Journal of Forecasting
J. Forecast. **23**, 173–196 (2004)
Published online in Wiley InterScience (www.interscience.wiley.com). **DOI**: 10.1002/for.912

Vector Smooth Transition Regression Models for US GDP and the Composite Index of Leading Indicators

MAXIMO CAMACHO*
Universidad de Murcia, Spain

ABSTRACT
In this paper, I extend to a multiple-equation context the linearity, model selection and model adequacy tests recently proposed for univariate smooth transition regression models. Using this result, I examine the nonlinear forecasting power of the Conference Board composite index of leading indicators to predict both output growth and the business-cycle phases of the US economy in real time. Copyright © 2004 John Wiley & Sons, Ltd.

KEY WORDS STAR models; turning points forecasting; leading indicators

INTRODUCTION

Much effort has been devoted to evaluating how well linear and nonlinear models use macroeconomic indexes in forecasting both output growth and business-cycle phases. On the one hand, the linear univariate specifications have followed extensions of the seminal analysis of Box and Jenkins (1976), and the most significant linear multivariate approaches have been the works of Auerbach (1982), Braun and Zarnowitz (1989), and Diebold and Rudebusch (1991). On the other hand, several recent studies have found evidence in favour of forecasting these features with nonlinear alternatives. First, many authors use univariate models for output growth, such as the Markov switching (MS) model of Hamilton (1989), the smooth transition regression (STR) models of Teräsvirta and Anderson (1992) and Teräsvirta (1995), and the threshold autoregressive (TAR) models of Tiao and Tsay (1994), Potter (1995), and Pesaran and Potter (1997). Second, other authors extend these univariate specifications to include economic indicators that may help in computing forecasts, for example, Filardo (1994, 1999), Granger *et al*. (1993), Hamilton and Perez-Quiros (1996), Krolzig (1997, 2000), Estrella and Mishkin (1998), Blix (1999), Warne (2000), Beine *et al*. (2002), and Camacho and Perez-Quiros (2002). Finally, recent developments try to characterize the business-cycle asymmetries by a dynamic factor model with regime switching as in Diebold and Rudebusch (1995), Kim and Nelson (1998), Chauvet (1998, 1999), Fukuda and Onodera (2001), Kim and Murray (2002), and Chauvet and Potter (2002).

In this paper, I develop both theoretical and empirical contributions to the previous literature. With respect to my theoretical contributions, I propose a vector autoregressive extension of the STR model

* Correspondence to: Maximo Camacho, Departmento Metoclos Cuantitativos, Facultad de Economia y Empresa, Universidad de Murcia, 30100 Murcia, Spain. E-mail: mcamacho@um.es

proposed by Granger and Teräsvirta (1993). By analogy, I call it the vector smooth transition regression (VSTR) model. The primary principle for estimation is maximum likelihood, which leads to simple linearity and model selection tests. In line with the univariate proposal of Eitrheim and Teräsvirta (1996), I also extend to the multiple-equation context the tests for examining the adequacy of VSTR models to the data. Finally, I consider the ability of recent model selection techniques in order to formally select one model from the family of VSTR, according to its output growth and business-cycle predictive performance.

With respect to my empirical contributions I find that a logistic-VSTR specification of US real gross domestic product (GDP) and the Conference Board composite index of leading indicators (CLI) is the best real-time forecasting VSTR model of output growth and business-cycle phases during the period 1978.1–2002.2. Note that each of the real-time forecasts is computed with the information that a forecaster would have had available at the time of the forecast, which requires the previous evaluation of two intriguing questions. The first one is related to the long-term relationship between GDP and CLI. Using the series of outputs ending in 1993.3 and 1999.4 and the respective indicator series, Hamilton and Perez-Quiros (1996) and Huh (2002) conclude that these series are cointegrated. However, Granger *et al*. (1993) and Camacho (2000) fail to detect cointegration between the series of outputs ending in 1989.2 and 1997.4 and the corresponding indicator series. In this paper, I find that these puzzling conclusions may be due to the main historical revisions of the CLI series. The second one refers to the 1984.1 structural break in the volatility output growth recently documented by Kim and Nelson (1999) and McConnell and Perez-Quiros (2000). These authors date the breakpoint using series of outputs ending in 1997.1 and 1999.2 respectively, but this information was not available at the time of each real-time forecast. Using a real-time exercise, I find that there is no strong evidence to consider 1984.1 as a breakpoint in the variance of output growth until 1995.3.[1]

The plan of the paper is as follows. The next section presents the baseline model and highlights the economic interpretation of the VSTR specification of GDP and CLI. The following section deals with the extension to the multiple-equation framework of linearity tests, model selection procedures and model adequacy tests. In addition, this section includes a brief discussion about the techniques used for comparing the forecasting performance of these nonlinear models. We then consider the empirical results, and a final section contains concluding remarks and suggests directions for future research.

THE BASELINE MODEL

Consider the following vector autoregressive generalization of the STR model proposed by Granger and Teräsvirta (1993):

$$
\begin{aligned}
y_t &= \beta_y' A_t + (\tilde{\beta}_y' A_t) F_y(D_{ty}) + \alpha_y e_{t-1} + u_{yt} \\
x_t &= \beta_x' A_t + (\tilde{\beta}_x' A_t) F_x(D_{tx}) + \alpha_x e_{t-1} + u_{xt}
\end{aligned} \tag{1}
$$

[1] Thus, I focus the multiple-equation STR models on an alternative view of the monetary policy analysis of Weise (1999) and the Granger causality from money to output of Rothman *et al*. (2001).

where y_t and x_t are the rates of growth of GDP and CLI, $A_t = (1, y_{t-1}, x_{t-1}, \ldots, y_{t-p}, x_{t-p})' = (1, X_t')'$, $\beta_y' = (\eta_y, a_1, b_1, \ldots, a_p, b_p)$, $\beta_x' = (\eta_x, c_1, d_1, \ldots, c_p, d_p)$, $\tilde{\beta}_y' = (\tilde{\eta}_y', \tilde{a}_1, \tilde{b}_1, \ldots, \tilde{a}_p, \tilde{b}_p)$, $\tilde{\beta}_x' = (\tilde{\eta}_x, \tilde{c}_1, \tilde{d}_1, \ldots, \tilde{c}_p, \tilde{d}_p)$. In case of cointegration, the equilibrium error $e_t = y_t - \rho_0 - \rho_1 x_t$ is included in the VSTR representation following Rothman *et al*. (2001). Finally, consider the serially uncorrelated series of errors

$$U_t = (u_{yt}, u_{xt})' \sim N[0, \tilde{\Omega}] \tag{2}$$

where the variance $\tilde{\Omega}$ is Ω_1 from the beginning of the sample until 1984.1 and Ω_2 since this date, reflecting the recently documented structural break in the variance of output.

The key component of a VSTR model is the *transition function F.* By convention, it is bounded between zero and one. If F is zero, then the baseline model becomes a linear VAR (VARa), with parameters β_y and β_x. On the other hand, if F is one, then the VSTR model becomes another linear VAR (VARb), with parameters $\beta_y + \tilde{\beta}_y$ and $\beta_x + \tilde{\beta}_x$. Hence, F may be interpreted as a filtering rule that locates the model between these two extreme regimes. This section presents a brief discussion about the economic interpretation of VSTR models, depending on the form of the transition function.

Logistic transition function

In this case, F is the following monotonically increasing function:

$$F_i(D_{ti}) = \frac{1}{1 + e^{-\gamma_i D_{ti}}} \tag{3}$$

where γ_i is the *smoothness parameter*, $i = y, x$. I refer to D_{ti} as a *switching expression* which may present two alternative forms. First, D_{ti} may be the difference between a proposed *transition variable* z_{ti}, which is usually a lagged value of y and x, and an estimated *threshold* g_i, that is

$$D_{ti} = z_{ti} - g_i \tag{4}$$

I call a logistic VSTR model with switching expression (4) logistic VSTR (LVSTR(z_{ty}, z_{tx})). Note that, as γ_i approaches infinity, F_i converges to the Heaviside function. In this extreme case, the baseline model generalizes to a VAR the SETAR model proposed by Tsay (1989). Second, D_{ti} may be the weighted average of the q_i lagged deviations from a linear path:

$$D_{ti} = \sum_{j=1}^{q_i} w_{ij} \hat{\vartheta}_{i,t-j} \tag{5}$$

where $\Sigma_{j=1}^{q} w_{ij} = 1$, and $\hat{\vartheta}_i$ is the estimated residual of the ith equation from a linear path. Similarly, a logistic VSTR model with D_{ti} as in (5) represents the LVSTR-deviated (LVSTR-D(q_y, q_x)) models.

Applied to GDP and CLI rates of growth, logistic models have a nice economic interpretation. Assume that $\tilde{\beta}$ and γ are both greater than zero. In logistic models, VARa (F close to zero) is interpreted as the linear path which models extreme recessionary periods, whereas VARb (F close to one) can be seen as the linear model associated with great expansions. To see this, note that in extreme contractions (expansions) the transition variable is lower (higher) enough than the threshold in

LVSTR models, and the actual GDP is lesser (greater) enough than a linear path in LVSTR-D models to keep the transition function close to zero (one). Thus, the transition function locates the model either near to or far from recessions, depending on the switching expression's values.

Exponential transition function
Consider the exponential transition function

$$F_i(D_{ti}) = 1 - e^{-\gamma_i D_{ti}} \tag{6}$$

where $i = y, x$. Assume the following alternative forms for the switching expression. First, let D_{ti} be the squared difference between the transition variable and the threshold:

$$D_{ti} = (z_{ti} - g_i)^2 \tag{7}$$

I denote an exponential model with switching expression (7) as exponential VSTR (EVSTR(z_{ty}, z_{tx})). Second, let D_{ti} be the weighted sum of the q lagged squared deviations from a linear path:

$$D_{ti} = \sum_{j=1}^{q_i} w_{ij} \hat{\vartheta}^2_{i,t-j} \tag{8}$$

where w_{ij} and $\hat{\vartheta}_i$ have been defined in (5). I refer to these models as EVSTR-deviated (EVSTR-D(q_y, q_x)).

Applied to GDP and CLI, exponential models have different economic interpretations to logistic models. Now, VARa can be associated with a middle ground, whereas troughs and peaks have similar dynamic structures associated with VARb. That is to say, if either the transition variable is different to the threshold in the EVSTR, or the model deviates from a linear path in the EVSTR-D, then F becomes different from zero, and the model smoothly approximates from the middle ground to any of the extreme situations represented by VARb ($F = 1$).

