Heikki Kauppi

Yield-Curve Based Probability Forecasts of U.S. Recessions: Stability and Dynamics

Aboa Centre for Economics

Discussion Paper No. 57
Turku 2010

ISSN 1796-3133

Printed in Uniprint
Turku 2010

Heikki Kauppi

Yield-Curve Based Probability Forecasts of U.S. Recessions: Stability and Dynamics

Aboa Centre for Economics
Discussion Paper No. 57
June 2010

ABSTRACT

Various papers indicate that the yield-curve has superior predictive power for U.S. recessions. However, there is controversial evidence on the stability of the predictive relationship and it has remained unclear how the persistence of the underlying binary recession indicator should be taken into account. We show that a yield-curve based probit model treating the binary recession series as a nonhomogeneous first-order Markov chain sufficiently captures the persistence of the U.S. business cycles and produces recession probability forecasts that outperform those based on a conventional static model. We obtain evidence for instability in the predictive content of the yield-curve that centers on a structural change in the early 1980s. We conclude that the simple dynamic model with parameters estimated using data after the breakpoint is likely to provide useful probability forecasts of U.S. recessions in the future.

JEL Classification: C22, C25, E32, E37

Keywords: recession forecast, yield curve, dynamic probit models, parameter stability

Contact information

Heikki Kauppi
Department of Economics
Assistentinkatu 7
FI-20014 University of Turku
Finland
Email: heikki.kauppi@utu.fi

Acknowledgements

Financial support from the Academy of Finland and the Okobank Group Research Foundation is gratefully acknowledged.

1 Introduction

Predicting recessions is an important task for business and policy makers that condition their decisions on their assessment of the future state of the economy. A number of papers suggest that a simple probit model using predictive information from the yield-curve, the spread between long and short-term interest rates, provides useful probability forecasts of U.S. recessions, at least up to one-year horizon.[1] However, there is controversial evidence on the stability of the predictive relationship. Using Bayesian techniques, Chauvet and Potter (2002, 2005) find evidence of instability and suggest structural breaks may explain why standard yield-curve based probit forecasts have given somewhat weak signals of specific recessions. The evidence is mixed by the analysis of Estrella, Rodrigues and Schich (2003) whose breakpoint tests (of classical statistics) indicate no instability in the predictive relationship. Also, while a few papers suggest that recession forecasts should take the serial dependence of the business cycle into account,[2] it is still common to obtain recessions forecasts based on static model specifications.[3] This paper provides new evidence on the question of stability and attempts to clarify the role of dynamics for yield-curve based forecasts of U.S. recessions.

The starting point of the paper is to incorporate dynamics to the standard yield-curve based probit model by adding as a regressor a lagged value of the underlying binary recession indicator. Thus, effectively, the state of the economy is modeled by a nonhomogeneous Markov chain of order one, with transition probabilities changing by the value of the yield-curve. We also examine a larger class of probit models with richer forms of dynamics, and introduce a model extension that has not been considered in the previous literature. However, due to the simplicity of the observed dynamics in the binary recession series, we find no reliable evidence in favor of models with high-order dynamic dependencies. Hence, we conclude that the simple dynamic specification is sufficient for capturing the persistence of the U.S. business cycles. The simple dynamic model is then chosen as the main target for more detailed analysis of parameter stability.

[1] E.g., Estrella and Hardouvelis (1991), Estrella and Mishkin (1998), and Estrella, Rodrigues and Schich (2003).

[2] E.g., Chauvet and Potter (2005) and Kauppi and Saikkonen (2008).

[3] E.g., Rudebusch and Williams (2009).

The stability analysis starts with tests for breakpoints at known and unknown dates. Unlike previous breakpoint tests that examine the case of 'pure' structural change, the present analysis conducts also 'partial' breakpoint tests that allow only a subset of the model parameters to change under the alternative hypothesis. Altogether, the applied breakpoint tests suggests that especially the coefficient of the yield-curve has changed, while there is no evidence of instability of the remaining parameters. Furthermore, various test results suggest that a structural break has occurred in the early 1980s, which is consistent with apriori expectations. We then examine the performance of model variants assuming specific types of parameter changes. We find clear differences in parameter estimates between samples before and after the estimated breakpoint in the early 1980s. On the other hand, models that allow parameter changes at specific business cycles indicate that the predictive content of the yield-curve may have experienced a temporal structural change in conjunction with exceptionally short recession and expansion periods around 1980. Further evaluation of the stability of the predictive content of the yield-curve is conducted in the context of an analysis of out-of-sample forecasts.

The first part of the analysis of out-of-sample forecasts illustrates issues in the workings of recession probability forecasts and demonstrates what kind of forecasts are likely to be useful in practice. We show how the static model may yield misleading or implausible recession probability forecasts due to the fact that it neglects the serial dependence of the business cycle phases of the economy. In particular, the static model tends to exaggerate the predictive content of the yield-curve so as to produce too prompt and too frequent recession signals. By contrast, it is shown that the simple dynamic probit model produces probability forecasts that are in line with the actual uncertainty that surround specific recessions.

The second part of the analysis of out-of-sample forecasts is concerned with different assumptions on structural changes in the predictive content of the yield-curve. It shows that over the last 25 years the performance of recession probability forecasts at one-year horizon depends on the estimation sample. The forecast performance is better the more recent data are applied in the estimation of the model. Altogether, the analysis supports the view that there has been a structural break in the early 1980s, but such a break does no longer cause practical harm to the forecast performance. We conclude that the simple dynamic model that is estimated using data after the breakpoint is likely to provide apt probability forecasts for U.S. recessions in the future.

The rest of the paper is organized as follows. Section 2 lays out the empirical setting and introduces the applied models and forecast procedures. Appendix shows how the models are estimated by maximum likelihood and how robust standard errors are obtained. Section 3 reports estimation results for baseline models and a few alternative dynamic specifications. Stability analyses are conducted in Section 4. Section 5 examines out-of-sample forecasts under alternative settings. Section 6 concludes.

2 Statistical Framework and Methodology

2.1 The Starting Point

We seek to forecast values of the binary time series y_t that indicates the presence ($y_t = 1$) or absence ($y_t = 0$) of a recession in the U.S. at month t. As is common in the literature, we define y_t by using the NBER business cycle turning points. Hence, a recession period starts from an NBER 'trough' month and lasts until the month preceding the subsequent NBER 'peak' month.[4] All those months that are not included in a recession period are classified as expansion months.

The key predictor is the yield-curve, x_t, the spread between long- and short-term interest rates. We apply the most common choices: the ten year Treasury bond rate (constant maturity) for the long and the three month Treasury bill rate (secondary market) for the short rate.[5] Estrella and Trubin (2006) find that this definition of the yield-curve is superior in comparison with various alternative long- and short-term interest rates for forecasting recessions.

Figure 1 depicts the data over the sample period from January 1955 through February 2009. The dashed area indicate recession months (with $y_t = 1$), while the solid line is the yield-curve, x_t. It is seen that the yield-curve tends to decline in advance to recession periods. This indicates that the yield-curve has predictive content for future recessions.

[4] For the dates of the peaks and troughs see http://www.nber.org/cycles/.

[5] The raw data are available at http://www.federalreserve.gov/releases/h15/data.htm.

Various papers give (theoretical) explanations for the predictive relationship (see Estrella et al. (2003) and Estrella (2003)). It is customary to apply a probit model to translate the yield-curve into recession forecasts. The next section discusses a class of candidate model specifications, while the section after that derives corresponding forecast procedures.

2.2 Models

Let $\mathcal{F}_t = \{(y_t, x_t), (y_{t-1}, x_{t-1}), ...\}$ denote the past values of (y_s, x_s) up to month t. We assume that conditional on $\mathcal{F}_{t-1}$, y_t has the probability function

$$P(y_t|\mathcal{F}_{t-1}) = \Phi(z_t)^{y_t}(1 - \Phi(z_t))^{1-y_t}, \; y_t \in \{0, 1\} \tag{1}$$

where $\Phi(\cdot)$ is a cumulative distribution function and z_t is a function of variables in $\mathcal{F}_{t-1}$. As in probit models, we assume $\Phi(\cdot)$ is the cumulative standard normal distribution. The corresponding density function is denoted by $\phi(\cdot)$.

We consider forecast models that differ by the specification of the series z_t in (1). The standard static yield-curve based probit model assumes

$$z_t = \alpha + \beta x_{t-k} \tag{2}$$

where k, the lag of the regressor, is typically set equal to the forecast horizon h. A number of papers have applied the specification in (2) for forecasting U.S. recessions a year ahead (the most cited papers are Estrella and Hardouvelis (1991) and Estrella and Mishkin (1998)).

The specification in (2) has the potential weakness that it does not take the apparent serial dependence of the recession series into account. The simplest possible dynamic extension to (2) is given by

$$z_t = \alpha + \beta x_{t-k} + \gamma y_{t-1} \tag{3}$$

This specification is analogous to one applied by Kauppi and Saikkonen (2008) for forecasting U.S. recessions at the quarterly frequency.[6] It is easy to see that under (3), the binary series y_t is governed by a first-order Markov chain, with transition probabilities

[6] Kauppi and Saikkonen (2008) suggest that it may be sometimes beneficial to set $k > h$ rather than assume $k = h$.

varying as a function of the regressor x_{t-k}, the lagged yield-curve. The model (3) is regarded as the baseline dynamic probit model in what follows.

It is of course possible to consider more complicated dynamic dependencies than is captured by the specification in (3). One possibility is to add more lags of the binary series on the right hand side of (3). Such extensions result in higher order Markov chains. For example, the specification

$$z_t = \alpha + \beta x_{t-k} + \gamma_1 y_{t-1} + \gamma_2 y_{t-2} + \gamma_3 y_{t-1} y_{t-2} \tag{4}$$

leads to a nonhomogeneous Markov chain of order two.[7] Interestingly, it turns out that there is not enough variation in the present data for reliable estimation of the specification in (4). Due to the strong persistence of the U.S. recession series y_t, the regressors y_{t-1}, y_{t-2} and $y_{t-1}y_{t-2}$ are highly correlated, hence, attempts to estimate (4) collapse to numerical difficulties, with specific matrices being singular to working precision.[8] The problem does not disappear even if one of the coefficients γ_j is set to zero. An alternative strategy for increasing the order of the process is to add lags of the series z_t on the right hand side of (3) (see Kauppi and Saikkonen (2008) and Rydberg and Shephard (2003)). Although such models can break the Markov property in a parsimonious manner, they are nevertheless rich enough in dynamics in that their estimation faces similar problems as the case of the second order Markov specification. Thus, we do not consider such extensions in this paper.

