to suffer the effects of the US recession. The main contribution of this paper is to propose a model for predicting the probability of recession that accounts for non-linearity and structural breaks when the spread between long- and short-term interest rates is the leading indicator.

The literature presents evidence that the spread, which represents the term structure of interest rates, is a good predictor of output growth (Estrella and Hardouvelis, 1991; Hamilton and Kim, 2000; and the surveys of Berk (1998) and Stock and Watson (2001)). The information

*The results of this paper are based on my doctoral research at University of Warwick. I would like to thank Mike Clements and Bill Russell for useful suggestions and Capes-Brazil for their financial support.

contained in the spread reflects not only monetary policy but future expected short rates and changes in the risk premium (Hamilton and Kim, 2000). In fact, the spread keeps its predictive power when other indicators of monetary policy (Anderson and Vahid, 2000) and oil prices (Hamilton and Kim, 2000) are included in a regression to explain output growth. The spread is also a good predictor of the probability of recession (Lahiri and Wang, 1996; Estrella and Mishkin, 1998).

However, Haubrich and Dombrosky (1996), Dotsey (1998) and Stock and Watson (2001) report that the predictive power of the spread between long- and short-term interest rates has decreased after 1985. The failure of the Stock and Watson (1989) indicator index to predict the 1990-91 recession has been attributed to the fact that the index gave too large weight to the spread (Dotsey, 1998). In contrast, employing Markov-switching models to obtain the probability of recession, Lahiri and Wang (1996) show that the spread predicted the last recession. Likewise, Dueker (1997) and Estrella and Mishkin (1998), using probit, demonstrate that the spread is still better than other leading indicators at predicting recessions in the US. The tests presented by Estrella et al. (2000) support the view that while there is no instability in the ability of the spread to predict recessions, the ability of the spread to predict the economic growth is unstable. Recently, Chauvet and Potter (2001) contest these results with findings of parameter instability in probit models.

The literature also presents evidence of non-linearities in models that use the spread to predict output growth (Galbraith and Tkacz, 2000; Anderson and Vahid, 2000). The inclusion of non-linearities improves the accuracy of predicting the probability of recession (Anderson and Vahid, 2000), while large spreads do not predict strong growth (Galbraith and Tkacz, 2000).

Regarding changes in the output growth series, an important stylized fact is that the variability of output growth decreased after 1984 (Kim and Nelson, 1999; McConnell and Perez-Quiros, 2000). Regarding interest rates, Watson (1999) suggests that the variability of the US long-term interest rate has been increasing while the short-term interest rate is smoothed by the monetary authority. However, the results of the tests applied by Sensier and Van Dijk (2001) indicate that while there is evidence of structural break in short- and long-term interest rates, the evidence of a structural break in their spread is not strong.

Therefore, the literature suggests that a linear model between output growth and spread is not a good representation of the dynamic responses between these variables because of parameter instability (Estrella et al., 2000; Stock and Watson, 2001), non-linearities (Galbraith and Tkacz, 2000; Anderson and Vahid, 2000) and changes in the variability of the output growth (Kim and Nelson, 1999; McConnell and Perez-Quiros, 2000). We show in this paper that a structural break threshold model can account for these characteristics and can be employed to generate better predictions of the probability of recession.

Section 2 reviews and estimates threshold and structural break models. Testing, modelling and estimation of structural break threshold models are presented in section 3. Section 4 shows how stochastic simulation can be employed to obtain probability of recessions from bivariate

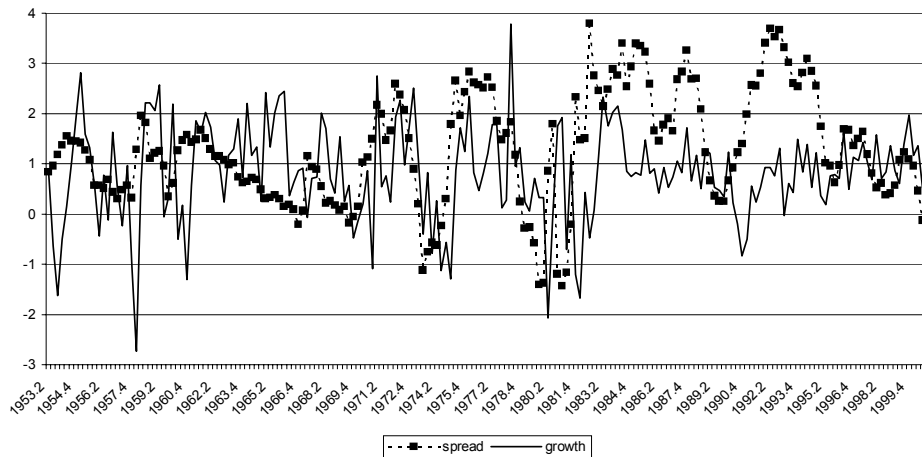


Figure 1 Spread (between the interests rates of a 10-year T-bond and a 3-month T-bill) and the growth rate of the real GDP.

systems, using two rules to define recessions. The evaluation of the in-sample probability forecasts is also presented in section 4. Section 5 presents the forecast of the probability of recession for 2001, compared with similar predictions from the Stock and Watson leading index, the Survey of Professional Forecasters and the probit models of Chauvet and Potter (2001). Section 6 summarizes the main findings of this paper and concludes.

2 Modelling Structural Breaks and Thresholds

Before proposing a structural break threshold model, we review threshold and structural break models, including results of their application to bivariate systems of spread and output growth. All the calculations of this paper are based on data taken from the Federal Reserve Bank at St Louis (<http://www.stls.frb.org/fred/index.html>). The spread is the difference between the interest rate of a 10-year T-bond and a 3-month T-bill, averaging to transform monthly into quarterly interest rates. The output growth is the first difference of the log of quarterly real GDP (seasonally adjusted) in chained 1996 prices (*100). The data set presented in Figure 1 regards the full sample from 1953:2 to 2000:4. For the models and tests discussed in this section, only data until 1999:4 are employed. Using information criteria and LR tests, we define the autoregressive order of each endogenous variable. The chosen autoregressive order of the VAR described in the Appendix is kept for the other models discussed in this paper.

Threshold non-linearity or structural breaks can be included in a bivariate VAR of the spread S_t and the output growth y_t as follows:

$$\begin{aligned}
 y_t &= x_{t-1}\beta_1(1 - I_1(\cdot)) + x_{t-1}\beta_2(I_1(\cdot)) + u_{1t} \\
 S_t &= x_{t-1}\beta_3(1 - I_2(\cdot)) + x_{t-1}\beta_4(I_2(\cdot)) + u_{2t},
 \end{aligned}$$

where $t = 1, \dots, T$; $x_{t-1} = (1, y_{t-1}, \dots, y_{t-p_1}, S_{t-1}, \dots, S_{t-p_2})'$; and $I_i(\cdot)$ is a transition function, which is an indicator function. If the non-linear functions $I_1(\cdot)$ and $I_2(\cdot)$ are the same for each equation of the model, we suppose that $cov(u_{1t}, u_{2t}) \neq 0$. Given that the explanatory variables of each equation are the same (x_{t-1}), the model is a generalization of a VAR model. However, to predict output growth using spread, and specifically recessions, we can relax the assumption that $I_1(s_{t-d_1}) = I_2(s_{t-d_2})$, given that $cov(u_{1t}, u_{2t}) = 0$. In this case, each equation can be seen as a non-linear regression.

2.1 Threshold Models

In the case of threshold models, the indicator function depends on the transition variable s , the delay d_i and the threshold r_i , i.e., $I_i(s_{t-d_i}) = 1$ when $s_{t-d_i} > r_i$ and $I_i(s_{t-d_i}) = 0$ when $s_{t-d_i} \leq r_i$. The threshold is estimated by grid search over all possible threshold values $r \in [r_L, r_U]$. The upper and the lower values of this interval are calculated as follows. The values ($t = 1, \dots, T$) of the transition variable are sorted and a proportion π of the observations is trimmed in each end with $0 < \pi < 1$. The delay is jointly estimated with the thresholds (by grid search) given that $d_L = 1$ and $d_U = d_{\max}$. Conditional on the threshold value, the delay and the autoregressive order, the threshold regressions are estimated by OLS, assuming that the residuals are not contemporaneously correlated. In the case that each equation has the same transition function, the model is estimated by multivariate least squares, conditional on the threshold value, using the grid search procedure described to estimate the threshold and the delay.

The test for linearity against the alternative hypothesis of a threshold regression does not have a conventional asymptotic distribution (Hansen, 1996) and the statistic is defined as the maximum of a set of F statistics which are calculated for comparison between the linear model and the threshold model for each possible value of the threshold¹. The *supF* statistic is the maximum F statistic obtained in a search over possible thresholds and delays. The p-value of this test is calculated using the asymptotic results of Hansen (1996). Hansen (2000a) argues that a bootstrap is a better approximation for finite samples, but his bootstrap approach implies being able to simulate values of y_t and S_t , which is not possible when we test each equation separately even though it is possible when a VAR is considered.

The threshold non-linearity test is also conducted in the VAR, applying a LR test. This test has been proposed by Clements and Galvão (2001) in the context of non-linear cointegrated systems and it is a multivariate extension of Hansen (1996; 2000a). Given the estimated variance-covariance matrices ($\hat{\Omega}_j$) of the residuals $(u_{1t}, u_{2t})'$, the supLR for testing non-linearity

¹Averages and exponentials, instead of the maximum have also been proposed by Hansen (1996). The results of the tests of this section do not change when other transformations of the F statistic are employed.

with a two-regime threshold model under alternative is

$$\sup LR_{12} = \max_{\substack{r_L \leq r \leq r_U \\ d_L \leq d \leq d_U}} (T(\det(\hat{\Omega}_1) - \det(\hat{\Omega}_2))), \quad (1)$$

where $j = 1$ is the index for the linear model, $j = 2$ is the index for a two-regime threshold model. A test of a two-regime against a three-regime threshold model is also performed using this same type of test (LR_{23}). The p-value of the LR test is calculated using heteroscedasticity corrected bootstrap².