SPECIFICATION OF VSTR MODELS

The aim of this section is to describe a battery of model selection rules in order to obtain one nonlinear specification from the set of VSTR models outlined in the previous section. Note that, since I base the estimation of VSTR upon the maximum likelihood principle, any test may be carried out through standard likelihood ratio tests except for the case of nuisance parameter problems. Additionally, I restrict the analysis to the case of $z_{ty} = z_{tx} = z_t$ and $q_y = q_x = q$.[2]

In the spirit of the seminal methodology of Tsay (1989), I describe in Figure 1 a stepwise procedure for modelling VSTR specifications. First, I specify a linear VAR and its maximum lag length using standard linear techniques. Second, I apply linearity and model selection tests for each candidate to be the switching expression. Third, I apply the model adequacy tests to the estimated models.

[2] In deviated models, this implies that the system is deviated from the linear path according to the same number of lagged deviations for both GDP and CLI. In the remaining cases, this implies that the same transition variable locates the entire system between regimes.

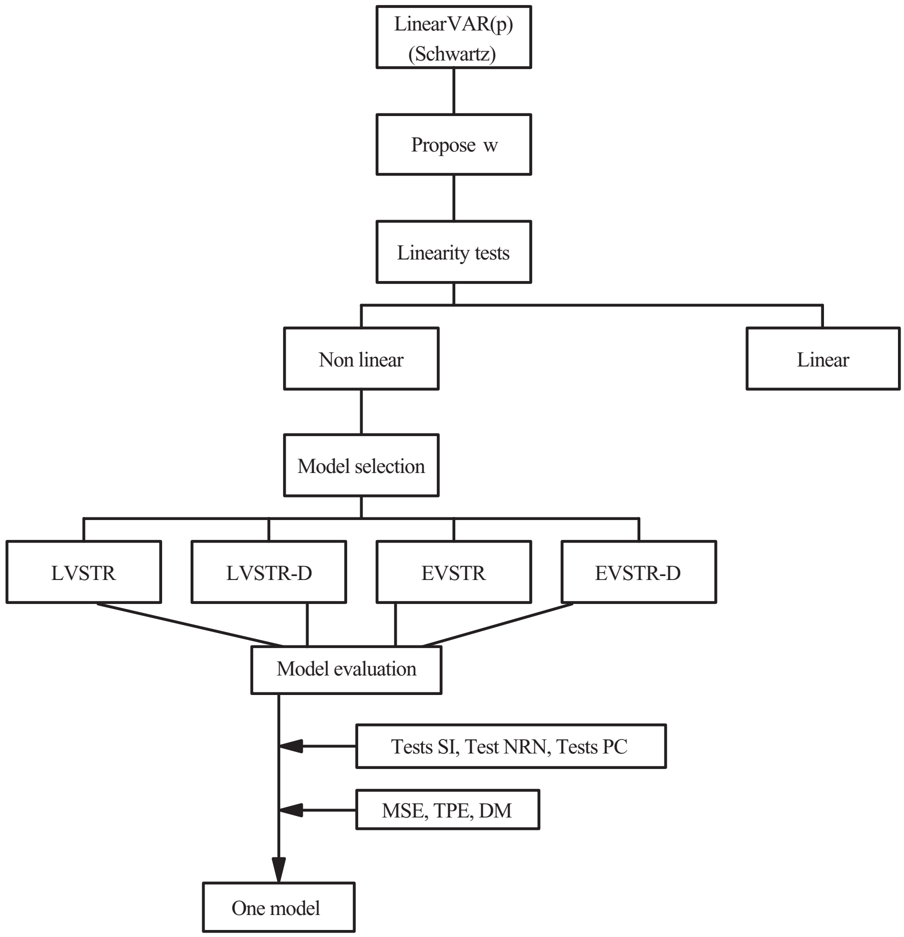

Figure 1. Description of VSTR selection

Note: This figure describes the specification of VSTR models in four steps. First, a linear VAR and its maximum lag length are specified. Second, linearity tests are applied for the *w* the researcher proposes. Third, for each *w* for which linearity was rejected, model selection tests are carried out. Finally, the validity of these models is evaluated by testing their adequacy (tests of serial independence of errors, test of no remaining nonlinearity and tests of parameter constancy), and by checking the accuracy of the resulting models at anticipating both output growth (MSE) and turning points (TPE), formally tested through Diebold and Mariano (DM) tests. The procedure concludes with the selection of one nonlinear specification.

Finally, since this procedure finds nonlinear models as rejections of linearity, I select the final model according to its ability to forecast output growth and business-cycle phases.

Linearity and model selection tests

The application of linearity and model selection tests requires the *a priori* selection of a set of variables to include in the switching expression. This implies selecting the value of q in deviated models,

Table I. Linear approximation of VSTR models

z belongs to X_t	
Logistic models	Exponential models
$y_t = \varepsilon_{y0} + \sum_{h=0}^{3} \xi'_{yh} X_t w^h + \upsilon_{yt}$	$y_t = \varepsilon_{y0} + \sum_{h=0}^{2} \xi'_{yh} X_t w^h + \upsilon_{yt}$
$x_t = \varepsilon_{x0} + \sum_{h=0}^{3} \xi'_{xh} X_t w^h + \upsilon_{xt}$	$x_t = \varepsilon_{x0} + \sum_{h=0}^{2} \xi'_{yh} X_t w^h + \upsilon_{xt}$
z does not belong to X_t and deviated models	
Logistic models	Exponential models
$y_t = \sum_{h=0}^{1} \left(\varepsilon_{yh} w^h + \xi'_{yh} X_t w^h\right) + \upsilon_{yt}$	$y_t = \sum_{h=0}^{2} \left(\varepsilon_{yh} w^h + \xi'_{yh} X_t w^h\right) + \upsilon_{yt}$
$x_t = \sum_{h=0}^{1} \left(\varepsilon_{xh} w^h + \xi'_{xh} X_t w^h\right) + \upsilon_{xt}$	$x_t = \sum_{h=0}^{2} \left(\varepsilon_{xh} w^h + \xi'_{xh} X_t w^h\right) + \upsilon_{xt}$

Note: This table applies to the case $z_y = z_x = z$ and $q_y = q_x = q$. Variable w is the (square) weighted deviation from the linear path in deviated models whereas it is the transition variable in other VSTR models.

and lagged values of y and x in the remaining cases. In addition, the natural way of choosing an appropriate maximum value of q and maximum lag of x and y is to base the decision upon the frequency of the data.[3]

Following Luukkoven *et al*. (1988), I base both linearity and model selection tests on suitable Taylor series expansions of the transition functions around the point $\gamma = 0$. Table I shows the different models' linearizations according to the different models' specifications. In the case of logistic models with the transition variable belonging to the set of explanatory variables, I avoid the identification problem by using a third-order linear approximation. In deviated models, I need a second-order approximation for discriminating between logistic and exponential models. In the remaining cases, I approximate the transition function with a first-order Taylor approximation.

A possible null hypothesis of linearity is $H_0: \gamma = 0$ and the alternative $H_0: \gamma > 0$. This choice leads to the nuisance parameters problem since the model is not identified under the null. As a consequence, the classical distribution theory does not work in this context. To overcome this problem, I use the linear approximations of the transition function to describe the linearity tests presented in Table II. These tests are based on standard LM-type tests on the auxiliary regressions depicted in the first column. The null of linearity proposed in the second column of this table consists of getting a linear VAR model under the null.

If linearity is rejected, the model selection tests must decide between logistic and exponential transition functions. In line with the univariate proposal of Granger and Teräsvirta (1993), Table III shows the sequence of nested hypothesis tests that should be applied to the auxiliary regressions of

[3] For example, with monthly (quarterly) data, it is convenient to try for a maximum value of q and a maximum lag of x and y of 12 (4).

Table II. Linearity tests

z belongs to X_t	
Auxiliary regressions	Null of linearity
$y_t = \varepsilon_{y0} + \sum_{h=0}^{3} \xi'_{yh} X_t w^h + \upsilon_{yt}$ $x_i = \varepsilon_{x0} + \sum_{h=0}^{3} \xi'_{xh} X_t w^h + \upsilon_{xt}$	$\xi_{i1} = \xi_{i2} = \xi_{i3} = 0$
z does not belong to X_t and deviated models	
Auxiliary regressions	Null of linearity
$y_t = \sum_{h=0}^{2} \left(\varepsilon_{yh} w^h + \xi'_{yh} X_t w^h\right) + \upsilon_{yt}$	$\varepsilon_{i1} = \varepsilon_{i2} = 0$
$x_i = \sum_{h=0}^{2} \left(\varepsilon_{xh} w^h + \xi'_{xh} X_t w^h\right) + \upsilon_{xt}$	$\xi_{i1} = \xi_{i2} = 0$

Note: See Table I for parameter definitions.

Table II. According to the results obtained by these tests, the last column of Table III shows the final decision about the nature of the transition function.[4]

Testing the adequacy of VSTR models

Eitrheim and Teräsvirta (1996) propose three kinds of tests for evaluating the adequacy of the estimated single-equation STR model. Specifically, they consider that a model with serially independent errors (test SI), with parameter constancy (test PC) and with no remaining Nonlinearity (test NRN) may be considered as adequate for fitting the data. This section extends these tests to a multiple-equation framework.

To derive the test SI, I consider an alternative representation of the baseline model that takes into account the possibility of serial dependence in the errors:

$$Y_t = G(A_t, \Psi) + U_t \tag{9}$$

where $U_t = (u_{yt}, u_{xt})'$, $Y_t = (y_t, x_t)'$, $G(\varphi_t, \Psi) = (G_y(\varphi_t, \Psi_y), G_x(\varphi_t, \Psi_x))'$, $G_i(\varphi_t, \Psi_i) = \beta_i' A_t + (\tilde{\beta}_i' A_t) F_i(D_{ti})$. In addition, $\Psi_i = (\beta_i, \tilde{\beta}_i, \gamma_i, g_i)$ is the $(\rho/2 \times 1)$ vector of unknown parameters contained in both the autoregressive lags and in the transition function, with $i = y, x$. Instead of (2), errors are assumed to evolve as

[4] For example, a proposed $z = x_{t-1}$ (which belongs to X_t) for which Test 1 and Test 2 are not rejected, but Test 3 is rejected, signals a logistic transition function.