In stead, we consider a new model formulation given by

$$z_t = \alpha + \gamma y_{t-1} + \upsilon_t \tag{5}$$

where

$$\upsilon_t = \lambda_1 \upsilon_{t-1} + ... + \lambda_p \upsilon_{t-p} + \beta x_{t-k}. \tag{6}$$

Suppose $\upsilon_t = 0$ for $t \leq 0$. Then it is easy to see that (5) and (6) yield

$$z_t = \alpha + \sum_{s=1}^{t} \rho_s \beta x_{t-k+1-s} + \gamma y_{t-1}, \text{ for } t \geq 1, \tag{7}$$

[7] Here the interaction term $y_{t-1}y_{t-2}$ is needed to obtain a fully saturated Markov chain.

[8] Due to the strong persistence of y_t, the regressors y_{t-1} and y_{t-2} take almost always the same value, that is, we have either $(y_{t-1}, y_{t-2}) = (0, 0)$ or $(y_{t-1}, y_{t-2}) = (1, 1)$ (See Figure 1). To capture the second-order Markov structure, we should observe the values $(y_{t-1}, y_{t-2}) = (1, 0)$ and $(y_{t-1}, y_{t-2}) = (0, 1)$ more frequently.

where $\rho_j = \lambda_1\rho_{j-1} + ... + \lambda_p\rho_{j-p}$, for $j > 1$, $\rho_1 = 1$, and $\rho_j = 0$ for $j < 1$.[9] In this specification, the dynamic impact of the regressor x_{t-k} is modeled in the fashion of an autoregressive distributed lag model. The specification allows a parsimonious modeling of the dynamic impact of x_{t-k}, while it maintains the simple first-order Markov property of the underlying binary series.

2.3 Forecast Procedures

Consider forecasting the value of y_t given that observations until date $t-h$, i.e., $\mathcal{F}_{t-h}$, are available.[10]

An optimal forecast of y_t in the mean square sense is the conditional expectation of y_t given $\mathcal{F}_{t-h}$:

$$E(y_t|\mathcal{F}_{t-h}) = P(y_t = 1|\mathcal{F}_{t-h}) \tag{8}$$

Clearly, when $h = 1$, we obtain (8) by setting $y_t = 1$ in (1).

Multiperiod ahead forecasts with $h \geq 2$ call for additional illustration. Define the vector notation

$$y_{t-m}^{t} = (y_{t-m}, y_{t-m+1}, ..., y_t) \text{ for } m = 0, 1, 2, ... \tag{9}$$

and the Cartesian product $B_m = \{1, 0\}^m$ for $m = 1, 2, ...$ That is, the set B_m contains all possible 2^m values that the m-vector y_{t-m}^{t} can take. Assume $k \geq h$ so that the value of the regressor x_{t-k} is known at the time of forecasting ($x_{t-k} \in \mathcal{F}_{t-h}$). Then, conditional on $\mathcal{F}_{t-h}$, the probability of the sequence y_{t-h+1}^{t-h+n} is

$$P(y_{t-h+1}^{t-h+n}|\mathcal{F}_{t-h}) = \prod_{j=1}^{n} P(y_{t-h+j}|\mathcal{F}_{t-h}, y_{t-h}^{t-h+j-1}), \quad n = 1, 2, ... \tag{10}$$

where $P(y_{t-h+j}|\mathcal{F}_{t-h}, y_{t-h}^{t-h}) = P(y_{t-h+j}|\mathcal{F}_{t-h})$, as $y_{t-h}^{t-h} = y_{t-h} \in \mathcal{F}_{t-h}$. Notice that the conditional probabilities $P(y_{t-h+j}|\mathcal{F}_{t-h}, y_{t-h}^{t-h+j-1})$ in (10) are readily obtained from (1) for a given specification of z_t. We have

$$P(y_t|\mathcal{F}_{t-h}) = \sum_{y_{t-h+1}^{t-1} \in B_{h-1}} P(y_{t-h+1}^{t-1}|\mathcal{F}_{t-h}) P(y_t|\mathcal{F}_{t-h}, y_{t-h+1}^{t-1}), \text{ for } h \geq 2 \tag{11}$$

[9] To ensure that the coefficients ρ_j in (6) decay to zero, as $j \to \infty$, one must assume that $\lambda_1, ..., \lambda_p$ in (6) are such that the roots of the characteristic equation $1 - \lambda_1 r - ... - \lambda_p r^p$ lie outside the unit circle.

[10] In practice, NBER business cycle turning points are announced with delay so that one is uncertain about whether the economy is currently in recession or not. This problem is discussed in Section 5.1.

Now, the optimal h-period ahead forecast in (8) is obtained by setting $y_t = 1$ in (11).

The formula in (11) expresses $P(y_t|\mathcal{F}_{t-h})$ as a probability weighted sum of conditional probabilities, each of which is conditional on a specific sequence of values $y_{t-h+1}, ..., y_{t-1}$ that can realize between periods $t-h$ and t. In the case of the dynamic specification in (3), the conditional probabilities $P(y_t|\mathcal{F}_{t-h}, y_{t-h+1}^{t-1})$ vary by y_{t-1}, that is, $P(y_t|\mathcal{F}_{t-h}, y_{t-h+1}^{t-1}) = P(y_t|\mathcal{F}_{t-h}, y_{t-1})$, while the weights $P(y_{t-h+1}^{t-1}|\mathcal{F}_{t-h})$ depend on the whole sequence $y_{t-h+1}, ..., y_{t-1}$. Thus, to obtain the optimal h-period ahead forecast for y_t one must compute the conditional probabilities $P(y_{t-h+1}^{t-1}|\mathcal{F}_{t-h})$ for all possible sequences (or paths) of $y_{t-h+1}, ..., y_{t-1}$. In the case of the static specification in (2) this is not needed, because we have $P(y_t|\mathcal{F}_{t-h}, y_{t-h+1}^{t-1}) = P(y_t|\mathcal{F}_{t-h})$, and thus, the optimal forecast is obtained directly (assuming $k \geq h$) as

$$P(y_t = 1|\mathcal{F}_{t-h}) = \Phi(\alpha + \beta x_{t-k})$$

Hence the formula for computing multiperiod ahead forecasts differs rather much between the dynamic and the static model. However, we will show below that the dynamics of the model need not matter that much for the actual empirical performance of this kind of multiperiod ahead forecasts. We will demonstrate that the dynamics of the model play a more significant role when the underlying multiperiod ahead forecast involves several future periods at the same time. Forecasts of the latter type are likely to be more interesting in practice.

For example, an investor or a policy maker may wish to forecast whether the current expansion will continue the following 12 months, say, rather than just forecast whether the economy is in a recession in a specific period in the future. Define the indicator e_{t-h}^{t} such that $e_{t-h}^{t} = 1$, if an expansion ongoing at time $t-h$ continues the next h periods (i.e., periods $t-h+1, ..., t$), and $e_{t-h}^{t} = 0$, otherwise. At time $t-h$, the optimal forecast of e_{t-h}^{t} is the conditional probability that the variables $y_{t-h+1}, ..., y_t$ are all zeros, that is,

$$P(e_{t-h}^{t} = 1|\mathcal{F}_{t-h}) = P(y_{t-h+1}^{t} = (0, ..., 0)|\mathcal{F}_{t-h}) \tag{12}$$

where we assume $y_{t-h} = 0$. The probability in (12) is easily computed by using the formula in (10). If the binary series is serially dependent, as it is in the present application, it makes a large difference for the forecast in (12) whether the applied forecast model is static or dynamic. This is illustrated in the empirical analysis below.

3 Baseline Estimation Results

This section reports estimation results for the baseline models and a few extensions. As in various previous papers, we focus on the situation where one wishes to forecast recessions at a one-year horizon. Accordingly, we apply the yield-curve x_{t-k} with $k = 12$ in our estimations. This choice of the lag of the yield-curve ensures that the one-year ahead forecast for y_t can be computed conditional on the yield-curve data observed until month $t - 12$.

Estimation results for the model in (3) with $k = 12$ are given in Table 1. The results here and below are obtained by using the maximum likelihood estimation procedures described in the appendix. The estimates of column (1) of Table 1 are for the static model that assumes (3) with the restriction $\gamma = 0$, while the results of column (2) are for the dynamic model without such restriction. In both columns, the parameter estimates are significantly different from zero at standard confidence levels.[11] A decrease in the yield-curve at month $t - 12$ increases the likelihood of a recession at month t. The estimation results of the dynamic probit model indicate positive serial dependence in the recession series: the likelihood of a recession at month t is much larger when the economy was in a recession at the previous month than it is otherwise. The pseudo R^2 reported in the table is a measure of the over-all fit of the model.[12] As the R^2 in an OLS regression, it lies between 0 and 1. According to the pseudo R^2, the dynamic probit yields more accurate in-sample predictions than the static model.

Figure 2 plots the estimated in-sample probabilities that the economy is in a recession state in a particular month from January 1955 to February 2009, for the models in columns (1) and (2) of Table 1. These are probabilities of recessions at t conditional on the value of the yield-curve at $t - 12$ and whether the economy is in recession or not at $t - 1$ (dynamic probit). Clearly, the figure shows that the dynamic probit model captures the

[11] The applied standard errors are robust to misspecification (see the appendix). The results are not sensitive to different choices of the kernel function and the bandwidth parameter applied in the computation of the covariance matrix estimator in (22).

[12] Denote by L_u the unconstrained maximum value of the likelihood function L and by L_c the corresponding maximum value under the constraint that all coefficients are zero except for the constant. The pseudo R^2 measure is defined as pseudo $R^2 = 1 - (\log(L_u)/\log(L_c))^{-2\log(L_c)/T}$, where T denotes the sample size (Estrella 1998).

recession series more accurately than the static probit model. However, it must be noted that Figure 2 does not yet illustrate how the models perform out-of-sample. In particular, multiperiod ahead forecasts based on the dynamic probit model cannot condition on the recession state at month $t-1$, and thus, the iterative forecast formulae of Section 2.3 must be applied. The performance of out-of-sample forecasts is analyzed in Section 5.