The results of the non-linearity tests are presented in Table 1, panel 1. Only the statistics correspondent to the transition variable that minimizes the sum of squared residuals (or the log of the determinant of the contemporaneous covariance matrix in the case of the TVAR) are presented, even though they are calculated for all the variables of the vector of explanatory variables $x_{t-1} = (1, y_{t-1}, y_{t-2}, S_{t-1}, S_{t-2}, S_{-3})'$ for the regressions and for $d_U = 5$ with the spread as transition variable for the system. The results support non-linearity, although the robust statistic for the output equation does not reject the null of non-linearity at 5% significance level but it does at 10%. We also use the LR_{23} for verifying the possibility of a third regime and the test cannot reject the null hypothesis (p-value of .22). Therefore, two different specifications are suggested by these tests: a two-regime threshold model (T) with different transition function for each equation and a two-regime threshold VAR model with the same transition function for each regimes (TVAR), which are described in the Appendix.

2.2 Structural Break Models

Structural break models have t as transition variable and the threshold is the change-point τ , so $I_i(t) = 1$ when $t > \tau$ and equal to zero otherwise. Even though the change-point can be employed to calculate the variance (and covariance) of the residuals $\sigma_{u_{it}}$ conditional on the subsamples, models with structural breaks that arise in the variance equation can be also defined. For example, assuming that the equation of the mean is linear, structural break in the equation of the variance can be written as:

$$\sigma_{u_{it}}^2 = \sigma_B^2 I(t) + \sigma_A^2 (1 - I(t)), \quad (2)$$

where $i = 1$ indicates the y_t equation and $i = 2$, the S_t equation, so u_{1t} are the residuals of the equation of the output growth.

Testing structural breaks in the mean and the variance for unknown break points implies the presence of a nuisance parameter under the null hypothesis. Hansen (2001) presents an up-to-date literature review on the advances in this area. The standard test for one structural

²For the LR_{12} test, the residuals of the linear model are standardized by the fitted values of a regression of the squared of the errors on x_{t-1}^2 . For the LR_{23} test, the bootstrap assume that the errors of the two-regime model are regime-dependent.

Table 1 Tests for structural breaks and non-linearities.

	Alternative Hypothesis	output	spread	both
1	Threshold non-linearities	22.28	46.53	61.35
		(0.021)	(0.000)	{0.002}
		[0.086]	[0.042]	
	Transition variable	S_{t-1}	S_{t-1}	S_{t-4}
2	SB in mean	26.12	25.88	28.33
		(0.006)	(0.007)	{0.153}
		[0.011]	[0.011]	
		{0.024}	{0.150}	
	Break Date	1959:2	1981:1	1980:4
3	SB in variance	23.47	12.40	
		(0.001)	(0.007)	
		[0.001]	[0.013]	
		{0.005}	{0.158}	
	Break Date	1983:2	1966:1	
4	SB in mean in threshold model	33.402	45.869	84.440
		(0.025)	(0.000)	(0.000)
		[0.043]		
		{0.038}		
	Break Date	1960:1	1980:4	1980:4
5	SB in variance in threshold model	19.194	13.513	
		(0.000)	(0.004)	
		[0.020]	[0.066]	
		{0.051}	{0.331}	
	Break Date	1984:1	1966:3	

Note: The statistics are supF when regressions and supLR for the system.

Test in panel (1) uses homoscedastic () and heteroscedastic []

asymptotic p-values. Tests in panels (2), (3), (4), (5) have p-values

calculated by an asymptotic distribution () and by fixed

regressors bootstrap assuming homoscedasticity[] and

heteroscedasticity { } of the residuals under the null.

$\pi=.1$. Number of bootstrap.: 1000.

break is the *supLR*, proposed by Andrews (1993). The idea is to calculate the LR statistic for all possible structural breaks, given that $\tau \in [\tau_L, \tau_U]$, where $\tau_L = \pi T$ and $\tau_U = (1 - \pi)T$ ³. Then the maximum LR value over $[\tau_L, \tau_U]$ is the test statistic⁴ and the associated $\hat{\tau}$ is the estimated change-point. The test is equivalent to the test for threshold non-linearity with the difference that x_{t-1} is ordered chronologically rather than ranked by the threshold value.

The p-values can be calculated numerically by the Hansen (1997) procedure, given the number of restrictions and the trimming factor π . Hansen (2000b) shows that a fixed regressor bootstrap works better than the asymptotic values in small samples. The residuals can be corrected for heteroscedasticity before they are bootstrapped to generate values of the endogenous variable. Therefore, in addition to the asymptotic p-values, fixed regressor bootstrap (Hansen, 2000b) is also employed to calculate p-values and heteroscedastic corrected p-values⁵.

In the case of testing for structural breaks jointly in both equations ($I_1(t) = I_2(t)$), we apply the *sup LR*₁₂ (eq. 1), employed to test non-linearity, using a time trend as the transition variable. Using this same approach, a test for a model with two structural breaks against one structural break is also applied (*sup LR*₂₃), which is similar to the test for multiple breaks proposed by Bai (1999). The p-values, as in the previous section, are calculated by bootstrap with a correction for heteroscedasticity.

The results of tests for parameter instability are presented in Table 1, panel 2. They suggest the presence of strongly dissimilar changing points depending whether either the spread or the output growth is treated as endogenous variable. The evidence of a structural break in the spread equation is weaker than in the output equation when the heteroscedasticity corrected bootstrap p-values are considered.

A similar procedure can be employed to test and to estimate a model with a structural break only in the variance equation (eq.2). Instead of seeking a break in a regression of, say, output growth against lagged values of spread and output growth, the break has to be located in regressions of the square of the residuals of each equation of the linear model against a constant. Therefore, the *sup F* is employed to test for a structural break in the variance: as above, p-values are calculated using the asymptotic distribution and homoscedastic and heteroscedastic fixed regressor bootstraps.

The tests for a structural break in the variance, shown in Table 1, panel 3, strongly support the existence of a break in the variance of the output growth equation. The same test when it is assumed that there is a structural break in the mean does not change this result (not

³Note that this may be necessary to adjust the values of πT and $(1 - \pi)T$ to the next integer to get feasible numbers for τ .

⁴Average and exponential test statistics have been proposed by Andrews and Ploberger (1994). The results of the structural break tests applied to the models of this work do not change when these transformations are employed instead of the *sup*.

⁵In the homoscedastic bootstrap, values are drawn from a normal distribution and regressed against the regressors under the null hypothesis and the regressors under the alternative hypothesis. The residual variances are calculated for both regressions and the F statistic is calculated. In the heteroscedastic bootstrap, the values are drawn from a normal distribution multiplied by the errors of the model under the alternative hypothesis.

shown). The break point estimated for the variance (1983:2) of the output equation is similar to the value (1984:1) of univariate models of output growth (Kim and Nelson, 1999; McConnell and Perez-Quiros, 2000). In the case of the spread equation, the heteroscedastic fixed regressor bootstrap suggests again that 1966:1 break in the variance is not statistically significant. This could be result of a poor estimated break point when changes are detected in the conditional mean and in the conditional variance. Therefore, we search for a joint structural break in mean and in variance. The structural break in mean and in variance model (SBMV) is defined as

$$z_{it} = \beta_1 x_{t-1} I(t) + \beta_2 x_{t-1} (1 - I(t)) + u_{it}$$

$$\sigma_{u_{it}}^2 = \sigma_B^2 I(t) + \sigma_A^2 (1 - I(t)).$$

Conditional on the break point τ , this regression is estimated by maximum likelihood. The break point is estimated by grid search, using $\pi = 0.20$. The point that gives the maximum value of the maximum likelihood for the spread is at 1981:1 and for the output at 1980:4. Although each equation was estimated separately, the break point is essentially the same, occurring in the period whereas a new monetary policy regime created strong interest rate volatility (Watson, 1999). Therefore, we suggest a structural break model with changes in the mean and the variance equations as good data representation (SBMV, described in the Appendix).

3 Structural Break Threshold Models

Although non-linear models can capture some characteristics of structural break models (Clements and Smith, 1999; Koop and Potter, 2000; Koop and Potter, 2001), it may be the case that the break also implies changes in the non-linear parameters. Univariate time-varying smooth transition models have been proposed by Lundbergh et al. (2000) and these have been applied to capture changes in seasonality in industrial production by Van Dijk et al. (2001). In contrast to time-varying smooth transition models, the model proposed in this section characterizes a discrete change in the parameters, including the transition function, that is, the threshold, the delay, the transition variable and the autoregressive order may change at a point in time.

Non-linearities and structural breaks in a bivariate system to predict output growth y_t using the spread S_t can be written as:

$$y_t = x_{t-1}\beta_1[(1 - I_1(s_{t-d_1}))(1 - I_1(t))] + x_{t-1}\beta_2[I_1(s_{t-d_1})(1 - I_1(t))] \\ + x_{t-1}\beta_3[(1 - I_2(s_{t-d_2}))I_1(t)] + x_{t-1}\beta_4[I_2(s_{t-d_2})I_1(t)] + u_{1t}$$

$$S_t = x_{t-1}\beta_5[(1 - I_3(s_{t-d_3}))(1 - I_2(t))] + x_{t-1}\beta_6[I_3(s_{t-d_3})(1 - I_2(t))] \\ + x_{t-1}\beta_7[(1 - I_4(s_{t-d_4}))I_2(t)] + x_{t-1}\beta_8[I_4(s_{t-d_4})I_2(t)] + u_{2t},$$

where $I_i(s_{t-d_i})$ is a transition (indicator) function which is equal to 1 when $s_{t-d_i} > r_i$ and equal to 0 when $s_{t-d_i} \leq r_i$; $I_i(t)$ is also a transition (indicator) function but it is equal to 1 when $t > \tau_i$. As before when $cov(u_{1t}, u_{2t}) = 0$, each equation is considered as a regression.