Table III. Model selection tests

z belongs to X_t				Choice
Hypoth.	Test 1	Test 2	Test 3	
H_0	$\xi_{i3} = 0$	$\xi_{ij} = 0$, $j = 2, 3$	$\xi_{ij} = 0$, $j = 1, 2, 3$	
H_α	$\xi_{i3} \neq 0$	$\xi_{i2} \neq 0$ $\xi_{i3} = 0$	$\xi_{i1} \neq 0$ $\xi_{ij} = 0$ $j = 2, 3$	
	Reject	. . .	. . .	Logistic
	Accept	Reject	Accept	Exponential
	Accept	Accept	Reject	Logistic
	Accept	Reject	Reject	No decision
z does not belong to X_t and deviated models				Choice
Hypoth.	Test 1			
H_0	$\varepsilon_{i2} = 0$, $\xi_{i2} = 0$			
H_α	$\varepsilon_{i2} \neq 0$, $\xi_{i2} \neq 0$			
	Accept			Logistic
	Reject			Exponential

Note: See Table I for parameter definitions

$$U_t = \Phi(L)U_t + \varsigma_t, \quad \varsigma_t \sim N[0, \tilde{\Gamma}] \tag{10}$$

where ς_t is serially independent, and $\tilde{\Gamma}$ is the (2×2) matrix $\Gamma 1$ from the beginning of the sample until the breakpoint 1984.1, and the (2×2) matrix Γ_2 since this date. Here, $\Phi(L) = (\Phi_1 L + \ldots + \Phi_r L^r)$ indicates a (2×2) matrix polynomial in the lag operator L. Under the null hypothesis of serial independence of errors, that is $H_0{:}\Phi_1 = \ldots = \Phi_r = 0$, the Lagrange multipliers (LM) test statistic

$$\mathrm{LM} = \frac{1}{T} m'_\Phi \left(M_{\Phi\Phi} - M_{\Phi\Psi} (M_{\Psi\Psi})^{-1} M'_{\Phi\Psi} \right)^{-1} m_\Phi \tag{11}$$

follows a χ^2 limiting distribution with $4r$ degrees of freedom. In the Appendix, I derive the following simple expression of this test. Let V_t be the $(2r \times 1)$ matrix $(v'_{yt}, v'_{xt})'$, where $v_{it} = (u_{i,t-1}, \ldots, u_{i,t-r})'$, and let Z_t be the $(\rho/2 \times 2)$ matrix (z_{yt}, z_{xt}), where z_{it} is $\partial G_i/\partial \Psi_i = \partial G_i(A_t, \Psi_i)/\partial \Psi_i$, with $i = y, x$. Thus, considering that a bar below any expression refers to its maximum likelihood estimate under the null, and that $\otimes$ denotes the Kronecker product, the LM test may be implemented with the estimates

$$\underline{m}_\Phi = \sum \left(\underline{\tilde{\Gamma}}^{-1} \underline{U}_t \otimes \underline{V}_t \right), \qquad M_{\Phi\Phi} = \frac{1}{T} \sum \left(\underline{\tilde{\Gamma}}^{-1} \otimes \underline{V}_t \underline{V}'_t \right)$$
$$\underline{M}_{\Phi\Phi} = \frac{1}{T} \sum \left(\underline{\tilde{\Gamma}}^{-1} \underline{Z}_t \otimes \underline{V}_t \right), \qquad M_{\Psi\Psi} = \frac{1}{T} \sum \left(\underline{Z}_t \underline{\tilde{\Gamma}}^{-1} \underline{Z}_t \right) \tag{12}$$

Table IV. Test of parameter constancy

Auxiliary regressions	Null of constant parameters
$y_t = \theta'_{y0}\alpha_{yt} + \theta'_{y1}\alpha_{yt}t + \ldots + \theta'_{yk}\alpha_{yt}t^k$ $\left[\tilde{\theta}_{y0}\tilde{\alpha}_{yt} + \tilde{\theta}'_{y1}\tilde{\alpha}_{yt}t + \ldots + \tilde{\theta}'_{yk}\tilde{\alpha}_{yt}t^k\right]F_y + \upsilon_{yt}$	$\theta'_{i1} = \ldots = \theta'_{ik} = 0$
$x_t = \theta'_{x0}\alpha_{xt} + \theta'_{x1}\alpha_{xt}t + \ldots + \theta'_{xk}\alpha_{xt}t^k$ $\left[\tilde{\theta}'_{x0}\tilde{\alpha}_{xt} + \tilde{\theta}'_{x1}\tilde{\alpha}_{xt}t + \ldots + \tilde{\theta}'_{xk}\tilde{\alpha}_{xt}t^k\right]F_x + \upsilon_{xt}$	$\tilde{\theta}'_{i1} = \cdots = \tilde{\theta}'_{ik} = 0$

Note: See Table I for parameter definitions.

Testing parameter constancy is an important way of checking the adequacy of VSTR models since they are estimated assuming constant parameters. The test PC is obtained under the assumption that the transition function has constant parameters, whereas both β_i and $\tilde{\beta}_i$ may change over time. I consider that the change may be possibly nonmonotonic and not necessarily symmetric, that is $\beta_i(t) = \beta_i + \lambda_{1i}H_i(t)$ and $\tilde{\beta}_i(t) = \tilde{\beta}i + \lambda_{2i}H_i(t)$, with $i = y$, x, and

$$H_i(t) = (1 + \exp\{-\gamma_i(t^k + s_{i(k-1)}t^{k-1} + \ldots + s_{i1}t + s_{i0})\})^{-1} - 0.5 \tag{13}$$

where substracting one-half is useful just in deriving the tests. After linear approximations of H_i, Table IV describes a simple LM-type test against the null of time-varying parameters. This test is based on imposing the null (second column) that the varying parameters are not significant in the auxiliary regressions described in the first column.

Finally, to obtain the test NRN it is useful to rewrite the baseline model allowing for additive misspecification as follows:

$$\begin{aligned} y_t &= \beta'_y A_{yt} + (\tilde{\beta}'_y A_{yt})F_y^1(D_{ty}^1) + (\tilde{\theta}'_y A_t)F_y^2(D_{ty}^2) + u_{yt} \\ x_t &= \beta'_x A_{xt} + (\tilde{\beta}'_x A_{xt})F_x^1(D_{tx}^1) + (\tilde{\theta}'_x A_t)F_x^2(D_{tx}^2) + u_{xt} \end{aligned} \tag{14}$$

where F is the transition function analysed in previous sections. After the linearization of F_i^2 described in the linearity tests, the method of implementing the test NRN is similar to the method of testing the null of linearity outlined in Table II. Following Eitrheim and Teräsvirta (1996), Table V generalizes the NRN test to have power not only against an omitted additive nonlinear component but also against omission of important lags from the estimated model. This is done by considering under the null the exclusion restrictions imposed to obtain the VSTR significant parameter estimates.

Examining the predictive accuracy

The sequence of tests described above finds nonlinear models as rejections of linearity. Teräsvirta (1994) suggests that in such a case the selected model should be the one with the smallest p-value in the linearity test. However, this procedure involves two drawbacks. First, one may find appropriate estimates and forecasts of the nonlinear model even if linearity is weakly rejected. Second, it is not clear what to do in case of similar p-values. The remainder of the section tries to guard against

Table V. Test of no remaining nonlinearity

z belongs to X_t	
Auxiliary regressions	Null of no remaining nonlinearity and no important omission
$y_t = \alpha_y' A_{yt} + \tilde{\alpha}_y' A_{yt} F_y^1 + \sum_{h=1}^{3} \xi_{yh}' X_t w^h + \bar{\alpha}_y' \bar{A}_{yt} + \bar{\alpha}_y' \bar{A}_{yt} F_y^1 + \upsilon_{yt}$	$\xi_{i1} = \xi_{i2} = \xi_{i3} = 0$
$x_t = \alpha_x' A_{xt} + \tilde{\alpha}_x' A_{xt} F_x^1 + \sum_{h=1}^{3} \xi_{xh}' X_t w^h + \alpha_x' A_{xt} + \tilde{\alpha}_x' A_{xt} F_x^1 + \upsilon_{xt}$	$\bar{\alpha}_i = 0, \tilde{\bar{\alpha}}_i = 0$
z does not belong to X_t and deviated models	
Auxiliary regressions	Null of no remaining nonlinearity and no important omission
$y_t = \alpha_y' A_{yt} + \tilde{\alpha}_y' A_{yt} F_y^1 + \sum_{h=1}^{2} \left(\varepsilon_{yh} w^h + \xi_{yh}' X_t w^h \right) + \bar{\alpha}_y' \bar{A}_{yt} + \tilde{\bar{\alpha}}_y \bar{A}_{yt} F_y^1 + \upsilon_{yt}$	$\varepsilon_{i1} = \varepsilon_{i2} = 0$ $\xi_{i1} = \xi_{i2} = 0$
$x_t = \alpha_x' A_{xt} + \tilde{\alpha}_x' A_{xt} F_x^1 + \sum_{h=1}^{2} \left(\varepsilon_{xh} w^h + \xi_{xh}' X_t w^h \right) + \alpha_x' A_{xt} + \tilde{\alpha}_x' A_{xt} F_x^1 + \upsilon_{xt}$	$\bar{\alpha}_i = 0, \tilde{\bar{\alpha}}_i = 0$

Note: Without loss of generalization, the elements in β_i and $\bar{\beta}_i$ are decomposed into those first k_i and $\tilde{k}_i$ nonzero elements (α_i and $\tilde{\alpha}_i$), and those last $\bar{k}_i$ and $\tilde{\bar{k}}_i$ elements that are assumed to be zero in the parameter estimation ($\bar{\alpha}_i$ and $\tilde{\bar{\alpha}}_i$). Consequently, the matrix of explanatory variables is decomposed into the A_{it} and $\bar{A}_{it}$ corresponding matrices. See Table I for further parameter definitions.

these drawbacks by basing the decision upon an *a posteriori* evaluation of the accuracy of the estimated nonlinear models at forecasting output growth and business-cycle phases.

As a first approximation to examine predictive accuracy of the estimated model, I propose the certain positive rate (CNR) and the certain negative rate (CPR) measures. The former (latter) signals the percentage of quarters that the models correctly anticipate GDP rises and NBER expansions (GDP falls and NBER recessions). In addition, I consider the following measures of the false signals provided by the forecasting models: the false positive rate (FPR), that measures the percentage of times of actual positive output growth and NBER expansions when the model predicts negative output growth and recessions, and the false negative rate (FNR), that measures the percentage of quarters of actual negative growth and official recessions when the model forecasts GDP rises and expansions.[5]

As a second approximation to the forecasting accuracy, I compute the following measures. The output growth forecasting accuracy may be checked using the well-known mean square error

$$\mathrm{MSE} = \frac{1}{T} \sum_{t=1}^{T} (y_t - \hat{y}_t)^2 \tag{15}$$

[5] In line with Stock and Watson (1993), I interpret an estimated probability of recession above 0.75 (below 0.25) as a signal of recession (expansion).

based upon the distance between actual (y) and estimated ($\hat{y}$) output growth (T is the sample size). The business-cycles forecasting accuracy may be investigated with the loss function turning points error

$$\text{TPE} = \frac{1}{T}\sum_{t=1}^{T}(d_t - \hat{d}_t)^2 \tag{16}$$

where d_t is an indicator variable taking value 1 at the official NBER recessions. Recall that logistic transition functions (in models with baseline parameters greater than zero) may be interpreted as probabilities of expansion. This leads us to define $\hat{d}_t = 1 - F_y(z_{ty})$.

