UNIVERSITY OF TÜBINGEN

SCHOOL OF BUSINESS AND ECONOMICS

FACULTY OF ECONOMETRICS, STATISTICS AND EMPIRICAL ECONOMICS

(Re-)Evaluating the Predictive Power of the Yield Curve: Evidence from the US and Germany

Master Thesis
Summer Semester 2015

Supervisor(s):	Prof. Dr. Joachim Grammig, Jantje Sönksen (MSc)
Author:	Andreas Lay
Degree program:	MSc Quantitative Economics
Matriculation no.:	3447330
Address:	Falkenweg 20 72076 Tübingen
Phone:	+49(0)1756165666
Email address:	andreas.lay@student.uni-tuebingen.de

Date of submission: December 24, 2015

Contents

List of Figures

List of Tables

Abstract

This paper uses US and German GDP growth as well as bond yield spread data to examine the predictive power of the yield curve. A simple linear VAR as well as non-linear VARs which include a structural break and a threshold (SBVAR, TSVAR, SBTVAR) are estimated by numerical methods to compute recession probabilities. Based on these probabilities, a threshold is defined on when to predict a recession and these prediction results are evaluated. We compare our results to those obtained by a static, dynamic, autoregressie and a dynamic-autoregressive probit model. We confirm the result that the yield spread has predictive power with respect to recession forecasting, and we find that this relationship still holds for latest recession of 2007/08 for both the US and Germany. Whereas including a structural break improves the forecasting performance for the US, this is not the case for Germany. In addition, autoregressive probits are shown to perform best at recession forecasting, however this is mainly due to the inclusion of a lagged recession dummy variable which results in a failure to correctly forecast the start and ending recession quarters.

1. Introduction

The latest financial crisis and the following recession of 2007/08 in the US and Europe have led to criticism towards economists that they did not see the crisis coming. Despite predictions being very difficult, especially if they are about the future (Niels Bohr), we will attempt in this paper to re-evaluate the predictive power of the spread between long-term and short-term bond yields using data for the US and Germany that include the most recent financial crisis. For both politicians and central banks, it is essential to be able to react to upcoming downturns in economic activity on time, e.g. by decreasing interest rates and conducting open market operations. Intensive research has been done regarding the forecasting power of the yield curve, given the fact that in the past, major recessions haven been preceded by an inverted yield curve, i.e. short-term bond yields which lie above long-term bond yields. Chinn and Kucko (2010) perform a cross-country analysis and find that European country models perform better when using more recent data. Duarte, Venetis, and Payac (2005) attempt to predict real growth and recession probabilities for Euro area countries using linear and non-linear regression models. They find that there are significant non-linearities and confirm the ability of the yield curve to anticipate recessions. An overview on recent literature in the field is provided by Wheelock and Wohar (2009): The general consensus is that the yield curve has good predictive power, however it is observed that this predictive power has decreased since the 1980s for the US. However, there is no generally accepted theory why the yield curve has this forecasting power.

When predicting recessions, we first of all need to define the event of a recession. On the one hand, for instance, one may use the recession dates provided by The National Bureau of Economic Research (NBER) which determines recession periods for the US ex-post. On the other hand, formal algorithms for recession dating may be applied: Chauvet and Hamilton (2005) provide an algorithm for real-time business cycle dating. They show that their dating method yields results mostly consistent with NBER dates, showing that it is possible to determine NBER business cycles not only ex-post, but as soon as GDP growth data for that quarter are available. Another definition states that a period of two consecutive quarters of negative growth marks times of recession. In this paper, we still stick to the NBER and ECRI recession periods for the probit models and any forecast evaluation and will use a criterion based on negative growth rates for the VAR models.

The structure of this paper is as follows: In section 2 we will take a look at two theories that have implications on the relationship between economic activity and the yield spread: First, we consider the Expectation Hypothesis which implies that when future short-term spot rates are expected to rise, the yield curve is inverted. Second, we will simulate a consumption-based asset pricing model and show that a stochastic production economy implies a positive correlation between spreads and economic growth.

Section 3 explains the estimation methods for the linear and non-linear VAR models as well as the probit models.

In section 4 we present our data basis, discuss the estimation results and illustrate our forecast evaluation methods.

Finally, section 5 summarises our results and points out problems that need to be approached in future research.

2. Theoretical Background: Relationship between Yields and Economic Activity

2.1. An Expectation Hypothesis Explanation

A simple textbook explanation for the slope of the yield curve can e.g. be found in Cochrane (2005): No abitrage requires that the τ-period return equals the return on consecutively investing at the prevailing one-period forward rates which can be formally expressed as

$$\exp(r_t^\tau \tau) = \exp(r_t^1 + f_t^{t+1} + \cdots + f_t^{\tau-1})$$

where f_t^{t+n} is the one-period forward rate at time t for investing 1 unit of currency from time $t+n$ to $t+n+1$. This can also be stated as

$$r_t^\tau = \frac{1}{\tau}(r_t^1 + f_t^{t+1} + \cdots + f_t^{\tau-1}).$$

Now assume that the the expectation hypothesis holds, i.e. forward rates are unbiased estimators of future spot rates: $f_t^{t+1} = \mathbb{E}_t(r_{t+1}^1)$. Then the no-arbitrage condition can be written as

$$r_t^\tau = \frac{1}{\tau}\mathbb{E}_t(r_t^1 + r_{t+1}^1 + \cdots + r_{t+\tau-1}^1)$$

Now assume that the yield curve is upward-sloping, i.e. $r_t^\tau > r_t^1$. Given that the long-term rate is an average of expected short-term rates, this can only be true if

$$\frac{1}{\tau-1}\mathbb{E}_t(r_{t+1}^1 + \cdots + r_{t+\tau-1}^1) > r_t^1$$

i.e. if future short-term spot rates are on average expected to increase.[1] The same argument can be made for downward-sloping curves: If the curve is inverted, this is because future spot rates are expected to decrease. If the monetary authority always lowers short-term interest rates in response to periods of macroeconomic and financial distress (”recession”) which is anticipated by market participants, then any recession would be preceded by an inverted yield curve. Note, however, that the reverse is not true: If short-term rates are expected to decline due to other factors than monetary operations, then an inverted curve need not be followed by a recession.

2.2. A Consumption Smoothing Explanation

Harvey (1988) uses the consumption-based asset pricing model to derive a relationship between expected future consumption growth and the slope of the yield curve. Suppose the household can decide on whether to consume its income or invest into τ-period real zero-coupon bonds of price p_t^τ, $\tau = 1, 2, \ldots, M$. The household faces the following intertemporal optimization problem

$$\begin{aligned} &\max_{\{c_t, b_t^\tau\}_{t=0}^\infty} \sum_{t=0}^{\infty} \beta^t \mathbb{E}(u(c_t)|\mathscr{F}_0) \\ &\text{s.t. } c_t + \sum_{\tau=1}^{M} p_t^\tau b_t^\tau = y_t + \sum_{\tau=1}^{M} b_{t-\tau}^\tau \;\; \forall \;\; t \end{aligned} \tag{2.2.1}$$

where b_t^τ is the amount of τ-period bonds purchased in time t. In the following, the conditional expectation $\mathbb{E}(\cdot|\mathscr{F}_t)$ will be denoted as $\mathbb{E}_t(\cdot)$. Furthermore let the log-income endowment follow an autoregressive process of order one:

$$\ln(y_{t+1}) = \rho \ln(y_t) + u_{t+1} \tag{2.2.2}$$

with $u_t \sim \mathcal{N}(0, \sigma_u^2)$. This results in the first-order conditions

$$\begin{aligned} p_t^\tau &= \mathbb{E}_t(m_t^\tau) \\ \forall \;\; \tau &= 1, 2, \ldots, M \end{aligned} \tag{2.2.3}$$

where $m_t^\tau \equiv \beta^\tau u'(c_{t+\tau})/u'(c_t)$. In particular, the following specification for the utility function $u(c) = (c^{1-\gamma} - 1)/(1-\gamma)$ yields the stochastic discount factor to discount future cash flows occuring at time $t+\tau$ back to time t as

$$m_t^\tau = \beta^\tau \left(\frac{c_{t+\tau}}{c_t}\right)^{-\gamma} \tag{2.2.4}$$

[1] The proof is straightforward and given in appendix C.1.

where γ is the Arrow-Pratt measure of relative risk aversion. Note that the price of a risk-free zero-coupon bond is $p_t^\tau = \exp(-r_t^\tau \tau)$ and therefore the annualized risk-free rate is given by $r_t^\tau = -\ln(p_t^\tau)/\tau$. Let us now define the τ-period spread as

$$\mathcal{S}_t^\tau \equiv r_t^\tau - r_t^1 \tag{2.2.5}$$

and τ-period annualized growth rates of income endowment and consumption as $\hat{y}_t^\tau \equiv \ln(y_{t+\tau}/y_t)/\tau$ and $\hat{c}_t^\tau \equiv \ln(c_{t+\tau}/c_t)/\tau$. We are now interested in the correlation between income growth (which is GDP growth in our setup) and the spread, i.e. $\mathrm{Corr}(\mathcal{S}_t^\tau, \hat{y}_t^\tau)$. It is shown by de Lint and Stolin (2003) that this this correlation is actually negative for $\tau > 1$ and $\rho < 1$ who also note that the literature occasionally misinterpretes that the indeed positive correlation $\mathrm{Corr}(r_t^\tau, \hat{y}_t^\tau)$ implies a positive correlation between spread and income growth. However, they suggest to overcome this contradiction with empirical data by including the production sector in their model such that the budget constraint in this stochastic production economy (SPE) is now

$$c_t + i_t + \sum_{\tau=1}^{M} p_t^\tau b_t^\tau = y_t + \sum_{\tau=1}^{M} b_{t-\tau}^\tau \quad \forall \ t \tag{2.2.6}$$

where

$$i_t = k_{t+1} - (1-d)k_t \tag{2.2.7}$$

and

$$y_t = \theta_t k_t^\alpha, \tag{2.2.8}$$

i.e. capital in the following period is given by the depreciated current-period capital plus investment and this capital is used in combination with technology θ_t to produce the output y_t. The log technology variable is again assumed to follow an autoregressive process of order one:

$$\ln(\theta_t) = \rho \ln(\theta_{t-1}) + u_t \tag{2.2.9}$$

with $u_t \sim \mathcal{N}(0, \sigma_u^2)$ and i.i.d. The FOC (2.2.3) still holds, however a second FOC determining the consumption-investment decision is now given by

$$c_t^{-\gamma} = \mathbb{E}_t\{\beta c_{t+1}^{-\gamma}\left(\alpha\theta_{t+1}k_{t+1}^{\alpha-1} + 1 - d\right)\} \tag{2.2.10}$$

Moreover, in equilibrium the household does not wish to hold any bonds, thus $b_t^\tau = 0$ for all τ.

Stochastic Production Economy: A Simulation Study

The SPE model as presented above cannot be solved analytically but needs to be simulated numerically for a given set of parameters. The approach used in this paper to solve the model is the Parameterizing Expectations approach as first introduced by den Haan and Marcet (1990) and which is also used by de Lint and Stolin (2003). The idea behind their algorithm is to start the simulation by approximating the integral in (2.2.10) by a polynomial in the state variables k_t and θ_t. The procedure as for example described in the appendix of den Haan (1995) goes as follows: First, fix an initial level for the capital stock $k_1 = \bar{k}$ and and write the right-hand side of (2.2.10) as an nth order polynomial, i.e.

$$\begin{aligned}\mathbb{E}_t\{\beta c_{t+1}^{-\gamma}\left(\theta_{t+1}\alpha k_{t+1}^{\alpha-1}+1-d\right)\} &= \mathrm{P}_n(k_t,\theta_t)\\ &= \exp(a_1 + a_2\ln(k_t) + a_3\ln(\theta_t) + a_4\ln(k_t)^2 + \dots)\end{aligned}$$

where the exponential function guarantees that consumption is positive for any parameter vector $\boldsymbol{a} \equiv (a_1, a_2, \dots)'$. In this paper we choose a second-order polynomial such that

$$\mathrm{P}_n(k_t,\theta_t) = \exp(a_1 + a_2\ln(k_t) + a_3\ln(\theta_t) + a_4\ln(k_t)^2 + a_5\ln(\theta_t)^2).$$

As a next step, simulate the (log) technology path over T_{sim} periods as given by equation (2.2.2) and fix an initial parameter vector $\boldsymbol{a}_0$. Given $\boldsymbol{a}_0$, $\{\theta_t\}_{t=1}^{t=T_{\text{sim}}}$ and $k_1 = \bar{k}$, the polynomial plus the budget constraint (2.2.6) in combination with equations (2.2.7) and (2.2.8) yield

$$\begin{aligned}c_1^{-\gamma} &= \mathrm{P}_2(k_1,\theta_1)\\ k_2 &= \theta_1 k_1^{\alpha} - c_1 + (1-d)k_1\end{aligned}$$

which finally gives the series $\{c_t, k_t\}_{t=1}^{t=T_{\text{sim}}}$. Define $z_{t+1} \equiv \beta c_{t+1}^{-\gamma}(\alpha\theta_{t+1}k_{t+1}^{\alpha-1} + 1-d)$ and note that the polynomial is supposed to approximate $\mathbb{E}_t(z_{t+1})$, then a new estimate for $\boldsymbol{a}$ is obtained through a nonlinear least squares regression of z_{t+1} on the polynomial $\mathrm{P}_2(k_t,\theta_t)$:

$$\hat{\boldsymbol{a}} = \arg\min_{\tilde{\boldsymbol{a}}} \sum_{t=t_{\text{init}}}^{T_{\text{sim}}} (z_{t+1} - \mathrm{P}_2(k_t,\theta_t;\tilde{\boldsymbol{a}}))^2$$

where t_{init} is the first value used in the regression as we would like to exclude the first $t_{\text{init}} - 1$ simulation values. In our application, we choose $T_{\text{sim}} = 26000$ and $t_{\text{init}} = 1000$ which is required to obtain robust parameter values in the nonlinear least-squares regressions. Note that in order to ensure convergence we have to use nonlinear least-squares and must not run a linear regression on the log-linearized form. We will denote the estimate for the first iteration by $\hat{\boldsymbol{a}}_1$. Using the same series $\{\theta_t\}_{t=1}^{t=T_{\text{sim}}}$ as simulated in the first step, repeat the procedure as described above using parameter vector

$$\boldsymbol{a}_2 = \phi\hat{\boldsymbol{a}}_1 + (1-\phi)\boldsymbol{a}_0$$

where a high $\phi, 1 \geq \phi > 0$ generally speeds up convergence but may for complex models result in divergence for parameter vector $\boldsymbol{a}$. Repeat this procedure until a convergence criterion as for example $\max_{a\in|\boldsymbol{a}_i-\boldsymbol{a}_{i-1}|}(|\boldsymbol{a}_i - \boldsymbol{a}_{i-1}|) < c$ is met.