Three different specifications of thresholds models with a structural break can be estimated. The first one assumes the threshold and the delay are known and a change-point is estimated. Let $S(\cdot)$ be the sum of square residuals of one of the threshold regressions, the change point is estimated by minimizing this criterion function:

$$\hat{\tau} = \min_{\tau_L \leq \tau \leq \tau_U} S(\tau, \hat{r}, \hat{d}).$$

In this case, the possibility of parameter instability in threshold models is tested using Hansen (2000b). Specifically, one equation (z_{it} could be output growth or spread) can be represented as the following regression under the null hypothesis:

$$z_{it} = \beta_1 x_{t-1}(1 - I(s_{t-d})) + \beta_2 x_{t-1}(I(s_{t-d})) + u_{it},$$

where $I(s_{t-d}) = 0$ when $s_{t-d} \leq r$ and $I(s_{t-d}) = 1$ when $s_{t-d} > r$. The model under the alternative hypothesis is:

$$\begin{aligned} z_{it} = & [\beta_1 x_{t-1}(1 - I(s_{t-d})) + \beta_2 x_{t-1}(I(s_{t-d}))](1 - I(t)) \\ & + [\beta_3 x_{t-1}(1 - I(s_{t-d})) + \beta_4 x_{t-1}(I(s_{t-d}))](I(t)) + \tilde{u}_{it}, \end{aligned} \quad (3)$$

where $I(t) = 0$ when $t \leq \tau$ and $I(t) = 1$ when $t > \tau$. Grid search is employed to estimate τ , given that s_{t-d} and r in $I(s_{t-d})$ are known. These values are employed to calculate a *supF* statistics with p-values calculated using the asymptotic distribution and fixed regressor bootstrap (homoscedastic and heteroscedastic)⁶. Using the residuals of each threshold regression, we also test for a structural break in the variance. Equivalently, in the case of a vector autoregressive system, the criterion function is the log of the determinant of the variance-covariance matrix of the residuals and *supLR* tests for structural breaks can be calculated as described in the last section. The evidence of structural break does not change by the inclusion of non-linearity, except for the structural break in the variance of the spread equation that is weaker, as the statistics in panels 4 and 5 of Table 1 indicate.

A second specification is to estimate r , d and τ jointly, assuming the same threshold and delay for each sub-sample determined by the change point:

$$\hat{\tau}, \hat{r}, \hat{d} = \min_{\substack{r_L \leq r \leq r_U \\ d_U \leq d \leq d_L \\ \tau_L \leq \tau \leq \tau_U}} S(r, d, \tau).$$

⁶For the spread equation and the TVAR, the sequential algorithm employed to calculate the p-values based on bootstrap does not work due to collinearity. Then, we report only asymptotic p-values.

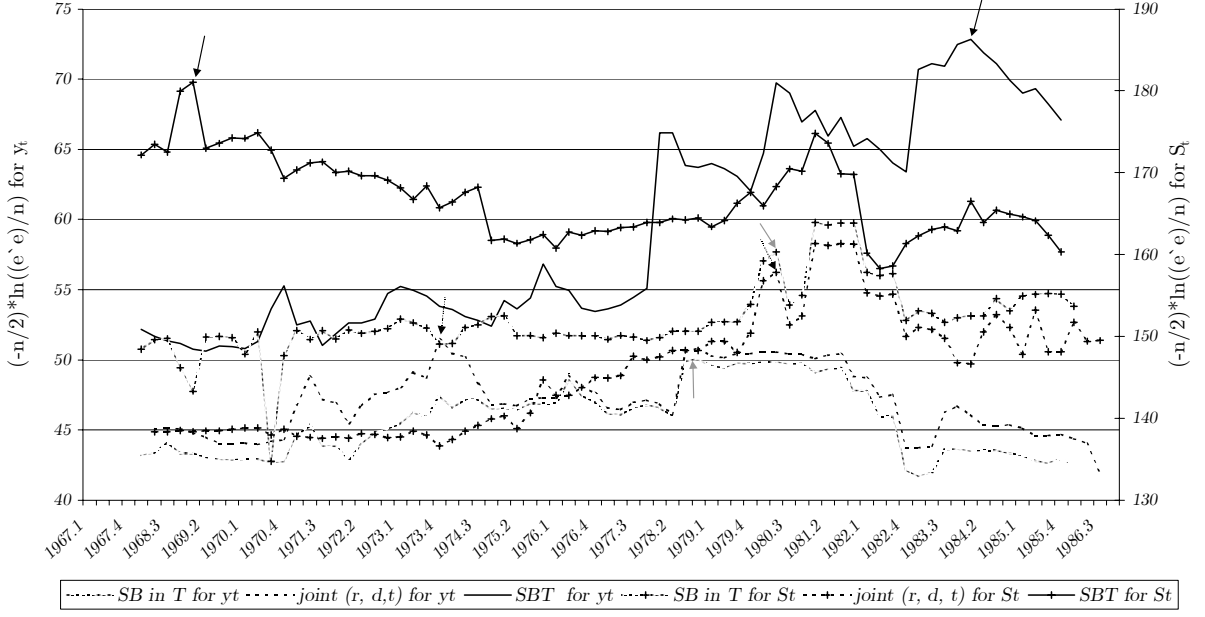


Figure 2 Comparing break-point estimation for different specifications of threshold structural break models of each equation.

Finally, the third specification relaxes the assumption that the transition function is the same before and after the changing-point by using a grid search to estimate the following model:

$$z_{it} = [\beta_1 x_{t-1}(1 - I_1(s_{t-d_1})) + \beta_2 x_{t-1}(I_1(s_{t-d_1}))](1 - I(t)) \\ + [\beta_3 x_{t-1}(1 - I_2(s_{t-d_2})) + \beta_4 x_{t-1}(I_2(s_{t-d_2}))](I(t)) + u_{it},$$

where $I_1(s_{t-d_1}) = 0$ when $s_{t-d_1} \leq r_1$ and $I_1(s_{t-d_1}) = 1$ when $s_{t-d_1} > r_1$; and $I_2(s_{t-d_2}) = 0$ when $s_{t-d_2} \leq r_2$ and $I_2(s_{t-d_2}) = 1$ when $s_{t-d_2} > r_2$. Conditional on the break-point τ , one threshold model is estimated in each sub-sample. The squares of the residuals in each sub-sample are summed and $\hat{\tau}$ is the value that minimizes this criterion. Specifically,

$$\hat{\tau}, \hat{r}_1, \hat{d}_1, \hat{r}_2, \hat{d}_2 = \min_{\tau_L \leq \tau \leq \tau_U} \left(\min_{\substack{r_{1L} \leq r_1 \leq r_{1U} \\ d_{1L} \leq d_1 \leq d_{1U}}} S_1(\tau, r_1, d_1) + \min_{\substack{r_{2L} \leq r_2 \leq r_{2U} \\ d_{2L} \leq d_2 \leq d_{2U}}} S_2(\tau, r_2, d_2) \right), \quad (4)$$

where S_1 is the sum of squared residuals of the threshold model estimated using a grid search over the delay and the threshold value for the first sub-sample and S_2 is the same for the second sub-sample. When the residuals of each equation are assumed to be contemporaneously correlated, these specifications can be estimated by grid search using the covariance of the residuals as criterion function.

The last two approaches cannot be formally tested using the available asymptotic theory. However, we can observe the sharpness of each type of model to estimate the change-point by

verifying the likelihood for each possible change-point assuming that $\pi = .3$, which is presented in Figure 2. All the specifications have the same number of autoregressive coefficients, while the SBT has additional transition function. The results show: (1) the break-point in the output equation depend on the specification, while in the case of the spread the break-point only changes when changing non-linearity is allowed; (2) the break-points estimated by the SBT regressions are sharper estimated than the other specifications, specially for the output equation in which the evidence of parameter instability is stronger. Therefore, we suggest a structural break threshold model (SBT) and a structural break threshold VAR (SBTVAR), which are described in the Appendix, as a good data representation.

Given the evidence of a structural break in the variance even when a threshold regression describes the mean equation, we analyze the ability of regime-dependent variances of structural break threshold models to describe changing variances over time in the output growth equation. In the case of the SBT model, for example, the standard deviations of the errors of the first sub-sample are twice the ones of the second sub-sample and the break point is 1983:4. As a result, the SBT model accounts for a structural break in the variance similar to the described in the univariate models of Kim and Nelson (1999) and McConnell and Perez-Quiros (2000).

The dynamics of the SBTVAR can explain the result of Stock and Watson (2001) that the spread does help to predict output growth in the period 1984-1999. The assumption of linearity does not allow to observe that the spread only helps to predict output growth when is $S_{t-5} \leq 1$. Therefore, for 77% of observations of the period 1981-1999, the inference that the spread lost its power to predict output growth is true, but not for the 23% of the observations in which the inequality $S_{t-5} \leq 1$ is valid. The dynamics of the structural break threshold model also explain why there is no instability in predicting recessions using spread but the relation output growth-spread is unstable (Estrella et al., 2000): the spread is a good predictor when it is a small or negative value, implying small or negative growth (recession), but not when the value is high, meaning that large spreads do not predict strong output growth. The latter dynamics is stronger in the more recent period (after 1981).