Note that the last two measures lead to a ranking of the competing models according to their forecasting performance. However, it is advisable to test whether the forecasts made with one of these models are significantly superior to the other model's forecasts. One interesting possibility is to test the null of no difference in the forecasting accuracy of these competing models using the following tests: the Diebold–Mariano (DM), modified Diebold–Mariano (MDM), Morgan–Granger–Newbold (MGN) and Meese–Rogoff (MR) tests, all of them described in Diebold and Mariano (1995) and Harvey *et al*. (1997). An additional possibility is to consider the forecast encompassing test that is based on testing the significance of α_1 in the OLS regression

$$l_t - \hat{l}_{t,j} = \alpha_0 + \alpha_1 \hat{l}_{t,j} + \varphi_t \tag{17}$$

where l_t is either y_t or d_t, and $\hat{l}_{t,i}$ and $\hat{l}_{t,j}$ are the forecasts computed from two competing models i and j.

EMPIRICAL RESULTS

The CLI is a weighted average of 10 macroeconomic leading variables which are expected to turn before the aggregate economy. In this section, I examine the effective real-time predictive power of the leading indicator series to forecast both output growth and business-cycle phases of the US economy using the VSTR models.

In-sample analysis

Even though this paper focuses on real-time forecasts, I consider a preliminary in-sample analysis of the 173 quarterly observations of GDP running from 1959.1 to 2002.1. Since the Conference Board issues the series of CLI monthly, in order to compare the leading indicator with the output series I transform the indicator into a quarterly series by selecting the data corresponding to the last month of each quarter. In addition, the indicator series is published one-month lagged, so the first CLI series with figures for March 2002 is released in April. However, these preliminary figures are dramatically changed in the following month's release. Consequently, in the in-sample analysis I consider the CLI series issued in May 2002.

Following the specification strategy outlined in Figure 1, I need to specify an appropriate linear VAR which is the base for the nonlinear models. In a preliminary analysis of data, the augmented Dickey–Fuller, Phillips–Perron, KPSS and Lobato–Robison tests detect unit roots in the log of both variables, which suggest the use of the stationary rate of growth transformation of GDP and CLI, hereafter y and x.[6] In addition, I apply the nonparametric cointegration test proposed by Bierens (1997), who considers that the number of cointegrating vectors r can be estimated as the argument that minimizes the function

$$g_m(r) = \begin{cases} \left(\prod_{k=I}^{T} \lambda_{km}\right)^{-1} & \text{if } r = 0 \\ \left(\prod_{k=I}^{T-r} \lambda_{km}\right)^{-1} \left(T^{2r} \prod_{k=n-r+1}^{T} \lambda_{km}\right) & \text{if } r = 1, \ldots, T-I \\ T^{2n} \prod_{k=1}^{T} \lambda_{km} & \text{if } r = T \end{cases} \tag{18}$$

where T is the sample size, λ_i is the ordered eigenvalue obtained from a nonparametric generalized eigenvalue problem in the same spirit as Johansen's method, and m is a parameter selected according to the values of r and T as stated in his paper.[7] Using this method, I obtain that the number of cointegrating vectors that minimizes $g_m(r)$ is zero, which indicates the absence of cointegration between the log of the series of GDP and CLI. Accordingly, I do not use the equilibrium errors to compute the in-sample estimates.

As documented by Kim and Nelson (1999) and McConnell and Perez-Quiros (2000), it is interesting to consider 1984.1 as a candidate for being a breakpoint in the variance in order to select the most appropriate VAR specification. According to these findings, I conduct a Chow test by imposing 1984.1 as a breakpoint in the variance of an AR(1) for both y and x, and in the variance–covariance matrix of a VAR(1) for a vector formed by these variables.[8] Since I obtain p-values that are always less than 0.01, my final specification considers the reduction in the volatility since 1984.1.

Linearity and model selection tests require the specification of a set of variables z and a set of values of q. For the former, I use lagged values of x and y within a year. For the latter, I use (square) weighted averages of the one to four lagged deviations from the linear path. The first column of Table VI reveals that the number of rejections of the null of linearity is large, which confirms the nonlinear nature of the relationships between GDP and CLI previously documented in the literature. In addition, columns two to five of this table present the results of the model selection tests, and the final nonlinear specification in those cases for which linearity was rejected. In order to reduce the number of VSTR models, I select the model presenting the strongest rejection of linearity within each family and postpone the decision of selecting one of them according to their forecasting ability. This leads us to consider one logistic ($z_t = y_{t-2}$), one exponential ($z_t = x_{t-2}$) and one logistic-deviated ($q = 1$) model, called LVSTR(y_{t-2}), EVSTR(x_{t-2}) and LVSTR-D(1), respectively. Table VII shows the maximum likelihood estimates of their significant parameters.[9] In illustrating how these nonlinear models work, Figure 2 plots the transition function of the CLI equation of LVSTR(y_{t-2}) and EVSTR(x_{t-2}). The former presents a smoothness parameter of 7.11, indicating that the transition

[6] These results are omitted but available from the author upon request.
[7] Note that, due to the nonparametric nature of this test, the Bierens results are independent of the data-generating process. Thus, even though the relationships between GDP and CLI were nonlinear, the test remains valid.
[8] I select the AR and VAR lag lengths using the Schwarz selection criterion.
[9] Note that, as Teräsvirta (1994) has emphasized, a precise joint estimation of the smoothness parameter and the threshold is usually a problem.

Table VI. Results of linearity and model selection tests

	Lin. Test	Test 1	Test 2	Test 3	Decision
$z = y_{t-1}$	40.70 (R)	11.13	5.42	11.13 (R)	LVSTR(y_{t-1})
$z = y_{t-2}$	62.73 (R)	2.75	...	...	LVSTR(y_{t-2})
$z = y_{t-3}$	51.90 (R)	52.21 (R)	...	...	EVSTR(y_{t-3})
$q = 1$	93.91 (R)	11.93	...	...	LVSTR-D(1)
$q = 2$	85.81 (R)	111.72 (R)	...	...	EVSTR-D(2)
$q = 3$	30.12 (R)	23.71	...	...	LVSTR-D(3)
$z = x_{t-1}$	101.18 (R)	37.22 (R)	8.94	55.02 (R)	LVSTR(x_{t-1})
$z = x_{t-2}$	254.92 (R)	111.24 (R)	...	...	EVSTR(x_{t-2})
$z = x_{t-3}$	170.68 (R)	54.97 (R)	...	...	EVSTR(x_{t-3})

Note: Tests are developed as described in text. Statistics are displayed only for models which reject linearity. Second column shows the results for linearity tests, whereas third to fifth columns present the results for model selection tests. Only when $z = y_{t-1}$ or x_{t-1} belongs to the set of explanatory variables (a three-stage testing procedure applies), since these tests refer to a vector autoregressive specification with lag length one. Results of the tests at 5% are in parentheses (R: reject).

between the two extreme regimes (characterized by $F = 0$ and $F = 1$) is relatively sharp. The estimated threshold is 0.12 and marks the halfway point between regimes. The latter model shows a much lower smoothness parameter (0.61), which implies smoother transitions between the middle ground ($F = 0$), marked by values of x_{t-2} near to −0.72, and the other extreme regime ($F = 1$).

In order to investigate the adequacy of these nonlinear models to the GDP and CLI data, Table VIII shows the p-values of the SI test for values of r from 1 to 4, the PC test for values of k from 1 to 3, and the NRN test. With regard to the logistic models, these entries show that there is no evidence of serial correlation of errors at any lag (p-values higher than 0.15), no empirical support for rejecting the null of parameter constancy at any value of k (p-values higher than 0.08), and no remaining nonlinearity (p-values higher than 0.90). With regard to the exponential model, these entries reveal that even though there is no strong evidence of remaining nonlinearity (p-value of 0.05), errors may be correlated (p-value of 0.01 for $r = 2$ and $r = 4$), and parameters may be nonconstant (p-values less than 0.05 for $k = 2$ and $k = 3$), showing that the exponential transition function may not be adequate for the data.

Table IX reveals that LVSTR(y_{t-2}) is the most accurate model to forecast output growth signs and business-cycle phases. With respect to the forecasting performance of output growth signs, this model presents the highest percentage of successes (CPR and CNR of 98.64 and 52.17) and the lowest percentage of failures (FPR and FNR of 14.28 and 7.00). With respect to the ability to forecast business-cycle phases, this model shows a higher percentage of successes (CPR and CNR of 59.25 and 87.50) and a lower percentage of false signals (FPR and FNR of 27.27 and 5.26) than the LVSTR-D(1) model.

In line with these results, Table X confirms the superior accuracy of the LVSTR(y_{t-2}) model to forecast growth and business cycles. This model presents much lower MSE than the LVSTR-D(1) and EVSTR(x_{t-2}) models, with relative MSE measures of 0.67 and 0.59 respectively. The p-values of the null of equal forecasting accuracy from DM, MDM, MGN and MR tests are always less than 0.001, which reveals that the output growth forecasts of the LVSTR(y_{t-2}) model are statistically superior to the other models' forecasts. In addition, the null that forecasts from this model encompass forecasts from the LVSTR-D(1) and EVSTR(x_{t-2}) models cannot be rejected at any standard significance level (p-values of 0.91 and 0.34). Finally, the relative TPE of the LVSTR(y_{t-2}) model over the

