Dynamic Factor Trees and Forests – A Theory-led Machine Learning Framework for Non-Linear and State-Dependent Short-Term U.S. GDP Growth Predictions

ETH zürich

KOF Swiss Economic Institute

Dynamic Factor Trees and Forests

A Theory-led Machine Learning Framework for Non-Linear and State-Dependent Short-Term U.S. GDP Growth Predictions

Daniel Wochner

KOF Working Papers, No. 472, February 2020

KOF

KOF

ETH Zurich
KOF Swiss Economic Institute
LEE G 116
Leonhardstrasse 21
8092 Zurich, Switzerland

Phone +41 44 632 42 39
Fax +41 44 632 12 18
www.kof.ethz.ch
kof@kof.ethz.ch

Dynamic Factor Trees and Forests*

A Theory-led Machine Learning Framework for Non-Linear and State-Dependent Short-Term U.S. GDP Growth Predictions

Daniel Wochner

ETH Zurich, KOF Swiss Economic Institute

this version:
January 31, 2020

Abstract

Machine Learning models are often considered to be "black boxes" that provide only little room for the incorporation of theory (cf. e.g. Mukherjee, 2017; Veltri, 2017). This article proposes so-called Dynamic Factor Trees (DFT) and Dynamic Factor Forests (DFF) for macroeconomic forecasting, which synthesize the recent machine learning, dynamic factor model and business cycle literature within a unified statistical machine learning framework for model-based recursive partitioning proposed in Zeileis, Hothron and Hornik (2008). DFTs and DFFs are non-linear and state-dependent forecasting models, which reduce to the standard Dynamic Factor Model (DFM) as a special case and allow us to embed theory-led factor models in powerful tree-based machine learning ensembles conditional on the state of the business cycle. The out-of-sample forecasting experiment for short-term U.S. GDP growth predictions combines three distinct FRED-datasets, yielding a balanced panel with over 375 indicators from 1967 to 2018 (FRED, 2019; McCracken & Ng, 2016, 2019a, 2019b). Our results provide strong empirical evidence in favor of the proposed DFTs and DFFs and show that they significantly improve the predictive performance of DFMs by almost 20% in terms of MSFE. Interestingly, the improvements materialize in both expansionary and recessionary periods, suggesting that DFTs and DFFs tend to perform not only sporadically but systematically better than DFMs. Our findings are fairly robust to a number of sensitivity tests and hold exciting avenues for future research.

Keywords: Forecasting, Machine Learning, Regression Trees and Forests, Dynamic Factor Model, Business Cycles, GDP Growth, United States

JEL classification: C45, C51, C53, E32, O47

*The views expressed in this paper are solely the ones of the author and do not necessarily reflect those of the KOF Swiss Economic Institute at ETH Zurich.

1. Introduction

Machine learning systems in general and deep learning systems in particular have achieved notable breakthroughs in predictive accuracy in recent years (cf. e.g. Esteva et al., 2017; LeCun, Bengio, & Hinton, 2015; McAfee & Brynjolfsson, 2017). Unfortunately, however, the predictive ability and interpretative feasibility of these systems stand usually in conflict with one another (Breiman, 2001b) as their inner working mechanisms are typically not amenable to human comprehension — an issue known as the "black box" problem (Appenzeller, 2017; Mittelstadt, Allo, Taddeo, Wachter, & Floridi, 2016; Mittelstadt & Floridi, 2016; Mukherjee, 2017; Veltri, 2017).

The *nature* of most machine learning algorithms is often fundamentally different from traditional data modelling techniques used in statistics or econometrics (cf. Breiman, 2001b; McAfee & Brynjolfsson, 2017; Mullainathan & Spiess, 2017; Veltri, 2017; Wochner, 2018). As discussed in Breiman (2001b), they shift the data modelling paradigm from a stochastic to an algorithmic one: Instead of assuming a stochastic data model (e.g. logistic regression) for the function that maps inputs to outputs, machine learning algorithms (e.g. neural networks) consider the functional form to be largely unknowable and instead seek to discover patterns that predict the response well by primarily building upon algorithmic properties (e.g. convergence speed) rather than stochastic properties (e.g. unbiasedness) (cf. Breiman, 2001b, for details; also see Athey & Imbens, 2019; Varian, 2014; Veltri, 2017; Wochner, 2018, for related discussions). A promising new field of research in *statistical* machine learning seeks to fruitfully merge these two paradigms so as to endow machine learning algorithms with more theory-led structure while maintaining their high predictive accuracy (cf. e.g. Athey & Imbens, 2017; Athey, Tibshirani, & Wager, 2019; Seibold, Zeileis, & Hothorn, 2016; Varian, 2016; Zeileis et al., 2008). To take an example outside the field of economics — which inspired this research — consider the benefits of treatment effects in medical applications, which are likely to differ between individuals and may depend, for instance, on their demographics (e.g. male/female) or health status (e.g. healthy/sick) (cf. Athey & Imbens, 2017, p. 10; Seibold et al., 2016). Seibold et al. (2016) promote advances in personalized medical healthcare by employing a segmented machine learning model that is able to autonomously identify distinct patient subgroups with heterogenous medical treatment effects based on a set of personal health conditions.