4 Evaluation of Probability Forecast

4.1 Extracting information on event probability

The majority of the models suggested by literature to predict the probability of recession consider recessions as unobserved components or a binary variable, implying that the model is a filter to extract recession probabilities from the data (see, e.g., the comparisons of Filardo, 1999 and Camacho and Perez-Quiros, 2000). However, the concept that predicting recession is nothing more than predicting a specific event implies that event probabilities are derived from density forecasts. Stock and Watson (1989) show how to obtain the probabilities of recession from a dynamic factor model of leading indicators using stochastic simulation. Stochastic simulation

Table 2 Comparing definitions of recession.

NBER recessions	Event A	Event B
1957:3-1958:1	1957:1-1957:4	1955:3-1957:4
1960:2-1960:4	—	1959:2-1960:2
1970:1-1970:4	1969:1-1969:4	1969:1-1970:1
1973:4-1975:1	1973:4-1974:4	1973:1-1974:1
1980:1-1980:2	1979:3-1980:2	1979:3-1982:1
1981:3-1982:4	1981:1-1981:4	—
1990:3-1991:1	1989:4-1990:4	1989:4-1990:4

Note: Event A: at least two consecutive quarters of negative growth in real GDP over next five quarters;
Event B: at least two quarters of negative growth in real GDP over next five quarters;

has been used to extract the probability of an event from macroeconomic models (Fair, 1993; Garratt et al., 2000) and non-linear models (Anderson and Vahid, 2000). Specifically in the case of non-linear models, the evaluation of probabilities or density forecasts does not add any computation burden over conditional expectations because non-linear models required these to be generated by stochastic simulation (Monte Carlo or Bootstrap, see, e.g., Granger and Teräsvirta (1993)). The analysis of the ability of a model to predict an event can be used for both out-of-sample (Garratt et al., 2000) or in-sample (Fair, 1993; Anderson and Vahid, 2000) evaluation.

In this work, the models estimated in the last sections will be evaluated according to their ability to predict in sample the probability of two events:

A. “At least two consecutive quarters of negative growth in real GDP over the next five quarters”;

B. “At least two quarters of negative growth in real GDP over the next five quarters”(Anderson and Vahid, 2000, p. 6).

The first event is a popular simple rule to define recessions without employing NBER turning points. It does not need to coincide with the NBER recessions for three reasons: the rule is based on a single series at a quarterly frequency, no censoring is applied, and the event is forward-looking. Because event A leads NBER recessions, models that predict event A with good performance can be employed to calculate leading recession probabilities indexes for the US economy, such as the experimental leading recession index (XRI) of Stock and Watson (1989). Observing Table 2, which includes NBER recessions and Event A, the forward-looking characteristic of the rule compared with the NBER turning points is evident. The rule to obtain event A fails to account for the recession in 1960, but the rule employed in event B can capture the 1960 recession. The evident drawback of the rule that calculates event B is that it does not differentiate the two recessions at the beginning of 1980’s. Therefore, these characteristics support the application of both events to evaluate the models.

The procedure to extract the probabilities of event A and B from the models is the same

as the one described by Anderson and Vahid (2000). Define \mathbf{x}_t as the vector of endogenous variables $(y_t, S_t)'$, $\mathbf{X}^{t-1} = \{\mathbf{x}_{t-1}, \mathbf{x}_{t-2}, \dots, \mathbf{x}_1\}$ as the history of \mathbf{x}_t and $\mathbf{x}_t = \mathbf{f}(\mathbf{X}^{t-1}; \boldsymbol{\beta}) + \boldsymbol{\epsilon}_t$ as the forecasting model where $\boldsymbol{\beta}$ is the matrix of parameters and $\boldsymbol{\epsilon}_t$ are iid with $Var(\boldsymbol{\epsilon}_t) = \Omega$. $\boldsymbol{\epsilon}_t$ is assumed to have a multivariate normal distribution. Given $\hat{\boldsymbol{\beta}}$ and $\hat{\Omega}$, the trial sequence of forecasts for $\{\mathbf{x}_t, \mathbf{x}_{t+1}, \mathbf{x}_{t+2}, \mathbf{x}_{t+3}, \mathbf{x}_{t+4}\}$ conditional on \mathbf{X}^{t-1} is built as follows. A random vector $\boldsymbol{\epsilon}_t$ is drawn from the distribution $\boldsymbol{\epsilon} \sim \mathbf{N}(\mathbf{0}, \hat{\Omega})$ and it is used to calculate $\hat{\mathbf{x}}_t$, given \mathbf{X}^{t-1} and $\hat{\boldsymbol{\beta}}$. $\hat{\mathbf{x}}_t$ is added to “history” to form $\hat{\mathbf{X}}^t$. Then a new draw ($\boldsymbol{\epsilon}_{t+1}$) is made from $N(0, \hat{\Omega})$ and it is employed to calculate $\hat{\mathbf{x}}_{t+1}$, given $\hat{\mathbf{X}}^t$ and $\hat{\boldsymbol{\beta}}$ and to form $\hat{\mathbf{X}}^{t+1}$. This procedure is continued until the sequence of forecasts is complete $\{\hat{\mathbf{x}}_t, \hat{\mathbf{x}}_{t+1}, \hat{\mathbf{x}}_{t+2}, \hat{\mathbf{x}}_{t+3}, \hat{\mathbf{x}}_{t+4}\}$. This sequence of forecasts can be called S_1 , and the same trial is repeated to obtain a set of 2000 forecast sequences. The probability of event A (B) is the proportion of these 2000 sequences in which the event A (B) occurs (P_t). Given the effective sample size T (1954:3-1999:4), a series of event probabilities P_t for $t = 1, \dots, T$ can be obtained.

In the case of threshold models, the forecasting model can be also written as $\mathbf{x}_t^j = \mathbf{f}^j(\mathbf{X}^{t-1}; \boldsymbol{\beta}^j) + \boldsymbol{\epsilon}_t^j$, where $j = 1, 2$ for models with two regimes and $j = 1, 2, 3, 4$ for structural threshold models. Therefore, $Var(\boldsymbol{\epsilon}_t^j)$ depends on the regime (defined by the threshold and the transition variable), so for each regime with different number of observations n^j ($T = \sum_{i=1}^s n_i$), there is a different Ω^j and $\boldsymbol{\epsilon}_t^j$ is supposed to be multivariate normal with variance $\hat{\Omega}^j$. In this framework, for each step to obtain the forecast sequences ($h = 0, \dots, 4$) for, say, a two-regime threshold model, either vector $\boldsymbol{\epsilon}_{t+h}^1$ is drawn from $\boldsymbol{\epsilon}^1 \sim N(0, \hat{\Omega}^1)$ or vector $\boldsymbol{\epsilon}_{t+h}^2$ is drawn from $\boldsymbol{\epsilon}^2 \sim N(0, \hat{\Omega}^2)$ depending on $\hat{s}_{T+h-1-d} < r$ or $\hat{s}_{T+h-1-d} > r$. The vector $\boldsymbol{\epsilon}_{t+h}^j$ is employed to compute $\hat{\mathbf{x}}_{t+h}$ that includes the transition variable that defines the regimes $\hat{s}_{T+h-1-d}$.

4.2 Accuracy Measures for Probability Forecasts

The probabilities of event A and B for each point in-sample for 1954:3 to 1999:4 are presented in Figures 3 and 4. To evaluate these probabilities, we employ the quadratic probability score (QPS), the log probability score (LPS) (Diebold and Rudebusch, 1989) and the Kuipers Score (Granger and Pesaran, 2000). The first one ranges from 0 to 2, with 0 being perfect accuracy. The second one ranges from 0 to ∞ . LPS and QPS imply different loss functions with large mistakes more heavily penalized under LPS. Let P_t be the prediction probability of the event A or B by the model for the next five periods starting at t and R_t is binary variable that is equal to 1 if the event occurs in the actual data and equal to 0 otherwise, then the Briers score (QPS) and the logarithm score (LPS) are written as:

$$QPS = \frac{1}{T} \sum_{t=1}^T 2(P_t - R_t)^2 \text{ and} \quad (5)$$

Table 3 Scores for the Prediction of Event Probabilities.

	1954:3-1999:4						1983:1-1999:4					
	Event A			Event B			Event A			Event B		
	QPS	LPS	KS	QPS	LPS	KS	QPS	LPS	KS	QPS	LPS	KS
VAR	0.178	0.310	0.078	0.268	0.426	0.318	0.117	0.225	0.000	0.135	0.262	0.000
T	0.152	0.274	0.430	0.240	0.386	0.333	0.118	0.222	0.000	0.123	0.242	0.000
TVAR	0.154	0.281	0.513	0.258	0.412	0.434	0.125	0.236	0.000	0.135	0.265	0.000
SBMV	0.184	0.317	0.352	0.272	0.430	0.504	0.123	0.236	0.000	0.157	0.297	0.034
SBT	0.149	0.243	0.419	0.230	0.347	0.530	0.112	0.151	0.494	0.111	0.156	0.494
SBTVAR	0.146	0.245	0.405	0.239	0.361	0.434	0.094	0.144	0.016	0.094	0.152	0.016

Note: The scores are calculated based on the probabilities of event A or B for each time in the period indicated. For each t, the event probabilities are calculated using 5-step-ahead forecasts generated assuming that the coefficients are known and equal to the one estimated for the full sample, but with data available until t-1. The forecasts are computed using Monte Carlo and regime-dependent variances. For LPS, see eq. 6; for QPS, see eq. 5; for KS, see eq. 7. The estimates of the models are in the Appendix.

$$LPS = -\frac{1}{T} \sum_{t=1}^T [(1 - R_t) \ln(1 - P_t) + R_t \ln(P_t)]. \quad (6)$$

The Kuipers score is based on the definition of two states as two different indications given by the model: the economy will be in recession or the economy will be in expansion. Suppose that the recession is imminent when the predicted probability is larger than 1/2. So one can calculate event forecasts (E_t): $E_t = 1$ when $P_t > 1/2$ and $E_t = 0$ when $P_t \leq 1/2$. Comparing these events forecasts with the actual outcomes (R_t), the following contingency matrix can be written:

		Actual Outcomes	
		recession ($R_t = 1$)	expansion ($R_t = 0$)
forecasts	recession ($E_t = 1$)	Hits	False Alarms
	expansion ($E_t = 0$)	Misses	Correct rejections

The Kuipers score is defined as the difference between the proportion of recessions that were correctly forecasted ($H = \frac{\text{hits}}{\text{hits} + \text{misses}}$) and the proportion of expansions that were incorrectly forecasted ($FA = \frac{\text{false alarms}}{\text{false alarms} + \text{correct rejections}}$):

$$KS = H - FA. \quad (7)$$

Large Kuipers scores mean that the model generates proportionally more hits than false alarms.