Table VII. Maximum likelihood estimates of parameters

	Model estimation
LVSTR(y_{t-2})	$\hat{y}_t = \underset{(0.34)}{1.08}\hat{F}_y + \underset{(0.04)}{0.33}x_{t-1}$
	$\hat{x}_t = \underset{(0.27)}{0.75} - \underset{(0.28)}{0.63}y_{t-1} + \underset{(0.08)}{0.38}x_{t-1} + -\underset{(0.22)}{0.51} + \underset{(0.32)}{0.57}\Big]\hat{F}_x$
	$\hat{F}_y = 1 + \exp\left(-\underset{(0.83)}{1.18}\left(y_{t-2} - \underset{(0.58)}{0.36}\right)\right)\Big]^{-1}$
	$\hat{F}_x = 1 + \exp\left(-\underset{(9.83)}{7.11}\left(y_{t-2} + \underset{(0.37)}{0.12}\right)\right)\Big]^{-1}$
	$\hat{\sigma}^1_{11} = \underset{(0.14)}{0.76} \quad \hat{\sigma}^1_{22} = \underset{(0.25)}{1.73} \quad \hat{\sigma}^1_{12} = \underset{(0.10)}{0.11}$
	$\hat{\sigma}^2_{11} = \underset{(0.05)}{0.28} \quad \hat{\sigma}^2_{22} = \underset{(0.14)}{0.81} \quad \hat{\sigma}^2_{12} = \underset{(0.06)}{0.19}$
LVSTR-D(1)	$\hat{y}_t = \underset{(0.04)}{0.33}x_{t-1} + \underset{(0.12)}{1.33}\hat{F}_y$
	$\hat{x}_t = \underset{(0.15)}{0.52} + \underset{(0.13)}{0.75}x_{t-1}\Big]\hat{F}_x$
	$\hat{F}_y = 1 + \exp\left(-\underset{(13.62)}{0.06}(y_{t-1} - \underset{(0.07)}{0.58} - \underset{(0.06)}{0.12}y_{t-2} - \underset{(0.04)}{0.32}x_{t-2}\right)\Big]^{-1}$
	$\hat{F}_x = 1 + \exp\left(-\underset{(0.23)}{1.61}(x_{t-1} - \underset{(0.12)}{0.38} - \underset{(0.10)}{0.21}y_{t-2} - \underset{(0.07)}{0.45}x_{t-2}\right)\Big]^{-1}$
	$\hat{\sigma}^1_{11} = \underset{(0.11)}{0.79} \quad \hat{\sigma}^1_{22} = \underset{(0.23)}{1.62} \quad \hat{\sigma}^1_{12} = \underset{(0.10)}{0.11}$
	$\hat{\sigma}^2_{11} = \underset{(0.05)}{0.30} \quad \hat{\sigma}^2_{22} = \underset{(0.14)}{0.84} \quad \hat{\sigma}^2_{12} = \underset{(0.06)}{0.14}$
EVSTR(x_{t-2})	$\hat{y}_t = \underset{(0.17)}{1.10} - \underset{(0.69)}{1.42}\hat{F}_y + \underset{(0.04)}{0.26}x_{t-1}$
	$\hat{x}_t = \underset{(0.25)}{0.55}\hat{F}_x + \underset{(0.07)}{0.29}x_{t-1}$
	$\hat{F}_y = 1 - \exp\left(-\underset{(0.06)}{0.06}\left(x_{t-2} - \underset{(1.12)}{2.66}\right)^2\right)$
	$\hat{F}_x = 1 - \exp\left(-\underset{(0.82)}{0.61}\left(x_{t-2} - \underset{(0.31)}{(-0.72)}\right)^2\right)$
	$\hat{\sigma}^1_{11} = \underset{(0.10)}{0.70} \quad \hat{\sigma}^1_{22} = \underset{(0.11)}{1.68} \quad \hat{\sigma}^1_{12} = \underset{(0.05)}{0.06}$
	$\hat{\sigma}^2_{11} = \underset{(0.04)}{0.26} \quad \hat{\sigma}^2_{22} = \underset{(0.14)}{0.80} \quad \hat{\sigma}^2_{12} = \underset{(0.05)}{0.13}$

Note: Parameter σ^1_{ij} (σ^2_{ij}) refers to the row i, column j element of the VARCOV matrix during the period 1959.1–1983.4 (1984.1–2002.1). Standard errors are in parentheses.

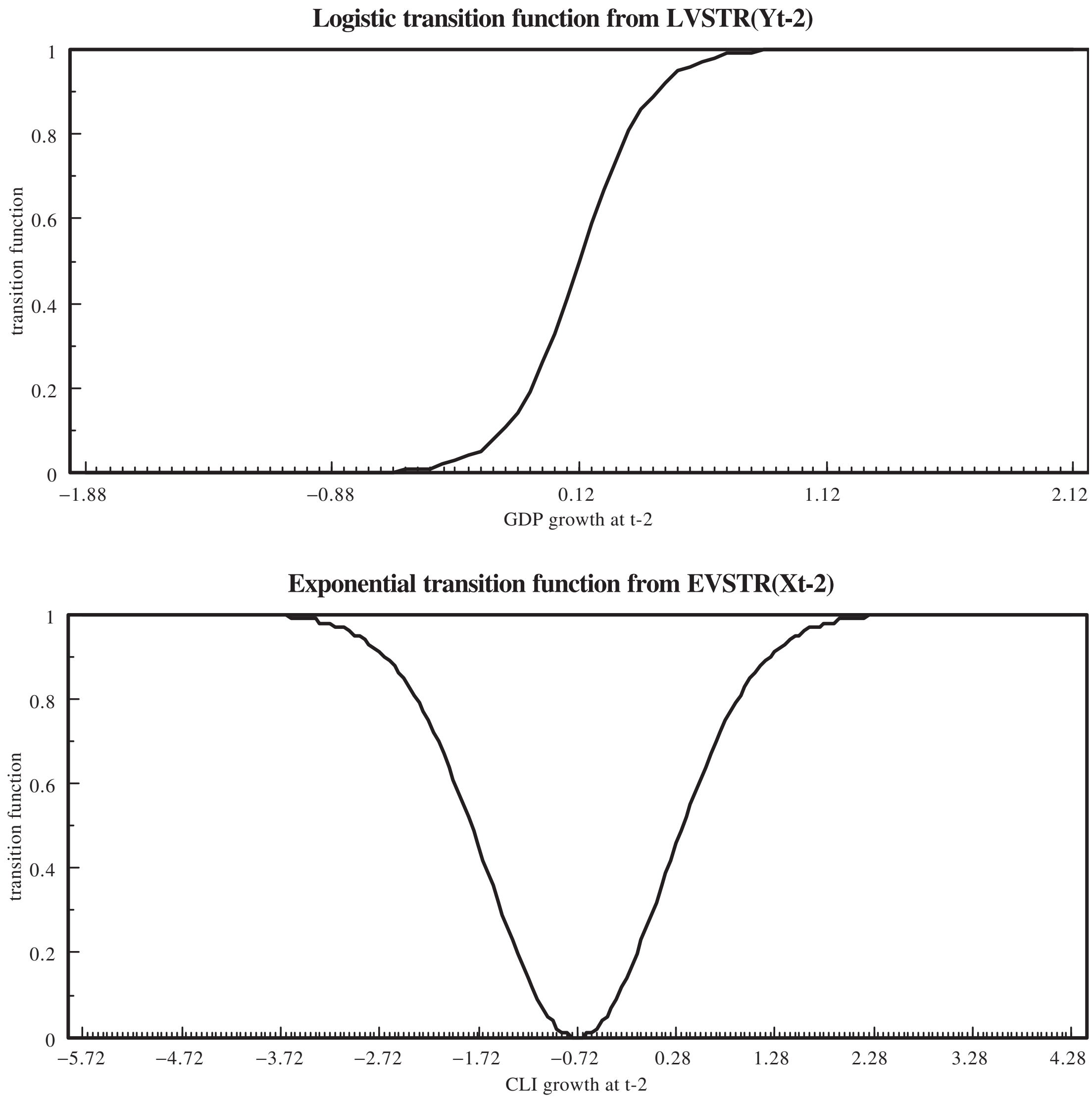

Figure 2. Estimated transition functions vs. transition variables

Note: This figure plots the transition functions of the CLI growth equation against y_{t-2} and x_{t-2}, respectively.

LVSTR-D(1) is 0.21 and the *p*-values of the equal forecast accuracy tests are always less than 0.001, revealing that the LVSTR(y_{t-2}) is also the best model to forecast the business-cycle phases.

Real-time forecasting

The prediction of output growth and business-cycle phases using leading indexes is usually evaluated with either in-sample or out-of-sample exercises. In both cases the forecasts are conducted with final revised values of the index: the former examines the forecasting accuracy using the entire series

Table VIII. Testing the adequacy of VSTR models

	Test SI				Test PC			Test NRN
	$r = 1$	$r = 2$	$r = 3$	$r = 4$	$k = 1$	$k = 2$	$k = 3$	
LVSTR(y_{t-2})	0.06	0.17	0.50	0.60	0.60	0.17	0.08	0.95
LVSTR-D(1)	0.15	0.33	0.59	0.17	0.31	0.08	0.25	0.92
EVSTR(x_{t-2})	0.31	0.01	0.05	0.01	0.01	0.03	0.06	0.05

Note: Each entry shows the p-values of serial independence of errors (SI), parameter constancy (PC) and no remaining nonlinearity (NRN) tests. Note that NRN tests have power against omission of important lags from the estimated model.

Table IX. (a) Certain positive and negative signal rates (CPR and CNR); (b) False positive and negative signal rates (FPR and FNR)

(a)	In-sample				Real-time			
	GDP		Business cycles		GDP		Business cycles	
	CPR	CNR	CPR	CNR	CPR	CNR	CPR	CNR
LVSTR(y_{t-2})	98.64	52.17	59.25	87.50	95.34	50.00	60.00	95.18
LVSTR-D(1)	97.97	26.08	14.81	29.86	93.02	58.33	0.00	9.63
EVSTR(x_{t-2})	96.62	34.78	...	...	91.86	50.00	...	...
(b)	**In-sample**				**Real-time**			
	GDP		Business cycles		GDP		Business cycles	
	FPR	FNR	FPR	FNR	FPR	FNR	FPR	FNR
LVSTR(y_{t-2})	14.28	7.00	27.27	5.26	40.00	6.81	33.76	4.81
LVSTR-D(1)	33.33	10.49	92.59	24.56	46.15	5.88	100.00	27.27
EVSTR(x_{t-2})	38.46	9.49	...	...	56.84	7.05	...	...

Note: 'In-sample' and 'real-time' refer to 1959.1–2002.1 and 1978.1–2002.2 respectively. Certain positive rates (certain negative rates) measure the percentage of quarters of estimated positive growth and expansions (negative growth and recessions) over the quarters of actual positive growth and NBER expansions (negative growth and recessions). False positive rates (false negative rates) measure the percentage of quarters that turn out to be actual periods of positive growth and expansions (negative growth and recessions) over the quarters of estimated negative growth and recessions (positive growth and expansions). Note that, in line with Stock and Watson (1992), an estimated probability of recession above 0.75 (below 0.25) is interpreted as a signal of recession (expansion).

and the latter evaluates the forecast accuracy using just a portion of the same indicator series. In contrast to these forecasting exercises, in this section I perform one-period-ahead real-time forecasts from 1978.1 to 2002.2, using a method that tries to mimic the information sets that were actually available at the historical date of each forecast. This implies that, prior to developing these forecasts, the following questions should be addressed: what series of CLI should be used to compute each forecast in real time, what is the order of cointegration between the real-time pairs of series of CLI and GDP, and at what time would a forecaster have recognized the volatility slowdown dated in the middle of the 1980s.