This article adopts and adapts these ideas for macroeconomic forecasting by employing and extending the unified statistical machine learning framework for segmented (or state-dependent) model estimation developed by Zeileis et al. (2008). Their "model-based recursive partitioning (MOB) algorithm" combines the well-known class of parametric models from statistics with tree-based partitioning algorithms from machine learning and derives a state-dependent model in four key steps: (1) a parametric model is estimated, (2) parameter instability tests over pre-determined partitioning variables are performed, (3) if instabilities are absent, stop; otherwise the initial parametric model is split in a tree-like fashion into two sub-states with regards to the partitioning variable that shows the most significant instabilities; (4) the procedure is recursively repeated in each sub-state until a stopping criterion is reached (see

Zeileis & Hothorn, 2015; Zeileis et al., 2008, for details). In the present case, we seek to predict US GDP growth and our state-dependent forecasting model consists of two key ingredients: First, to take advantage of big macroeconomic datasets, we extract a set of dynamic factors and employ factor-augmented autoregressive processes as parametric models in the MOB-framework (Stock & Watson, 2006, 2016, 2017; Siliverstovs & Wochner, 2019). Second, instead of choosing the health conditions of patients as in Seibold et al. (2016), we align more closely with the business cycle literature and choose the health conditions of the economy (proxied via recession probability indices) as partitioning variable (e.g. Chauvet & Potter, 2013; Doz & Fuleky, 2019). In short: Provided that there are sufficient instabilities over the estimation period, this design results in a state-dependent dynamic factor model with, for instance, two states — one dynamic factor model for time periods in which the economy is in a bad state ("sick") and another one for time periods in which it is in a good state ("healthy"). If no instabilities can be detected, there is only a single state and the standard dynamic factor model is estimated for all observations. The proposed modelling design therefore merges dynamic factor models with regression trees and we shall call the resulting model "dynamic factor trees" (DFT). Moreover, Garge et al. (2013) extended Zeileis et al. (2008)'s model-based recursive partitioning to tree ensembles and we extend dynamic factor trees in a similar spirit to, what we call, "dynamic factor forests" (DFF) by combining hundreds of dynamic factor trees with each tree being grown from a (block-)bootstrapped sample (also see Breiman, 1996, 2001a; Hastie, Tibshirani, & Friedman, 2009; Canty, 2002; see Section 2.4 for details).

Why may we expect this to be a sensible modelling strategy? Based on Stock and Watson (2017)'s views in their summary article on time series econometrics, we see both conceptual and empirical advantages. On the conceptual side, Stock and Watson (2017) expect that "the next steps towards exploiting additional information in large datasets will need to use new statistical methods guided by economic theory." (ibid., p. 83). Model-based recursive partitioning may provide a suitable framework in this regard (cf. Veltri, 2017; Zeileis et al., 2008, for a general discussion): Standard regression tree algorithms split the feature space into non-overlapping subgroups and fit a constant model, such as the sample mean, to each of these (cf. e.g. Hastie et al., 2009; James, Witten, Hastie, & Tibshirani, 2013). The proposed dynamic factor trees and forests, instead, allow us to fit dynamic factor models to each subgroup (Zeileis et al., 2008), which not only enjoy great empirical performance but are also well-grounded in macroeconomic equilibrium theories (cf. e.g. Forni, Giannone, Lippi, & Reichlin, 2009; Stock & Watson, 2006, 2011, 2016, 2017; also e.g. Diebold & Rudebusch, 1996, p. 69ff.).[1] On the empirical side, Stock and Watson (2017) refer to the forecasting models' repeated inability to capture the severity of economic downturns as the "Mother of All Forecast Errors" (ibid., p. 82) and question in the context of dynamic factor models "whether there is exploitable nonlinear structure [...] that could perhaps be revealed by modern machine learning methods." (ibid., p. 83). The proposed state-dependent machine learning approach may contribute along these lines in that they allow for non-linearities by fitting distinct models to autonomously detected subsets of the data (cf. Seibold et al., 2016; Zeileis et al., 2008) and thereby equip the models with the flexibility to be more adaptive in states in which the economic situation

deteriorates and less adaptive when it improves (see K. Kim & Swanson, 2016, for a similar argument; also see discussions in Siliverstovs & Wochner, 2019).

This article contributes to the current state of research by synthesizing three nascent streams of the macroeconomic forecasting literature within a novel statistical machine learning framework. First, there is a growing interest among macroeconomists to employ machine learning algorithms for more accurate assessments and predictions of macroeconomic dynamics (e.g. Coulombe et al., 2019; Garcia et al., 2017; H. H. Kim & Swanson, 2014, 2018; Kock & Teräsvirta, 2016; Stock & Watson, 2017; Wochner, 2018, among many others). Among these, there is an increasing number of studies reporting that standard tree-based ensembles, such as random forests, can have great predictive power for a range of macroeconomic and financial indicators (e.g. J. C. Chen et al., 2019; Coulombe et al., 2019; Khaidem et al., 2016; Medeiros et al., 2019; Wochner, 2018; also see Garcia et al., 2017 who find satisfactory performance). These findings are relevant because the MOB-framework embeds parametric models within tree-based structures (Zeileis et al., 2008) and is thus closely related but clearly different from conventional tree-based models, which belong to the class of non-parametric methods (cf. X. Chen & Ishwaran, 2012).