4.3 Evaluating Predictions

The scores are calculated for the whole in-sample period and also for the 1983-1999 sub-sample and are presented in Table 3. The inclusion of non-linearities improves the overall calibration of the probability forecasts for both events and dramatically improves the Kuipers score for the event A in the full sample. The inclusion discrete parameter changes (SBMV) also makes

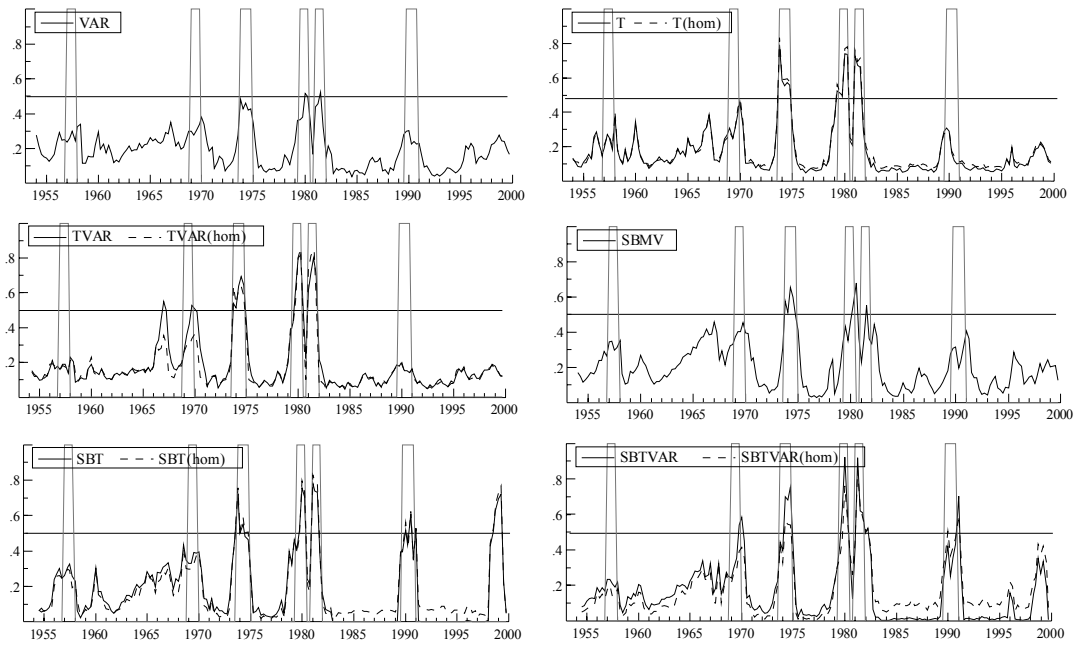


Figure 3 Predictive Probabilities for Event A.

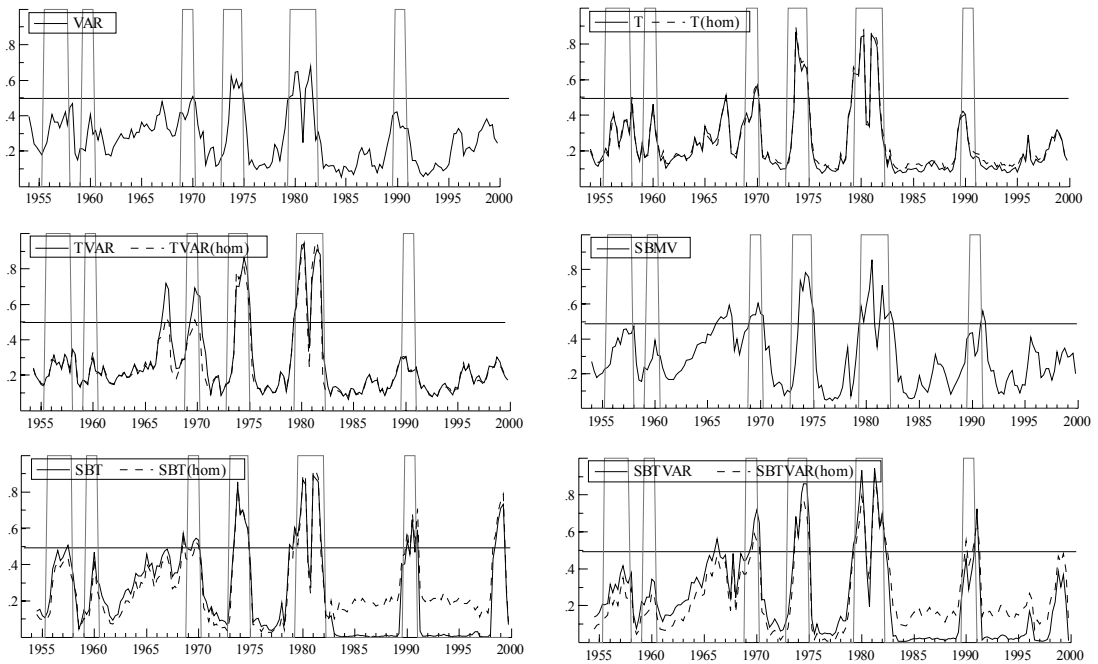


Figure 4 Predictive Probabilities for Event B.

better the predictions of the system but they are inferior to the ones derived from the threshold models. The models with a structural break and non-linearities have the best fit of the predictive probabilities for both events. None of the models can account for the events in the beginning of the sample, which is a potential drawback of the models evaluated.

Figures 3 and 4 and Table 3 show that the recession of 1990/91 can be only predicted when structural breaks and non-linearities are both present in the model. An important characteristic of the structural break threshold models that may have helped to predict the event A is that they allow for regime-dependent variance. The plots of the predicted probabilities of the SBT and the SBTVAR models (Figures 3 and 4) are generated with regime-dependent variances and without (hom). The presence of this type of heteroscedasticity improves significantly the probability fit after 1983.

However, employing the SBT may not be done without a cost: the model generates a strong false alarm in the end of 1999. This occurs because the spread during 1998 was smaller than 0.8 that is the threshold for the recession regime of the second sub-sample, which has a large negative constant and large impact of the spread on the output growth. This effect is milder in the SBTVAR because this model has coefficients with smaller size in similar regime although the threshold (1) is larger. This false alarm is derived from a monetary tightening by the Fed to control inflation at the beginning of 1998 that was followed by a financial crisis generated by the Russian default in August 1998. As described by Marshall (2001), the crisis was characterized by rapid increases in uncertainty, which implies decreasing values of the spread. The inclusion of this information in the SBT model implies that the predictions are based on the recession regime because that is the type of regularity found previously in the historical data and incorporated into the model. However, the Fed policy action of cutting interest rates, switching the policy from controlling inflation to fighting against an imminent financial crisis, was enough to calm the markets after a couple of months and to lead the US economy to the longest period of expansion in history (Marshall, 2001). Thus, the recession predicted for 1999 by model SBT did not materialize, because of a credible action by the Fed.

Allowing for different non-linear behaviors for the periods before and after the break, the structural break threshold models have scored better after 1983 (Table 3). The fact that the power of the spread to predict the events is asymmetric over the cycle does not override the characteristic that the standard deviations of the shocks after the break (around 1981) are half that of the previous period.

5 Predicting the Probability of Recession for 2001

The values of the spread in the last quarter of 2000 are negative, suggesting a future recession, when one take into account the regularity described by Estrella and Hardouvelis (1991). In this section we evaluate how this information is translated into probability of recessions by some of the models evaluated in the last section compared with other sources of predictions for recession

Table 4 Comparing 2001 probability forecasts (per cent).

		XRI	SPF	T	SBT	SBTVAR
Pr(A _{2001:1})				56	55	81
Pr(B _{2001:1})				67	56	85
Pr($y_{2001:1} < 0$ and $y_{2001:2} < 0$)		7		34	0	7
Risk of negative growth	2001:1		37	50	1	7
	2001:2		32	59	1.5	75
	2001:3		23	47	47	86
	2001:4		18	34	56	73
	2002:1		13	24	59	72

Notes: ST, T, SBT and SBTVAR are described in the Appendix; SPF is the Survey of Professional Forecasters, with information until Jan 2001; XRI is the Stock and Watson experimental leading recession index, with information until Dec 2001

probabilities presented in the literature. Specifically, probability forecasts are generated from the T, SBT and SBTVAR models.