Table 1: Parameters for the polynomial used to approximate the consumption function obtained as described in the text.

a_1	a_2	a_3	a_4	a_5
0.9566	-0.9314	-1.0920	-0.0304	-0.2019

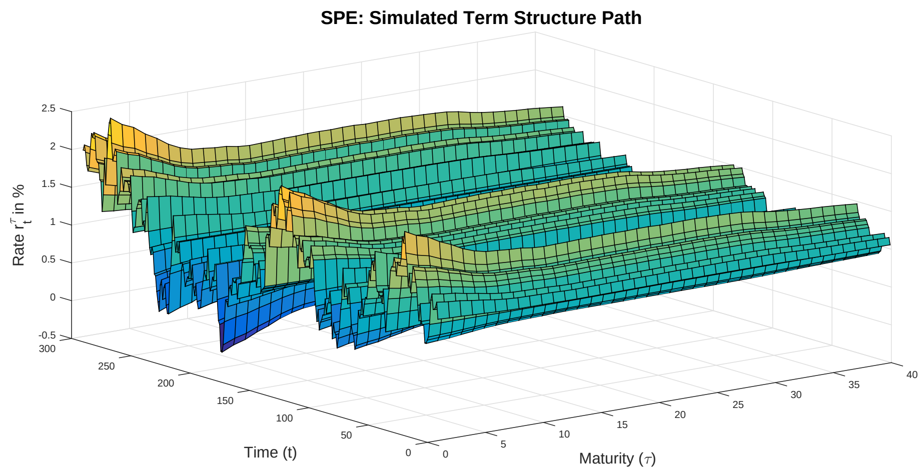


Figure 1: A simulated sample path of the term structure of interest rates in the stochastic production economy. On average, the SPE implies an inverted term structure with long-term rates lying below short-term rates.

Note that good starting values can consiberably speed up the iterative algorithm, therefore obtained parameter values are printed in table 1 so the reader may use these values as a starting point when implementing this model herself. Based on this final sample, to obtain bond prices another polynomial is used to approximate equation (2.2.3) for any maturity. First compute the M series of stochastic discount factors according to (2.2.4) and regress the resulting series of m_t^τ on the polynomial $\mathrm{B}_n(k_t, \theta_t, c_t)$. In particular, we follow den Haan (1995) and choose the polynomial

$$\begin{aligned} \mathrm{B}_n(k_t, \theta_t, c_t) = \exp(&b_1 + b_2 \ln k_t + b_3 ln\theta_t + b_4 \ln c_t^{-\gamma} + \\ &b_5 \ln k_t^2 + b_6 \ln \theta_t^2 + b_6 \ln c_t^{-2\gamma} + b_8 \ln \theta_t^3) \end{aligned} \tag{2.2.11}$$

taking into account the high persistency in the stochastic discount factors such that higher-order polynomials are required to approximate their expected values. Note that equation (2.2.11) results in M nonlinear least-square regressions that yield one polynomial to price each bond with a specific maturity. For our simulation, we use the same parameter values as in de Lint and Stolin (2003):

$$\beta = 0.99, \gamma = 3, d = 0.025, \rho = 0.95, \alpha = 0.33, \sigma_u = 0.018$$

which are corresponding to US quarterly data. Based on these parameters, $N_{\text{sim}} = 500$ simulations with each consisting of $T_{\text{sim}} = 300$ observations are simulated which can be interpreted as a simulation over a period of 75 years on a quarterly frequency. Interest rates are backed out up to 40 periods, i.e. up to 10 years. One simulated term structure path is illustrated in figure 2: It can be seen that the simulated interest rate series show high persistency which is in line with empirical facts as unit-root tests on real interest rate series often fail to reject the hypothesis that interest rates follow a unit-root process. It can also be seen from figure 2 that short-term rates are more volatile than long-term rates with the volatility slowly declining with the bond maturity. However, there are also some drawbacks to the model: First of all, as pointed out by den Haan (1995), the model term structure will on average be downward-sloping. According to den Haan (1995), this may be "corrected" by introducing transaction costs that are increasing in the bond maturity. Apart from that, our polynomial approximation guarantees that prices are positive, however there is nothing that prevents zero-bond prices from being larger than 1, or equivalently interest rates from becoming negative. In table 2 several correlations of model variables are printed. The table also includes the results from the regressions of τ-period consumption and income growth rates on the τ-period spread:

$$\hat{c}_t^\tau = a_{\hat{c}}^\tau + b_{\hat{c}}^\tau \mathcal{S}_t^\tau + \epsilon_{\hat{c}}^\tau$$

and

$$\hat{y}_t^\tau = a_{\hat{y}}^\tau + b_{\hat{y}}^\tau \mathcal{S}_t^\tau + \epsilon_{\hat{y}}^\tau .$$

It can be seen that the correlation between the spread and consumption growth Corr($\mathcal{S}_t^\tau, \hat{c}_t^\tau$) as well as the correlation between the spread and output growth Corr($\mathcal{S}_t^\tau, \hat{y}_t^\tau$) are positive and increasing with maturity. Whereas there are few simulations where the former correlation is lower than or equal to zero, the latter correlation is positive for all simulations.

Table 2: Correlations between different model variables in the SPE based on $N_{\text{sim}} = 500$ simulations, each consisting of $T_{\text{sim}} = 300$ periods. The values are averages over all simulations.

	$\tau = 2$	$\tau = 4$	$\tau = 8$	$\tau = 20$	$\tau = 40$
Corr($\mathcal{S}_t^\tau, \hat{c}_t^\tau$)	0.16	0.20	0.25	0.36	0.44
Percent Corr($\mathcal{S}_t^\tau, \hat{c}_t^\tau$) > 0	98.40	99.40	99.00	98.80	99.60
$b_{\hat{c}}^\tau$	3.07	1.03	0.44	0.33	0.20
$\sigma(b_{\hat{c}}^\tau)$	1.99	0.64	0.26	0.16	0.08
Corr($\mathcal{S}_t^\tau, \hat{y}_t^\tau$)	0.23	0.28	0.36	0.47	0.55
Percent Corr($\mathcal{S}_t^\tau, \hat{y}_t^\tau$) > 0	100.00	100.00	100.00	100.00	100.00
$b_{\hat{y}}^\tau$	7.18	2.38	0.98	0.67	0.37
$\sigma(b_{\hat{y}}^\tau)$	0.07	0.08	0.10	0.13	0.13
Corr($\hat{c}_t, \hat{y}_t$)	1.00	0.99	0.99	0.99	0.98
Corr($\hat{c}_t, u_t$)	-0.03	-0.04	-0.05	-0.07	-0.10
Corr($\hat{y}_t, u_t$)	-0.06	-0.07	-0.09	-0.12	-0.15
Corr($\mathcal{S}_t^\tau, u_t$)	-0.33	-0.33	-0.33	-0.34	-0.34

Furthermore, the regression coefficients $b_{\hat{c}}^\tau$ and $\sigma(b_{\hat{y}}^\tau)$ are positive and decreasing in maturity τ.

3. Econometric Models for Recession Forecasting

3.1. Nonlinear (Vector-)Autoregressive Models

3.1.1. Structural Break Threshold VAR and Nested Models

Define the $m \times 1$ vector of endogenous variables as $\boldsymbol{y}_t \equiv (y_{1t}, y_{2t}, \ldots, y_{mt})'$ and the $m(p+1) \times 1$ vector $\boldsymbol{Y}_{t-1} \equiv (\mathbf{1}', \boldsymbol{y}_{t-1}', \ldots, \boldsymbol{y}_{t-p}')'$, then the Structural Break Threshold VAR (SBTVAR) can be written as

$$\begin{aligned} \boldsymbol{y}_t = & \left[\boldsymbol{\Phi}^{(1)}\boldsymbol{Y}_{t-1}I_{1,t-d_1}(r_1) + \boldsymbol{\Phi}^{(2)}\boldsymbol{Y}_{t-1}(1 - I_{1,t-d_1}(r_1))\right] I_t(\tau) + \\ & \left[\boldsymbol{\Phi}^{(3)}\boldsymbol{Y}_{t-1}I_{2,t-d_2}(r_2) + \boldsymbol{\Phi}^{(4)}\boldsymbol{Y}_{t-1}(1 - I_{2,t-d_2}(r_2))\right] (1 - I_t(\tau)) + \boldsymbol{u}_t \end{aligned} \tag{3.1.1}$$

where $I_{i,t-d_i}(r_i), i \in \{1, 2\}$ is a threshold indicator function which is either 1 or 0 according to

$$I_{i,t-d_i}(r_i) = \begin{cases} 1 & \text{if } z_{t-d_i} \leq r_i \\ 0 & \text{otherwise} \end{cases}$$

and $I_t(\tau)$ is a structural break indicator function that is

$$I_t(\tau) = \begin{cases} 1 & \text{if } t \leq \tau \\ 0 & \text{otherwise.} \end{cases}$$

Galvao (2006) uses (3.1.1) to model real GDP growth and the slope of the yield curve. Hence, in our recession forecasting framework, we choose $m = 2$ and let the first element of $\boldsymbol{y}_t$ denote real GDP growth and its second element denote the slope of the yield curve. Moreover, the threshold variable z will be observable and given by real GDP growth, i.e. there are four regimes and the level of GDP growth rates determines the regime both before and after the structural break point in time τ. Note that $\boldsymbol{\Phi}^{(j)} \equiv (\boldsymbol{\Phi}_0^{(j)}, \boldsymbol{\Phi}_1^{(j)}, \ldots, \boldsymbol{\Phi}_p^{(j)})$, then the data-generating process in regime $j \in \{1, 2, 3, 4\}$ is given by the standard VAR:

$$\boldsymbol{y}_t = \boldsymbol{c} + \boldsymbol{\Phi}_1^{(j)}\boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Phi}_p^{(j)}\boldsymbol{y}_{t-p} + \boldsymbol{u}_t$$

with $c \equiv \boldsymbol{\Phi}_0^{(j)}\mathbf{1}$. Furthermore, we will assume that the error terms $\{\boldsymbol{u}_t\}$ are i.i.d. normally distributed and their variance-covariance matrix is allowed to be regime-dependent: $\boldsymbol{u}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}(r_1, r_2, \tau))$.

The SBTVAR nests the structural break VAR (SBVAR), the threshold VAR (TVAR) and the usual linear VAR as special cases: Assuming that the threshold variable is bounded from above, $z < \infty$, and setting the thresholds r_i infinitely large yields the SBVAR as

$$\boldsymbol{y}_t = \boldsymbol{\Phi}^{(1)}\boldsymbol{Y}_{t-1}I_t(\tau) + \boldsymbol{\Phi}^{(2)}\boldsymbol{Y}_{t-1}(1 - I_t(\tau)) + \boldsymbol{u}_t \tag{3.1.2}$$

while just allowing for a threshold but no structural break gives the TVAR:

$$\boldsymbol{y}_t = \boldsymbol{\Phi}^{(1)}\boldsymbol{Y}_{t-1}I_{1,t-d_1}(r) + \boldsymbol{\Phi}^{(2)}\boldsymbol{Y}_{t-1}(1 - I_{1,t-d_1}(r)) \tag{3.1.3}$$

3.1.2. Maximum Likelihood Estimation of Nonlinear VARs and Bootstrapped Standard Errors

Galvao (2006) employs a grid search estimation approach that involves looping over a grid of structural break dates $\tilde{\tau}$, the thresholds in each structural break regime $\tilde{r}_i$ and the threshold delays in each structural break regime $\tilde{d}_i$. For a given combination of $\tilde{\tau}$, $\tilde{r}_i$ and $\tilde{d}_i$ the VAR parameters $\mathbf{\Phi}^i$ can be consistently estimated by a simple linear regression. As the assumption of Gaussian error terms gives rise to a normal likelihood function, one can focus on maximizing this likelihood function by minimizing the (log) determinant of the estimated error variance-covariance matrix $\hat{\mathbf{\Sigma}} \equiv \widehat{\mathrm{Var}}(\hat{\boldsymbol{u}}_t)$. Specifically, the following procedure is applied:

1. *Grid determination:* Let $H_v(\cdot)$ denote the empirical cumulative distribution function (c.d.f.) of variable v. The grid search bounds of the sets $[\tilde{\tau}_{lb}, \tilde{\tau}_{ub}]$ and $[\tilde{r}_{i,lb}, \tilde{r}_{i,ub}]$ are chosen in a way such that there are at least $\alpha_t T$ observations in each subsample (for SBTVAR and SBVAR), $\alpha_{g_1} T$ observations in each regime (for TSVAR) and at least $\alpha_{g_2} T_j$ observations in each regime for subsample $j = 1, 2$ (for SBTVAR). Formally, we write

$$
\begin{aligned}
\tilde{\tau}_{\text{lb}} &= H_t^{-1}(\alpha_t), \tilde{\tau}_{\text{ub}} = H_t^{-1}(1-\alpha_t) \;\; [\textit{SBTVAR, SBVAR}] \\
\tilde{r}_{lb} &= H_g^{-1}(\alpha_{g_1}), \tilde{r}_{ub} = H_g^{-1}(1-\alpha_{g_1}) \;\; [\textit{TVAR}] \\
\tilde{r}_{1,lb} &= H_g^{-1}(\alpha_{g_2}|t \leq \tilde{\tau}), \tilde{r}_{1,ub} = H_g^{-1}(1-\alpha_{g_2}|t \leq \tilde{\tau}) \;\; [\textit{SBTVAR}] \\
\tilde{r}_{2,lb} &= H_g^{-1}(\alpha_{g_2}|t > \tilde{\tau}), \tilde{r}_{2,ub} = H_g^{-1}(1-\alpha_{g_1}|t > \tilde{\tau}) \;\; [\textit{SBTVAR}]
\end{aligned}
$$

   where $H_t^{-1}(\cdot)$ and $H_g^{-1}(\cdot)$ are the inverse empirical c.d.f. of the time index variable and real GDP growth rate variable, respectively. Furthermore it is assumed that $d_i \in \{1, \ldots, 4\}$.