The above analysis shows that the simple dynamic model in (3) provides better in-sample performance than the standard static probit model. It is reasonable to ask whether alternative and more general dynamic specifications might yield even better in-sample performance than the simple first-order Markov chain specification. In the previous section, we noted that various model extensions that imply more complicated serial dependencies in the binary series than the simple first-order Markov structure cannot be estimated reliably (if at all) from the present data. This conclusion is supported by simulation experiments that indicate the applied estimation procedures produce reliable estimation results for more complicated models when the data are truly generated by models with higher order dynamics. To save space we do not report these results; they are available upon request.

To this end, we consider the in-sample performance of models where the dynamic impact of the yield-curve is formulated in an autoregressive manner (see equations (5) and (6)), as these types of models allow parsimonious modeling of the impact of the regressor, but simultaneously can maintain the simple first-order Markov dynamics in the binary series. Table 2 reports estimation results for a model that assumes (5) with $\gamma = 0$ and (6) with $p = 1$. The estimate of the autoregressive parameter is positive and statistically significant. This fact and the values of the Schwarz (1978) Bayesian information criterion (BIC) suggest that the model of column (1) of Table 2 may be a useful alternative to the baseline static model of column (1) of Table 1. However, in terms of the pseudo R^2, the two models do not differ a lot. Column (2) of Table 2 reports estimation results for a model that assumes (5) with $\gamma = 0$ and (6) with $p = 2$. Now, the coefficient of the second order autoregressive term is significant, while the first order term is not. Again, the in-sample fit is not markedly different from that of the baseline static model in column (1) of Table 1.

It is of interest to see whether the autoregressive formulation plays any role when the lagged recession series is allowed in (5). Columns (3) and (4) of Table 2 show estimation results for models that specify (6) with $p = 1$ and $p = 2$, respectively. In both cases, the autoregressive coefficients are no longer statistically significant. Otherwise, the remaining coefficient estimates are similar to the corresponding ones in Table 1. These observations indicate that the autoregressive terms do not improve the performance of the baseline dynamic model. The significant autoregressive terms that appear in the models of columns (1) and (2) of Table 2 may reflect the fact that these specifications do not have the lagged response y_{t-1} as a regressor. We conclude that the simple dynamic baseline model is superior in terms of its in-sample performance compared with a large number of alternative dynamic specifications. We will consider its out-of-sample performance later on, but now turn to examining its stability over time.

4 Stability Analysis

Various recent papers address the question whether the predictive content of the yield-curve for U.S. recessions has been stable over time. The parameter stability of the simple static probit model considered above is examined by Chauvet and Potter (2002), using Bayesian techniques, and Estrella, Rodrigues and Schich (2003), using classical statistical techniques. The former paper finds evidence for breakpoints, while the latter paper does not find evidence for parameter instability. Chauvet and Potter (2005) consider a dynamic probit model formulated through an autoregressive latent variable with business cycle specific error variances. Using Bayesian techniques, they find that the predictive content of the yield-curve for U.S. recessions is subject to structural breaks. This section contributes to these studies by examining the stability of the simple dynamic model of the previous section and by conducting breakpoint tests that help to see whether structural changes concern only a subset of parameters. We start by breakpoint tests.

4.1 Breakpoint Tests

We consider breakpoint tests for pure and partial structural change. The former case assumes that all parameters may change, while the latter case assumes that only a subset of the parameters may change.

To set up our tests, decompose the vector of parameters of the model as $\theta = (\delta', \eta')'$, where δ may be subject to a structural change, while η is regarded as unchanged throughout. Under this setting, the null hypothesis is

$$H_0\colon\ \delta = \delta_0 \text{ for all } t \geq 1 \tag{13}$$

and the alternative hypothesis of a one-time structural change is

$$H_1(\pi)\colon\ \delta = \begin{cases} \delta_1 & \text{for } t = 1, ..., \tau \\ \delta_2 & \text{for } t = \tau + 1, ... \end{cases}$$

where $\pi \in (0, 1)$ is related to the breakpoint τ by $\pi = \tau/T$. In the case of pure structural change, we have $\theta = \delta$ and there is no η.

Following techniques developed by Andrews (1993), our breakpoint tests are based on the Lagrange multiplier (LM) type test statistic (see equation (4.4) in Andrews (1993))

$$LM(\pi) = \frac{T}{\pi(1-\pi)} \overline{d}_\pi' \widehat{S}^{-1} \widehat{U}_{\theta\delta} \left(\widehat{U}_{\delta\theta} \widehat{S}^{-1} \widehat{U}_{\theta\delta} \right)^{-1} \widehat{U}_{\delta\theta} \widehat{S}^{-1} \overline{d}_\pi \tag{14}$$

where $\pi = \tau/T$ indicates the proportion of the data before the breakpoint. Here, the matrix $\widehat{S}$ is given in (22) in the appendix, while

$$\begin{aligned} \overline{d}_\pi &= \frac{1}{T} \sum_{t=1}^{\pi T} \frac{\partial l_t(\widehat{\theta})}{\partial \theta} \\ \widehat{U}_{\theta\delta} &= \frac{1}{T} \sum_{t=1}^{T} \frac{\partial^2 l_t(\widehat{\theta})}{\partial \theta \partial \delta'}, \widehat{U}_{\delta\theta} = \widehat{U}_{\theta\delta}' \end{aligned}$$

where the derivatives $\partial l_t(\widehat{\theta})/\partial\theta$ and $\partial^2 l_t(\widehat{\theta})/\partial\theta\partial\delta'$, respectively, are obtained from (17) and (18) in the appendix with θ replaced by the (full sample) ML estimator $\widehat{\theta}$.[13] The statistic in (14) is convenient in that it only entails computing the full sample estimate of θ.

[13] Notice that $\partial^2 l_t(\theta)/\partial\theta\partial\delta'$ is found from (18) by choosing the columns that correspond to the parameters in δ.

The LM statistics in (14) can be used to test whether a structural break has occurred at a known date. By the general results of Andrews and Fair (1998), it has asymptotic chi-squared distribution for fixed π, with degrees of freedom equal to the number of parameters in δ. To test for a break when the break date is unknown one can apply the sup of $LM(\pi)$:

$$\sup_{\pi \in \Pi} LM(\pi) \tag{15}$$

where the sup is taken over an interior portion of the full sample that excludes observations (a nonzero fraction of the total observations) at each end (that is, Π is chosen with closure in $(0, 1)$). The theory of Andrews (1993) shows that under the null hypothesis in (13) the statistic in (15) converges in distribution to the square of a standardized tied-down Bessel process. Critical values for this distribution can be obtained by simulation as in Andrews (1993), or by methods of Estrella (2003). In what follows, p-values are computed by the simulation procedure in Andrews (1993).

We turn to applying the above defined breakpoint tests to examine the stability of the parameters of the baseline models. In particular, we seek to examine whether the predictive content of the yield-curve has changed over time. Thus, testing the stability of the coefficient of the yield-curve is of particular interest. On the other hand, we note that potential changes in the predictive content of the yield-curve can well result in structural changes in other parameters as well. Hence, to gain as many insight as possible, we consider tests on (13) for all possible choices of δ, including the case of pure structural change $\delta = \theta$.

We report test results for known and unknown breakpoints. As to known breakpoint dates, we refer to Estrella et al. (2003) who argue that October 1979 and October 1982, both associated with specific shifts in the Federal Reserve's monetary policy practices, are plausible candidates for breakpoints in a yield-curve based forecasting model for U.S. recessions. The test results under different settings are given in Table 3. Panel (a) of the table is concerned with the static model, while panel (b) is concerned with the dynamic model. The rows vary by the composition of δ. In both panels, the first row reports results on tests for a pure structural change, while the remaining rows consider different types of partial structural change.

The results of Table 3 yield at least four interesting observations. First, all those tests that reject the null hypothesis of stability at 10% percent level involve the coefficient on the yield-curve. In other words, no test rejects the null hypothesis unless the coefficient on the yield-curve is allowed to change under the alternative hypothesis. Second, the strongest rejections (tests with the lowest p-values) occur in cases where the coefficient of the yield-curve alone can change under the alternative hypothesis. Third, the LM tests for the known breakpoint date of October 1982 tend to reject the null of stability, while the implied breakpoint date associated with the sup LM tests is usually very close to this date, in most cases November 1982. These observations suggest that the coefficient of the yield-curve may have changed, while the remaining parameters may have been stable. Also, the results support the idea that there is a single breakpoint in the early 1980s which is consistent with apriori expectations.

The next section tries to obtain a more detailed picture of the possible break in the predictive content of the yield-curve.

4.2 Models with Shifts in Parameters

In this section, we try to examine whether the baseline models could be extended to account for potential structural changes in the parameters. We start by considering models that are most closely in line with the above test results on a one-time change in the coefficient of the yield-curve. However, it must be noted that even if the applied test procedures are designed for detecting a one-time change, they have power against various other forms of structural changes. In particular, the sup LM statistic can be regarded as a general model stability test, as it has power against gradual or temporal changes in the parameters as well as structural changes at multiple breakpoints, which can locate at any date of the full sample (see Andrews (1993)). Given that at least theoretically our test outcomes might result from various types of breaks, we will also consider model extensions consistent with alternative forms of structural changes.

Table 4 reports estimation results for the baseline models based on two subsamples obtained by cutting the full sample in December 1982, the breakpoint date implied by the sub LM test statistic of the previous section.[14] Clearly, the coefficient estimate of the yield-curve is larger (in absolute value) in the second subsample, while the other parameter estimates are more or less similar in size across the two subsamples. It is interesting that the coefficient estimate of the yield-curve is not statistically significant in the model of column (3). This suggests that the predictive content of the yield-curve for recessions prior to early 1980s is rather weak, at least when twelve-month-ahead forecasts are considered. The fact that the yield-curve coefficient is significant in the static model (in column (1)) may reflect the lack of dynamics in the model.