Table X. Predictive accuracy analysis

	In-sample			Real-time		
	LVSTR(y_{t-2})	LVSTR-D(1)	EVSTR(x_{t-2})	LVSTR(y_{t-2})	LVSTR-D(1)	EVSTR(x_{t-2})
			Output growth			
MSE	0.354	0.527	0.597	0.524	0.751	0.588
RMSE	1.000	0.671	0.592	1.000	0.698	0.891
DM	...	<0.001	<0.001	...	0.016	0.036
MDM	...	<0.001	<0.001	...	0.019	0.039
MGN	...	<0.001	<0.001	...	0.006	0.072
MR	...	<0.001	<0.001	...	0.002	0.052
Enc	...	0.913	0.366	...	0.151	0.077
			Business cycles			
TPE	0.088	0.418	...	0.088	0.295	...
RTPE	1.000	0.210	...	1.000	0.295	...
DM	...	<0.001	...	...	<0.001	...
MDM	...	<0.001	...	...	<0.001	...
MGN	...	<0.001	...	...	<0.001	...
MR	...	<0.001	...	...	0.001	...
Enc	...	0.001	...	...	0.180	...

Note: 'In-sample' and 'real-time' refer to 1959.1–2002.1 and 1978.1–2002.2. MSE and TPE refer to the mean squared error and turning point error measures. RMSE and RTPE are the relative MSE and TPE of the forecasting models over the LVSTR(y_{t-2}). Entries in rows three to six in both output growth and business-cycle analysis show the p-values of the following tests of equal forecast accuracy: DM (Diebold–Mariano), MDM (modified DM), MGN (Morgan–Granger–Newbold), and MR (Meese–Rogoff), all of them described in Diebold and Mariano (1995) and Harvey *et al*. (1997). The last rows in these studies present the p-values of the forecast encompassing test that is based upon the significance test of α_1 in the OLS regression

$$l_t - \hat{l}_{t,\text{LVSTR}} = \alpha_0 + \alpha_1 \hat{l}_{t,j} + \varphi_t$$

where l_t is either actual output growth or a dichotomous variable with ones at the official recessions, and $\hat{l}_{t,\text{LVSTR}}(\hat{l}_{t,j})$ is either its one-step-ahead forecast or the probability of recession computed from the LVSTR(y_{t-2}) (one of the other nonlinear models).

First, in line with Diebold and Rudebusch (1991), the real-time forecasts should reproduce the CLI data vector available at the quarter of each forecast. Towards the end of each month of the real-time exercise, the forecasters face the new issue of the leading index containing the provisional estimate of the previous month, the revisions of the preceding months, and the historical series from the beginning of the sample. However, the last figure is very preliminary and usually subject to large modifications in the first revision of the index issued in the following month. In line with Hamilton and Perez-Quiros (1996), I base the forecast of quarter $t + 1$ on the estimates of a VSTR that uses the GDP series with data until quarter t, and the CLI published two months after the end of this quarter that is transformed into quarterly observations by selecting the data corresponding to the last month of each quarter.[10]

Second, the analysis of the long-term relationship between GDP and CLI presents puzzling results in the literature. Using the series of outputs ending in 1993.3 and 1999.4 and the corresponding series

[10] In the real-time analysis the CLI series issued prior to 1988.10 starts in 1948.01, the series issued from this date until 1993.03 starts in 1952.11, and the series issued since 1993.04 starts in 1959.01.

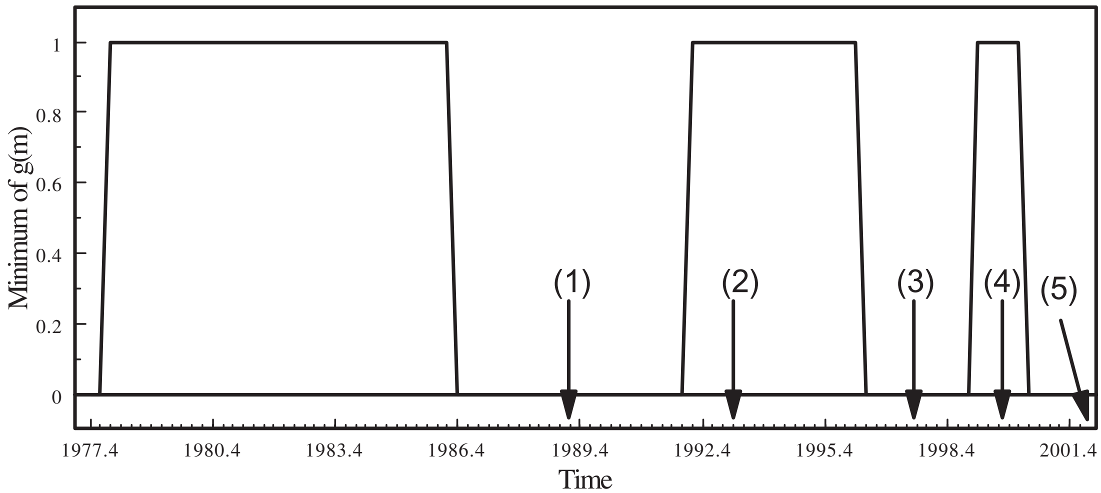

Figure 3. Real-time contegration analysis

Note: At any quarter k during the period 1997.4–2002.1, this figure plots the degree of cointegration from the Bierens (1997) nonparametric cointegration test applied to the GDP series ending in quarter k and the CLI data vector issued two months after this quarter. The numbers show the last observation of the GDP series used in the cointegration analysis of (1) Granger *et al*. (1993), (2) Hamilton and Perez-Quiros (1996), (3) Camacho (2000), (4) Huh (2002) and (5) the in-sample analysis of this paper.

of the leading index, Hamilton and Perez-Quiros (1996) and Huh (2002) conclude that GDP and CLI are cointegrated. This contrasts with Granger *et al*. (1993) Camacho (2000) and the in-sample analysis of this paper, where we fail to detect cointegration using the GDP series ending in 1989.2, 1997.4 and 2002.1 and their corresponding transformations of the CLI series respectively. Figure 3 tries to shed some light on these puzzling results by investigating the number of cointegrating relationships of GDP and CLI computed in real time. At any quarter, this figure plots the number of cointegrating vectors r that minimize the function $g_m(r)$ evaluated with GDP data until this quarter and the corresponding quarterly transformation of the CLI series issued two months after the end of this quarter. This figure confirms the switches in the order of cointegration previously documented by the literature. It is interesting to note that many of the switches of the order of cointegration coincide with the historical revisions in the CLI series, which may be a possible explanation of this puzzling phenomenon.[11] This is the case for the changes in the CLI definition produced in the middle of the 1980s and in the early 1990s, and the trend adjustment changes occurring the middle of the 1990s.[12]

Third, as outlined in the in-sample analysis, both Kim and Nelson (1999) and McConnell and Perez-Quiros (2000) have documented a reduction in the output volatility since 1984.1. However, these authors analyse the series of outputs ending in 1997.1 and 1999.2 so they have information sets that were not available during the first quarters of the real-time forecasting period. In order to compute the real-time forecasts, it is interesting to know at what time a forecaster would have

[11] The historical revisions of the CLI series are either statistical revisions (due to revisions in the components) or definitional revisions (index components are reselected and reweighted).
[12] A deeper analysis of this puzzling phenomenon is beyond the scope of this paper and is left for further research.

realized the breakpoint in the variance. For this attempt, at any quarter t of the real-time analysis, I compute the GMM estimates of the system

$$y_\iota = \mu + \phi y_{\iota-1} + \varepsilon_{1\iota} \tag{19}$$

$$\frac{\sqrt{\pi}}{2}|\hat{\varepsilon}_{1\iota}| = \alpha_1 D_{1\iota} + \alpha_2 D_{2\iota} + \varepsilon_{2i} \tag{20}$$

where ι refers to data of the period from 1959.1 to t. In this system, a constant, y_ι, $D_{1\iota}$ and $D_{2\iota}$ are the instruments for each period ι, and the dummies are

$$D_{1\iota} = \begin{cases} 0 \text{ if } \iota \le N \\ 1 \text{ if } \iota > N \end{cases} \quad \text{and} \quad D_{2\iota} = \begin{cases} 1 \text{ if } \iota \le N \\ 0 \text{ if } \iota > N \end{cases}$$

where N is the estimated breakpoint. For any quarter of the real-time analysis, Figure 4 uses the approximation suggested by Hansen (1997) to plot the p-values of the supremum test defined in Andrews (1993) and the exponential and average tests developed in Andrews and Ploberger (1994) of the null that $\alpha_1 = \alpha_2$. This figure reveals that even though there is some evidence to consider 1984.1 as a breakpoint in variance by the end of the 1980s (the p-values of the supremum and exponential tests are less than 0.05 during the period 1988.4–1990.3), a clear signal of the structural break does not appear until 1995.3 since only from this date do the three tests present p-values definitely below the standard critical value of 0.05.

Summarizing, at any quarter t of the period 1997.4–2002.1, I estimate the VSTR models with the series of GDP ending in this quarter and the corresponding quarterly transformation of the CLI data

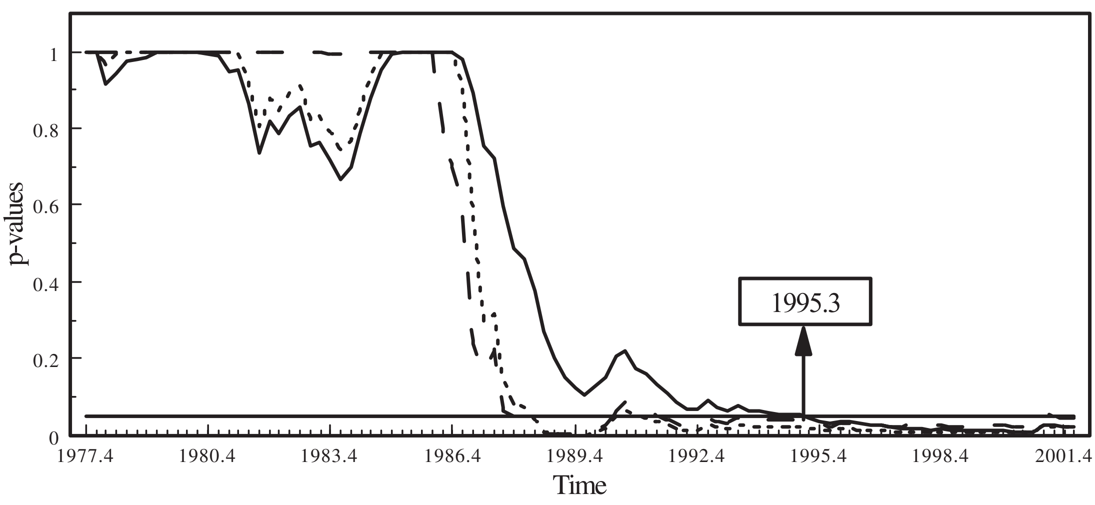

Figure 4. Real-time structural break analysis

Note: Using the approximation of Hansen (1997),this figure plots the p-values of the supremum (dashedline) test developed by Andrews (1993) and the exponential (dotted line) and average (straight line) tests suggested by Andrews and Ploberger (1994) applied to the GDP growth rate enlarged with one additional observation during the period 1997.4–2002.1.

vector issued two months after the end of the quarter with the following considerations. First, I compute the Bierens nonparametric cointegration test and I include the corresponding cointegrating errors in the VSTR specification in case of detecting cointegration. Second, I assume that since 1995.3 there was enough evidence to believe that any forecaster should realize the reduction in the volatility dated in 1984.1. Consequently, I consider this quarter as a breakpoint in the variance of the real-time VSTR estimates of any quarter since 1995.3. Finally, using the estimates available at quarter t, I compute the VSTR forecasts of both the output growth and the probability of recession for quarter $t + 1$.