Second, our work relates to the rich literature on dynamic factor models in general and the factor-based literature on structural instabilities in particular (cf. e.g. Stock & Watson, 2006, 2016, 2017; Doz & Fuleky, 2019; Rossi, 2013, for reviews). Structural instabilities can arise from policy changes, preference shifts or technological advances (B. Chen & Hong, 2012; Yousuf, 2019) and this line of research typically distinguishes structural instabilities in factor loadings from non-linear dynamics in factor processes (Bai & Wang, 2016; Doz & Fuleky, 2019). As to the former, Banerjee et al. (2008), Stock and Watson (2009) and Breitung and Eickmeier (2011) provide experimental and empirical evidence for instabilities in factor loadings. Recent developments show that the factors can be consistently estimated via principle components provided that the instabilities are limited (see Bates et al., 2013, for precise definitions; Bai & Han, 2016; Stock & Watson, 2016). For example, in case of a single small break, Pesaran and Timmermann (2007) show that the use of full sample data (pre- and post-break data) can yield lower MSFE than only using post-break data as the inclusion of pre-break may greatly reduce the variance at a slight increase in bias. In a similar vein, Stock and Watson (2009) find full sample factor estimates determined via principal components (PC) to provide stable estimates of the factors. Their best forecasting model was therefore not obtained by allowing for instabilities in factor loadings but for instabilities in the (time-varying) coefficients of the factors in their forecasting equations and they find this to be in accord with changing dynamics of the factor processes. Our work is related to Stock and Watson (2009) in that we also employ full sample PC-factor estimates and allow for time-varying coefficients of factors in the forecasting equations.

Third, a related strand allows the dynamics of factor processes to depend on the state of the business cycle (Doz & Fuleky, 2019). For example, Diebold and Rudebusch (1996) ranked among the first who combined linear dynamic factor models with Hamilton's (1989) Markov regime-switching approach so as to treat expansions and recessions as distinct stochastic entities (cf. Diebold, 2003; see Doz & Fuleky, 2019, for

turning point detections). A growing literature documents systematic differences in predictive power across states of the business cycle (cf. e.g. Chauvet & Potter, 2013; Fossati, 2018; Siliverstovs, 2019; Siliverstovs & Wochner, 2019). Building upon this fact, Kim and Swanson (2016), for example, propose a state-dependent model that switches between autoregressive benchmarks during expansions and (mixed-frequency) factor models during recessions to predict output growth and inflation. Likewise, Del Negro et al. (2016) find their DSGE model with financial frictions to perform well during recessions but less so during expansions. Similarly, Chauvet and Potter (2013) show in their comprehensive assessment of leading forecasting models that predictability differs across business cycle states and find that their Markov-switching DFMs rank among the best models in terms of out-of-sample forecasting performance, especially during recessions. Our forecasting experiment is closely related to Chauvet and Potter (2013)'s regime-switching dynamic factor model, but clearly distinct in that our regime-switching is not governed probabilistically via Markov processes but rather algorithmically via recent advances in statistical machine learning (Zeileis et al., 2008). Moreover, a desirable property of our proposed dynamic factor trees over Markov switching models is the fact that they have a clear interpretation (James et al., 2013; Zeileis et al., 2008) whereas Markov-switching models are typically more difficult to interpret as the hidden state-variables are not observed (Kock & Teräsvirta, 2011; Kuan, 2002).

This article aims to assess the empirical performance of the proposed dynamic factor trees and forests against standard dynamic factor models as well as conventional benchmarks, such as autoregressive and distributed lag processes. The main analyses are based on three distinct FRED datasets (FRED-MD, FRED-QD, FRED; see Section 3), which can be combined to a balanced panel with real GDP growth as dependent variable, 375 explanatory variables plus a recession probability indicator from October 1967 to September 2018 (e.g. McCracken & Ng, 2016, 2019a, 2019b). The analyses make two key modelling assumptions: First, to cope with mixed frequencies, we follow Kim and Swanson (2014)'s approach in the Journal of Econometrics and interpolate quarterly GDP to a monthly frequency. We will show the robustness of our results by means of two alternative interpolation methods (see Section 4.2) and run the experiment also in quarterly frequency (cf. Foroni & Marcellino, 2013). Second, as a consequence, we also interpolate the recession probability index to a monthly frequency and assume the series to be released together with all other monthly indicators in FRED-MD. Under this interpolation and publication scheme, we consider to have an accurate monthly proxy for the recession probability index available. This assumption also allows us to make our results more comparable with the existing Markov-switching literature, which typically derives the recession probabilities from a contemporaneous set of available predictors (cf. Kuan, 2002). We assess the sensitivity of this assumption by predicting the missing values on the current edge (cf. e.g. Bulligan, Marcellino, & Venditti, 2015) and show robustness for two alternative recession probability indices (see Section 4.2).

We find strong empirical evidence in favor of the proposed dynamic factor trees and forests. Our out-of-sample forecasting experiment shows that they may significantly improve upon standard dynamic factor models and yield MSFE-improvements of al-

most 20%. Interestingly, their state-dependent model design allows to systematically improve upon DFMs in *both* expansionary and recessionary periods, which suggests that DFTs and DFFs perform not only sporadically but systematically better than DFMs. In light of the generally strong performance of DFMs (Chauvet & Potter, 2013; Stock & Watson, 2017), we perceive these additional improvements as notable gains in predictive accuracy. Moreover, the results qualify as fairly robust against a large number of robustness tests.

The remainder of this article is organized as follows: Section 2 sets out the modelling environment and formalizes dynamic factor trees and forests. Section 3 describes the data. Section 4 evaluates the models' forecasting performance and their robustness. Section 5 concludes with directions for future research.