2. *Regression parameter estimation:* Start the grid search at $\tilde{\tau}_{lb}$ (for SBTVAR and SBVAR) and $\tilde{r}_{lb}$ (for TVAR). For the SBTVAR, use the given $\tilde{\tau}$ from the current structural break loop to determine the threshold bounds $\tilde{r}_{i,lb}, \tilde{r}_{i,ub}$ for both subsamples $i = 1, 2$. For both threshold models, add a loop over the threshold lag variable(s) d_i. For this set of parameters, determine for each point in time in which state of the world – e.g. one of the four subsample/regime combinations in the SBTVAR – we are. The state variable j at time t for the SBTVAR is thus defined as

$$
j_t \equiv \begin{cases} 1 & \text{if } I_{1,t-d_1}(\tilde{r}_1) = 1 \wedge I_t(\tilde{\tau}) = 1 \\ 2 & \text{if } I_{1,t-d_1}(\tilde{r}_1) = 0 \wedge I_t(\tilde{\tau}) = 1 \\ 3 & \text{if } I_{2,t-d_2}(\tilde{r}_2) = 1 \wedge I_t(\tilde{\tau}) = 0 \\ 4 & \text{if } I_{2,t-d_2}(\tilde{r}_2) = 0 \wedge I_t(\tilde{\tau}) = 0 \end{cases}
$$

   Let T_j denote the set of time indeces including all points in time the data was generated by state j: $T_j := \{t : j_t = j\}$. Then the regression parameters are consistently estimated by

$$\hat{\boldsymbol{\Phi}}^{(j)} = \left(\sum_{t \in T_j} \boldsymbol{y}_t \boldsymbol{Y}_{t-1}' \right) \left(\sum_{t \in T_j} \boldsymbol{Y}_{t-1} \boldsymbol{Y}_{t-1}' \right)^{-1}$$

for all states j.

3. *Choice of Maximum Likelihood estimates:* Having calculated the regression parameters for the given combination $\tilde{\tau}$, $\tilde{r}_i$ and $\tilde{d}_i$, back out the residuals for the whole sample by

$$\hat{\boldsymbol{u}}_t = \boldsymbol{y}_t - \hat{\boldsymbol{\Phi}}^{(j_t)} \boldsymbol{Y}_{t-1}$$

and compute the state-specific error variance-covariance matrices by

$$\hat{\boldsymbol{\Sigma}}_j = \frac{1}{|T_j|} \sum_{t \in T_j} \hat{\boldsymbol{u}}_t \hat{\boldsymbol{u}}_t'$$

where $|T_j|$ denotes the cardinality of set T_j.[2] Repeat this step for all parameter combinations on the grid and choose the set of parameter estimates that maximizes the log-likelihood, or equivalently minimizing

$$\hat{\tau}, \hat{r}_1, \hat{r}_2 = \min_{\substack{\tilde{\tau}_{lb} \leq \tilde{\tau} \leq \tilde{\tau}_{ub} \\ \tilde{r}_{1,lb} \leq \tilde{r}_1 \leq \tilde{r}_{1,ub} \\ \tilde{r}_{2,lb} \leq \tilde{r}_2 \leq \tilde{r}_{2,ub}}} \frac{1}{2} \sum_{j=1}^{4} |T_j| \log(\det(\hat{\boldsymbol{\Sigma}}_j))$$

taking into account that the error-variance-covariance matrix is state-dependent.

To obtain standard errors, a simple residual-based time series bootstrapping method as for example described in Lütkepohl (2005) is applied: Collect the initial p observation in the vector $\boldsymbol{Y}_0 = (\boldsymbol{1}', \boldsymbol{y}_0', \ldots, \boldsymbol{y}_{-p+1}')'$ and determine the state j_1 the first observation is generated by based on the estimated parameters and the initial observations. Then draw a random residual vector $\boldsymbol{u}_1^*$ from $\{\hat{\boldsymbol{u}}_t\}_{t \in T_{j_1}}$ and compute the first observation of the bootstrap sample as

$$\boldsymbol{y}_1^* = \hat{\boldsymbol{\Phi}}^{(j_1)} \boldsymbol{Y}_0 + \boldsymbol{u}_1^*$$

and define $\boldsymbol{Y}_1^* \equiv (\boldsymbol{1}', \boldsymbol{y}_1^{*\prime}, \ldots, \boldsymbol{y}_{-p+2}')'$. Proceed to the second generated observation and determine by which state j_2 it is generated. Draw $\boldsymbol{u}_2^*$ from $\{\hat{\boldsymbol{u}}_t\}_{t \in T_{j_2}}$, compute

$$\boldsymbol{y}_2^* = \hat{\boldsymbol{\Phi}}^{(j_2)} \boldsymbol{Y}_1^* + \boldsymbol{u}_2^*$$

and repeat the steps above for $t = 3, 4. \ldots T$. Note that there are also alternative bootstrapping methods as for example the stationary bootstrap by Politis and Romano (1994) which is based on drawing subsamples of random length from the original time series $\boldsymbol{y}_t$, recombining them and conducting estimation based on this sample. However, residual-based bootstrapping still is a widely used method when dealing with vectorautoregressive model and we therefore follow this approach as described above.

[2]The cardinality of a set S measures the number of elements contained in this set.

3.1.3. Backing out Recession Probabilities in Nonlinear VARs

To back out recession probabilities from our VAR models, we first need to define the event "recession". In this paper we stick to definition of Galvao (2006) who defines quarter t to be in a recession if there are two consecutive periods of negative growth in the period from t to $t+4$. This definition implies that whether a quarter t is in a recession can just be identified ex-post when growth rates for dates $t+1$ to $t+4$ are known (unless for example t and $t+1$ are both negative and growth rates are observable instantly, then quarter t is already identified as in recession at date $t+1$). It follows that the probability of recession in period $t+1$ conditional on observing time t information is the probability

$$\begin{aligned}
\mathrm{P}(R_{t+1} = 1|\mathscr{F}_t) = \mathrm{P}(&[y_{1,t+1} < 0 \quad \wedge \quad y_{1,t+2} < 0] \ \vee \\
&[y_{1,t+2} < 0 \quad \wedge \quad y_{1,t+3} < 0] \ \vee \\
&\dots \\
&[y_{1,t+4} < 0 \quad \wedge \quad y_{1,t+5} < 0]|\mathscr{F}_t)
\end{aligned}$$

which cannot be calculated analytically in our VAR model framework but has to be computed numerically. The method to compute recession probabilities as described in the appendix of Galvao (2006) is based on drawing from the estimated residuals and simulating K series $\tilde{\boldsymbol{y}}_{t+1}, \tilde{\boldsymbol{y}}_{t+2}, \dots, \tilde{\boldsymbol{y}}_{t+5}$ for all t. As with the bootstrapping method, it is important to draw the residuals for the $t+s$ simulated value from $\{\hat{\boldsymbol{u}}_t\}_{t \in T_{j_{t+s}}}$.

3.2. Probit Models

In contrast to the (vector-)autoregressive models as described above, Probit models do not model the stochastic dynamics of the growth and spread series but instead take a look at the series of recessions whose probability of occurrence is assumed to be a function of exogenous variables as for example the slope of the yield curve.

3.2.1. Static, Dynamic and Autoregressive Probit Models

Probit models attempt to describe the dynamics in the time series of a binary variable R_t which is assumed to be Bernoulli distributed, i.e.

$$R_t|\mathscr{F}_{t-k} \sim \mathscr{B}(\mathrm{P_t}) \tag{3.2.1}$$

where the filtration $\mathscr{F}_{t-k}$ is the information set at time $t-k$, and k our forecasting horizon. Time series probit models assume the general stucture for the conditional probability P_t takes the form

$$\mathrm{P}(R_t = 1|\mathscr{F}_{t-k}) = \Phi(\pi_t) \tag{3.2.2}$$

where $\Phi(\cdot)$ is a strictly monotonically increasing function that takes values between 0 and 1 – in our case the normal cumulative distribution function – and π_t is linear function of forecasting variables. Following Kauppi and

Saikkonen (2008) and Ratcliff (2011), probit models can generally be seperated into four kind of models depending on the choice on which variables to include in π_t time series that allow different degrees of time series dynamics. Let k denote our forecasting horizon, then the models we consider look as follows:

1. *Static:* $\pi_t = c + \alpha \mathcal{S}_{t-k}$ (model (a_k))

2. *Dynamic:* $\pi_t = c + \alpha \mathcal{S}_{t-k} + \beta R_{t-1}$ (model (b_k))

3. *Autoregressive:* $\pi_t = c + \alpha \mathcal{S}_{t-k} + \gamma \pi_{t-1}$ (model (c_k))

4. *Dynamic Autoregressive:* $\pi_t = c + \alpha \mathcal{S}_{t-k} + \beta R_{t-1} + \gamma \pi_{t-1}$ (model (d_k))

The dynamic models (b_k) and (d_k) account for the fact that we observe strong autocorrelation in the time series of recession dummies R_t. To interprete the lagged value of π_t, note that

$$\pi_t = \Phi^{-1}[\mathrm{P}(R_t = 1|\mathscr{F}_{t-1})]$$

and we can therefore interprete models incorporating this dynamic structure as including an autoregressive structure in the time series of conditional recession probabilities. The autoregressive models c_k and d_k allow yield curve spreads at time $\tau, \tau < t - k$ to influence the probability of period t being in a recession indirectly through π_{t-1}.

3.2.2. Maximum Likelihood Estimation of Probit Models and Asymptotic Results

Parameter estimation of Probit models can be carried out using standard maximum likelihood estimation techniques as for example described by Kauppi and Saikkonen (2008). Let us denote the observed time series of recession dummies by $\boldsymbol{R}_{1:T} \equiv (R_1, R_2, \ldots, R_T)'$, the series of spreads by $\boldsymbol{\mathcal{S}}_{(1-k):T} \equiv (\mathcal{S}_{1-k}, \ldots, \mathcal{S}_1, \ldots, \mathcal{S}_T)'$ and the series of π by $\boldsymbol{\pi}_{0:T} \equiv (\pi_0, \pi_1, \ldots, \pi_T)'$. The likelihood function is then given by

$$\mathrm{P}(\boldsymbol{R}'_{1:T}|\boldsymbol{\mathcal{S}}'_{(1-k):T}, \boldsymbol{\pi}'_{0:T}) = \prod_{t=1}^{T} \mathrm{P}(R_t|\mathcal{S}_{t-k}, R_{t-1}, \pi_{t-1})$$

and taking logs as well as using equation (3.2.2) yields the log-likelihood as

$$\begin{aligned} \mathscr{L}(\tilde{\boldsymbol{\theta}}) &= \sum_{t=1}^{T} l_t(\tilde{\boldsymbol{\theta}}) \\ &= \sum_{t=1}^{T} \{R_t \log \Phi(\pi_t(\tilde{\boldsymbol{\theta}})) + (1 - R_t) \log(1 - \Phi(\pi_t(\tilde{\boldsymbol{\theta}})))\}. \end{aligned} \tag{3.2.3}$$

Note that in order to back out the whole series $\boldsymbol{\pi}_{0:T}$, we need to assume an initial value of π_0. We follow Kauppi and Saikkonen (2008) and choose π_0 to be the the unconditional mean of π_t, i.e.

$$\pi_0 = \frac{c + \alpha \bar{\mathcal{S}} + \beta \bar{R}}{1 - \gamma}.$$

Kauppi and Saikkonen (2008) also provide asymptotic results they applied following parameter estimation which account for possible model misspecification. A possible misspecification would be to estimate a Probit model although the underlying process follows a logit process or not including the appropriate number of lags. Let $\boldsymbol{\theta}^*$ be a value in the parameter space that maximizes the probability limit of the likelihood contribution $T^{-1}l(\tilde{\boldsymbol{\theta}})$, then the limiting distribution of $\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)$ is given by

$$\mathcal{N}(0, \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}). \tag{3.2.4}$$

If the model is correctly specified, we have $\boldsymbol{\theta}^* = \boldsymbol{\theta}$, i.e. the parameter vector maximizing the likelihood function in the limit is the true parameter vector. Matrix $\boldsymbol{A}$ can be consistently estimated for both a miss- and correctly specified model by

$$\hat{\boldsymbol{A}}(\hat{\boldsymbol{\theta}}) = \frac{1}{T}\sum_{t=1}^{T} \frac{\partial^2 l_t(\tilde{\boldsymbol{\theta}})}{\partial\tilde{\boldsymbol{\theta}}\partial\tilde{\boldsymbol{\theta}}'}\bigg|_{\tilde{\boldsymbol{\theta}}=\hat{\boldsymbol{\theta}}}$$

and an estimate for matrix $\hat{\boldsymbol{B}}$ is obtained as

$$\hat{\boldsymbol{B}}(\hat{\boldsymbol{\theta}}) = \frac{1}{T}\left(\sum_{t=1}^{T}\hat{\boldsymbol{d}}_t\hat{\boldsymbol{d}}_t' + \sum_{j=1}^{T-1} w_{Tj} \sum_{t=j+1}^{T}\{\hat{\boldsymbol{d}}_t\hat{\boldsymbol{d}}_{t-j}' + \hat{\boldsymbol{d}}_{t-j}\hat{\boldsymbol{d}}_t'\}\right)$$

where $\hat{\boldsymbol{d}}_t \equiv \partial l_t(\tilde{\boldsymbol{\theta}})/\partial\tilde{\boldsymbol{\theta}}|_{\tilde{\boldsymbol{\theta}}=\hat{\boldsymbol{\theta}}}$ and $w_{Tj} = k(j/m_T)$. The function $k(x)$ is a kernel function and m_T the corresponding bandwith which we choose as in Kauppi and Saikkonen (2008) $m_t = \lfloor 4(T/100)^{2/9} \rfloor$. For the function $k(x)$ a Gaussian kernel is chosen:

$$k(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$$

4. Empirical Evidence from Germany and the United States

4.1. Data

Interest rate and real GDP data used in our models for the US are taken from the Federal Reserve Bank of St. Louis:[3] The 10-Year Treasury Constant Maturity Rates range from April 1, 1953, to March March 1, 2015, on a monthly basis which makes 744 observations. The same source also supplies 3-Month Treasury Bill Market Rates available from January 1, 1934, to March 1, 2015. Based on this data we construct the series of monthly spreads as the difference between 10-year and 3-month rates starting in April 1953. Quarterly data on real GDP range from Quarter 1, 1947, to Quarter 4, 2014. To make the monthly spread data fit the quarterly frequency of GDP, arithmetic averages over the last three months are taken and the end-of-quarter spread averages are kept for further analysis, i.e. the average interest rate

[3] `https://research.stlouisfed.org/fred2/categories`, last access on July 4, 2015, 4 pm CET.

in for example quarter 4, 2014, is observed on January 1, 2015 and is the average of beginning-of-month spreads of October, November and December 2014. Real GDP growth rates are computed on a quarter-to-quarter basis as the log difference between real GDP in levels. The US Department of the Treasury also supplies Treasury yield curves on a daily basis from January 2, 1990, to July 2, 2015, for maturities of from 1 month to 30 years.[4] This data is not used in our analysis but only for illustrative purposes for certain plots. Moreover missing values are linearly interpolated. Interest rate and real GDP data for Germany are obtained from the German Bundesbank:[5] Svensson-fitted yield curves on listed federal securities are available from September 1972 to April 2015. The data for Germany are processed in the same way as the data for the US.