As was noted above, the breakpoint tests of the previous section tend to have power against various forms of parameter changes. If there are multiple breakpoints, one possibility is that they are associated with business cycles as in the model of Chauvet and Potter (2005). A business cycle starts at the first month of an expansion period and lasts until the final month of the subsequent recession period. Following this definition, let D_{ct} denote business cycle specific indicator functions such that $D_{ct} = 1$ for the months of the business cycle c and $D_{ct} = 0$ otherwise. Table 5 reports the actual dates of the business cycles of the sample. Using the corresponding dummies we can augment the baseline models with interaction terms that allow specific coefficients to change by business cycle. Given the results from the above breakpoint tests, we focus on examining business cycle specific shifts in the coefficient of the yield-curve.

Table 6 reports estimation results for specifications where the coefficient of the yield-curve is allowed to change at one business cycle at the time. The estimate of the coefficient of the interaction term $x_{t-12} \cdot D_{ct}$ is not significant except for the business cycle from August 1980 to November 1982. We note that the final month of this particular business cycle is precisely the breakpoint month implied by the results on the sup LM test of the previous section. Also, the known breakpoint of October 1982 is within this business cycle. Recall again that even if the above applied breakpoint tests are designed for testing parameter stability against a one-time (permanent) change in parameters, they have power against temporal parameter changes. Hence the estimation results of Table 6 and the above breakpoint tests are in agreement.

[14] We obtain qualitatively similar results with alternative breakpoints that are close to December 1982.

The estimates in column (6) of Table 6 suggest that the coefficient of the yield-curve is positive during the business cycle from August 1980 to November 1982. This reverse sign of the estimated coefficient calls for an explanation. Notice that the business cycle in question is associated with an expansion period of only twelve months, the shortest one in the sample. Moreover, the preceding recession period (of the previous business cycle) happens to last only six months, and is also the shortest in the sample. Consider a twelve-month-ahead recession forecast made for August 1980, the first (expansion) month of the 1980-82 business cycle. This forecast is conditional on the value of the yield-curve in August 1979, which is only six months in advance to the previous recession starting in February 1980. Figure 1 shows that the yield-curve is at a low level in August 1979, so that it signals the recession in early 1980. But, given that the recession in 1980 lasts only until July 1980, the signal is wrong for August 1980, which is an expansion month. Altogether, a close inspection of the data (or Figure 1) indicates that during the business cycle 1980-82 the state of the economy at month t is often theoretically in disagreement with the value of the yield-curve at month $t-12$. These observations explain the estimation result in column (6) of Table 6.

The above notes brought up the fact that both the recession in early 1980 and the subsequent expansion were exceptionally short lived. It must be recalled that these recession and expansion periods are determined by the NBER business cycle committee. The short expansion period in 1980 appears to be controversial. The following quote from the general statement of the NBER business cycle committee is illustrative:[15]

> "The Committee applies its judgment .. and has no fixed rule to determine whether a contraction is only a short interruption of an expansion, or an expansion is only a short interruption of a contraction. The most recent example of such a *judgment that was less than obvious* was in 1980-1982, when the Committee determined that the contraction that began in 1981 was not a continuation of the one that began in 1980, but rather a separate full recession."

[15] Source: http://www.nber.org/cycles/general_statement.html

Given this statement it is likely that at least some of the contemporary market participants have had mixed assertions as to whether the U.S. economy experienced the short expansion period in 1980-81 or whether the period was a part of long recession. Such a confusion may also explain why the observed association between the lagged yield-curve and the NBER dated recession dummy is reversed in the early 1980s. These points suggest that the above breakpoint test results might derive from a temporary break rather than a permanent change in the predictive relationship between the yield-curve and the U.S. economy. We suspect that there may be both a temporary and a permanent structural break around early 1980s.

5 Forecast Performance

This section illustrates the forecast performance of the baseline models and models that assume breakpoints. We start by discussing issues that arise when recession forecasts are made out-of-sample.

5.1 Out-of-sample Forecasts in Practice

An issue with recession forecasting has to do with the fact that recession dating from NBER is typically available with a lag of six months or more. This means that one may be uncertain whether the economy is currently in an expansion or not. To illustrate different situations, suppose one wishes to predict whether the economy is turning into a recession at any month from $t-h-d$ to $t-d$ $(d \geq 0, h > 0)$ conditional on yield-curve data through month $t-h$ and knowing the state of the economy through month $t-h-d$. Here $t-h$ may be regarded as the month where the forecast is made, d as the information lag in recession dating, and h as the forecast horizon.

In practice, the forecast is made under an assumption about d. For example, in February 5, 2008, in a discussion at *Econbrowser*, Michael Dueker (from Federal Reserve) says that one can be reasonably certain that the NBER will not classify the fourth quarter of 2007 as a recessionary period and thus one can condition out-of-sample forecasts accordingly.[16] In this case, $d = 2$ (for monthly data). Later, it turned out that Michael Dueker was (just) right, as the NBER business cycle committee declared in December 2008 that the expansion period ended in December 2007, that is, the first (full) recession month was January 2008. Chauvet and Potter (2005) consider simulated out-of-sample forecasts made in advance to the recession that started in April 2001. For one of their cases, they argue that in March 2000 the public was certain that the economy was still in an expansion in December 1999, while there was lots of uncertainty about the state of the economy from January 2000 on. They then analyzed various forecasts under the assumption that $d = 3$ and $h = 15$.

The above examples indicate that the actual "information lag" of the forecaster tends to be shorter than the "publication lag" of the NBER business cycle dating. The NBER business cycle committee determines the business cycle turning points on the basis of various economic indicators. To avoid later revisions of the business cycle turning points, the committee tends to delay its decisions until the final figures of the most relevant economic indicators become available. By contrast, market participants make judgements about the current state of the economy using preliminary figures of various economic indicators. While the preliminary figures may be subject to later revisions, the forecasters may be rather successful in determining the state of the economy in real time. The risk that the forecaster makes a wrong judgement about the current state of the economy varies over time. In what follows, we abstract from this uncertainty and assume a situation where the forecaster knows the state of the economy at the time of forecasting (i.e., $d = 0$). The analysis is not sensitive to this assumption in that similar results hold under reasonable alternative settings such as $d = 3$.

5.2 Baseline Forecasts

This section considers forecasts based on the baseline models under the assumption that the model parameters are stable. We first illustrate the performance of standard one-year-ahead recession probability forecasts.

Figure 3 depicts twelve-month-ahead probability forecasts based on the dynamic model

[16] See http://www.econbrowser.com/archives/2008/02/predicting_rece.html.

(panel (a)) and the static model (panel (b)). These are "simulated" out-of-sample forecasts. That is, a recession forecast for month t is made conditional on observations on the yield-curve and the binary recession indicator through month $t-12$. Hence, for each month t, the applied forecast model is estimated using data through month $t-12$. Then, given the estimated model, the recession forecast for month t is computed by using the formula in (11) with $h = 12$. These simulated twelve-month-ahead forecasts are made for the last 25 years (January 1985 through February 2009) of the full sample. This period covers the three most recent recessions.

The predicted recession probabilities in Figure 3 are all below 0.5 for both models. Some of the predicted recession probabilities for actual recession months are smaller than those for some expansion months. The figure illustrates the fact that the yield-curve based twelve-month-ahead recession probability forecasts tend to be difficult to apply for making very sharp forecasts of the timing of a coming recession. This is not surprising given that the yield-curve evolves smoothly rather than in a discrete manner (See Figure 1). It is natural to assume that the yield-curve carries predictive power for the overall risk that the economy is turning into a recession, while it cannot pinpoint the precise date at which a recession realizes. Another interesting observation from Figure 3 is that the twelve-month-ahead recession probability forecasts are pretty similar across the two models. This observation might suggest that it is not so important to take the serial dependence of the binary series into account, but such conclusion is wrong. To see this, one must look at forecasts that involve several future periods at the same time.

One possibility is to forecast the probability that an expansion continues twelve months (by applying the formula in (12)). This type of forecast is likely to be more useful in practice than the month-by-month forecasts considered above (c.f., Chauvet and Potter (2005)). Figure 4 plots such probabilities over the period that is considered in Figure 3. At each month t in Figure 4, the line indicates the probability that the economy stays in an expansion from month $t+1$ to month $t+12$ conditional on being in an expansion at month t. The forecasts in Figure 4 differ clearly between the dynamic model (panel (a)) and the static model (panel (b)). The static model tends to produce very sharp recession calls years in advance to actual recessions, while the dynamic model produces more moderate forecasts. Consider the period in advance to the 2001 recession. Based on the static model (panel (b)), the predicted probability of continued expansion next twelve months is close to 0.1 already in early 1996 and again about 0.05 in 1998. Such forecasts are likely to prompt false or too early calls of recessions and may therefore give rise to adverse economic decisions. The dynamic model produces more moderate recession forecasts prior to the 2001 recession and seems to reflect the fact that at that time there was considerable uncertainty as to the future state of the economy. Indeed, various authors argue that the 2001 recession was very difficult to anticipate well in advance. Similar notes apply to the period in advance to the 1990-1991 recession. The static probit produces very sharp recession calls already in early 1987, while the forecast of the dynamic model is more moderate. The dynamic forecast is again consistent with the fact that the 1990-1991 recession is commonly regarded as difficult to forecast early in advance. The case of the most recent recession is also interesting. Both of the models seem to give stronger signals for this recession than they did for the preceding two recessions. Again, the static model gives its warning a year too early, while the signal of the dynamic model is more in line with the actual timing of the recession.

The above comparisons suggest that forecasts based on the dynamic model are superior to those based on the static model. However, one must recall that the considered forecasts are conditional probabilities. Basically, it is not clear how the underlying probability forecasts should be translated into actual zero-one recession forecasts. One can figure out various threshold rules that determine whether a given probability forecast is 1 or 0, but such thresholds are arbitrary. Thus, comparing probability forecasts based on such rules is problematic. In stead, it is common to assess probability forecasts by applying specialized measures. The most common one is the (half) Brier's (1950) quadratic probability score

$$QPS = \frac{1}{P}\sum(p_t - r_t)^2$$

where r_t is the realized value of an underlying binary series, p_t is the probability forecast for the event $r_t = 1$ and the summation is over P forecasts. The QPS varies between 0 and 1, with 0 implying perfect accuracy. The QPS is the probability-forecast analog of mean square error (MSE). It is motivated here, because the considered probability forecasts are derived so as to approximate the conditional probability that minimizes the population MSE.