5.1 Forecasts of the probability of events A and B for 2001

The probability that two quarters of negative growth will occur in the next five quarters (event A) is 56% using the T model and 81% using the SBTVAR, as presented in Table 4. The different probabilities among models depend on how the negative spreads of the last quarters of 2000 affect the non-linearity of the model. The SBT model has the strongest negative growth reaction which generates a false alarm in 1999. The average probability in predicting event A is 62%, which means that is quite likely that a recession will happen by 2002:1, while the average probability of event B is slightly higher (71%), as expected.

5.2 Comparing with Stock and Watson Recession Index

Table 4 also shows the probabilities of negative growth in the first two quarters of 2001. These probabilities are calculated to compare with the (monthly) Stock and Watson XRI index. The Stock and Watson experimental leading recession index (XRI)⁷ extracts the probability of recession using eight components (after the 1997 revision) in a dynamic factor model (Stock and Watson, 1989, 1993). Using different series such as interest rates and manufacturers' unfilled orders, the authors try to explore comovement between economic variables and to detect recession when a downturn is signaled by different sectors of the economy. A monthly period is said to be in recession if that month is either in a sequence of six consecutive declines of the composite index below some boundary or in a sequence of nine declines below the boundary with no more than one increase during the middle seven months (Stock and Watson, 1989,

⁷Definition of the index and historical values are at <http://ksghome.harvard.edu/~JStock.Academic.Ksg/xri/INDEX.HTM>.

p. 357). This recession definition is employed to identify recession periods in the observed data and also to calculate the leading recession index. To compare with the probabilities extracted from the models presented in this work, we suppose that this recession definition is equivalent to obtaining two negative output growth predictions in a two-quarter horizon. Given that the last calculation of the experimental leading index was still above the boundary of the recession period, the second definition of recession can be neglected.

The experimental leading recession index calculated with information until 2000:12 is presented in Table 4. The probability that a recession would happen in the first two quarters of 2001 is larger (11 and 34%) than the Stock and Watson XRI (7%) when it is not assumed a structural break in the model. Because the effect of the negative spread is more delayed in the SBT, this model is still predicting growth, while the threshold model indicates a probability of 34% for this definition of recession. This is an evidence that the presence of structural breaks in non-linear models affects the timing of the predictions of recession.

5.3 Comparing with the Survey of Professional Forecasters

The Survey of Professional Forecasters (SPF) collects information on forecasts of economic series, such as output, unemployment, interest rates and inflation, made by private sector forecasters (34 of them), organized by the Federal Reserve Bank of Philadelphia (Croushore, 1993). The information is collected every three months and the results are published on <http://www.phil.frb.org/econ/spf>. The SPF publishes the mean (over forecasters) of the estimated probability of negative output growth for each one of the next 5 quarters. The mean of the risk of negative growth for the period 2001:1 to 2002:1, published by the SPF in February 2001, using information until January 2001, is in the lower panel of Table 4. By way of comparison, we compute the probability of negative output growth for the 2001:1-2002:1 period from our models. The models are more optimistic than the SPF predictions for the first quarter of 2001 and more pessimistic for the last two quarters of 2001, when the probability of negative growth is on average greater than 50% for both quarters. The results of the threshold model, however, are an exception because they predict negative growth for 2001:2 but give a small probability for 2002:1, showing differences in the dynamics of the evaluated models.

5.4 Comparing with the Probit Models of Chauvet and Potter (2001)

Chauvet and Potter (2001) show that depending on the supposition on serial correlation of the residuals or variance changes, the predictions for 2001 from a probit model using the spread change. The authors, however, employ a definition of recession that cannot be extracted directly from the autoregressive models described in this work, because it implies predictions for a binary variable. Therefore, we can only compare our results in terms of the timing and strength of a possible recession. As discussed before, the models with a structural break and thresholds predict a recession for the second semester of 2001, whereas the chosen model by Chauvet and

Potter (2001) (model 4) gives a predictive probability of 18%. Both the SBT and the model 4 of Chauvet and Potter (2001) take into account possible parameter instability and use the same leading indicator. Then, the main reason for these dissimilar predictions is the strong serial correlation included in the probit predictions. This serial correlation implies high persistence of the actual state of the binary variable because it takes into account the long expansion that follows the 1990/91 recession. This is seen as an advantage of the probit model because this imply that the model does not give a false alarm in 1999, as the SBT.

6 Conclusions

The presence of structural breaks and non-linearities is needed for the prediction of the 1990-91 recession. Koop and Potter (2000) suggest that non-linearity tests may indicate threshold models when the data is generated from a structural break model. Our results suggest that non-linearity and structural breaks are *both* necessary for a model be able to capture the dynamics between spread and output growth in the period 1955-1999. We show how a structural break threshold model may account for two reported characteristics of the output growth-spread relationship: it is non-linear (Galbraith and Tkacz, 2000; Anderson and Vahid, 2000) and it is unstable over time (Haubrich and Dombrosky, 1996; Stock and Watson, 2001). The structural break threshold model can also account for further feature of the output growth series: its variability has decreased after 1984 (McConnell and Perez-Quiros, 2000).

The predictions of the models with a structural break and threshold non-linearity agree with the Stock and Watson (1989) experimental leading recession index but they are too pessimistic when compared to the predictions of the Survey of Professional Forecasters and Chauvet and Potter (2001). While the threshold model predicts that a recession is more likely to occur in 2001:2-2001:3, the structural break threshold model predicts a recession in 2001:4-2002:1. Therefore, the presence of structural break changes the timing of the recession.

References

- Anderson, H. M. and Vahid, F. (2000). Predicting the probability of a recession with nonlinear autoregressive leading indicator models, *Monash University, Department of Econometrics and Business Statistics Working Paper 3/2000*.
- Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point, *Econometrica* **61**: 821–56.
- Andrews, D. W. K. and Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative, *Econometrica* **62**: 1393–1414.
- Bai, J. (1999). Likelihood ratio tests for multiple structural break, *Journal of Econometrics* **91**: 299–323.

- Berk, J. M. (1998). The information content of the yield curve for monetary policy: A survey, *De Economist* **146**: 303–20.
- Camacho, M. and Perez-Quiros, G. (2000). That is what the US leading indicators lead, *European Central Bank, Working Paper* **27**.
- Chauvet, M. and Potter, S. M. (2001). Forecasting recessions using the yield curve, *University of California at Riverside, Department of Economic, Working Paper* **01-17**.
- Clements, M. P. and Galvão, A. B. C. (2001). A comparison of tests of non-linear cointegration with an application to the predictability of the US term structure of interest rates, *University of Warwick (mimeo)* .
(www.warwick.ac.uk/fac/soc/Economics/clements/asyecmvp.pdf).
- Clements, M. P. and Smith, J. (1999). Testing self-exciting threshold autoregressive models against structural change models, *University of Warwick (mimeo)* .
- Croushore, D. (1993). Introducing: The Survey of Professional Forecasters, *Federal Reserve Bank of Philadelphia, Business Review* **nov./dec.**: 3–13.
- Diebold, F. X. and Rudebusch, G. D. (1989). Scoring the leading indicators, *Journal of Business* **62**: 369–91.
- Don't Mention the R-Word: Are Economic Forecasters Wishful Thinkers or Wimps?* (2001).
The Economist, printed edition **March 1**.
- Dotsey, M. (1998). The predictive content of the interest rate term spread for future economic growth, *Federal Reserve of Richmond, Economic Quarterly* **84**: 31–51.
- Dueker, M. J. (1997). Strengthening the case for the yield curve as a predictor of US recessions, *Federal Reserve Bank of St. Louis Review* **mar./apr.**: 41–51.
- Estrella, A. and Hardouvelis, G. A. (1991). The term structure as a predictor of real economic activity, *Journal of Finance* **46**: 555–76.
- Estrella, A. and Mishkin, F. S. (1998). Predicting US recessions: Financial variables as leading indicators, *Review of Economics and Statistics* **80**: 45–61.
- Estrella, A., Rodrigues, A. P. and Schich, S. (2000). How stable is the predictive power of the yield curve? Evidence from Germany and the United States, *Federal Reserve of New York, Staff Report* **113**.
- Fair, R. C. (1993). Estimating event probabilities from macroeconomic models using stochastic simulation, in J. H. Stock and M. W. Watson (eds), *Business Cycles, Indicators, and Forecasting*, NBER, University of Chicago Press, Chicago, pp. 157–78.
- Filardo, A. J. (1999). How reliable are recession prediction models?, *Federal Reserve Bank of Kansas City, Economic Review* **second quarter**: 35–55.
- Galbraith, J. W. and Tkacz, G. (2000). Testing for asymmetry in the link between the yield spread and output in the G-7 countries, *Journal of International Money and Finance* **19**: 657–672.
- Garratt, A., Lee, K., Pesaran, M. H. and Shin, Y. (2000). Forecast uncertainties in macroeconomic modelling: An application to the UK economy, *Trinity College, Cambridge*

(*mimeo*) .