Table IX analyses the models' accuracy to anticipate the output growth sign and the business-cycle phases. With respect to the output growth sign forecasting accuracy, this table shows inconclusive results: even though the LVSTR(y_{t-2}) is the model with the highest CPR (95.34) and lowest FPR (40.00), the model with highest CNR (58.33) and lowest FNR (5.88) is the LVSTR-D(1) model. With respect to the business-cycle forecasting accuracy, this table reveals that the LVSTR(y_{t-2}) model is unequivocally the best since this model presents more certain signals (CPR and CNR of 60.00 and 95.18) and less false signals (FPR and FNR of 33.76 and 4.81). In addition, Table X considers a more formal analysis of the forecasting performance of the selected nonlinear models. The ability of the LVSTR(y_{t-2}) model to forecast GDP growth is superior to the LVSTR-D(1) and EVSTR (x_{t-3}) models, with relativeMSE of 0.69 and 0.89 respectively. The null of equal forecasting accuracy between the LVSTR(y_{t-2}) model and the other competing models is rejected with any test for any model (p-values never greater than 0.05), with the exception of the MGN test for comparing forecasts of the LVSTR(y_{t-2}) and EVSTR(x_{t-2}) models with p-value of 0.07. Moreover, we cannot reject the null that the forecast of the LVSTR(y_{t-2}) model encompasses the other models' forecasts (p-values of 0.16 and 0.08 respectively). On the other hand, the LVSTR(y_{t-2}) model produces much lower forecast errors than the LVSTR-D(1) model to anticipate the NBER business-cycle phases (relative TPE of 0.29). The tests of equal forecast accuracy confirm that the forecasting improvements with the LVSTR(y_{t-2}) model are statistically significant (p-values less than or equal to 0.001). Moreover, the forecast encompassing test shows that the LVSTR(y_{t-2}) forecasts encompass the LVSTR-D(1) forecasts of the business-cycle fluctuations (p-value of 0.18).

The NBER has been dating US expansions and recessions for the last 50 years based on careful deliberations of the members of its Business Cycle Dating committee. The procedure requires the examination of numerous *ex post* data series that are believed to be coincidental with the aggregate economy. This approach implies that the NBER procedure cannot be used to forecast the future direction of the economy in real time since the committee's decisions about the business-cycle turning points are usually slow in coming.[13] As Boldin (1994) points out, the delay in the availability of the NBER schedule has motivated the use of several filtering rules to transform changes in the business-cycle leading indicators (that may capture the dampening of fluctuations over the cycle) into turning point predictions. In this context, Figure 5 assesses the degree to which the LVSTR(y_{t-2}) model anticipates the US business-cycle fluctuations in real time. As observed, the probabilities of recession (values of $1 - F$) tend to increase at the beginning of recessions (peaks) and to decrease around the end of recessions (troughs). The high correspondence with the NBER *ex post* dating procedure includes the last 2001 recession. In this recession, the NBER announces the March peak in November, with a delay of 8 months. However, in 2000.4 the logistic model forecast a probability of recession of 0.4 for 2001.1 and in 2001.1 forecast a probability of recession of almost 1 for 2001.2. In

[13] For example, in December 1992 the NBER announced that the 1990s recession trough occurred in March 1991, which implies a delay of 20 months.

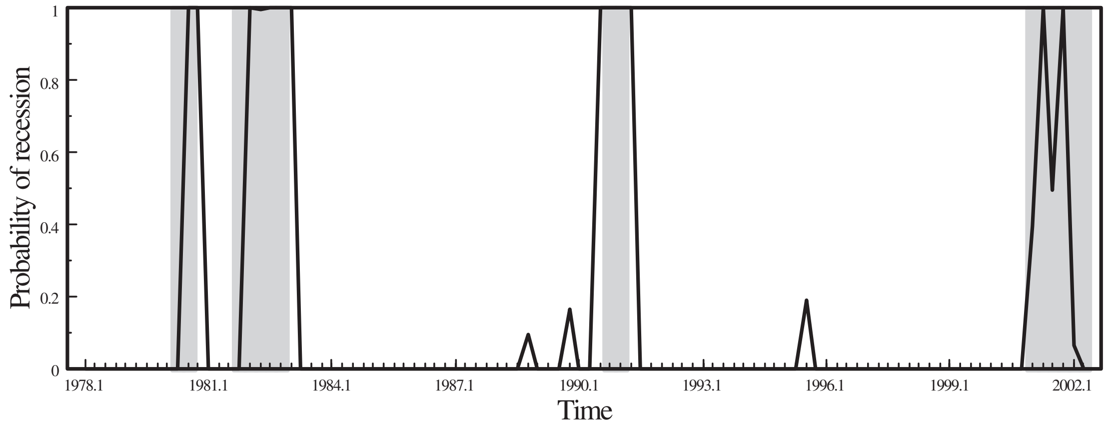

Figure 5. Real-time probabilities of recession

Note: This figure shows the real-time (period 1978.1–2002.2) probabilities of recession from the LVSTR(y_{t-2}). Shaded areas correspond to the NBER recessions.

addition, the NBER has not announced the trough yet but the LVSTR(y_{t-2}) model, using the information available in the last quarter of 2001, predicts that the last recession ended in the first quarter of 2002 (the probability of recession predicted for this quarter is 0.06), which coincides with the trough in January 2002 proposed by Chauvet (2002) using a dynamic factor model with regime switching methodology. This confirms that the VSTR models may be used as an additional filter to forecast the US business-cycle turning points in real time.

CONCLUSION

Prediction or even timely recognition of GDP growth rates and business-cycle turning points continues to be an exciting problem in econometrics, especially due to the current 2001 recession. In this paper I provide both theoretical and empirical support to consider an extension of the STAR models as an alternative filter to convert the Conference Board leading indicator movements into predictions of output growth and probabilities of recession.

From the theoretical point of view, I provide a vector autoregressive extension of the STAR models advocated by Granger and Teräsvirta (1993). First, following the maximum likelihood principle to estimate these models, I adapt to the multiple-equation framework the linearity and model selection tests. Second, I extend the single-equation tests of serial independence of errors, the tests of parameter constancy and the test of no remaining nonlinearity proposed by Eitrheim and Teräsvirta (1996).

From the empirical point of view, I examine with several model evaluation techniques the usefulness of the proposed models to forecast output growth and business-cycle phases. In order to compute real-time forecasts, I consider two preliminary questions. First, I find that the puzzling switches in the order of cointegration between the output and leading indicator series documented in the literature may be due to the historical redefinitions of the leading indicator series. Second, despite the widespread acceptance of the reduction in the volatility of output growth since the middle of the 1980s, I show that there are no strong signals to consider the structural break in real time until the mid-1990s. Finally, I find that a logistic nonlinear model whose transition variable is output growth with two periods of lag is the best model to forecast output growth and business-cycle phases.

APPENDIX: LM TEST OF SERIAL INDEPENDENCE OF ERRORS IN VSTR MODELS

It is well known that the test statistic (11) follows a limiting χ^2 distribution with as many degrees of freedom as the number of parameters which are assumed to be zero under the null. Thus, our target is to find an explicit definition for expressions appearing in this test.

Following the notation used in the text, let the $(1 \times r)$ vector $\Phi'_{ij} = (\Phi^1_{ij}, \ldots, \Phi^r_{ij})$ be the block ij of the matrix Φ' such that $U_t = \Phi' V_t$, with $i, j = y, x$. Let us collect the $4r$ elements of Φ in the column vector $\overline{\Phi} = (\Phi'_{yy}, \Phi'_{yx}, \Phi'_{xy}, \Phi'_{xx})'$, and let us define the vector $\vartheta = (\overline{\Phi}', \psi')'$. Hence, the null of serially uncorrelated errors may be expressed as $H_0: \overline{\Phi} = 0$. To derive the test, it is useful to left-multiply the model (9) by $I - \Phi(L)$, which leads to the likelihood function

$$l_t = C - \frac{1}{2}\ln|\Gamma| - \frac{1}{2}(\varsigma^2_{yt}\Gamma^{yy} + 2\varsigma_{yt}\varsigma_{xt}\Gamma^{yx} + \varsigma^2_{xt}\Gamma^{xx}) \tag{A.1}$$

where Γ^{ij} is block ij of the symmetric matrix Γ^{-1}, with $i, j = y, x$. On the one hand, to derive the estimates of the score under the null, it is useful to note that $\partial \underline{l}_t/\partial\Phi_{yj} = (\Gamma^{yy}\underline{\varsigma}_{yt} + \Gamma^{xy}\underline{\varsigma}_{xt})\underline{v}_{jt}$ and $\partial \underline{l}_t/\partial\Phi_{xj} = (\Gamma^{yx}\underline{\varsigma}_{yt} + \Gamma^{xx}\underline{\varsigma}_{xt})\underline{v}_{jt}$, with $j = y, x$. This leads to the $(4r \times 1)$ vector

$$\underline{m}_{\Phi} = \sum (\partial l_t/\partial\Phi'_{yy}, \partial\underline{l}_t/\partial\Phi'_{yx}, \partial\underline{l}_t/\partial\Phi'_{xy}, \partial\underline{l}_t/\partial\Phi'_{xx})' = \sum\left(\underline{\Gamma}^{-1}\underline{\varsigma}_t \otimes \underline{V}_t\right) \tag{A.2}$$

On the other hand, let us consider the expressions related to the Hessian matrix

$$\underline{M} = \frac{1}{T}\sum \partial^2 \underline{l}_t \Big/ \partial\vartheta\,\partial\vartheta' = \begin{pmatrix} \underline{M}_{\overline{\Phi}\overline{\Phi}} & \underline{M}_{\overline{\Phi}\Psi} \\ \underline{M}_{\Psi\overline{\Phi}} & \underline{M}_{\Psi\Psi} \end{pmatrix} \tag{A.3}$$

First, the upper left block is formed of 16 matrices $\underline{M}(\Phi_{ij}, \Phi_{hk}) = \frac{1}{T}\sum \partial^2 \underline{l}_t/\partial\Phi_{ij}\partial\Phi'_{hk}$ that may be estimated as $\underline{M}(\Phi_{ij}, \Phi_{hk}) = \underline{\Gamma}^{jk}\underline{v}_{it}\underline{v}'_{ht}$, with $i, j, h, k = x, y$. This leads us to consider that $\underline{M}_{\overline{\Phi}\overline{\Phi}} = \frac{1}{T}\sum\left(\underline{\Gamma}^{-1} \otimes \underline{V}_t \underline{V}'_t\right)$. Second, the upper right block $\underline{M}_{\overline{\Phi}\Psi} = \frac{1}{T}\sum \partial^2 \underline{l}_t \big/ \partial\overline{\Phi}\,\partial\Psi'$ may be approximated by $\frac{1}{T}\sum\left(\underline{\Gamma}^{-1}\underline{Z}'_t \otimes \underline{V}_t\right)$, and the lower left block by $\underline{M}'_{\overline{\Phi}\Psi}$. Third, the estimates of the lower right block $\underline{M}_{\Psi\Psi} = \frac{1}{T}\sum \partial^2 l_t/\partial\Psi\partial\Psi'$ are $\frac{1}{T}\sum\left(\underline{Z}_t \underline{\Gamma}^{-1} \underline{Z}'_t\right)$.