2. Modelling Framework

2.1. Setup, Notation and Environment

For the definition of our modelling framework, we follow the notational and methodical conventions in the relevant literature (e.g. Elliott & Timmermann, 2016; Stock & Watson, 2006, 2016; Siliverstovs & Wochner, 2019; Wochner, 2018). Let $t \in \{1, \dots, T\}$ denote the time index in monthly frequency and $h \in \{1, 3\}$ corresponds to the monthly forecasting horizon. The timeline is divided into an estimation window (Dec. 1967 – Dec. 1997) and forecasting window (Jan. 1998 – Sep. 2018). Let Q denote the last time period of the first estimation window (Dec. 1997), so that the recursive estimation window is given as $\mathscr{S}_e = \{1 + h, ..., \tau\}$ with $\tau \in \{Q, \dots, T - h\}$, and the forecasting window as $\mathscr{S}_f = \{Q + 1, \dots, T - h + 1\}$ (e.g. Siliverstovs & Wochner, 2019).

Denoting vectors and matrices in bold letters, the dataset consists of four different types of stationary variables: The dependent variable, $Y_t^{(h)} \in \mathbb{R}$, the set of K mean-zero and unit-variance standardized explanatory variables (including their 1st and 2nd-order lags; cf. H. H. Kim & Swanson, 2014), $\boldsymbol{X}_t^{(1)} \in \mathbb{R}^K$, from which we extract the vector of R factors $\boldsymbol{F}_t^{(1)} \in \mathbb{R}^R$ (Stock & Watson, 2006, 2016, for details) as well as the partitioning variables, $\boldsymbol{Z}_t^{(1)} \in \mathbb{R}^P$, which are used in the MOB-framework to segment the sample space into S distinct subsets (cf. Zeileis et al., 2008). The bracketed super-indices indicate the h-period ahead stationarity transformation of dependent and independent variables (cf. e.g. Stock & Watson, 2012). Further, let $\boldsymbol{W}_t^{(1)} = \left(1, Y_t^{(1)}, \dots, Y_{t-L+1}^{(1)}\right)'$ denote additional controls that include an intercept constant and autoregressive dependent variables up to lag L, such that the relevant dataset for model estimation is given as $\boldsymbol{D}_t = \left(Y_t^{(h)}, \boldsymbol{W}_t^{(1)\prime}, \boldsymbol{F}_t^{(1)\prime}, \boldsymbol{Z}_t^{(1)\prime}\right)$.

2.2. Regression Trees

The macroeconomic forecasting literature managed to successfully employ tree-based methods such as random forests (e.g. Medeiros et al., 2019). To appreciate the difference between this literature and Zeileis et al. (2008)'s more recent advances put forth in here, we shall briefly review conventional regression trees (also see Strobl, Wickelmaier, & Zeileis, 2011).

A regression tree seeks to minimize some objective function (e.g. the sum of squared residuals) by repeatedly sub-dividing the predictor space spanned by the variables in $\boldsymbol{Z}_t^{(1)}$ into mutually exclusive regions and fitting a constant model (e.g. sample mean) to each of these subregions until some stopping criterion is reached (e.g. minimum quantity of observations per region) (James et al., 2013). Unfortunately, trying to find the first-best partition of the space requires to consider every possible partition, which is generally infeasible in practice (Hastie et al., 2009, p. 307; James et al., 2013). However, binary recursive partitioning algorithms are a greedy and iterative search procedure that make the estimation of regression trees computationally feasible by determining the "best" partitioning variable k and partitioning point ζ for space division as the pair (k, ζ) that yields the largest reduction of the objective function in the current iteration (Hastie et al., 2009; James et al., 2013). Using a least squares objective, the algorithm seeks to divide region $\mathcal{R}$ into two rectangular sub-regions $\mathcal{R}_1 = \{\boldsymbol{Z}_t^{(1)} | Z_{k,t}^{(1)} \leq \zeta\}$ and $\mathcal{R}_2 = \{\boldsymbol{Z}_t^{(1)} | Z_{k,t}^{(1)} > \zeta\}$ and chooses (k, ζ) such that they solve the following minimization problem,

$$(k^\star, \zeta^\star) = \arg\min_{k,\zeta} \left[\min_{\varphi_1} \sum_{t:\boldsymbol{Z}_t^{(1)} \in \mathcal{R}_1} \left(Y_{t+h}^{(h)} - \varphi_1\right)^2 + \min_{\varphi_2} \sum_{t:\boldsymbol{Z}_t^{(1)} \in \mathcal{R}_2} \left(Y_{t+h}^{(h)} - \varphi_2\right)^2 \right]$$

and repeats this procedure until the tree is grown to full depth (cf. Hastie et al., 2009, p. 307ff.; James et al., 2013; also Wochner, 2018). Two points are worth noting: First, the procedure reveals the algorithmic nature of regression trees in that they do not presume any theory-led stochastic model but instead rely on algorithmic search procedures to discover complex patterns that closely approximate the response in a purely data-driven manner (Breiman, 2001b; Veltri, 2017; Wochner, 2018). Second, this recursive procedure lends to a hierarchical representation ("dendrograms") in which the first binary splits are at the top and subsequent splits are attached to the resulting regions of the previous splits and the bottom layer constitutes the leaf nodes of the tree (James et al., 2013).