Figure 2 shows the evolution of the term structure of interest rates for the US and Germany based on the US Treasury Department and Bundesbank data: Especially for the US it is clear that short-term (3 months) rates are more volatile than long-term (30 and 10 years) rates (e.g. the standard deviation for 3-month rates is 2.34 whereas the standard deviation for 10-year rates is 1.82). It can be seen that the drecrease in the spread between long- and short-term rates preceding the 2007 recession is almost only driven by an increase in short-term rates and not by a decrease in long-term rates. Whereas short-term rates drop to zero during the recession as a consequence of monetary policy, the long-term rates only decrease slightly. We also note the strong persistency in interest rates and may wonder if interest rates actually follow a stationary process. However, even if we believe that interest rates follow a unit-root process, i.e. are non-stationary, this poses no problem for our model estimation as we are using the spread between long- and short-term rates, and could in this case reasonably argue that the two rates are cointegrated.

In order to evaluate our forecasts and implement our Probit model, we will have to define what a recession actually is. For the US, the National Bureau of Economic Research (NBER) Business Cycle Dating Committee officially determines the business cycle peaks and troughs, i.e. determine the months in which a recession starts and ends.[6]

[4] http://www.treasury.gov/resource-center/data-chart-center/interest-rates/Pages/TextView.aspx?data=yield, last access on July 4, 2015, 4 pm CET.

[5] http://www.bundesbank.de/Navigation/EN/Statistics/Time_series_databases/Macro_economic_time_series/its_list_node.html?listId=www_s140_it03a, last access on July 5, 2015, 3 pm CET.

[6] http://www.nber.org/cycles.html, last access on July 5, 2015, 3 pm CET.

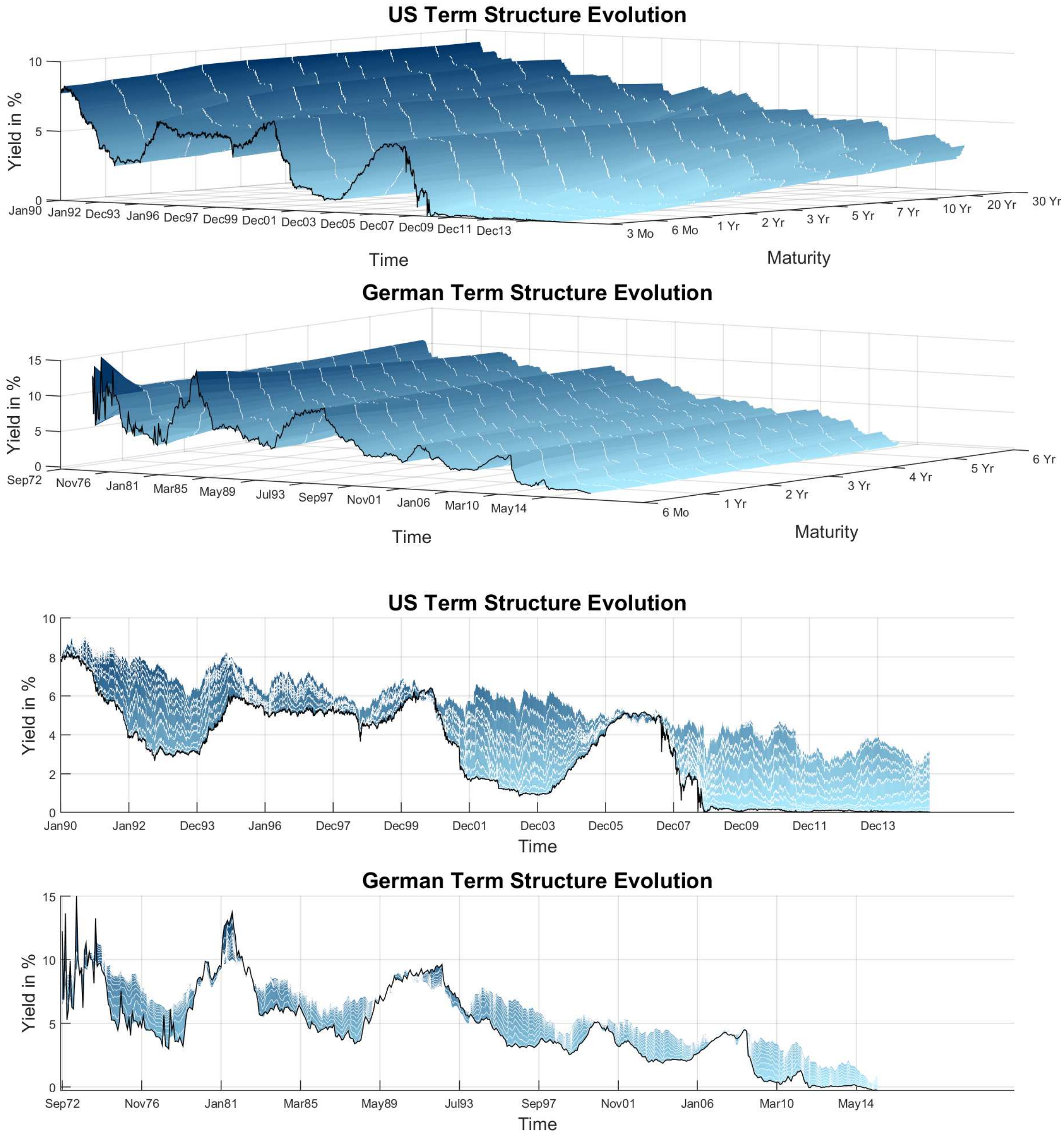


Figure 2: The plots show the term structure of interest rates evolution for the US and Germany based on the US Treasury Department and Bundesbank data. Daily US data range from Januar 2, 1990, to July 2, 2015, monthly German data range from September 1972 to June 2015. The two lower plots show the upper two plots in the X-Z-space.

Table 3: Business Cycle dates for the US as determined by the NBER Business Cycle Dating Committee. The quarter in which the business cycle reaches its peak/trough is given in brackets.

Peak	Through	Duration in Months
July 1953 (II)	May 1954 (II)	10
August 1957 (III)	April 1958 (II)	11
April 1960 (II)	February 1961 (I)	16
December 1969 (IV)	November 1970 (IV)	6
November 1973 (IV)	March 1975 (I)	16
January 1980 (I)	July 1980 (III)	6
July 1981 (III)	November 1982 (IV)	16
July 1990 (III)	March 1991(I)	8
March 2001 (I)	November 2001 (IV)	8
December 2007 (IV)	June 2009 (II)	18

Table 3 gives the NBER recession dates from 1953 up to 2009. For our analysis, the quarter in which the peak month occurs is still coded as an expansion period, i.e. non-recession quarter whereas the quarter in which the trough occurs is coded as a recession period. For Germany, there are no official recession dates as for the US. However, the Economic Cycle Research Institute (ECRI) offers business cycle dates for a variety of counties, among others Germany.[7] Comparing the recession dates of Germany and the US it

Table 4: Business Cycle dates for Germany as determined by the ECRI. The quarter in which the business cycle reaches its peak/trough is given in brackets.

Peak	Through	Duration in Months
March 1966 (I)	May 1967 (II)	14
August 1973 (III)	July 1975 (III)	23
January 1980 (I)	October 1982 (IV)	33
January 1991 (I)	April 1994 (II)	39
January 2001 (I)	August 2003 (III)	31
April 2008 (II)	January 2009 (I)	9

can be seen that whereas for the US, the most recent recession following the financial crisis is the most severe one in terms of duration, this is not true for Germany for which the so far most serious recession was the one at the beginning of the 90s with a duration of 39 months with compared to just 9 months of the 2007/2008 recession. A look at figure 3 makes clear how declines in the yield spread between long- and short-term rates (mostly due to an increase in short-term rates as mentioned above) historically precede and coincide with recession periods for both the US and Germany.

[7] https://www.businesscycle.com/ecri-business-cycles/international-business-cycle-dates-chronologies, last access on July 5, 2015, 5 pm CET.

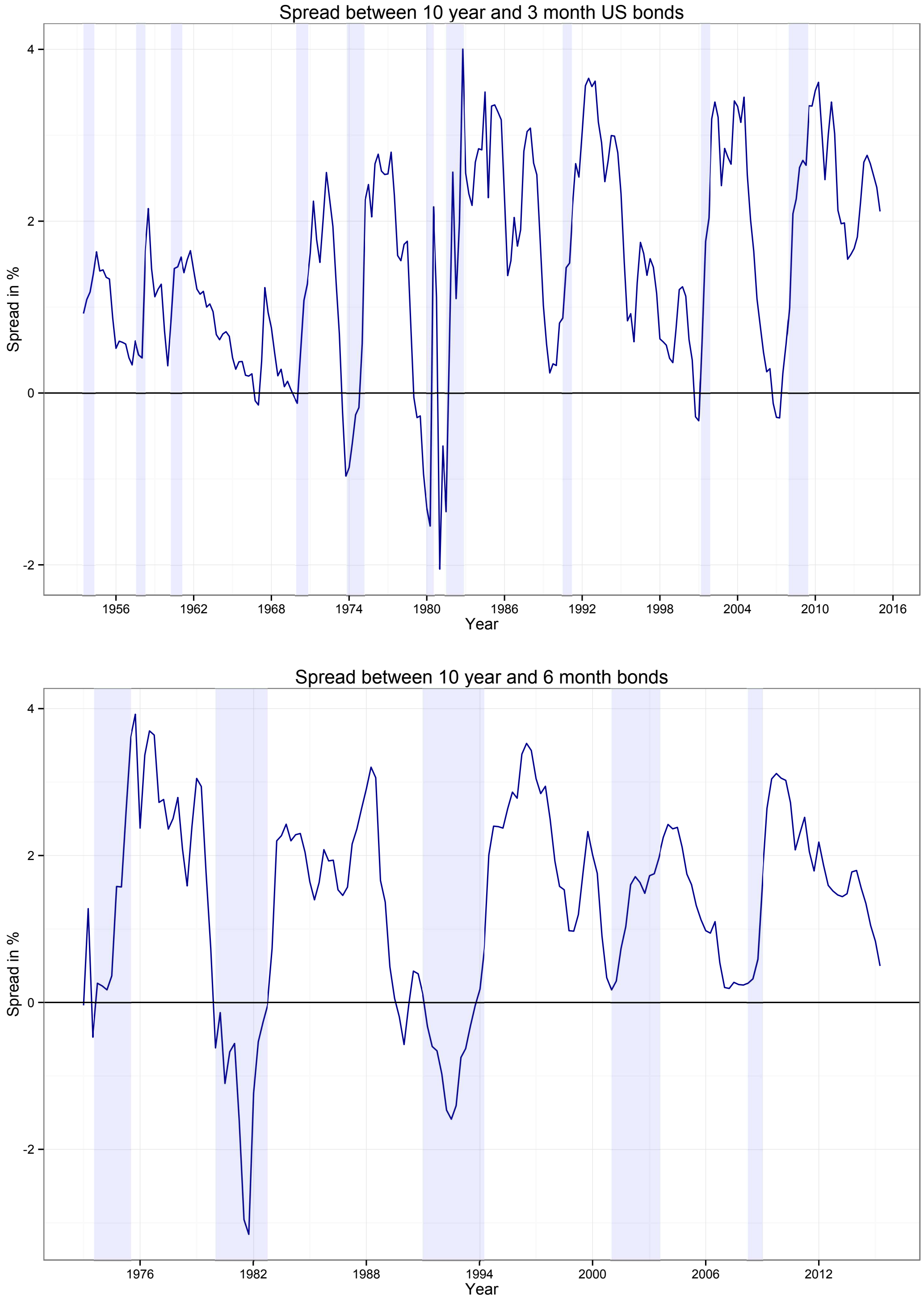


Figure 3: The upper plot shows the yield spread between 10 year and 3 month US treasury bonds, the lower plot the yield spread between 10 year and 6 month German Government bonds. The shaded areas mark recession periods as given in table 3 and 4.

4.2. Estimation Results

Estimation results for the Probit models using the full sample US data are printed in table 5: The spread parameter estimate for α is mostly highly statistically significant and negative, where significance is generally lower for the autoregressive models and not significant only for the autoregressive model using the one-quarter recession lag (model c_1). Thus a higher spread is associated with lower recession probabilities. The estimate for the recession parameter β is positive and statistically significant on the 1% level for all models as we have expected given the strong serial correlation in the recession time series. Finally, the parameter estimate for γ in the autoregressive models is positive when the lagged recession variable is excluded (model c) and negative when included (model d) for all spread lags. That is for model c periods of high recession probabilities are ceteris paribus followed by periods of high recession probabilities. The negative value for γ in model d is hard to interprete sensibly and may occur due to some interaction between the lagged recession variable and lagged recession probabilities. These results are consistent with the estimates reported by Kauppi and Saikkonen (2008) whose parameter estimates correspond in sign and size to the ones printed in this paper.

We additionally computed values for the Akaike Information Criterion (AIC) as well as the Bayesian Information Criterion (BIC): For both the static and autoregressive probit models, adding a lagged recession dummy variable greatly decreases the values for both information criteria. Using the AIC, the dynamic models b_k are preferred to the dynamic autoregressive models for $k = 1, 2$, the contrary is true for lags $k = 3, 4$. On the other hand, the BIC prefers the dynamic model for any lag which becomes clear when looking at the value of the log-likelihood function: Allowing for the lagged recession probability hardly increases the value of the log-likelihood.

The pseudo R^2 for the dynamic models is considerably higher than for the models without dynamics. The best fit is reached for model b_1 with a value of 39%. The difference in pseudo R^2 for the dynamic and the dynamic autoregressive model is only 1 percentage point for $k = 1$ and zero for the other lags. Overall we may conclude that the in-sample fit for model b_1 is best which is also supported by both information criteria.