In the case of the probability for continued expansion next twelve months, we have $r_t = e_t^{t+12}$ (see section 2.3) and p_t is given by the formula in (12). For the forecasts considered in Figure 4, the QPS is 0.09 for the dynamic and 0.21 for the static model. Hence, the dynamic model performs better in terms of the QPS. It is of interest to note that the corresponding QPS for the month-by-month (point) forecasts in Figure 3 is 0.077 for the dynamic model and 0.071 for the static model. Hence, the dynamic and the static model are essentially equal in accuracy when month-by-month forecasts are considered, the slight difference in favor of the static model is due to the jumps of the dynamic forecasts right after recessions (see Figure 3). Nevertheless, these measures illustrate that the dynamic model outperforms the static model when the probability of continued expansion is considered. Hence it matters that the serial dependence of the binary series is taken properly into account.

5.3 Forecasts Based on Different Estimation Samples

Above, in section 4, we obtained evidence that the predictive content of the yield-curve may have changed in the early 1980s. Specifically, various breakpoint tests suggest that the coefficient of the yield-curve may have changed in December 1982. Table 4 shows that it indeed makes a difference for the coefficient estimate of the yield-curve whether the model is estimated with data before or after the breakpoint in December 1982. Here we consider forecasts based on these different model estimates. We focus on forecasts for the probability of continued expansion next twelve months, and make these for the period from January 1985 to February 2009, as in Figure 4. Figure 5 shows forecasts based on models in columns (1) and (2) of Table 4, while Figure 6 shows forecasts based on models in columns (3) and (4) of Table 4. While the forecasts in Figure 6 are not actual out-of-sample forecasts, they help to assess whether the estimation sample matters for the forecast performance.

First, compare forecasts based on the static model (panel (b) in Figures 4, 5 and 6). The forecasts in Figures 4 and 5 are fairly similar, while the forecast in Figure 6 seem to be somewhat sharper than the two ones. This is consistent with the fact that the estimated coefficient of the yield-curve is larger in absolute value in column (2) than in column (1) of Table 4. In terms of the breakpoint tests above, this difference is not statistically significant. We do not make further analysis of the stability of the static model in this paper, because the model has been analyzed elsewhere and it nevertheless has the weakness that it does not account for serial dependence in the recession series.

Next, consider forecasts based on the dynamic model (panel (a) in Figures 4, 5 and 6). Now, there are larger differences between the forecasts, which is consistent with the breakpoint test results and the estimation results in Table 4. The forecasts in Figure 5 seem to be inferior to those in Figures 4 and 6. In particular, the forecasts in Figure 5 give rather low probabilities of continued expansion for periods where there should be no marked risk of a recession. This observation is not surprising given that the coefficient estimate of the yield-curve in column (3) of Table 4 is rather small (in absolute value) and is not statistically significant. For the dynamic forecasts, the QPS is 0.11 in Figure 5 and 0.06 in Figure 6. Hence, the dynamic forecasts in Figure 4 (with QPS 0.09) are less accurate that those in Figure 6, but more accurate than those in Figure 5. This suggests that the standard out-of-sample forecasting procedure in which the forecast model is estimated using data until the last available observation may reduce a part of potential forecast error deriving from changes in the predictive content of the yield-curve. To allow for more flexibility, one can drop observations in the distant past and apply estimation samples with a fixed number of the most recent observations. Using this 'rolling sample' approach to generate simulated out-of-sample forecasts that correspond to those in Figure 4 one obtains a QPS value of 0.08 and 0.07 when the sample size is fixed to 200 and 150, respectively. This suggests that rather simple procedures may deliver recession forecasts that adapt to potential changes in the strength of the predictive content of the yield-curve.

The above considerations support the view that the predictive content of the yield-curve has experienced a one-time change in the early 1980s. Looking at Figure 1, it seems plausible that the predictive content of the yield-curve has remained stable over the last 25 years. If this view holds, then the simple dynamic model estimated using the last 25 years' data should be useful for making probability forecasts for U.S. recessions in the future as well. To further test this view, one could still examine models that allow for more complicated structural changes. In section 4.2, we considered model specifications in which the coefficient of the yield-curve can differ at one business cycle at the time. To add more variation across business cycles, one can allow different coefficients at several (or even at all) business cycles. However, such specifications are estimated with considerable uncertainty. Basically, if one allows the coefficient of the yield-curve to change at several business cycles, the corresponding estimates turn out to have very large standard errors, usually none is statistically significant. Hence it is difficult to make conclusive statistical inferences about multiple breakpoints using this route. An alternative strategy is to apply Bayesian techniques that offer flexibility in the modeling of multiple breaks and provide ways to incorporate prior information to parameter estimation (see Geweke and Whiteman (2004) and Chauvet and Potter (2005)). It is an interesting topic for future research to investigate whether alternative approaches could refine the picture on structural changes in the dynamic model analyzed here.

6 Conclusion

Recent research provides mixed evidence on the stability, the dynamics and the overall performance of yield-curve based probit forecasts of U.S. recessions. To contribute to this literature, this paper analyzed the predictive performance and the stability of a simple dynamic probit model that treats the underlying recession indicator as a nonhomogeneous first-order Markov chain with transition probabilities changing as a function of the yield-curve. The analysis of the paper shows that the simple dynamic specification is successful in capturing the apparent serial dependence of the U.S. recession indicator and it provides more plausible recession probability forecasts than the static yield-curve based probit model that is commonly applied in the previous literature.

The stability analysis of the paper conducted tests for breakpoints at known and unknown dates. In contrast to previous studies, these tests examined the possibility of a structural change involving only a subset of model parameters. Interestingly, altogether, the test results indicate that the coefficient of the yield-curve alone may be subject to structural changes, while there is no evidence against the stability of the remaining model parameters. Furthermore, the evidence suggests a one-time break in the early 1980s. As the applied breakpoint tests are known to have power against various forms of structural changes, we examined the performance of model variants that allow alternative forms of structural changes. These analyses give additional support for a one-time structural change, but there is also evidence for the presence of a temporal break around the 1980-81 recession and the preceding expansion period, both of which were exceptionally short lived.

Finally, the paper conducted an analysis of out-of-sample performance of selected model specifications. The first part of the analysis showed how the static probit model tends to exaggerate the predictive content of the yield-curve so as to produce false or too prompt recession signals and that the dynamic probit model produces probability forecasts that are in line with the actual uncertainty that surround specific recessions. In particular, the results are consistent with the assessment that the 1990-1991 and 2001 recessions were inherently uncertain and thus difficult to forecast early in advance. The second part of the analysis was concerned with recession probability forecasts under different assumptions about the presence of a structural break in the predictive relationship. The out-of-sample results give additional support for the view that the predictive content of the yield-curve has changed in the early 1980s, while there is no evidence for further instability in the predictive relationship in the recent decades. Hence, the simple dynamic model that is estimated with a rolling sample scheme should produce apt probability forecasts of U.S. recessions in the future.

Appendix: Estimation Procedures

This section shows how the parameters of the models considered in the empirical analysis are estimated by maximum likelihood (ML) and how corresponding robust standard errors are obtained. The estimated models are nested in the specification given by equations (5) and (6).

One observes the series y_t and x_{t-k} for $t = 1, ..., T$ and the initial value y_0 is available.

Let $\theta = (\alpha, \beta, \gamma, \lambda_1, ..., \lambda_p,)'$. Then the log-likelihood function (conditional on the initial value) is

$$l(\theta) = \sum_{t=1}^{T} l_t(\theta) = \sum_{t=1}^{T} y_t \log \Phi(z_t(\theta)) + (1 - y_t) \log(1 - \Phi(z_t(\theta))) \tag{16}$$

where z_t is given in (7). The first derivative of the log likelihood function (the score function) is

$$\frac{\partial l(\theta)}{\partial \theta} = \sum_{t=1}^{T} \frac{\partial l_t(\theta)}{\partial \theta} = \sum_{t=1}^{T} \frac{[y_t - \Phi(z_t)]\, \phi(z_t)}{\Phi(z_t)[1 - \Phi(z_t)]} \frac{\partial z_t}{\partial \theta} \tag{17}$$

and the second derivative (the Hessian matrix) is

$$\begin{aligned} \frac{\partial^2 l(\theta)}{\partial \theta \partial \theta'} &= \sum_{t=1}^{T} \frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta'} \\ &= \sum_{t=1}^{T} \left[\frac{\phi(z_t) z_t}{(-\Phi(z_t))^{y_t} (1 - \Phi(z_t))^{1-y_t}} - \left(\frac{\phi(z_t)}{\Phi(z_t)^{y_t} (1 - \Phi(z_t))^{1-y_t}} \right)^2 \right] \frac{\partial z_t}{\partial \theta} \frac{\partial z_t}{\partial \theta'} \end{aligned} \tag{18}$$

Here $\partial z_t / \partial \theta$ is the vector of derivatives

$$\frac{\partial z_t}{\partial \theta} = \begin{bmatrix} \partial z_t / \partial \alpha \\ \partial z_t / \partial \beta \\ \partial z_t / \partial \gamma \\ \partial z_t / \partial \lambda \end{bmatrix} = \begin{bmatrix} 1 \\ \sum_{s=1}^{t} \rho_s x_{t-s+1} \\ y_{t-1} \\ \sum_{s=1}^{t} \frac{\partial \rho_s}{\partial \lambda} \beta x_{t-k+1-s} \end{bmatrix}$$

where $\partial \rho_s / \partial \lambda$ is the vector of derivatives $(\partial \rho_s / \partial \lambda_1, ..., \partial \rho_s / \partial \lambda_p)$ with

$$\frac{\partial \rho_s}{\partial \lambda_i} = \lambda_1 \frac{\partial \rho_{s-1}}{\partial \lambda_i} + ... + \lambda_p \frac{\partial \rho_{s-p}}{\partial \lambda_i} + \rho_{s-i}, \quad \frac{\partial \rho_j}{\partial \lambda_i} = 0, \ j \leq 1.$$

The ML estimator $\hat{\theta}$ of θ is obtained by maximizing the log-likelihood function in (16), or equivalently, by solving the first order conditions $\partial l(\theta) / \partial \theta = 0$ by applying standard algorithms (e.g., the Newton-Raphson). To enforce that $\lambda_1, ..., \lambda_p$ are such that the roots of the characteristic equation $1 - \lambda_1 r - ... - \lambda_p r^p$ lie outside the unit circle, it is convenient to reparametrize $\lambda_1, ..., \lambda_p$ in terms of partial correlations and then restrict these to lie within the interval $[-1, 1]$ (see Barndorff-Nielsen and Schou (1973) and Monahan (1984)).