Formally, we may describe the forecasting equation of regression trees as a piecewise constant model as,

$$Y_{t+h}^{(h)} = \sum_{s=1}^{S} \left(\varphi_s + \epsilon_{s,t+h}^{(h)}\right) \mathbb{1}(\boldsymbol{Z}_t^{(1)} \in \mathcal{R}_s) \tag{1}$$

with $\mathcal{R}_s$ denoting region $s \in \{1, \ldots, S\}$, $\epsilon_{t+h}^{(h)}$ the error term and $\mathbb{1}(\cdot)$ is the indicator function, which evaluates to unity whenever the condition is satisfied and to zero else (Hastie et al., 2009; Kock & Teräsvirta, 2011; Medeiros et al., 2019; Wochner, 2018). The best estimator for φ_s in a least squares sense is the sample average, $\hat{\varphi}_s = \frac{1}{T_s} \sum_{t \in \mathcal{R}_s} Y_{t+h}^{(h)}$, with $T_s = |\{t : \boldsymbol{Z}_t^{(1)} \in \mathcal{R}_s\}|$ many observations in region $\mathcal{R}_s$ and $|\cdot|$ designating the size of the set (Hastie et al., 2009, p. 307f.).

2.3. Dynamic Factor Trees (DFT)^2

2.3.1. Formal Model

Simply put, Zeileis et al. (2008)'s model-based recursive partitioning algorithm is a general statistical framework that allows to fit parametric models to distinct subsets of the data. A key advantage of their framework is that these subsets (or states) are autonomously detected through the recursive application of parameter instability tests over the space spanned by the partitioning variables in a tree-based fashion (ibid.). Their algorithm is therefore capable to autonomously detect nonlinearities arising from interactions of variables and fits, in case of their presence, local models to subsets of the data that yield better fit than a single global model for all observations (cf. Zeileis & Hothorn, 2015; Zeileis et al., 2008).

More precisely, let $\mathcal{M}(\boldsymbol{\psi}_s, \boldsymbol{\phi}_s; \boldsymbol{D})$ denote a parametric model with parameter coefficients $\boldsymbol{\psi}_s$ and $\boldsymbol{\phi}_s$ in state $s \in \{1, \dots, S\}$ for dataset $\boldsymbol{D}$, the MOB algorithm seeks to minimize the objective function, $\sum_{t \in \mathscr{S}_e} \Omega(\boldsymbol{\psi}_s, \boldsymbol{\phi}_s; \boldsymbol{D}_t)$, for a given loss function $\Omega(\cdot)$ (Zeileis et al., 2008). As detailed in Zeileis et al. (2008) and Zeileis and Hothorn (2015) (also see Kopf et al., 2013), the algorithm starts in the first state with all observations and derives a state-dependent model in four key steps:

(1) estimate a parametric model of type $\mathcal{M}(\cdot)$ in the current state by minimizing the objective function (e.g. via OLS if $\Omega(\cdot)$ is the squared residual loss);

(2) run score-based fluctuation tests for parameter instability over all partitioning variables in $\boldsymbol{Z}_t$ and if instabilities are present (at significance level α), determine the variable k with the most significant instability; else, stop;

(3) determine the optimal split point ζ for $Z_{k,t}^{(1)}$ such that the objective function, given as the sum of local objective functions in the two resulting sub-states, $\sum_{t \in \{t|Z_{k,t} \leq \zeta\}} \Omega(\boldsymbol{\psi}_{s1}, \boldsymbol{\phi}_{s1}; \boldsymbol{D}_t) + \sum_{t \in \{t|Z_{k,t} > \zeta\}} \Omega(\boldsymbol{\psi}_{s2}, \boldsymbol{\phi}_{s2}; \boldsymbol{D}_t)$, is minimized;

(4) split the current state according to $(k^\star, \zeta^\star)$ into two sub-states, namely $\mathscr{S}_e^{(s_1)} = \{t|Z_{k^\star,t} \leq \zeta^\star\}$ and $\mathscr{S}_e^{(s_2)} = \{t|Z_{k^\star,t} > \zeta^\star\}$, and repeat the procedure in each substate until stability is achieved or the minimum node size, η, is reached.

A few points deserve further attention: First, a partitioning variable can either be categorical or numerical (Zeileis et al., 2008). The former provides the possibility to exogenously determine different states (via manual discretization of a numeric variable) whereas the latter allows to endogenously determine different states (based on automated detection mechanisms) (cf. Strobl et al., 2011; Zeileis & Hornik, 2007; Zeileis et al., 2008). Second, the empirical parameter instability tests are based on generalized M-fluctuation tests that assess whether or not the scores of estimated objective functions $\hat{\boldsymbol{\omega}}_{s,t} = \partial\big(\Omega(\hat{\boldsymbol{\psi}}_s, \hat{\boldsymbol{\phi}}_s; \boldsymbol{D}_t)\big) / \partial\big((\hat{\boldsymbol{\psi}}_s', \hat{\boldsymbol{\phi}}_s')\big)$ deviate systematically from zero over some $Z_{k,t}$ and use robust standard errors to account for possibly heteroscedastic and autocorrelated errors (Zeileis et al., 2008, p. 496ff.; Zeileis & Hornik, 2007). These fluctuation or instability tests comprise several well-known tests employed in the macroeconomic literature on structural breaks, such as Nyblom's (1989) maximum likelihood score tests for constant parameters or Andrews's (1993) Lagrange multiplier

test for single discrete breaks (see Zeileis et al., 2008, p. 497f.; Rossi, 2013, p. 1236f.; Zeileis, 2005). Breitung and Eickmeier (2011) show, for example, that Andrews's (1993) test can be used to detect structural breaks in factor models. Third, as the recursive testing procedure entails multiple testing, the p-values are Bonferroni-adjusted (Zeileis & Hothorn, 2015, p. 4).