Table 5: Probit estimation results for the US plus information criteria. Standard errors are given in parentheses, stars indicate statistical significance. ***: Significance at 1% level, **: Significance at 5% level, *: Significance at 10% level.

	$k=1$				$k=2$			
	a_k	b_k	c_k	d_k	a_k	b_k	c_k	d_k
Constant: $\hat{c}$	-0.68***	-1.23***	-0.04	-1.45***	-0.58***	-1.25***	-0.11	-1.65***
	(0.18)	(0.2)	(0.28)	(0.3)	(0.2)	(0.24)	(0.28)	(0.36)
Spread: $\hat{\alpha}$	-0.41***	-0.62***	-0.22	-0.65***	-0.57***	-0.51***	-0.32*	-0.58***
	(0.13)	(0.14)	(0.19)	(0.14)	(0.16)	(0.17)	(0.23)	(0.16)
Recession: $\hat{\beta}$		2.6***		2.93***		2.17***		2.75***
		(0.3)		(0.37)		(0.33)		(0.32)
Probability: $\hat{\gamma}$			0.76***	-0.14			0.63***	-0.26**
			(0.22)	(0.12)			(0.24)	(0.14)
Log-likelihood	-75.8	-40.87	-66.84	-40.57	-69.59	-42.79	-64.82	-42.01
Pseudo R^2	0.07	0.39	0.14	0.38	0.12	0.37	0.16	0.37
AIC	155.6	87.74	139.68	89.14	143.18	91.58	135.65	92.01
BIC	162.49	98.07	150.01	102.91	150.06	101.91	145.97	105.78

	$k=3$				$k=4$			
	a_k	b_k	c_k	d_k	a_k	b_k	c_k	d_k
Constant: $\hat{c}$	-0.53***	-1.23***	-0.25	-1.66***	-0.54***	-1.29***	-0.44*	-1.74***
	(0.21)	(0.24)	(0.28)	(0.29)	(0.22)	(0.23)	(0.31)	(0.28)
Spread: $\hat{\alpha}$	-0.66***	-0.5***	-0.47**	-0.6***	-0.65***	-0.41***	-0.58**	-0.51***
	(0.17)	(0.14)	(0.25)	(0.15)	(0.15)	(0.12)	(0.28)	(0.13)
Recession: $\hat{\beta}$		2***		2.63***		1.96***		2.55***
		(0.31)		(0.28)		(0.29)		(0.26)
Probability: $\hat{\gamma}$			0.4*	-0.31***			0.14	-0.32***
			(0.28)	(0.13)			(0.38)	(0.12)
Log-likelihood	-66.11	-43.89	-64.67	-42.45	-67.07	-46.06	-66.92	-44.31
Pseudo R^2	0.15	0.36	0.16	0.36	0.14	0.34	0.14	0.34
AIC	136.21	93.77	135.34	92.91	138.14	98.12	139.83	96.62
BIC	143.1	104.1	145.67	106.68	145.02	108.44	150.16	110.39

Estimation results based on German data are printed in table B.1 of appendix B: As for the US data, the spread parameter α is highly significant for all models and all lags, the same is true for the lagged recession parameter β in the dynamic models. Furthermore the parameter γ of the autoregressive models c_k has a significant positive sign for $k = 1, 2$ and is insignificant for $k = 3, 4$. As we have already seen above using US data, the sign of γ is negative for models d_k which raises the same questions in terms of interpretation. Both AIC and BIC prefer the dynamic models b_k for all lags k which is supported by the pseudo R^2 for model b_k which is always higher than or equal to the pseudo R^2 of model d_k. As for the US, the best model for Germany in terms of pseudo R^2 – which decreases in k for model b_k – is model b_1.

4.3. Forecast Evaluation and Threshold Choice

4.3.1. Forecast Evaluation Methodology: Metrics

Both the VAR models and the probit models enable us to compute recession probabilities once having obtained parameter estimates. However, the forecaster may want to translate these recession probabilties to a simple yes/no forecast based on these probabilities. To do so, one can define a threshold for the recession probability p^* such that

$$\hat{R}_{t+1} = \begin{cases} 1 & \text{if } \mathrm{P}(R_{t+1} = 1 | \mathscr{F}_t) \geq p^* \\ 0 & \text{otherwise} \end{cases}$$

where $\hat{R}_{t+1}$ is our one-period-ahead recession forecast variable. For our in-sample analysis, we can choose p^* in a way such that hits and correct rejections are maximized for a given level of misses and false alarms as defined in table 6.

Table 6: This matrix shows all four possible Forecast/Realization combinations (forecast recession – actual recession, forecast recession – actual expansion, forecast expansion – actual recession and forecast expansion – actual expansion).

Forecast	**Realization**		**Sum**
	Recession	**Expansion**	
Recession	Hits (11)	False alarms (10)	(11) + (10)
Expansion	Misses (01)	Correct rejections (00)	(01) + (00)
Sum	(11) + (01)	(10) + (00)	T

Let the number of hits for a given threshold be denoted by h, the number of false alarms by f, the number of misses by m and the number of correct rejections by c. To evaluate the goodness of our forecasts for different threshold levels p^*, three forecast evaluation metrics for binary forecast variables as among others applied by Ratcliff (2011) are used:

1. Equitable Threat Score (ETS): The ETS is defined as

$$\mathrm{ETS} \equiv h - h_r/(h + m + f - h_r)$$

   where the number of "random hits" h_r is defined as $h_r \equiv (h + f)(h + m)/T$. The ETS gives the same negative weight to misses and false alarms and accounts for hits that are expected to have occured just due to chance. Using this criterion, a higher ETS is favourable.

2. Bias: The bias is defined as

$$\text{bias} \equiv \frac{h + f}{h + m}$$

and measures if our model is biased towards overforecasting ($\text{bias} > 1$) or towards underforecasting ($\text{bias} < 1$) recession. The best model has a bias close to 1 (if $\text{bias} = 1$ then $f = m$ and our model is unbiased in a sense that the number false alarms equals the number of misses). We will redefine the bias by substracting one, such that our optimal model has a bias that is close to zero.

3. Hit Rate minus False Alarm: Finally, this measure is defined as

$$\text{hmf} \equiv \frac{h}{h+m} - \frac{f}{f+c}$$

where the perfect model with $f = m = 0$ has an hmf of 1.

Each of the 16 probit models (model a_k to d_k for spread lags $k = 1, \ldots, 4$) is evaluated at the thresholds of $p^* = 0.1, 0.2, \ldots, 0.9$.

4.3.2. Out-of-Sample Forecast Evaluation Using Optimal Thresholds: Results

The US and German sample are both split into an in-sample and an out-sample period. The US in-sample period ends at Q1 1980, the German in-sample period at Q4 1982. Model parameters are re-estimated every four quarters on a rolling basis and the new parameter estimates used to compute recession probabilities.

Probit Models for the US

The optimal threshold for each criterion and each probit model using US data is printed in table 8. Note that it may be the case that several consecutive thresholds yield the same criterion values as the categorization for quarters into expansion and recession periods does not change (this is obvious for very fine grids, e.g. in general it does not matter if one uses a threshold of $p^* = 0.5$ or $p^* = 0.51$). When we do find K optimal thresholds sorted in increasing order, $p_i^*, p_{i+1}^*, \ldots, p_{i+K-1}^*$, threshold $p_j^*, j = \lceil K/2 \rceil$ is chosen. The first column named k gives the spread lag used for model a_k, b_k, c_k and d_k. The optimal thresholds over all criteria are generally below 0.5, lying around 0.1 to 0.4 with the exception of the ETS for model d_4 which is the only model-criterion combination that yields an optimal threshold of 0.5. The left part of the table prints the results for models a and c, i.e. our models that do not include a lagged recession dummy whereas the right part of the table contains results for the models b and d which do include a lagged recession dummy variable. As expected, including the lagged recession variable improves the out-of-sample forecasting power for all models and all criteria (ETS and HMF criterion values increase whereas bias decreases). This mostly comes from an improved hit rate: Including the lagged recession variable, the hit rate of model a_1 increases from 35.71% to 64.29% whereas the share of correct rejections increases as well from 82.76% to 96.06%. When quarter t is in recession, the lagged-recession variable models will generally assign a probability larger than 50% that the next quarter is in recession as well, i.e. the model implication is to expect the next quarter still to be in a recession whenever the previous quarter is in recession. Therefore we will generally fail to forecast the end of a recession period. Moreover, these dynamic models also do not improve the model's ability to forecast the start of a recession.

Taking a closer look at the ETS criterion, we see that the performance of the dynamic and the dynamic autoregressive models b and d is the same for recession lags $k = 1, 2, 3$ and the performance for the dynamic autoregressive model is actually worse than for the simple dynamic model for recession lag $k = 4$, meaning that the additional autoregressive feature does not actually improve the forecasting performance. Whenever criterion values for the simple dynamic and the more complex dynamic autoregressive model are equal, we are going to opt for the dynamic model considering the values of the information criteria in table 5 that tend to favor the dynamic compared to the dynamic autoregressive models. Thus the best model according to the ETS is the dynamic model with one lag, b_1, and a threshold of 0.2, followed by the same model with a lag of $k = 4$. The hmf criterion prefers the dynamic model with a low threshold of 0.1 and lag $k = 2$ (model b_2). Last but not least the bias criterion favours dynamic model b_4 with a threshold of 0.3.

Probit Models for Germany

The results using German data are printed in table B.2 of appendix B: The share of overall correct forecasts ranges from 81% to 90% for non-dynamic models a_k and c_k and from 94% to 97% for dynamic models b_k and d_k, i.e. dynamic models perform strictly better than non-dynamic models in terms of overall correct forecasts. We also observe that criterion values generally improve when including dynamics.

Note that the evaluation results for the dynamic models b_k and d_k are especially for $k = 3, 4$ quite insensitive to changes in the threshold as the models assign either large recession probabilities close to 1 or very low probabilities close to 0 to individual quarters. This is due to the fact that the effect of the lagged recession dummy taking a value of 1 is larger for Germany than for the US, which is in turn caused by a higher "marginal" lagged recession effect $\hat{a}$. For example in case of the US, during recession periods model b_4 assigns probabilities between 70% and 90% whereas for Germany, recession probabilities between 70% and 100% are assigned. Recession probability plots for Germany are printed in figure A.3 of appendix A. Be aware that – as mentioned above for the US – including dynamics improves the evaluation metrics by exploiting the high serial correlation in the recession dummy time series and generally leads to a failure when it comes to forecasting the end of a recession as the first quarter in an expansion period will falsely be declared as a recession period.

Therefore, despite the conclusion that dynamic models seem to outperform the simple static and autoregressive models so clearly, we may not yet fully reject the latter ones depending on the aim of the forecaster: When it comes to forecasting the beginning of a recession, the models without dynamics tend to show a relatively smooth increase in recession probabilities in line with the decrease in spreads preceding recessions. To make this point clearer, compare the German recession probability plots for models a_1 and b_1 in figure A.3 of appendix A: Before the 2008 recession, recession probability constantly increases from a low level of about 2.5% in 2005 to almost 50% at the end of 2007 for model a_1. However for model b_1, there is an increase of the recession probability in line with the decrease in the yield curve spread, however this increase starts from a level of 0% in 2005 and just goes to a level of about 12% at the end of 2007. This difference is due to a much lower marginal effect of spread changes on the recession probability due to differences in the regression constant: The regression constant is 0.13 for model a_1 and -0.91 for model b_2. Noting that the standard normal c.d.f. has its maximum slope at 0, this difference causes the difference in marginal effects of spread changes. This effect is also in place for the US, however less strong as the difference in regression constants is smaller.

Table 7: This table contains the optimal thresholds for the VAR, SBVAR, TSVAR and SBTVAR and each of the three criteria based on US data. Column CF contains the share of overall correct forests (recession and expansion), column HR the hit rate (share of correct recession forecasts) and column CR the share of correct rejections (predict expansion when expansion occurs). Column CritVal contains the criterion value at the optimal threshold in %.

Criterion	Treshold	CF in %	HR in %	CR in %	CritVal
			VAR		
ETS	0.2	77.88	57.14	80.81	15.70
Bias	0.3	87.17	25.00	95.96	46.43
hmf	0.2	77.88	57.14	80.81	37.95
			SBVAR		
ETS	0.3	83.63	64.29	86.36	25.14
Bias	0.4	86.73	39.29	93.43	14.29
hmf	0.2	77.43	75.00	77.78	52.78
			TSVAR		
ETS	0.4	88.50	25.00	97.47	17.50
Bias	0.3	81.86	35.71	88.38	17.86
hmf	0.1	62.83	100.00	57.58	57.58
			SBTVAR		
ETS	0.2	78.32	78.57	78.28	22.16
Bias	0.3	84.51	46.43	89.90	17.86
hmf	0.2	78.32	78.57	78.28	56.85

VAR Models for the US

Threshold choices and forecast evaluation results for the VAR, SBVAR, TSVAR and SBTVAR are printed in table 7 for the US in in table B.3 for German data. For the US, the simple VAR is clearly outperformed by the SBVAR, the TSVAR and SBTVAR as all three evaluation criteria are worse for the VAR than for the other models, indicating that including non-linear dynamics in the form of a structural break and/or a threshold improves forecasting performance. Using the ETS, the SBVAR is best, followed by the SBTVAR and the TSVAR. However, taking into account the both the bias and hmf criterion, the picture is not so clear anymore: Whereas the bias favours the SBVAR, the hmf takes its highest value for the TSVAR, followed by the SBT-VAR. Taking a look at the optimal threshold values, we see they range from 0.1 to 0.4 and are thus similar to the range of the Probit thresholds for the US. It is thus optimal to choose a threshold below 0.5, no matter which model we opt for.

VAR Models for Germany

Let us now turn to the results for Germany which are reported in table B.3 of appendix B: For any of the three criteria, the simple VAR with an optimal threshold of 0.3 outperforms the SBVAR, TSVAR and SBTVAR and has a share of overall correct forecasts of 87.25% which is higher than for any other model. Just looking at the SBVAR, TSVAR and SBTVAR, on the other end, the SBVAR performs worst for all three criteria, i.e. for each criterion there is a model which that shows a better criterion value, even after excluding the simple VAR: E.g. considering the ETS, model TSVAR performs best, closely followed by the SBTVAR, which also has the lowest bias. The highest hmf criterion value is for the TSVAR model.