Finally, the estimates of $Z_t = (z_{yt}, z_{xt})$, where z_{it} is $\partial G_i/\partial\Psi_i = \partial G_i(A_t, \Psi_i)/\partial\Psi_i$, are defined as follows. It is easy to check that $\partial G_i/\partial\beta_i = A_t$ and that $\partial G_i/\partial\tilde{\beta}_i = A_t F_i(D_{ti})$. Expressions for $\partial G_i/\partial\gamma_i$ are $[1 + e^{-\gamma_i D_{ti}}]^{-2} D_{ti} e^{-\gamma_i D_{ti}} \tilde{\beta}'_i A_t$ and $D_{ti} e^{-\gamma_i D_{ti}} \tilde{\beta}'_i A_t$ in the case of logistic and exponential models, respectively. However, expressions of $\partial G_i/\partial g_i$ are zero for deviated models, and $-[1 + e^{-\gamma_i D_{ti}}]^{-2}\gamma_i e^{-\gamma_i D_{ti}}\tilde{\beta}'_i A_t$ (for logistic) and $-2\gamma_i D^{1/2}_{ti} e^{-\gamma_i D_{ti}}\tilde{\beta}'_i A_t$ (for exponential) nondeviated models.

ACKNOWLEDGEMENTS

I wish to express my gratitude to the seminar participants at the 20th International Symposium on Forecasting. I am indebted to Michael Boldin and Ataman Ozyildirim for so graciously sharing the data, and to the Spanish Ministry of Education for financial support (CICYT SEC2001-0855). I thank Gabriel Perez-Quiros for everything. Software (written in GAUSS) used in this paper can be obtained at www.um.es/econometria/Maximo. Errors are entirely my own responsibility.

REFERENCES

Andrews D. 1993. Test for parameter instability and structural change with unknown change point. *Econometrica* **61**: 821–856.

Andrews D, Ploberger W. 1994. Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* **62**: 1383–1414.

Auerbach A. 1982. The index of leading indicators: measurement without theory thirty years later. *Review of Economics and Statistics* **69**: 589–595.

Beine M, Candelon B, Sekkat K. 2003. EMU membership and business cycle phases in Europe: a Markov switching VAR analysis. *Journal of Economic Integration* **18**: 214–242.

Bierens H. 1997. Nonparametric cointegration analysis. *Journal of Econometrics* **77**: 379–404.

Blix M. 1999. Forecasting Swedish inflation with a Markov switching approach. Sveriges Riksbank Working Paper No. 76.

Boldin M. 1994. Dating turning points in the business cycle. *Journal of Business* **67**: 97–131.

Box G, Jenkins G. 1976. *Time Series Analysis: Forecasting and Control*, revised edn. Holden-Day: San Francisco.

Braun P, Zarnowitz V. 1989. Major macroeconomic variables and leading indexes: some estimates of their interrelations, 1886–1982. NBER Working Paper No. 2812.

Camacho M. 2000. Vector smooth transition regression models for US GDP and the composite index of leading indicators. Manuscript, Universitat Autònoma de Barcelona.

Camacho M, Perez-Quiros G. 2002. This is what the US leading indicators lead. *Journal of Applied Econometrics* **17**: 61–80.

Chauvet M. 1998. An econometric characterization of business cycles with factor structure and Markov switching. *International Economic Review* **39**: 969–996.

Chauvet M. 1999. Stock market fluctuations and the business cycle. *Journal of Economic and Social Measurement* **25**: 235–257.

Chauvet M. 2002. Is the U.S. recession over yet? CREFC Economic Letter No. 2002-01.

Chauvet M, Potter S. 2002. Markov switching in disaggregate unemployment rates. *Empirical Economics* **27**: 205–232.

Diebold F, Mariano R. 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* **13**: 253–263.

Diebold F, Rudebusch A. 1991. Forecasting output with the composite leading index: a real-time analysis. *Journal of the American Statistical Association* **86**: 603–610.

Diebold F, Rudebusch A. 1995. Measuring business cycles: a modern approach. *Review of Economics and Statistics* **78**: 67–77.

Eitrheim Ø, Teräsvirta T. 1996. Testing the accuracy of smooth transition autoregressive models. *Journal of Econometrics* **74**: 59–75.

Estrella A, Mishkin F. 1998. Predicting US recessions: financial variables as leading indicators. *Review of Economics and Statistics* **80**: 45–61.

Filardo A. 1994. Business cycle phases and their transitional dynamics. *Journal of Business and Economics Statistics* **12**: 279–288.

Filardo A. 1999. How reliable are recession prediction models. *Reserve Bank of Kansas City Quarterly Review* **Q2**: 35–55.

Fukuda S, Onodera T. 2001. A new composite index of coincident indicators in Japan: how can we improve the forecast performance? University of Tokyo Working Paper No. 101.

Granger C, Teräsvirta T. 1993. *Modelling Nonlinear Economic Relationships.* Oxford University Press: New York.

Granger C, Teräsvirta T, Anderson H. 1993. Modelling nonlinearities over the business cycle. In *Business Cycles, Indicators and Forecasting*, Stock J, Watson M (eds). University of Chicago Press: Chicago.
Hamilton J. 1989. A new approach to the economic analysis of nonstationary time series and the business cycles. *Econometrica* **57**: 357–384.
Hamilton J, Perez-Quiros G. 1996. What do the leading indicators lead? *Journal of Business* **69**: 27–49.
Hansen B. 1997. Approximate asymptotic p-values for structural change tests. *Journal of Business and Economic Statistics* **15**: 60–67.
Harvey D, Leybourne S, Newbold P. 1997. Testing equality of prediction mean squared errors. *International Journal of Forecasting* **13**: 287–291.
Huh H. 2002. GDP growth and the composite leading index: a nonlinear causality analysis for eleven countries. *Economics Letters* **77**: 93–99.
Kim C, Murray C. 2002. Permanent and transitory components of recessions. *Empirical Economics* **27**: 163–183.
Kim C, Nelson C. 1998. Business cycle turning points, a new coincident index and tests of duration dependence based on a dynamic factor model with Markov switching. *Review of Economics and Statistics* **80**: 188–201.
Kim C, Nelson C. 1999. Has the U.S. economy become more stable? A Bayesian approach based on a Markov-switching model of the business cycle. *The Review of Economics and Statistics* **81**: 608–616.
Krolzig HM. 1997. *Markov-Switching Vector Autoregressions. Modelling, Statistical Inference and Application to Business Cycle Analysis.* Lecture Notes in Economics and Mathematical Systems, Vol. 454. Springer: Berlin.
Krolzig HM. 2000. Predicting Markov-switching vector autoregressive processes. Nuffeld College Economics Discussion Paper No. 31.
Layton AP. 1998. A further test of the influence of leading indicators on the probability of US business cycle phase shifts. *International Journal of Forecasting* **14**: 63–70.
Luukkoven R, Saikkoven P, Teräsvirta T. 1988. Testing linearity against smooth transition autoregressive models. *Biometrika* **75**: 491–499.
McConnell M, Perez-Quiros G. 2000. Output fluctuations in the United States: what has changed since the early 80s? *American Economic Review* **90**: 1464–1476.
Pesaran M, Potter S. 1997. A floor and ceiling model of US output. *Journal of Economics, Dynamics and Control* **21**: 661–695.
Potter S. 1995. A nonlinear approach to US GNP. *Journal of Applied Econometrics* **10**: 109–125.
Rothman P, van Dijk D, Hans Franses P. 2001. A multivariate STAR analysis of the relationship between money and output. *Macroeconomic Dynamics* **5**: 506–532.
Stock J, Watson M. 1993. A procedure for predicting recessions with leading indicators: econometrics issues and recent experience. In *Business Cycles, Indicators and Forecasting*, Stock J, Watson M (eds). University of Chicago Press: Chicago.
Teräsvirta T. 1994. Specification, estimation and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association* **89**: 208–218.
Teräsvirta T. 1995. Modelling nonlinearity in U.S. gross national product 1889–1987. *Empirical Economics* **20**: 577–597.
Teräsvirta T, Anderson H. 1992. Characterizing nonlinearities in business cycles using smooth transition autoregression models. *Journal of Applied Econometrics* **7**: S119–136.
Tiao G, Tsay R. 1994. Some advances in nonlinear and adaptive modelling in time series analysis. *Journal of Forecasting* **13**: 109–131.
Tsay R. 1989. Testing and modelling threshold autoregressive processes. *Journal of American Statistical Association* **84**: 245–292.
Warne A. 2000. Causality and regime inference in a Markov switching VAR. Manuscript, Sveriges Riskbank Research Department.
Weise Ch. 1999. The asymmetric effects of monetary policy: a nonlinear vector autoregression approach. *Journal of Money, Credit and Banking* **31**: 85–108.

Author's biography:
Maximo Camacho completed his PhD in Economics at the Universitat Autonoma de Barcelona (advisor: Gabriel Perez-Quiros). His research has been published in various journals including Journal of Applied Econometrics, Journal of Economic Dynamics and Control, and Spanish Economic Review.

Author's address:
Maximo Camacho, Facultad de Economia y Empresa, Universidad de Murcia, 30100 Murcia, Spain.