2.3.2. Empirical Implementation

Our empirical implementation proceeds in two steps: First, we extract the primary R factors, $\boldsymbol{F}_t^{(1)}$, from the set of explanatory variables contained in $\boldsymbol{X}_t^{(1)}$ via principal components estimation (e.g. Stock & Watson, 2006, 2016; Siliverstovs & Wochner, 2019). In addition to these plain factors, we consider targeted factors in the spirit of Bai and Ng (2008), where the factors are not extracted from the full set of original explanatory variables, $\boldsymbol{X}_t^{(1)}$, but the subset of variables $\mathring{\boldsymbol{X}}_t^{(1)} \subseteq \boldsymbol{X}_t^{(1)}$ that have a statistically significant relation with the response at the 5% level (also see Stock & Watson, 2012; Wochner, 2018). Second, we employ factor-augmented autoregressive processes as parametric model (similar to Siliverstovs & Wochner, 2019) and use, in the spirit of the business cycle and structural breaks literature, recession probability indices as partitioning variables within the MOB-framework (cf. Chauvet & Potter, 2013; Doz & Fuleky, 2019).

This procedure yields a non-linear and state-dependent dynamic factor model, which we call "Dynamic Factor Trees" (DFT). Building upon the discussions above, we may formalize this model as,

$$Y_{t+h}^{(h)} = \sum_{s=1}^{S} \left(\boldsymbol{\psi}_s' \boldsymbol{W}_t^{(1)} + \boldsymbol{\phi}_s' \boldsymbol{F}_t^{(1)} + \epsilon_{s,t+h}^{(h)} \right) \mathbb{1}\left(\boldsymbol{Z}_t^{(1)} \in \mathcal{R}_s(\alpha, \eta) \right) \tag{2}$$

where $\boldsymbol{\psi}_s$ and $\boldsymbol{\phi}_s$ denote the regression coefficients and $\mathcal{R}_s(\alpha, \eta)$ refers to the set of observations in state s (cf. Kock and Teräsvirta's, 2011, for closely related non-linear regime-switching models; and e.g. Stock & Watson, 2016, for conventional DFMs).[3] This notation also highlights that the determination of states depends on two key hyper-parameters: the statistical level of the instability tests α as well as the minimum state size η (see Section 2.3.1). While conventional significance levels are chosen for the former, we determine the latter via blocked cross-validation that divides the sample into ten approximatively equally sized blocks and leaves out a dozen observations in the training set from either side of every held-out test set (Racine, 2000; cf. discussion in Wochner, 2018). This allows us to attain an optimal tree size via pre-pruning to limit the risk of overfitting (Zeileis & Hothorn, 2015, p. 4). Finally, equation (2) formally reveals the hybrid nature of dynamic factor trees as a blend between regression trees (in equation (1)) and dynamic factor models (in equation (4)).

2.3.3. Stylized Example

To appreciate the difference between the proposed dynamic factor trees, regression trees, dynamic factor models as well as other well-known regression models, consider the following stylized example: Figure 1 displays scatterplots and fitted models of

equation (1) and (2) for 100 simulated realizations of the dependent variable $Y_{t+h}^{(h)}$, the first estimated factor $\hat{F}_t^{(1)}$ and partitioning variable $Z_t^{(1)}$, which is assumed to be binary for simplicity (see James et al., 2013, p. 303ff. and Zeileis et al., 2008, p. 501ff. for related illustrations).