Asymptotic theory for $\hat{\theta}$ is studied by Fokianos and Kedem (1998). They prove existence, consistency and asymptotic normality of $\hat{\theta}$ under regularity conditions. When the model is correctly specified, we have the result

$$T^{1/2}(\hat{\theta}-\theta) \xrightarrow{d} N(0, S(\theta)^{-1}), \tag{19}$$

where $S(\theta) = \text{plim}_{T\to\infty} T^{-1} \sum_{t=1}^{T} d_t d_t'$, with $d_t = \partial l_t(\theta)/\partial\theta$.

In practice, the applied forecasting model may be misspecified. For example, the applied lag of the yield-curve, x_{t-k}, may be wrong under the restriction $k \geq h$. Also, the dynamics of the model may not capture precisely the true form of serial dependence of the binary series. Alternative dynamic binary response models include the autoregressive latent variable formulation of Chauvet and Potter (2005). Finally, the distribution function $\Phi(\cdot)$ needs not be normal; it might be logistic or some other distribution. Given that there are various possibilities for model misspecification, it is useful to consider the standard extension of (19) given by

$$T^{1/2}(\hat{\theta}-\theta_*) \xrightarrow{d} N(0, U(\theta_*)^{-1}S(\theta_*)U(\theta_*)^{-1}), \tag{20}$$

where

$$U(\theta_*) = -\underset{T\to\infty}{\text{plim}}\, T^{-1}\sum_{t=1}^{T} D_t$$

with $D_t = \partial^2 l_t(\theta)/\partial\theta\partial\theta'$, and θ_* is a value in the parameter space of θ assumed to maximize the probability limit of $T^{-1}l(\theta)$ (for details, see Section 9.3 of Davidson (2000)). In the case of a correctly specified model $S(\theta) = U(\theta)$ and consistent estimators of this matrix are given by both $T^{-1}\sum_{t=1}^{T} \hat{d}_t \hat{d}_t'$, where $\hat{d}_t = \partial l_t(\widehat{\theta})/\partial\theta$, and

$$\widehat{U} = \widehat{U}(\hat{\theta}) = T^{-1}\sum_{t=1}^{T} \widehat{D}_t \tag{21}$$

where $\widehat{D}_t = \partial^2 l_t(\widehat{\theta})/\partial\theta\partial\theta'$. In the case of a misspecified model, the estimator $\widehat{U}(\hat{\theta})$ still estimates the matrix $U(\theta_*)$ consistently but consistent estimation of the matrix $S(\theta)$ must account for potential serial dependence in the derivatives d_t. A general estimator is given by

$$\widehat{S} = \widehat{S}(\hat{\theta}) = T^{-1}\sum_{t=1}^{T} \hat{d}_t \hat{d}_t' + T^{-1}\sum_{j=1}^{T-1} w_{Tj} \sum_{t=j+1}^{T} (\hat{d}_t \hat{d}_{t-j}' + \hat{d}_{t-j}\hat{d}_t'), \tag{22}$$

where $w_{Tj} = k(j/m_T)$ for an appropriate function $k(x)$ referred to as a kernel function. The quantity m_T is the so-called bandwidth which for consistency is assumed to tend to infinity with T but at a slower rate. In the empirical application, the Parzen kernel function (see Davidson (2000, p. 227)) is applied and, following the suggestion of Newey and West (1994), m_T is selected according to the rule $m_T = int(4(T/100)^{2/9})$, where $int(x)$ returns the integer part of x.

Using the estimators $\widehat{U}$ and $\widehat{S}$ in conjunction with the asymptotic results (19) and (20) one can construct standard Wald tests for hypotheses on the parameter vector θ. In particular, approximate standard errors for the components of the ML estimator $\hat{\theta}$ can be obtained in the usual way from the diagonal elements of the matrix $\widehat{U}^{-1}\widehat{S}\widehat{U}^{-1}$ or, if a correct specification is assumed, from the diagonal elements of the matrix $\widehat{U}^{-1}$.

References

Andrews, D. W. K. (1993) "Tests for parameter instability and structural change with unknown change point," *Econometrica*, 61, 821–856.

Andrews, D. W. K. (2003) "Tests for parameter instability and structural change with unknown change point: Corrigendum," *Econometrica*, 71, 395–397.

Andrews, D. W. K., and R. C. Fair (1988) "Inference in non-linear econometric models with structural change," *Review of Economic Studies*, 55, 615–640.

Barndorff-Nielsen, O., and G. Schou (1973) "On the parametrization of autoregressive models by partial autocorrelations," *Journal of Multivariate Analysis*, 3, 408–419.

Chauvet, M., and S. Potter (2002) "Predicting a recession: evidence from the yield curve in the presence of structural breaks," *Economics Letters*, 77, 245–253.

Chauvet, M., and S. Potter (2005) "Forecasting recessions using the yield curve," *Journal of Forecasting*, 24, 77–103.

Davidson, J. (2000) "Econometric theory," Oxford, UK: Blackwell Publishers Ltd.

Estrella, A. (1998) "A new measure of fit for equations with dichotomous dependent variables," *Journal of Business & Economic Statistics*, 16, 198–205.

Estrella, A. (2003) “Critical values and p values of Bessell process distributions: computation and application to structural break tests,” *Econometric Theory*, 19, 1128–1143.

Estrella, A., and G. Hardouvelis (1991) “The term structure as a predictor of real economic activity,” *Journal of Finance*, 46, 555–576.

Estrella, A., and F. S., Mishkin (1998) “Predicting U.S. recessions: financial variables as leading indicators,” *Review of Economics and Statistics*, 80, 45–61.

Estrella, A., A. P. Rodrigues, and S. Schich (2003) “How stable is the predictive power of the yield curve? Evidence from Germany and the United States,” *Review of Economics and Statistics*, 85, 629–644.

Estrella, A., and M. R. Trubin (2006) “The yield curve as a leading indicator: some practical issues,” Federal Reserve Bank of New York, *Current Issues in Economics and Finance*, 12, 1–7.

Fokianos, K., and B. Kedem (1998) “Prediction and classification of non-stationary categorical time series,” *Journal of Multivariate Analysis*, 67, 277–296.

Geweke, J., and C. Whiteman (2006): “Bayesian forecasting,” in Elliott, G., Granger, W. J., and Timmerman, A. (eds.) Handbook of Economic Forecasting, North-Holland.

Kauppi, H., and P. Saikkonen (2008): “Predicting U.S. recessions with dynamic binary response models,” *Review of Economics and Statistics*, 90, 777–791.

Monahan, J. F. (1984) “A note on enforcing stationarity in autoregressive -moving average models,” *Biometrica*, 71, 403–404.

Newey, W. K., and K. D. West (1994) “Automatic lag selection in covariance matrix estimation,” *Review of Economic Studies*, 61, 631–653.

Rudebusch, G. D., and J. C. Williams (2009) “Forecasting recessions: the puzzle of the enduring power of the yield curve,” *Journal of Business and Economic Statistics*, 27, 492–503.

Rydberg, T. H., and N. Shephard (2003) "Dynamics of trade-by-trade price movements: decomposition and models," *Journal of Financial Econometrics*, 1, 2–25.

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

Table 1. Estimation Results for Baseline Probit Models

	(1) Static		(2) Dynamic	
Predictor	coeff.	s.e.	coeff.	s.e.
Constant	$-.37$	.15	-1.76	.17
Yield-curve, x_{t-12}	$-.80$	.12	$-.33$	.13
Recession, y_{t-1}	—		3.23	.20
Pseudo R^2	.23		.69	
Log-likelihood	-201.6		-69.2	
BIC	208.0		78.9	

Notes: The models are estimated using monthly data from January 1955 through February 2009 (650 observations). The reported standard errors (s.e.'s) are robust to misspecification and are computed with procedures described in the appendix.

Table 2. Estimation Results for Probit Models with Autoregressive Terms

	(1)		(2)		(3)		(4)	
Predictor	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.
Constant	$-.30$	.19	$-.28$	.20	-1.8	.18	-1.8	.18
Recession, y_{t-1}	—		—		3.2	.20	3.2	.20
Yield-curve, x_{t-12}	$-.38$	.20	$-.54$	.19	$-.37$	.16	$-.44$	.15
Autoreg. lag 1, υ_{t-1}	.57	.25	$-.05$	.20	$-.14$	.45	$-.52$	.42
Autoreg. lag 2, υ_{t-2}	—		.46	.22	—		.16	.24
Pseudo R^2	.24		.24		.69		.69	
Log-likelihood	-198.1		-197.4		-69.2		-69.2	
BIC	207.8		210.4		82.2		85.4	

Notes: The models are given by equations (5) and (6), and are estimated using monthly data from January 1955 through February 2009. The reported standard errors (s.e.'s) are robust to misspecification and are computed with procedures described in the appendix.