Model Forecasting Performance Comparison: Probit against VARs, US against Germany

We have so far determined which models within each model "family" – VAR and Probit models – can be regarded as the best and worst when it comes to forecasting applying different forecast evaluation criteria. We now want to compare 1. how for both the US and Germany, the best Probit model compares to the best VAR model and 2. how the best models' forecasting performance differs between the two countries.

First of all, for the US, we may want to compare the dynamic probit model b_1 with a threshold of 0.2 to the SBVAR with a threshold of 0.3. The probit model appears to perform better than the SBVAR with a share of overall correct forecasts of 92.64% for the probit against 83.63% for the SBVAR. For the probit, both the hit rate (78.57% against 64.29%) and share of correct rejections (94.58% against 86.36%) are better than for the SBVAR.

For Germany, we will compare the dynamic probit model b_1 with a threshold of 0.6 to the VAR with a threshold of 0.3. As for the US, we observe that the probit with a share of overall correct forecasts of 96.75% performs better than the VAR with a share of overall correct forecasts of 87.25%: Again, we see that for the Probit both the hit rate (89.19% against 70.27%) and the share of correct rejections (99.15% against 92.86%) are higher than for the VAR.

Looking at the numbers above, it becomes clear that both models, VAR and Probit, perform better for Germany than for the US. This may be due to the fact that our sample for Germany is smaller than for the US and there are fewer recessions in our forecasting period for Germany. Whereas including a structural break when modelling US growth and spread dynamics improves forecasting performance, this is not the case for Germany. However, there are also similarities: The probit appears to perform better than the VARs for both countries and dynamic probits have proven best in terms of our forecast evaluation criteria.

Table 8: This table contains the optimal thresholds p^* for each of the 16 probit models (model a_k to d_k for spread lags $k = 1, \ldots, 4$) and each of the three criteria based on US data. Column CF contains the share of overall correct forecasts (recession and expansion), column HR the hit rate (share of correct recession forecasts) and column CR the share of correct rejections (predict expansion when expansion occurs). Column CritVal contains the criterion value at the optimal threshold in %.

				a_1					b_1		
k	**Criterion**	**Threshold**	**CF in %**	**HR in %**	**CR in %**	**CritVal**	**Threshold**	**CF in %**	**HR in %**	**CR in %**	**CritVal**
	ETS	0.3	88.31	17.86	98.03	12.65	0.2	92.64	78.57	94.58	51.43
	Bias	0.2	77.06	35.71	82.76	60.71	0.4	92.21	64.29	96.06	7.14
	hmf	0.1	58.01	71.43	56.16	27.59	0.1	87.88	85.71	88.18	73.89
1				c_1					d_1		
	ETS	0.3	84.85	39.29	91.13	17.62	0.2	92.64	78.57	94.58	51.43
	Bias	0.3	84.85	39.29	91.13	3.57	0.4	93.07	71.43	96.06	0.00
	hmf	0.1	68.83	92.86	65.52	58.37	0.1	88.74	85.71	89.16	74.88
				a_2					b_2		
	Criterion	**Threshold**	**CF in %**	**HR in %**	**CR in %**	**CritVal**	**Threshold**	**CF in %**	**HR in %**	**CR in %**	**CritVal**
	ETS	0.3	87.45	35.71	94.58	20.45	0.2	92.64	75.00	95.07	50.35
	Bias	0.3	87.45	35.71	94.58	25.00	0.4	92.64	67.86	96.06	3.57
	hmf	0.1	64.94	85.71	62.07	47.78	0.1	90.04	85.71	90.64	76.35
2				c_2					d_2		
	ETS	0.3	86.58	46.43	92.12	23.43	0.2	92.64	75.00	95.07	50.35
	Bias	0.3	86.58	46.43	92.12	3.57	0.4	92.64	67.86	96.06	3.57
	hmf	0.1	68.83	92.86	65.52	58.37	0.1	89.61	82.14	90.64	72.78
				a_3					b_3		
	Criterion	**Threshold**	**CF in %**	**HR in %**	**CR in %**	**CritVal**	**Threshold**	**CF in %**	**HR in %**	**CR in %**	**CritVal**
	ETS	0.2	81.82	67.86	83.74	23.21	0.3	92.64	71.43	95.57	49.23
	Bias	0.3	87.01	39.29	93.60	14.29	0.3	92.64	71.43	95.57	3.57
	hmf	0.1	69.70	92.86	66.50	59.36	0.1	88.74	78.57	90.15	68.72
3				c_3					d_3		
	ETS	0.3	87.45	42.86	93.60	23.62	0.3	92.64	71.43	95.57	49.23
	Bias	0.3	87.45	42.86	93.60	10.71	0.3	92.64	71.43	95.57	3.57
	hmf	0.1	70.56	92.86	67.49	60.34	0.1	89.61	78.57	91.13	69.70

Table 8 continued

				a_4					b_4		
	Criterion	**Treshold**	**CF in %**	**HR in %**	**CR in %**	**CritVal**	**Treshold**	**CF in %**	**HR in %**	**CR in %**	**CritVal**
	ETS	0.2	82.25	67.86	84.24	23.82	0.4	93.07	71.43	96.06	50.93
	Bias	0.3	85.28	32.14	92.61	14.29	0.4	93.07	71.43	96.06	0.00
	hmf	0.1	68.40	89.29	65.52	54.80	0.1	89.18	78.57	90.64	69.21
4				c_4					d_4		
	ETS	0.2	80.95	67.86	82.76	22.06	0.5	93.07	67.86	96.55	49.76
	Bias	0.3	85.71	32.14	93.10	17.86	0.4	92.21	67.86	95.57	0.00
	hmf	0.1	70.13	89.29	67.49	56.77	0.1	90.04	78.57	91.63	70.20

5. Concluding Remarks

In this paper, we have estimated differently specified Probit and VAR models and seen that simple dynamic probits, i.e. probit models include lagged recession dummies, appear to perform best in recession forecasting. Whereas for the US, including a structural break in the VAR (SBVAR) improves forecasting performance, this is not the case for Germany where as simple VAR performs better than more complex VARs that include non-linearities. We also included a spread-threshold in the VAR model (i.e. the SBVAR and SBTVAR) but found that neither for Germany nor the US, there is a clear benefit when it comes to recession forecasting. Our probit models have the benefit that the parameter estimates are easy to interprete and quite stable over time, i.e. robust to including new observations in the estimation, which is not the case for the VARs. Despite that fact that dynamic probits seem to perform so well by just looking at the numbers, there are some drawbacks to mention: First of all, the inclusion of a lagged recession dummy leads to a failure of the probit model to correctly forecast the end of a recession. In our samples, the dynamic probits always incorrectly forecasted the first expansive quarter after a recession as a recession period. In fact, much of the probit forecasting performance comes from the simple fact that when we already are in a recession period, the following quarter is assigned a very high recession probability as well. Many papers on this topic only refer to the good forecasting performance measures but do not mention this problematic fact. When it comes to evaluating the predictive power of the yield spread itself, probit models without dynamics may be more informative.

We may also note that the computation of non-linear VARs through grid-search is computationally quite expensive and a drawback of these kind of models compared to probit models. The researcher may not be sure if the true optimum is actually found when an optimum on the grid is being determined. As we have seen for Germany, a simple VAR can actually perform better than more complex specifications and therefore the estimation of non-linear VARs may not be worthwhile. A clear benefit of VARs is that no information on previous recession periods are required but only the time series of spreads and growth rates. The VAR models are also less prone to overfitting as might be the case with probit models that include a lagged recession variable. Thus, from a practical perspective – e.g. monitoring of economic activity to register possible recession signals – a simple VAR or a probit without lagged recession variables might be the best choice.

It has been discussed in the literature whether the predictive power of the yield curve has decreased since the 90s. The financial crisis of 2007/08 has provided a new recessive period that enabled us to re-evaluate the yield spread. Our data shows that this most recent recession so far has been preceded by a steady increase in short-term bond yields, resulting in a decline in the spread. Given that the probit models directly map the yield spread to a value (probability) between 0 and 1, this translates into a steadily increasing recession probability before the outbreak of the actual recession. However, it is also true that the declines in interest rate spreads preceding recessions has been less strong for the 2001 and 2007 recessions than in previous decades (although spreads still became negative for the US).

Last but not least, a convincing theoretical explanation for the yield curve's predictive power is still missing. It has been discussed if US monetary policy has actually been the cause of the 2007 financial crisis due to its low interest rate policy. An increase in short-term rates by the FED may then have triggered subprime borrowers to default due to contracts that specified variable interest payments depending on short-term rates (as e.g. the LIBOR). If this was true, then the observed predictive power of the yield curve would just be an artifact caused by an underlying variable, namely monetary policy.

References

CHAUVET, M., AND J. D. HAMILTON (2005): "Dating Business Cycles Turning Points," *NBER Working Paper Series*.

CHINN, M. D., AND K. J. KUCKO (2010): "The Predictive Power of the Yield Curve across Countries and Time," *NBER*.

COCHRANE, J. H. (2005): *Asset Pricing*. Princeton University Press.

DE LINT, C. R., AND D. STOLIN (2003): "The Predictive Power of the Yield Curve: A Theoretical Assessment," *Journal of Monetary Economics*, 50, 1603–1622.

DEN HAAN, W. J. (1995): "The Term Structure of Interest Rates in Real and Monetary Economies," *Journal of Economic Dynamics and Control*, 19, 909–940.

DEN HAAN, W. J., AND A. MARCET (1990): "Solving the Stochastic Growth Model by Parameterizing Expectations," *Journal of Business and Economics Statistics*, 8(1), 31–34.

DUARTE, A., I. A. VENETIS, AND I. PAYAC (2005): "Predicting real growth and the probability of recession in the Euro Area using the yield spread," *International Journal of Forecasting*, 21, 261–277.

GALVAO, A. B. C. (2006): "Structural Break Threshold VARs for Predicting US Recessions using the Spread," *Journal of Applied Econometrics*.

HARVEY, C. R. (1988): "The Real Term Structure and Consumption Growth," *Journal of Financial Economics*, 22, 305–333.

KAUPPI, H., AND P. SAIKKONEN (2008): "Predicting US Recessions with Dynamic Binary Response Models," *The Review of Economics and Statistics*, 90(4), 777–791.

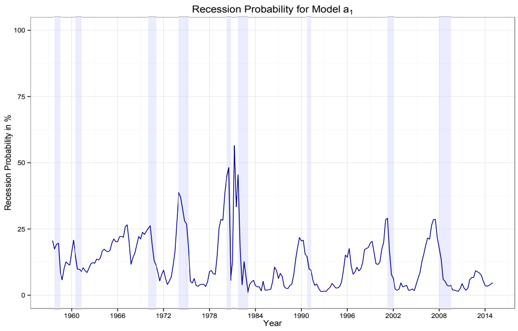
LÜTKEPOHL, H. (2005): *New Introduction to Multiple Time Series Analysis*. Springer.

Politis, D. N., and J. P. Romano (1994): "The Stationary Bootstrap," *Journal of the American Statistical Association*.

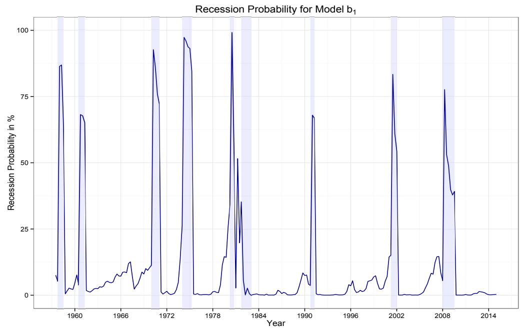
Ratcliff, R. (2011): "The "probability of recession": Evaluating probabilistic and non-probabilistic forecasts from probit models of U.S. recessions," *Economics Letters*, 121, 311–315.

Wheelock, D. C., and M. E. Wohar (2009): "Can the Term Spread Predict Output Growth and Recessions? A Survey of the Literature," *Federal Reserve Bank of St. Louis Review*, 91(5), 419–40.