- Granger, C. W. J. and Pesaran, M. H. (2000). Economic and statistical measures of forecast accuracy, *Journal of Forecasting* **19**: 537–60.
- Granger, C. W. J. and Teräsvirta, T. (1993). *Modelling Nonlinear Economic Relationships*, Oxford University Press, Oxford.
- Hamilton, J. D. and Kim, D. H. (2000). A re-examination of the predictability of economic activity using the yield spread, *NBER Working Papers* **7954**.
- Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis, *Econometrica* **64**: 413–30.
- Hansen, B. E. (1997). Approximate asymptotic p-values for structural-change tests, *Journal of Business and Economic Statistics* **15**: 60–67.
- Hansen, B. E. (2000a). Testing for linearity, in D. A. R. George, L. Oxley and S. Potter (eds), *Surveys in Economic Dynamics*, Blackwell, Oxford, pp. 47–72.
- Hansen, B. E. (2000b). Testing for structural change in conditional models, *Journal of Econometrics* **97**: 93–115.
- Hansen, B. E. (2001). The new econometrics of structural change: Understanding and dating changes in US productivity, *University of Wisconsin (mimeo)* .
- Haubrich, J. G. and Dombrosky, A. M. (1996). Predicting real growth using the yield curve, *Federal Reserve Bank of Cleveland, Economic Review* **32**: 26–34.
- Kim, C.-J. and Nelson, C. R. (1999). Has the US economy become more stable? A Bayesian approach based on a Markov-switching model of the business cycle, *Review of Economics and Statistics* **81**: 608–16.
- Koop, G. and Potter, S. M. (2000). Nonlinearity, structural breaks, or outliers in economic time series?, in W. A. Barnett, D. F. Hendry, S. Hylleberg, T. Teräsvirta, D. Tjostheim and A. Würtz (eds), *Nonlinear Econometric Modeling in Time Series*, Cambridge University Press, Cambridge, pp. 61–78.
- Koop, G. and Potter, S. M. (2001). Are apparent findings of nonlinearity due to structural instability in economic time series?, *Econometrics Journal* **4**: 37–55.
- Lahiri, K. and Wang, J. G. (1996). Interest rate spreads as predictors of business cycles, in G. S. Maddala and C. R. Rao (eds), *Handbook of Statistics, Vol. 14*, Elsevier, Amsterdam, pp. 297–315.
- Lundbergh, S., Teräsvirta, T. and Van Dijk, D. (2000). Time-varying smooth transition autoregressive models, *SSE/EFI Working Papers in Economics and Finance* **376**.
- Marshall, D. A. (2001). The crisis of 1998 and the role of the Central Bank, *Federal Reserve Bank of Chicago, Economic Perspectives* **First Quarter**: 2–23.
- McConnell, M. M. and Perez-Quiros, G. (2000). Output fluctuations in the United States: What has changed since early 1980s?, *American Economic Review* **90**: 1464–76.
- Sensier, M. and Van Dijk, D. (2001). Short-term volatility versus long-term growth: Evidence in US macroeconomic time series, *Centre for Growth and Business Cycle Research, University*

of Manchester, Discussion Paper 8.

Stock, J. H. and Watson, M. W. (1989). New indexes of coincident and leading economic indicators, *NBER Macroeconomics Annual* 4: 351–94.

Stock, J. H. and Watson, M. W. (1993). A procedure for predicting recessions with leading indicators: Econometric issues and recent experience, in J. H. Stock and M. W. Watson (eds), *Business Cycles, Indicators, and Forecasting*, NBER, University of Chicago Press, Chicago, pp. 95–156.

Stock, J. H. and Watson, M. W. (2001). Forecasting output and inflation: The role of asset prices, *NBER Working Paper* 8180.

Van Dijk, D., Strikholm, B. and Teräsvirta, T. (2001). The effects of institutional and technological change and business cycle fluctuations on seasonal patterns in quarterly industrial production series, *SSE/EFI Working Paper Series in Economics and Finance* 429.

Watson, M. W. (1999). Explaining the increased variability in long-term interest rates, *Federal Reserve Bank of Richmond, Economic Quarterly* 85: 71–96.

A Appendix: Description of Models

The effective sample employed to estimated the following models is from 1954:3 to 1999:4. y_t is the output growth and S_t is the spread between a 10-year T-bond and a 3-month T-Bill. SSE is the sum of squared errors and σ_{u_i} is the standard error of the residuals u_{it} .

1. *Vector Autoregressive model* (VAR):

$$y_t = 0.305 + 0.245y_{t-1} + 0.081y_{t-2} - 0.013S_{t-1} + 0.322S_{t-2} - 0.108S_{t-3}$$

[0.13]
[0.08]
[0.08]
[0.12]
[0.15]
[0.13]

$$SSE_1 = 137.23, \sigma_{u_1} = 0.88$$

$$S_t = 0.280 - 0.076y_{t-1} - 0.103y_{t-2} + 1.034S_{t-1} - 0.351S_{t-2} + 0.215S_{t-3}$$

[0.09]
[0.04]
[0.06]
[0.09]
[0.13]
[0.13]

$$SSE_2 = 49.65, \sigma_{u_2} = 0.53$$

2. *Threshold Model* (T)

$$y_t = (-0.045 + 0.255y_{t-1} + 0.409y_{t-2} - 0.187S_{t-1} + 1.001S_{t-2} - 0.523S_{t-3})(I_1(S_{t-1}))$$

[0.18]
[0.14]
[0.15]
[0.27]
[0.31]
[0.26]

$$+ (0.851 + 0.278y_{t-1} - 0.033y_{t-2} - 0.121S_{t-1} + 0.024S_{t-2} - 0.085S_{t-3})(1 - I_1(S_{t-1}))$$

[0.19]
[0.08]
[0.07]
[0.15]
[0.19]
[0.13]

$$I_1(S_{t-1}) = 1(S_{t-1} \leq 0.667)$$

$$\sigma_{u_1} = 0.916I_1(S_{t-1}) + 0.807(1 - I_1(S_{t-1})) \quad SSE_1 = 121.98$$

$$\begin{aligned}
S_t &= \left(\underset{[0.07]}{-0.411} - \underset{[0.04]}{0.025}y_{t-1} - \underset{[0.05]}{0.211}y_{t-2} + \underset{[0.09]}{1.233}S_{t-1} - \underset{[0.11]}{0.494}S_{t-2} + \underset{[0.08]}{0.103}S_{t-3} \right) (I_1(S_{t-1})) \\
&+ \left(\underset{[0.42]}{-0.522} + \underset{[0.09]}{0.022}y_{t-1} + \underset{[0.08]}{0.048}y_{t-2} + \underset{[0.17]}{0.961}S_{t-1} - \underset{[0.17]}{0.072}S_{t-2} + \underset{[0.12]}{0.233}S_{t-3} \right) (1 - I_1(S_{t-1})) \\
I_1(S_{t-1}) &= 1(S_{t-1} \leq 1.79) \\
\sigma_{u_2} &= 0.445I_1(S_{t-1}) + 0.360(1 - I_1(S_{t-1})) \quad SSE = 39.63
\end{aligned}$$

3. Threshold Vector Autoregressive model (TVAR):

$$\begin{aligned}
y_t &= \left(\underset{[0.23]}{-0.016} - \underset{[0.25]}{0.006}y_{t-1} + \underset{[0.26]}{0.295}y_{t-2} + \underset{[0.24]}{0.032}S_{t-1} + \underset{[0.34]}{0.829}S_{t-2} - \underset{[0.33]}{0.004}S_{t-3} \right) I_1(S_{t-4}) \\
&+ \left(\underset{[0.14]}{0.515} - \underset{[0.07]}{0.253}y_{t-1} + \underset{[0.07]}{0.040}y_{t-2} + \underset{[0.16]}{0.076}S_{t-1} + \underset{[0.21]}{0.127}S_{t-2} - \underset{[0.14]}{0.095}S_{t-3} \right) (1 - I_1(S_{t-4})) \\
\sigma_{u_1} &= 1.313I_1(S_{t-4}) + 0.841(1 - I_1(S_{t-4})) \quad SSE_1 = 125.54 \\
S_t &= \left(\underset{[0.19]}{0.798} - \underset{[0.20]}{0.149}y_{t-1} - \underset{[0.21]}{0.424}y_{t-2} + \underset{[0.20]}{0.882}S_{t-1} - \underset{[0.28]}{1.027}S_{t-2} + \underset{[0.27]}{1.021}S_{t-3} \right) I_1(S_{t-4}) \\
&+ \left(\underset{[0.07]}{0.162} - \underset{[0.04]}{0.081}y_{t-1} - \underset{[0.04]}{0.049}y_{t-2} + \underset{[0.07]}{1.098}S_{t-1} - \underset{[0.07]}{0.186}S_{t-2} + \underset{[0.06]}{0.023}S_{t-3} \right) (1 - I_1(S_{t-4})) \\
I_1(S_{t-4}) &= 1(S_{t-4} \leq 0.1933) \\
\sigma_{u_2} &= 1.076I_1(S_{t-4}) + 0.391(1 - I_1(S_{t-4})) \quad SSE_2 = 37.33 \\
\sigma_{u_1}\sigma_{u_2} &= -0.637(S_{t-4}) + 0.024(1 - I_1(S_{t-4}))
\end{aligned}$$

4. Structural Break in Mean and in Variance model (SBMV)

$$\begin{aligned}
y_t &= \left(\underset{[0.16]}{0.186} + \underset{[0.11]}{0.197}y_{t-1} + \underset{[0.09]}{0.043}y_{t-2} + \underset{[0.20]}{0.497}S_{t-1} - \underset{[0.34]}{0.206}S_{t-2} + \underset{[0.25]}{0.211}S_{t-3} \right) I(t) \\
&+ \left(\underset{[0.18]}{0.192} + \underset{[0.12]}{0.290}y_{t-1} + \underset{[0.12]}{0.247}y_{t-2} - \underset{[0.10]}{0.305}S_{t-1} + \underset{[0.14]}{0.593}S_{t-2} - \underset{[0.14]}{0.189}S_{t-3} \right) (1 - I(t)) \\
I(t) &= 1(t \leq 1980 : 4) \quad \sigma_{u_1} = \underset{[0.13]}{0.950}I(t) + \underset{[0.05]}{0.574}(1 - I(t)) \quad SSE_1 = 122.42 \\
S_t &= \left(\underset{[0.10]}{0.195} - \underset{[0.04]}{0.096}y_{t-1} - \underset{[0.04]}{0.075}y_{t-2} + \underset{[0.09]}{1.097}S_{t-1} - \underset{[0.24]}{0.622}S_{t-2} + \underset{[0.24]}{0.449}S_{t-3} \right) I(t) \\
&+ \left(\underset{[0.17]}{0.553} - \underset{[0.12]}{0.179}y_{t-1} - \underset{[0.15]}{0.314}y_{t-2} + \underset{[0.94]}{0.934}S_{t-1} - \underset{[0.17]}{0.075}S_{t-2} - \underset{[0.12]}{0.078}S_{t-3} \right) (1 - I(t)) \\
I(t) &= (t \leq 1981 : 1) \quad \sigma_{u_2} = \underset{[0.06]}{0.473}I(t) + \underset{[0.04]}{0.501}(1 - I(t)) \quad SSE_2 = 43.16
\end{aligned}$$