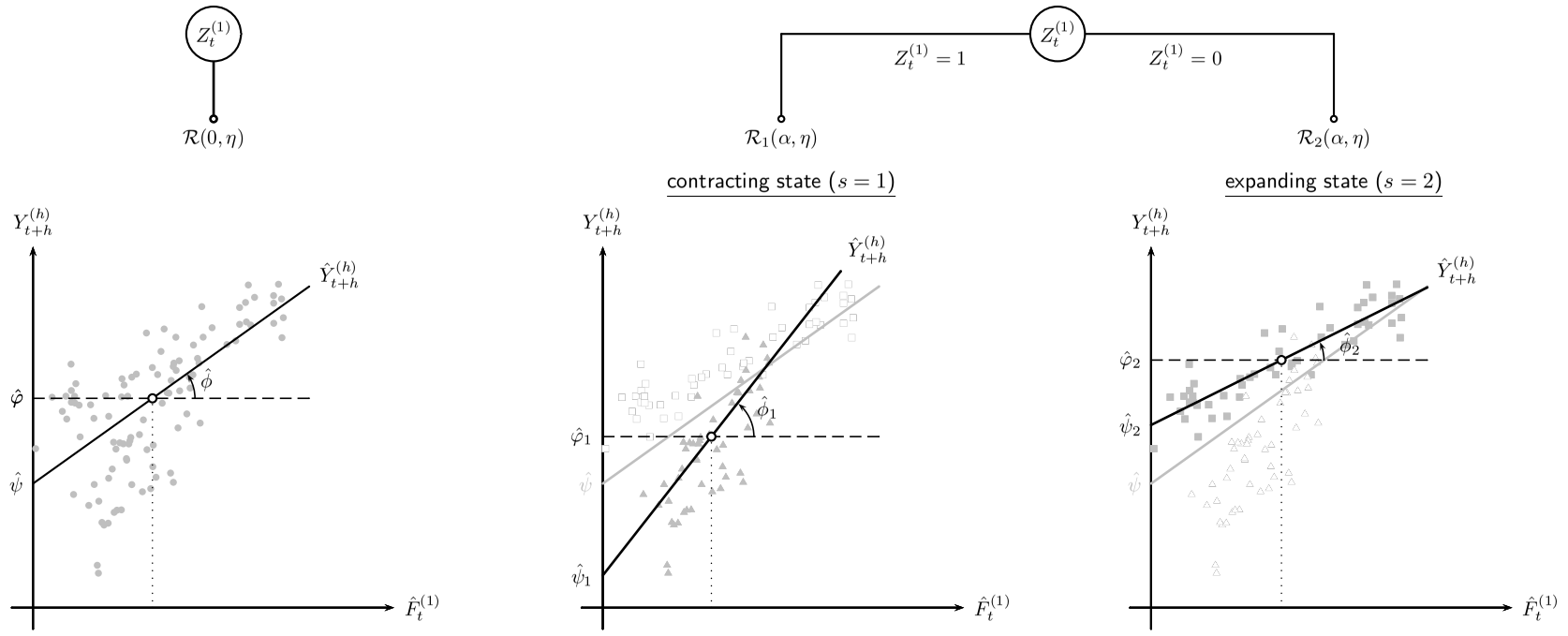
Equation (2) nests several models as special cases. For example, when partitioning is not allowed ($\alpha = 0$), a single-state model arises ($S = 1$) and the reduced form corresponds exactly to the standard dynamic factor model. This situation is depicted in Figure 1.a, where a single factor model (without autoregressive processes; AR0-DFM1) is fitted to all observations. By contrast, if partitioning is enforced ($\alpha = 1$), a multi-state model arises ($S > 1$) in which we have state-dependent dynamic factor models that are fitted to subsets of the data, i.e. "expanding" and "contracting" states. This situation is illustrated in Figure 1.b and depicts a situation in which the model reacts more sensitively to changes in the underlying factors during contractive periods ($Z_t^{(1)} = 1$) and less sensitive during expansionary periods ($Z_t^{(1)} = 0$). For the remaining values of $\alpha \in (0, 1)$, either one of the two cases may emerge depending on the outcome of the statistical parameter instability tests (see Section 2.3.1).

Furthermore, Figure 1 also depicts the fitted models of regression trees as horizontal dashed lines in all possible final states considered. Model-based recursive partitioning allows to overcome at least two limitations of conventional regression trees: First, while regression trees fit constant models to subsets of the data (James et al., 2013) that leave only little room for theory-led modelling, model-based recursive partitioning enables the incorporation of theory-driven parametric models within tree structures (Zeileis et al., 2008; Veltri, 2017). Second, while regression trees seek to partition the space based on plain differences in the dependent variable, model-based recursive partitioning builds upon a rigorous statistical foundation and partitions the space based on significant differences in model parameters (Strobl et al., 2011; Zeileis et al., 2008).

Finally, model-based recursive partitioning is closely related yet clearly different from interaction models. The estimated model displayed in Figure 1.b can also be obtained by fully interacting the partitioning variable with the standard dynamic factor model displayed in Figure 1.a. However, the MOB-framework is more general in that fully interacted models are enforced to apply splits whereas MOB-based models only do so if there are statistically significant parameter differences (Strobl et al., 2011; Zeileis et al., 2008). Moreover, we may also imagine more sophisticated partitions and sub-group detections than those in Figure 1. For instance, in case of a continuous partitioning variable, either one of the substates in Figure 1 could be further partitioned into two new substates, yielding a more complex partition of the space. In real-world cases it is harder to know à-priori how the relevant interactions look like in order to "hard-code" them into an interacted model and one may thus prefer to resort to MOB's automated detection mechanisms (see Section 2.3.1).

2.4. Dynamic Factor Forests (DFF)

Our discussions above, have so far only considered a single dynamic factor tree. While single tree models enjoy a high degree of interpretability, they may overfit the data



1.a: Single State Model

1.b: Multi-State Model

Notes: Figure 1 shows the tree-structure and distinct state-dependent model estimations for 100 simulated realizations of the dependent variable $Y_{t+h}^{(h)}$, the first estimated factor $\hat{F}_t^{(1)}$ and partitioning variable $Z_t^{(1)}$, which is assumed to be binary for simplicity (see James et al., 2013, p. 303ff. and Zeileis et al., 2008, p. 501ff. for related illustrations). Figure 1.a shows the case in which partitioning is not allowed ($\alpha = 0$) and gives rise to a standard dynamic factor model (here shown without autoregressive processes but a single factor, AR0-DFM1). Figure 1.b, instead, shows the effects when partitioning into contractive states (triangles) and expansive states (squares) is allowed ($\alpha \in (0, 1]$) or enforced ($\alpha = 1$), which results in a multi-state model that estimates different relations between explanatory and dependent variables in each state (the single-state regression is displayed in gray; the white dot on the regression line indicates the average point $(\bar{Y}_{t+h}^{(h)}, \bar{F}_t^{(1)})$). Moreover, the dashed lines indicate the values, φ, that a conventional regression tree would estimate in each of the three modelling scenarios (cf. James et al., 2013). See Section 2 for details.

Figure 1: Dynamic Factor Models, Regression Trees and Dynamic Factor Trees

(particularly when they are grown at full depth) and can consequentially be very sensitive to minor changes in the data (Garge et al., 2013; James et al., 2013; Zeileis et al., 2008). Bagged or randomized forests solve the overfitting problem at the cost of interpretability through averaging among a multitude of trees, each of which is grown from bootstrapped samples (Breiman, 1996, 2001a; Hastie et al., 2009; James et al., 2013).