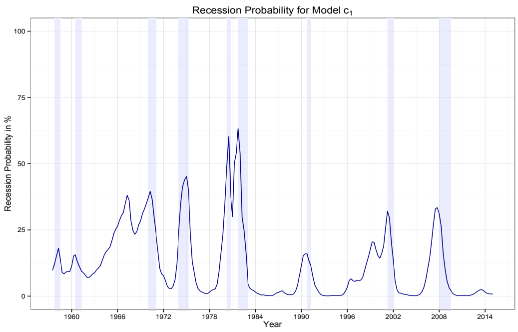
A. Supplementary Figures



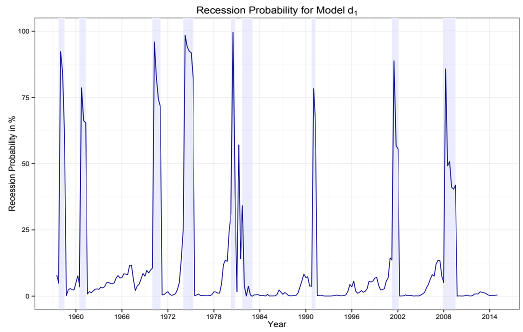
Model a_1, US data



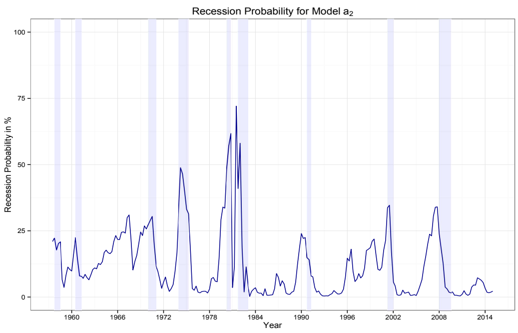
Model b_1, US data



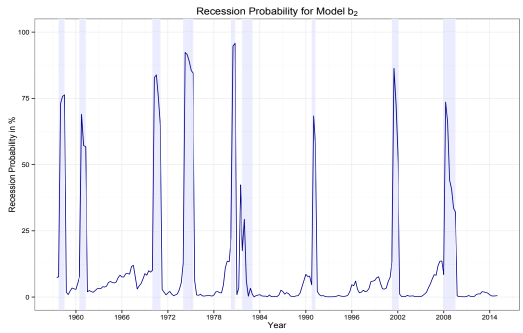
Model c_1, US data



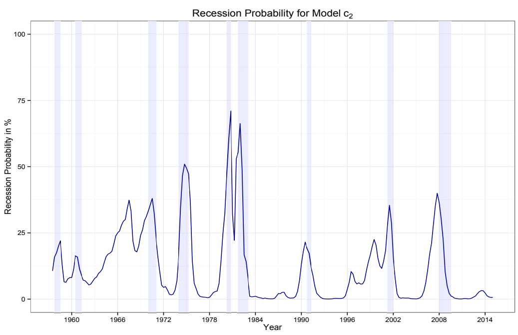
Model d_1, US data



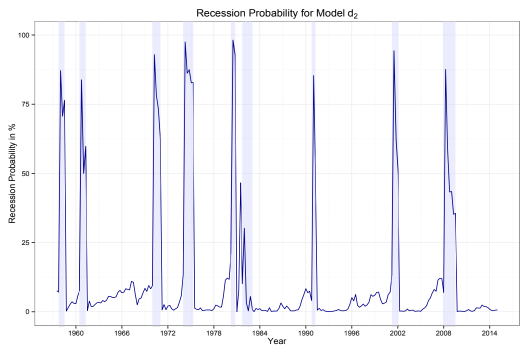
Model a_2, US data



Model b_2, US data

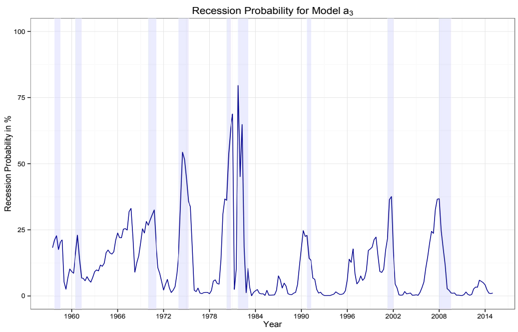


Model c_2, US data

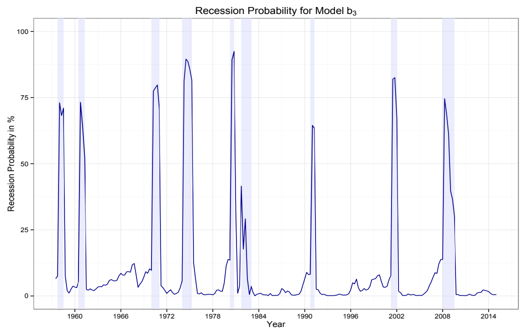


Model d_2, US data

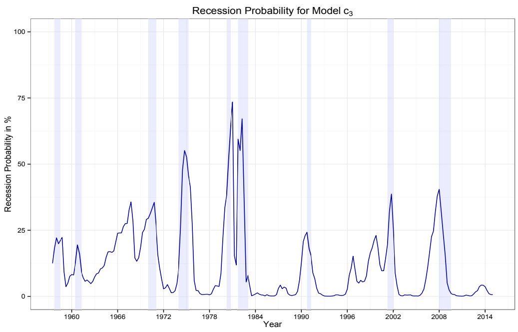
Figure A.1: Recession probability plots for probit models a_k–d_k for $k = 1, 2, 3, 4$ using US data.



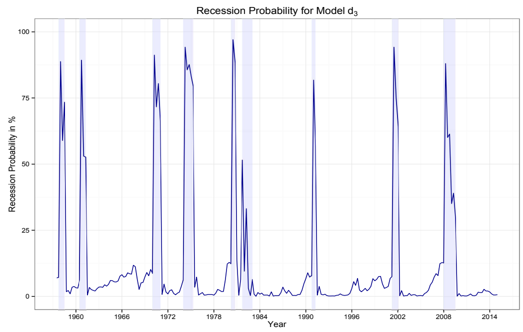
Model a_3, US data



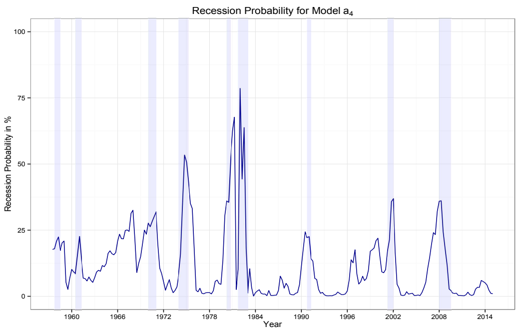
Model b_3, US data



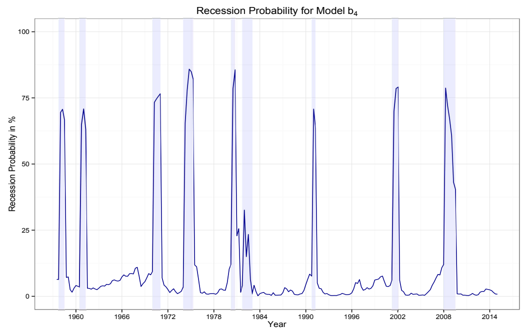
Model c_3, US data



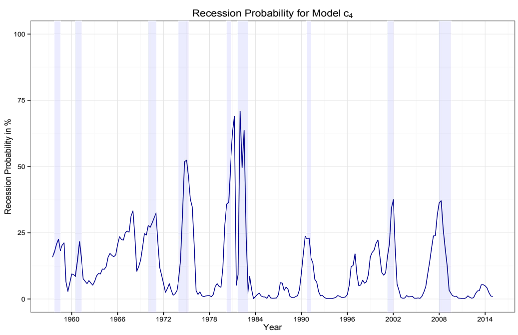
Model d_3, US data



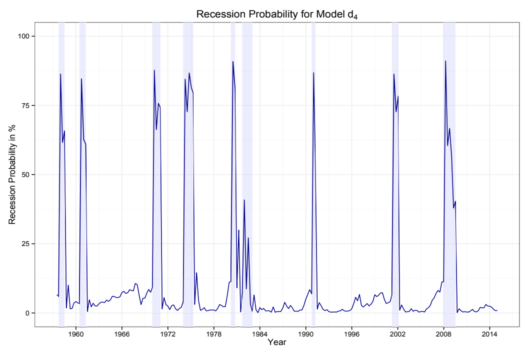
Model a_4, US data



Model b_4, US data

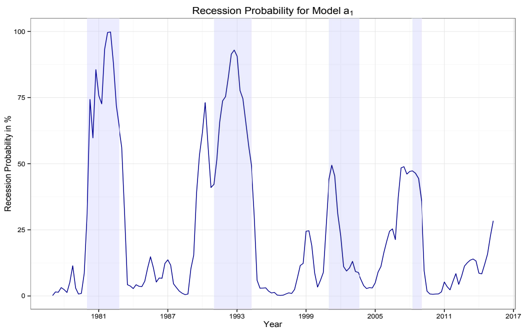


Model c_4, US data

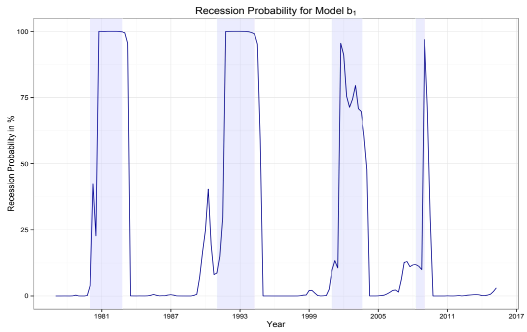


Model d_4, US data

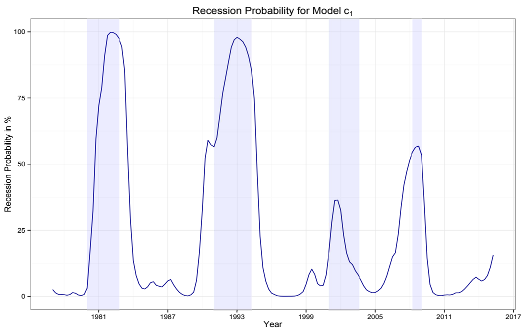
Figure A.1 continued



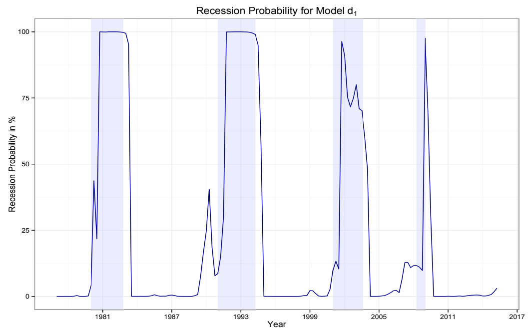
Model a_1, German data



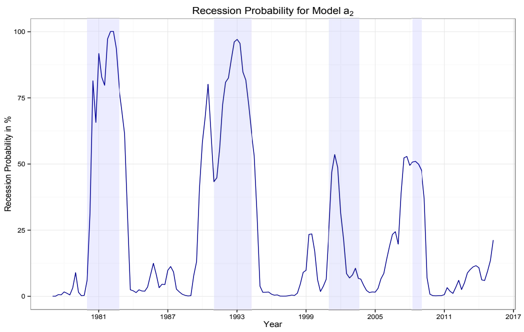
Model b_1, German data



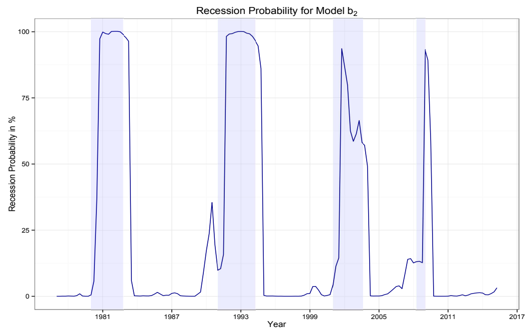
Model c_1, German data



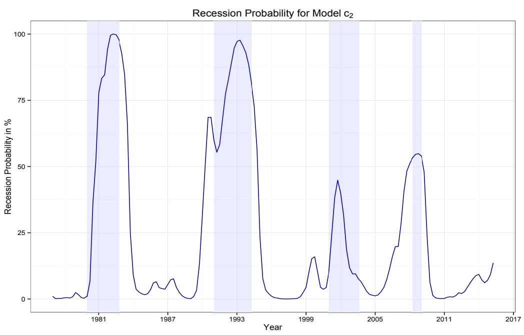
Model d_1, German data



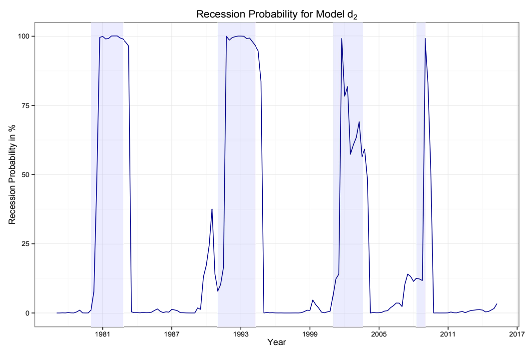
Model a_2, German data



Model b_2, German data

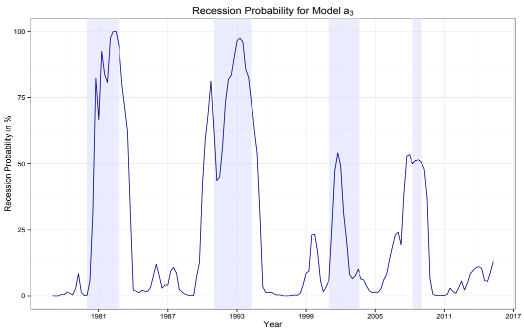


Model c_2, German data

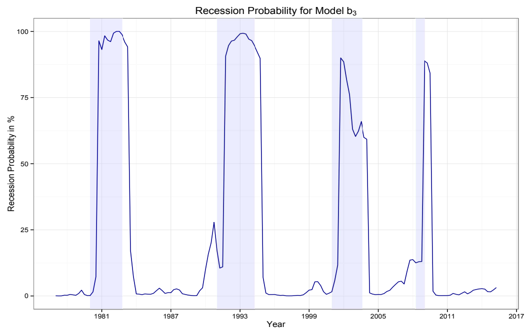


Model d_2, German data

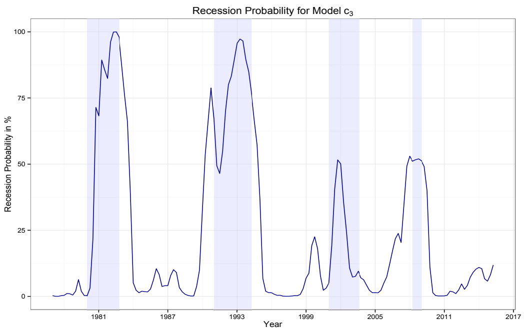
Figure A.3: Recession probability plots for probit models a_k–d_k for $k = 1, 2, 3, 4$ using German data.