5. Structural Break Threshold Model (SBT)

$$\begin{aligned}
y_t &= \left(\underset{[0.15]}{0.088} + \underset{[0.11]}{0.148}y_{t-1} + \underset{[0.11]}{0.218}y_{t-2} + \underset{[0.21]}{0.113}S_{t-1} + \underset{[0.26]}{0.452}S_{t-2} - \underset{[0.21]}{0.076}S_{t-3} \right) (1 - I_1(S_{t-1}))I(t) \\
&\quad + \left(\underset{[0.68]}{2.266} + \underset{[0.14]}{0.057}y_{t-1} - \underset{[0.13]}{0.332}y_{t-2} - \underset{[0.30]}{0.662}S_{t-1} - \underset{[0.29]}{0.003}S_{t-2} + \underset{[0.21]}{0.507}S_{t-3} \right) I_1(S_{t-1})I(t) \\
&+ \left(\underset{[0.90]}{-2.530} + \underset{[0.24]}{1.187}y_{t-1} + \underset{[0.43]}{0.565}y_{t-2} + \underset{[1.03]}{2.684}S_{t-1} - \underset{[1.21]}{1.307}S_{t-2} + \underset{[0.86]}{0.481}S_{t-3} \right) (1 - I_2(S_{t-5}))(1 - I(t)) \\
&\quad + \left(\underset{[0.21]}{0.918} - \underset{[0.14]}{0.135}y_{t-1} + \underset{[0.13]}{0.225}y_{t-2} + \underset{[0.15]}{0.036}S_{t-1} + \underset{[0.23]}{0.047}S_{t-2} - \underset{[0.14]}{0.137}S_{t-3} \right) I_2(S_{t-5})(1 - I(t)) \\
\sigma_{u_1} &= 0.98(1 - I_1(S_{t-1}))I(t) + 0.84I_1(S_{t-1})I(t) + 0.61(1 - I_2(S_{t-5}))(1 - I(t)) + 0.40I_2(S_{t-5})(1 - I(t)) \\
I_1(S_{t-1}) &= 1(S_{t-1} > 1.58) \quad I_2(S_{t-5}) = 1(S_{t-5} > 0.817) \quad I(t) = 1(t \leq 1983 : 4) \quad SSE_1 = 107.67
\end{aligned}$$

$$\begin{aligned}
S_t &= \left(\underset{[0.29]}{0.527} - \underset{[0.16]}{0.333}y_{t-1} + \underset{[0.28]}{0.219}y_{t-2} + \underset{[0.34]}{0.335}S_{t-1} + \underset{[0.50]}{0.047}S_{t-2} - \underset{[0.46]}{0.156}S_{t-3} \right) (1 - I_1(S_{t-4}))I(t) \\
&\quad + \left(\underset{[0.07]}{0.057} + \underset{[0.03]}{0.002}y_{t-1} - \underset{[0.03]}{0.083}y_{t-2} + \underset{[0.11]}{1.260}S_{t-1} - \underset{[0.15]}{0.819}S_{t-2} + \underset{[0.11]}{0.516}S_{t-3} \right) I_1(S_{t-4})I(t) \\
&+ \left(\underset{[0.09]}{0.556} - \underset{[0.09]}{0.076}y_{t-1} - \underset{[0.08]}{0.379}y_{t-2} + \underset{[0.11]}{1.254}S_{t-1} - \underset{[0.14]}{0.506}S_{t-2} + \underset{[0.11]}{0.142}S_{t-3} \right) (1 - I_2(S_{t-1}))(1 - I(t)) \\
&\quad + \left(\underset{[0.44]}{-0.599} + \underset{[0.09]}{0.033}y_{t-1} + \underset{[0.09]}{0.091}y_{t-2} + \underset{[0.17]}{0.968}S_{t-1} - \underset{[0.17]}{0.062}S_{t-2} + \underset{[0.12]}{0.222}S_{t-3} \right) I_2(S_{t-1})(1 - I(t)) \\
\sigma_{u_2} &= 0.62(1 - I_1(S_{t-4}))I(t) + 0.19I_1(S_{t-4})I(t) + 0.52(1 - I_2(S_{t-1}))(1 - I(t)) + 0.56I_2(S_{t-1})(1 - I(t)) \\
I_1(S_{t-4}) &= 1(S_{t-4} > 0.313) \quad I_2(S_{t-1}) = 1(S_{t-1} > 1.79) \quad I(t) = 1(t \leq 1968 : 3) \quad SSE_2 = 34.40
\end{aligned}$$

6. Structural Break Threshold VAR (SBTVAR)

$$\begin{aligned}
y_t &= \left(\underset{[0.18]}{-0.191} - \underset{[0.19]}{0.121}y_{t-1} + \underset{[0.19]}{0.841}y_{t-2} + \underset{[0.23]}{0.123}S_{t-1} + \underset{[0.37]}{1.567}S_{t-2} - \underset{[0.34]}{1.216}S_{t-3} \right) (1 - I_1(S_{t-3}))(1 - I(t)) \\
&\quad + \left(\underset{[0.22]}{0.501} + \underset{[0.11]}{0.144}y_{t-1} - \underset{[0.11]}{0.037}y_{t-2} + \underset{[0.22]}{0.180}S_{t-1} + \underset{[0.31]}{0.159}S_{t-2} + \underset{[0.24]}{0.044}S_{t-3} \right) I_1(S_{t-3})(1 - I(t)) \\
&\quad + \left(\underset{[0.38]}{-0.641} + \underset{[0.25]}{0.791}y_{t-1} + \underset{[0.20]}{0.211}y_{t-2} - \underset{[0.22]}{0.092}S_{t-1} + \underset{[0.28]}{1.063}S_{t-2} - \underset{[0.28]}{0.386}S_{t-3} \right) (1 - I_2(S_{t-5}))I(t) \\
&\quad + \left(\underset{[0.21]}{0.678} + \underset{[0.14]}{0.204}y_{t-1} + \underset{[0.13]}{0.146}y_{t-2} - \underset{[0.16]}{0.054}S_{t-1} + \underset{[0.20]}{0.158}S_{t-2} - \underset{[0.13]}{0.119}S_{t-3} \right) I_2(S_{t-5})I(t) \\
\sigma_{u_1} &= 0.61(1 - I_1(S_{t-3}))I(t) + 1.02I_1(S_{t-3})I(t) + 0.63(1 - I_2(S_{t-5}))(1 - I(t)) + 0.50I_2(S_{t-5})(1 - I(t)) \\
S_t &= \left(\underset{[0.25]}{0.416} + \underset{[0.26]}{0.340}y_{t-1} - \underset{[0.26]}{0.361}y_{t-2} + \underset{[0.31]}{1.604}S_{t-1} - \underset{[0.50]}{2.436}S_{t-2} + \underset{[0.47]}{1.470}S_{t-3} \right) (1 - I_1(S_{t-3}))(1 - I(t)) \\
&\quad + \left(\underset{[0.07]}{0.126} - \underset{[0.04]}{0.097}y_{t-1} - \underset{[0.04]}{0.070}y_{t-2} + \underset{[0.07]}{1.071}S_{t-1} - \underset{[0.10]}{0.248}S_{t-2} + \underset{[0.08]}{0.147}S_{t-3} \right) I_1(S_{t-3})(1 - I(t)) \\
&\quad + \left(\underset{[0.18]}{1.216} - \underset{[0.12]}{0.252}y_{t-1} - \underset{[0.09]}{0.377}y_{t-2} + \underset{[0.11]}{0.386}S_{t-1} - \underset{[0.14]}{0.405}S_{t-2} - \underset{[0.14]}{0.676}S_{t-3} \right) (1 - I_2(S_{t-5}))I(t) \\
&\quad + \left(\underset{[0.18]}{0.295} - \underset{[0.12]}{0.058}y_{t-1} - \underset{[0.10]}{0.014}y_{t-2} + \underset{[0.13]}{1.224}S_{t-1} - \underset{[0.17]}{0.202}S_{t-2} - \underset{[0.11]}{0.143}S_{t-3} \right) I_2(S_{t-5})I(t)
\end{aligned}$$

$$\sigma_{u_2} = 0.83(1 - I_1(S_{t-3}))I(t) + 0.33I_1(S_{t-3})I(t) + 0.31(1 - I_2(S_{t-5}))(1 - I(t)) + 0.42I_2(S_{t-5})(1 - I(t))$$

$$I_1(S_{t-3}) = 1(S_{t-3} > 0.1566) \quad I_2(S_{t-5}) = 1(S_{t-5} > 0.98) \quad I(t) = 1(t < 1981 : 1)$$

$$\sigma_{u_1}\sigma_{u_2} = -0.24(1 - I_1(S_{t-3}))I(t) - 0.02I_1(S_{t-3})I(t)$$

$$-0.04(1 - I_2(S_{t-5}))(1 - I(t)) + 0.07I_2(S_{t-5})(1 - I(t))$$

$$SSE_1 = 110.31 \quad SSE_2 = 28.21$$