Table 3. Tests for Breakpoints at Unknown and Known Dates

	supLM			LM (79:10)		LM (82:10)	
(a) Static Probit							
$\delta = (\alpha, \beta)$	4.97	[.58]	(69:12)	3.52	[.17]	1.04	[.59]
$\delta = \beta; \eta = \alpha$	4.16	[.36]	(80:07)	**3.51**	[.06]	.59	[.44]
$\delta = \alpha; \eta = \beta$	4.61	[.30]	(69:12)	.69	[.41]	.10	[.76]
(b) Dynamic Probit							
$\delta = (\alpha, \beta, \gamma)$	11.23	[.15]	(82:11)	1.07	[.79]	**8.90**	[.03]
$\delta = (\alpha, \beta); \eta = \gamma$	**11.16**	[.06]	(82:11)	.95	[.62]	**8.67**	[.013]
$\delta = (\alpha, \gamma); \eta = \beta$	2.07	[.98]	(69:12)	.58	[.75]	.09	[.96]
$\delta = (\beta, \gamma); \eta = \alpha$	**10.76**	[.08]	(82:11)	.28	[.87]	**8.58**	[.014]
$\delta = \beta; \eta = (\alpha, \gamma)$	**9.90**	[.03]	(82:11)	.19	[.67]	**7.40**	[.007]
$\delta = \gamma; \eta = (\alpha, \beta)$	1.27	[.94]	(80:08)	.005	[.94]	.08	[.78]
$\delta = \alpha; \eta = (\beta, \gamma)$	1.10	[.97]	(81:07)	.33	[.57]	.07	[.60]

Notes: The results are obtained for the sample from January 1955 through February 2009. The model is given by $z_t = \alpha + \beta x_{t-12} + \gamma y_{t-1}$ with restriction $\gamma = 0$ in panel (a), and no restriction in panel (b). The first column indicates parameters δ that are allowed to change under the alternative hypothesis, and parameters η that are assumed constant throughout (see (13) in the text). "sup LM" refers to the test statistic in (15) for one breakpoint with unknown date, computed with $\Pi = (.15, .85)$ so that 15% of the sample is dropped at each end. The p-value of the test is given in square brackets, while the implied breakpoint date is given in parentheses. "LM (Oct-79)" and "LM (Nov-82)" refer to the LM statistics in (14) with known breakpoint dates (denoted in parentheses). The p-value of the LM tests are given in square brackets. In the case of the sup LM statistic, the p-values are obtained by simulating the null distribution by applying the procedure of Andrews (1993), with the exception that 100000 repetitions are used as in Andrews (2003). The p-values of the LM statistics are from χ^2_{df}-distribution, where $df = \dim(\delta)$. Statistics that are significant at 10% level are written in bold face.

Table 4. Sub-Sample Estimation Results for Baseline Probit Models

	Static				Dynamic			
	(1)		(2)		(3)		(4)	
Predictor	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.
Constant	$-.43$	.17	$-.05$	.31	-1.90	.23	-1.40	.28
Yield-curve, x_{t-12}	$-.67$	.13	-1.25	.35	$-.11$	.29	$-.91$	.41
Recession, y_{t-1}	—		—		3.33	.19	3.22	.40
Pseudo R^2	.18		.26		.73		.64	
Sample	Jan55-Nov82		Dec82-Feb09		Jan55-Nov82		Dec82-Feb09	

Notes: The models are estimated as in Table 1, but using the sub-sample observations indicated at the final row.

Table 5. Business Cycles

Business cycle	Expansion		Recession	
	First month	Last month	First month	Last month
55-58	Jan-55	Aug-57	Sep-57	Apr-58
58-61	May-58	Apr-60	May-60	Feb-61
61-70	Mar-61	Dec-69	Jan-70	Nov-70
70-75	Dec-70	Nov-73	Dec-73	Mar-75
75-80	Apr-75	Jan-80	Feb-80	Jul-80
80-82	Aug-80	Jul-81	Aug-81	Nov-82
82-91	Dec-82	Jul-90	Aug-90	Mar-91
91-01	Apr-91	Mar-01	Apr-01	Nov-01
01-09	Dec-01	Dec-07	Jan-08	Feb-09

Notes: The month of the first (the last) business cycle is given by the first (the last) month of the sample period. The actual starting month of the first business cycle is June 1954. The ending month of the last business cycle is sometimes after February 2009, but the NBER Business Cycle Committee has not determined the actual month by May 2010.

Table 6. Estimation Results for Probit Models with Temporary Shifts in Parameters

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Business cycle	55-58	58-61	61-70	70-75	75-80	80-82	82-91	91-01	01-09
Regressor									
Constant	-1.77	-1.78	-1.69	-1.77	-1.76	-1.59	-1.78	-1.71	-1.72
	(.18)	(.17)	(.18)	(.17)	(.18)	(.18)	(.17)	(.17)	(.18)
Recession, y_{t-1}	3.23	3.22	3.22	3.24	3.22	3.16	3.28	3.20	3.22
	(.21)	(.20)	(.22)	(.21)	(.21)	(.23)	(.22)	(.21)	(.21)
Yield-curve, x_{t-12}	$-.33$	$-.35$	$-.31$	$-.35$	$-.32$	$-.56$	$-.27$	$-.30$	$-.33$
	(.14)	(.14)	(.14)	(.14)	(.15)	(.19)	(.16)	(.14)	(.15)
$x_{t-12} \cdot D_{ct}$	.08	.44	-2.6	.22	$-.06$	.81	$-.54$	-1.07	.02
	(.97)	(.36)	(3.2)	(.39)	(.71)	(.28)	(.99)	(.98)	(1.06)
Pseudo R^2	.686	.689	.701	.687	.686	.702	.690	.693	.686
Log-likelihood	-69.20	-68.45	-65.37	-68.95	-69.21	-65.11	-68.11	-67.49	-69.22
BIC	82.16	81.40	78.32	81.90	82.16	78.06	81.07	80.44	82.17

Notes: The models are estimated using monthly data from January 1955 through February 2009 (650 observations). The applied business cycle specific indicator (D_{ct}) varies by column. The second row indicates the starting and ending years of the given business cycle. The numbers in parentheses are misspecification robust standard errors of the coefficient estimates and are computed with procedures described in the appendix.

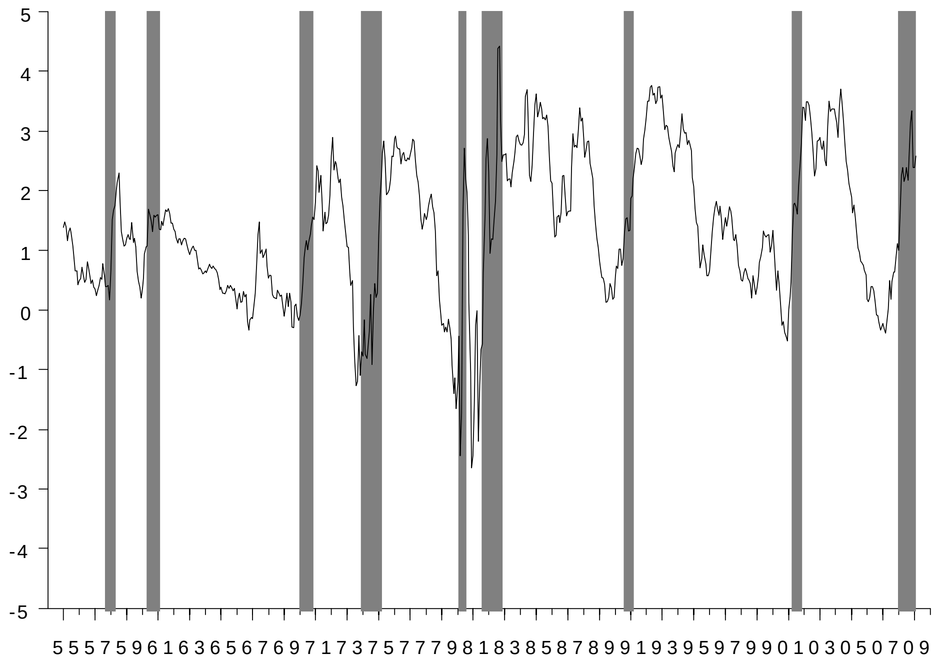

Figure 1: The Yield-Curve (the shaded area indicate NBER-dated recessions)

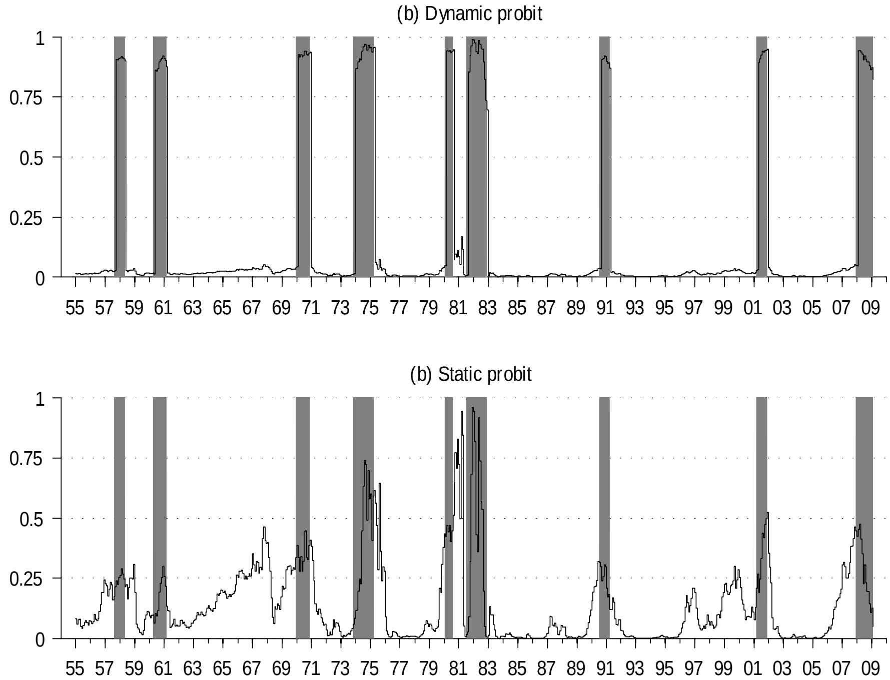

Figure 2: Probability of Recession, In-sample Prediction (the shaded area indicate NBER-dated recessions)