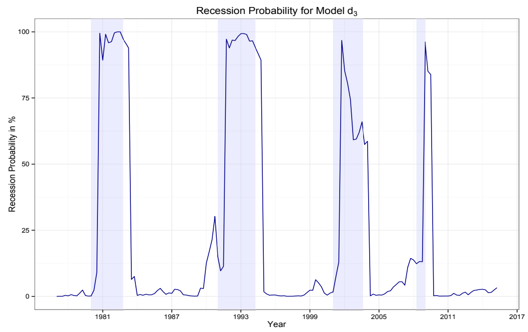
Model a_3, German data



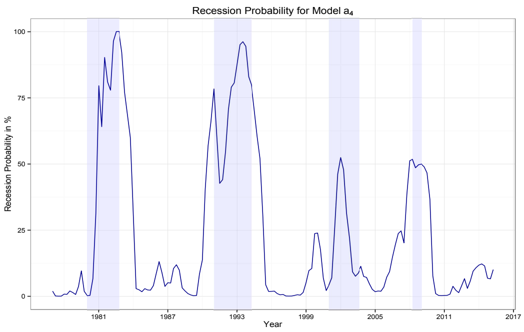
Model b_3, German data



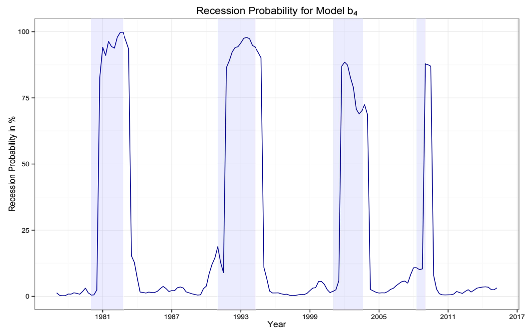
Model c_3, German data



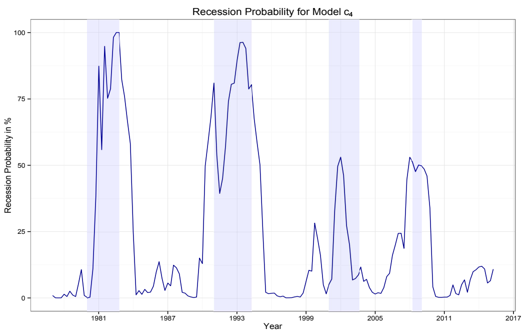
Model d_3, German data



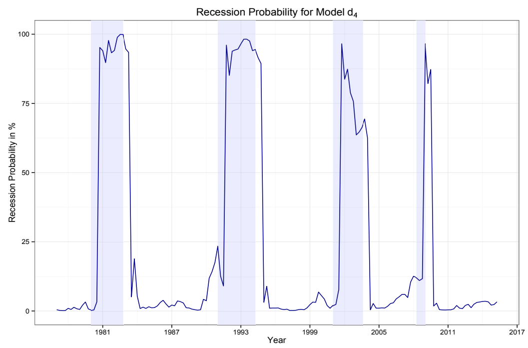
Model a_4, German data



Model b_4, German data

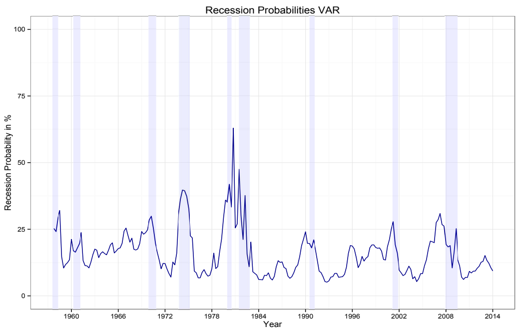


Model c_4, German data

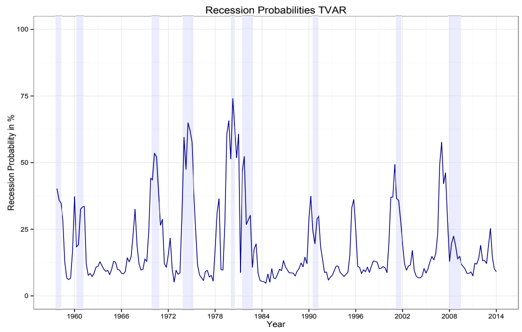


Model d_4, German data

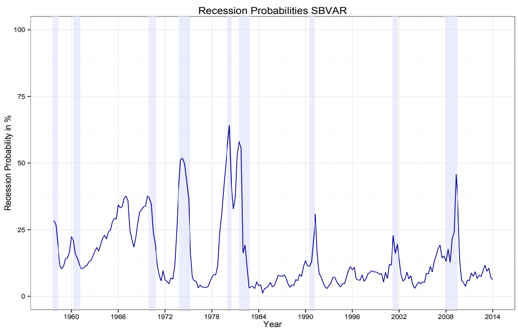
Figure A.3 continued



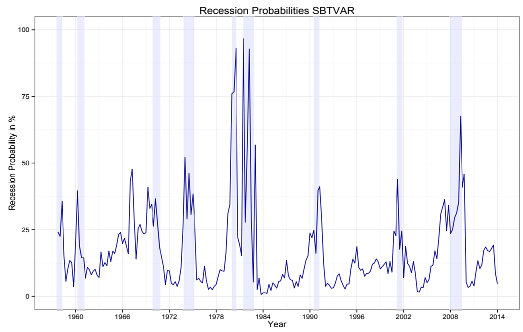
VAR, US data



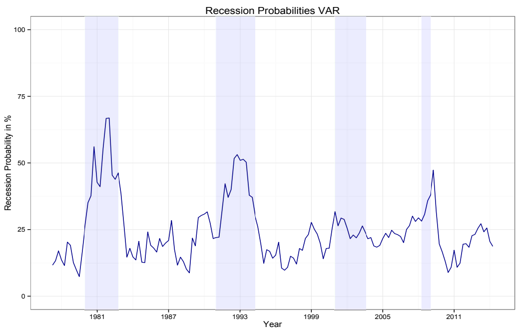
TSVAR, US data



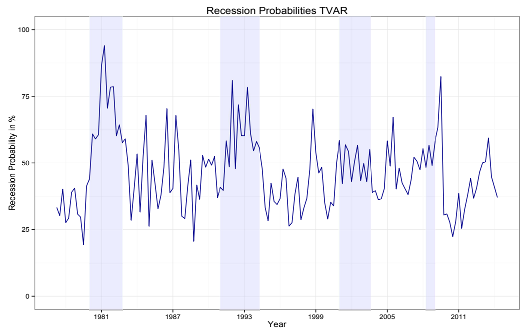
SBVAR, US data



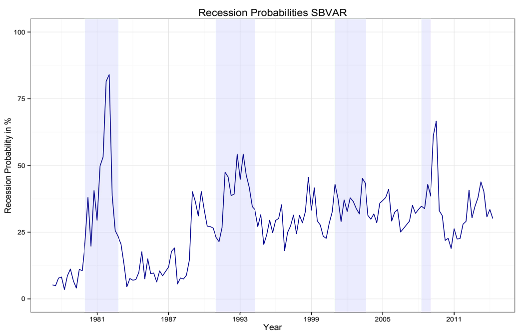
SBTVAR, US data



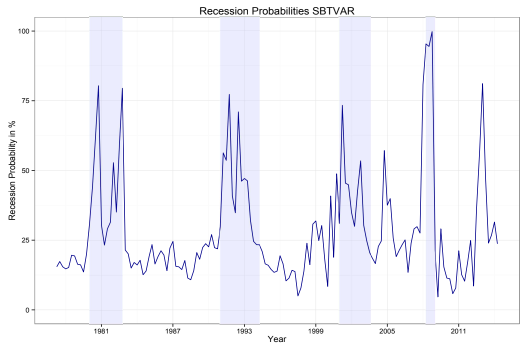
VAR, German data



TSVAR, German data



SBVAR, German data



SBTVAR, German data

Figure A.5: Recession probability plots for VAR models. The upper four plots show the results for the US, the lower four plots the results for Germany.

B. Supplementary Tables

Table B.1: Probit estimation results for Germany plus information criteria. Standard errors are given in parentheses, stars indicate statistical significance. ***: Significance at 1% level, **: Significance at 5% level, *: Significance at 10% level.

	k=1				k=2			
	a_1	b_1	c_1	d_1	a_2	b_2	c_2	d_2
Constant: $\hat{c}$	0.13	-0.91***	0.17	-0.94*	0.26	-0.90***	0.19	-1.22***
	(0.22)	(0.32)	(0.23)	(0.65)	(0.22)	(0.34)	(0.23)	(0.43)
Spread: $\hat{\alpha}$	-0.84***	-1.16***	-0.36**	-1.19***	-1.01***	-0.91***	-0.50**	-1.2***
	(0.16)	(0.36)	(0.19)	(0.33)	(0.19)	(0.26)	(0.25)	(0.35)
Recession: $\hat{\beta}$		3.45***		3.57***		2.68***		3.61***
		(0.76)		(1.14)		(0.48)		(0.91)
Probability: $\hat{\gamma}$			0.72***	-0.03			0.57***	-0.31*
			(0.16)	(0.35)			(0.19)	(0.22)
Log-likelihood	-53.4	-18.76	-42.06	-18.75	-45.92	-20.71	-42.35	-20.09
Pseudo R^2	0.38	0.78	0.5	0.77	0.47	0.76	0.50	0.75
AIC	110.80	43.53	90.13	45.51	95.84	47.41	90.70	48.19
BIC	116.88	52.64	99.24	57.66	101.91	56.52	99.81	60.34
	k=3				k=4			
	a_3	b_3	c_3	d_3	a_4	b_4	c_4	d_4
Constant: $\hat{c}$	0.28	-0.97***	0.24	-1.18***	0.23	-1.15***	0.29	-1.38***
	(0.23)	(0.35)	(0.24)	(0.31)	(0.23)	(0.32)	(0.25)	(0.30)
Spread: $\hat{\alpha}$	-1.05***	-0.66***	-0.82**	-0.83***	-0.97***	-0.46***	-1.25***	-0.67***
	(0.18)	(0.17)	(0.36)	(0.21)	(0.14)	(0.14)	(0.41)	(0.17)
Recession: $\hat{\beta}$		2.36***		2.91***		2.42***		3.00***
		(0.43)		(0.40)		(0.38)		(0.44)
Probability: $\hat{\gamma}$			0.22	-0.23**			-0.27	-0.29**
			(0.26)	(0.13)			(0.33)	(0.17)
Log-likelihood	-44.52	-24.44	-44.24	-23.88	-47.48	-27.04	-47.04	-26.04
Pseudo R^2	0.48	0.71	0.48	0.71	0.45	0.68	0.44	0.68
AIC	93.04	54.87	94.48	55.77	98.97	60.08	100.08	60.09
BIC	99.12	63.99	103.6	67.92	105.04	69.20	109.19	72.24

Table B.2: This table contains the optimal thresholds p^* for each of the 16 probit models (model a_k to d_k for spread lags $k = 1, \ldots, 4$) and each of the three criteria based on German data. Column CF contains the share of overall correct forecasts (recession and expansion), column HR the hit rate (share of correct recession forecasts) and column CR the share of correct rejections (predict expansion when expansion occurs). Column CritVal contains the criterion value at the optimal threshold in %.

				a_1					b_1		
k	**Criterion**	**Treshold**	**CF in %**	**HR in %**	**CR in %**	**CritVal**	**Treshold**	**CF in %**	**HR in %**	**CR in %**	**CritVal**
	ETS	0.5	86.36	59.46	94.87	42.11	0.60	96.75	89.19	99.15	83.24
	Bias	0.4	81.82	67.57	86.32	10.81	0.40	94.16	89.19	95.73	2.70
	hmf	0.3	81.17	75.68	82.91	58.58	0.20	94.16	94.59	94.02	88.61
1				c_1					d_1		
	ETS	0.6	89.61	59.46	99.15	50.73	0.60	96.75	89.19	99.15	83.24
	Bias	0.4	83.77	67.57	88.89	2.70	0.40	94.16	89.19	95.73	2.70
	hmf	0.3	85.71	81.08	87.18	68.26	0.20	94.16	94.59	94.02	88.61
				a_2					b_2		
	Criterion	**Treshold**	**CF in %**	**HR in %**	**CR in %**	**CritVal**	**Treshold**	**CF in %**	**HR in %**	**CR in %**	**CritVal**
	ETS	0.7	88.31	54.05	99.15	45.38	0.50	95.45	89.19	97.44	77.67
	Bias	0.5	86.36	67.57	92.31	8.11	0.40	94.81	89.19	96.58	0.00
	hmf	0.3	83.77	81.08	84.62	65.70	0.30	94.81	91.89	95.73	87.62
2				c_2					d_2		
	ETS	0.3	87.01	83.78	88.03	50.23	0.40	95.45	91.89	96.58	78.04
	Bias	0.5	85.06	67.57	90.60	2.70	0.40	95.45	91.89	96.58	2.70
	hmf	0.3	87.01	83.78	88.03	71.82	0.40	95.45	91.89	96.58	88.47
				a_3					b_3		
	Criterion	**Treshold**	**CF in %**	**HR in %**	**CR in %**	**CritVal**	**Treshold**	**CF in %**	**HR in %**	**CR in %**	**CritVal**
	ETS	0.4	86.36	78.38	88.89	47.38	0.40	94.81	89.19	96.58	75.09
	Bias	0.5	85.71	67.57	91.45	5.41	0.40	94.81	89.19	96.58	0.00
	hmf	0.3	85.06	83.78	85.47	69.25	0.40	94.81	89.19	96.58	85.77
3				c_3					d_3		
	ETS	0.4	87.01	78.38	89.74	48.91	0.40	94.81	89.19	96.58	75.09
	Bias	0.5	86.36	70.27	91.45	2.70	0.40	94.81	89.19	96.58	0.00
	hmf	0.2	83.12	86.49	82.05	68.54	0.40	94.81	89.19	96.58	85.77

Table B.2 continued

	Criterion	Treshold	CF in %	HR in %	CR in %	CritVal	Treshold	CF in %	HR in %	CR in %	CritVal
				a_4					b_4		
	ETS	0.4	85.06	75.68	88.03	43.78	0.40	94.81	89.19	96.58	75.09
	Bias	0.4	85.06	75.68	88.03	13.51	0.40	94.81	89.19	96.58	0.00
	hmf	0.2	81.82	86.49	80.34	66.83	0.40	94.81	89.19	96.58	85.77
4				c_4					d_4		
	ETS	0.3	85.06	81.08	86.32	45.17	0.40	94.81	89.19	96.58	75.09
	Bias	0.5	83.12	59.46	90.60	10.81	0.40	94.81	89.19	96.58	0.00
	hmf	0.3	85.06	81.08	86.32	67.41	0.40	94.81	89.19	96.58	85.77

Table B.3: This table contains the optimal thresholds for the VAR, SBVAR, TSVAR and SBTVAR and each of the three criteria based on German data. Column CF contains the share of overall correct forests (recession and expansion), column HR the hit rate (share of correct recession forecasts) and column CR the share of correct rejections (predict expansion when expansion occurs). Column CritVal contains the criterion value at the optimal threshold in %.

Criterion	Treshold	CF in %	HR in %	CR in %	CritVal
			VAR		
ETS	0.3	87.25	70.27	92.86	48.03
Bias	0.3	87.25	70.27	92.86	8.11
hmf	0.3	87.25	70.27	92.86	63.13
			SBVAR		
ETS	0.4	79.2	43.2	91.1	23.5
Bias	0.4	79.2	43.2	91.1	29.7
hmf	0.3	66.4	78.4	62.5	40.9
			TSVAR		
ETS	0.5	76.51	75.68	76.79	29.42
Bias	0.6	80.54	40.54	93.75	40.54
hmf	0.5	76.51	75.68	76.79	52.46
			SBTVAR		
ETS	0.3	77.85	67.57	81.25	29.15
Bias	0.4	78.52	45.95	89.29	21.62
hmf	0.3	77.85	67.57	81.25	48.82

C. Derivations

C.1. Expectation Hypothesis: Upward-sloping Curve Implies increasing Short Rates

The arithmetic mean of sample $\{x_1, x_2, \ldots, x_N\}$ is given as

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i.$$

Assume that $\bar{x} > x_k$, then

$$\frac{x_k}{N} + \frac{1}{N} \sum_{i \neq k} x_i > x_k$$

which can equivalently be stated as

$$x_k < \frac{1}{N-1} \sum_{i \neq k} x_i.$$