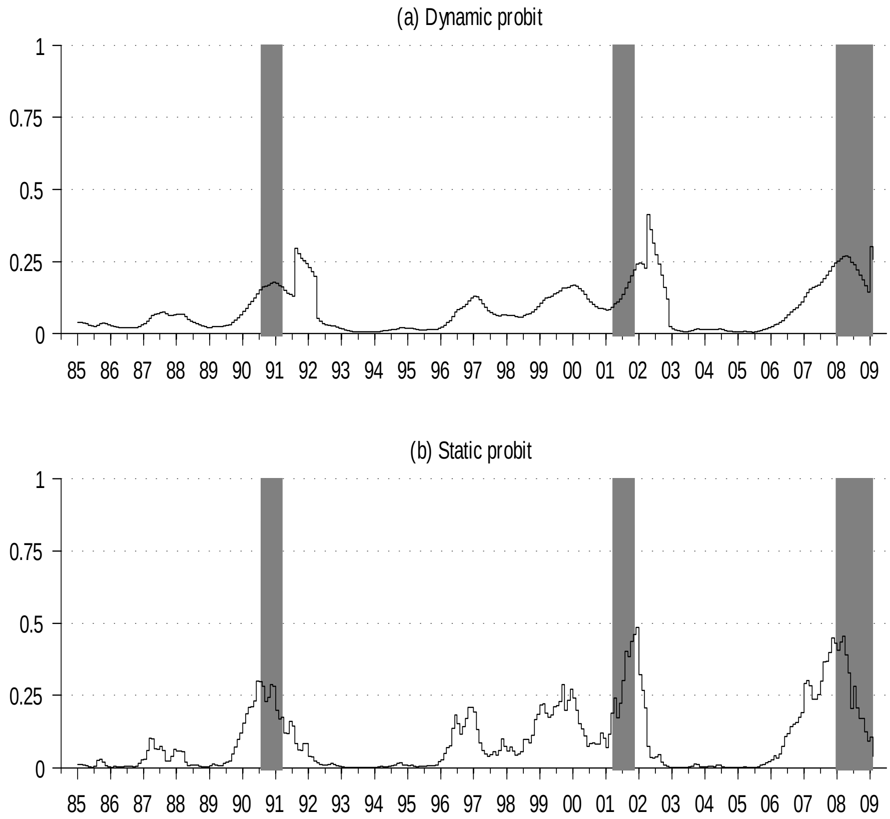

Figure 3: Probability of Recession, Out-of-sample Prediction Twelve Months Ahead (the shaded bars indicate NBER-dated recession months)

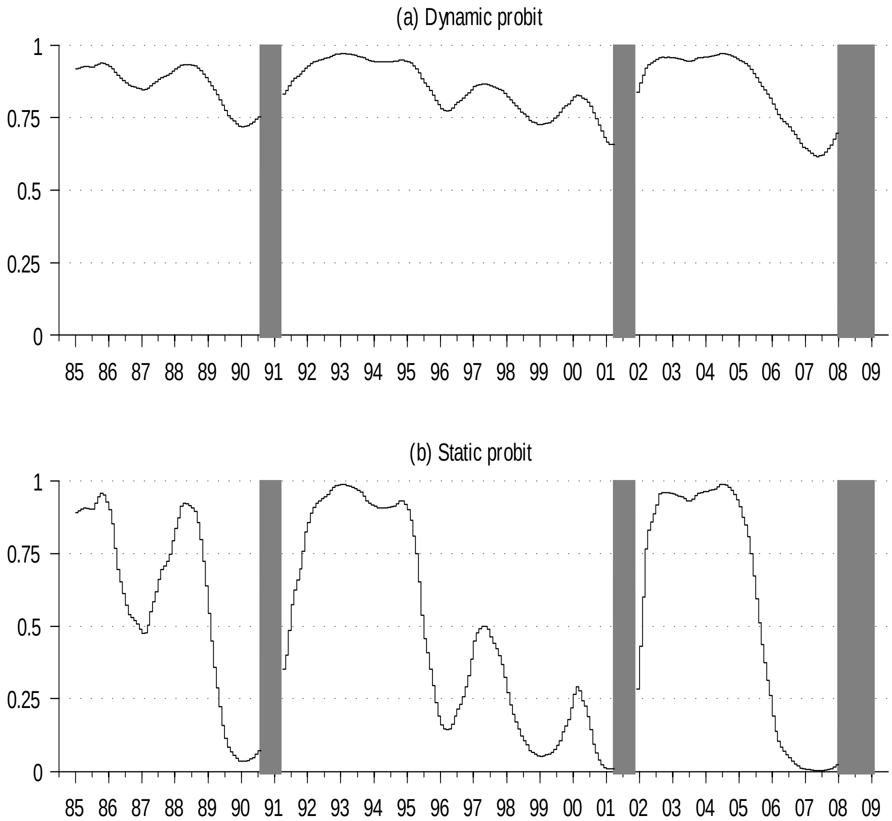

Figure 4: Probability of Continuing Expansion Next 12 Months, Rolling Out-of-sample Prediction (the shaded bars indicate NBER-dated recession months)

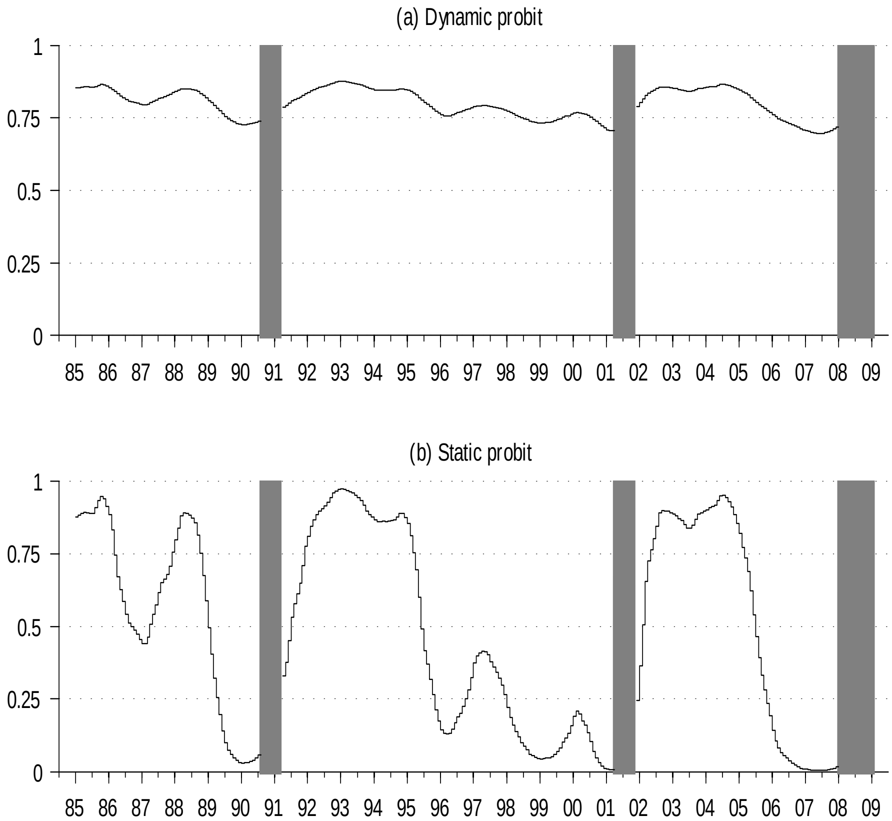

Figure 5: Probability of Continuing Expansion Next 12 Months, Forecasts Based on Models Estimated with Data From January 1955 to December 1982 (the shaded bars indicate NBER-dated recession months)

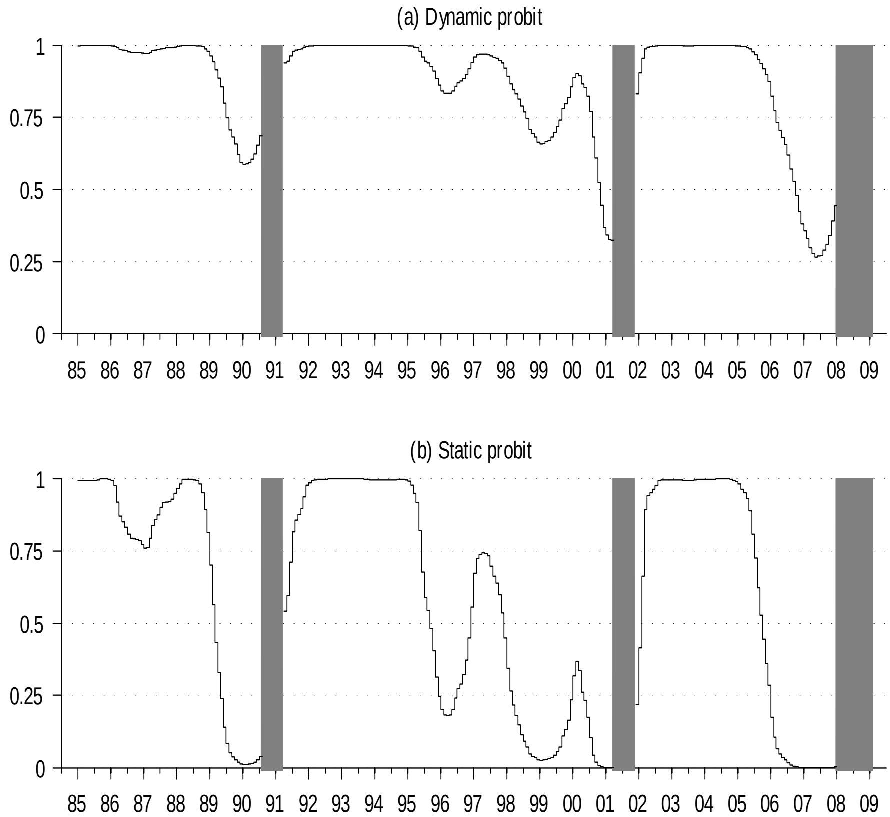

Figure 6: Probability of Continuing Expansion Next 12 Months, Predictions Based on Models Estimated with Data From December 1982 to February 2009 (the shaded bars indicate NBER-dated recession months)

Aboa Centre for Economics (ACE) was founded in 1998 by the departments of economics at the Turku School of Economics, Åbo Akademi University and University of Turku. The aim of the Centre is to coordinate research and education related to economics in the three universities.

Contact information: Aboa Centre for Economics, Turku School of Economics, Rehtorinpellonkatu 3, 20500 Turku, Finland.

Aboa Centre for Economics (ACE) on Turun kolmen yliopiston vuonna 1998 perustama yhteistyöelin. Sen osapuolet ovat Turun kauppakorkeakoulun kansantaloustieteen oppiaine, Åbo Akademin nationalekonomi-oppiaine ja Turun yliopiston taloustieteen laitos. ACEn toiminta-ajatuksena on koordinoida kansantaloustieteen tutkimusta ja opetusta Turun kolmessa yliopistossa.

Yhteystiedot: Aboa Centre for Economics, Kansantaloustiede, Turun kauppakorkeakoulu, 20500 Turku.

www.ace-economics.fi

ISSN 1796-3133