Garge et al. (2013) extended Zeileis et al. (2008)'s model-based recursive partitioning to tree ensembles. We extend dynamic factor trees in a similar vein and propose Dynamic Factor Forests (DFF) in two steps: First, we will estimate B dynamic factor trees, each of which was grown from a (stationary) block-bootstrapped sample, $\tilde{\boldsymbol{D}}_{b,t} = \left(\tilde{Y}_{b,t}^{(h)}, \tilde{\boldsymbol{W}}_{b,t}^{(1)\prime}, \tilde{\boldsymbol{F}}_{b,t}^{(1)\prime}, \tilde{\boldsymbol{Z}}_{b,t}^{(1)\prime}\right)$ for $b \in \{1, \dots, B\}$ with $B = 500$ (Politis & Romano, 1994; Canty, 2002; Wochner, 2018). Second, in the spirit of conventional bagging (cf. Hastie et al., 2009, p. 282ff.), we subsequently average among all fitted dynamic factor trees and may characterize dynamic factor forests as,

$$Y_{t+h}^{(h)} = \frac{1}{B} \sum_{b=1}^{B} \sum_{s=1}^{S} \left(\boldsymbol{\psi}_{b,s}' \boldsymbol{W}_t^{(1)} + \boldsymbol{\phi}_{b,s}' \boldsymbol{F}_t^{(1)} + \epsilon_{s,t+h}^{(h)} \right) \mathbb{1}\left(\boldsymbol{Z}_t^{(1)} \in \mathcal{R}_{b,s}(\alpha, \eta)\right) \tag{3}$$

where the parameters $\boldsymbol{\psi}_{b,s}$ and $\boldsymbol{\phi}_{b,s}$ are the state-dependent parameters for the b-th bootstrap and $\mathcal{R}_{b,s}(\alpha, \eta)$ denotes the s-th sub-state of the b-th dynamic factor tree. The minimum node size η is derived via blocked cross-validation for the original sample and is subsequently used for each block-bootstrapped sample.

2.5. Benchmarks

We assess the proposed dynamic factor trees and dynamic factor forests against two key benchmarks and five conventional benchmarks typically employed in the relevant literature (e.g. H. H. Kim & Swanson, 2014; Siliverstovs & Wochner, 2019; Stock & Watson, 2012).

2.5.1. Key Benchmarks

Dynamic Factor Models (DFM)

Our first benchmark model of interest is the standard dynamic factor model,

$$Y_{t+h}^{(h)} = \boldsymbol{\psi}' \boldsymbol{W}_t^{(1)} + \boldsymbol{\phi}' \boldsymbol{F}_t^{(1)} + \epsilon_{t+h}^{(h)} \tag{4}$$

where parameter coefficients are estimated from the original sample (OS) and use $L = 2$ autoregressive lags as well as $R = 5$ factors (cf. Stock & Watson, 2006, 2012, 2016; Siliverstovs & Wochner, 2018). In the spirit of tree-based forests, we will also consider a bootstrapped version, where we estimate a standard dynamic factor model for each bootstrapped sample (BS) (yielding $\boldsymbol{\psi}_b$ and $\boldsymbol{\phi}_b$) and subsequently average among all bootstrapped model coefficients. These models will be abbreviated as OS-DFM and BS-DFM, respectively.

Recession Probability Augmented Dynamic Factor Models (DFM-RP)

Our second key benchmark is closely related to the regime-switching dynamic factor model of Chauvet and Potter (2013),

$$Y_{t+h}^{(h)} = \boldsymbol{\psi}' \boldsymbol{W}_t^{(1)} + \boldsymbol{\phi}' \boldsymbol{F}_t^{(1)} + \boldsymbol{\vartheta}' \boldsymbol{Z}_t^{(1)} + \epsilon_{t+h}^{(h)} \tag{5}$$

where we directly include our partitioning variable as an explanatory variable into the DFM model (ibid., see their equation (16) on p. 164). While Chauvet and Potter (2013) use Markov-switching processes, we employ similar and exogenously provided recession probability indices $\boldsymbol{Z}_t^{(1)}$ (see Section 3 and 4). In analogy to the previous section, we will also consider a bootstrapped version, where we estimate equation (5) for each bootstrapped sample (yielding $\boldsymbol{\psi}_b$, $\boldsymbol{\phi}_b$, and $\boldsymbol{\vartheta}_b$) and then average among all bootstrapped model coefficients. We shall abbreviate the corresponding models as OS-DFM-RP and BS-DFM-RP, respectively.

2.5.2. Common Benchmarks

Similar to Siliverstovs and Wochner (2019) and Wochner (2018), we employ the following five common benchmarks: Historic mean (HMN), $Y_{t+h}^{(h)} = \varphi + \epsilon_{t+h}^{(h)}$, autoregressive processes with either a two, four or BIC-based lag order L (AR2, AR4, ARL), $Y_{t+h}^{(h)} = \boldsymbol{\psi}' \boldsymbol{W}_t^{(1)} + \epsilon_{t+h}^{(h)}$, as well as combined autoregressive distributed lag models (CADL). The CADL model first estimates for each explanatory variable in $\boldsymbol{X}_t^{(1)}$ an autoregressive distributed lag model, $Y_{t+h}^{(h)} = \boldsymbol{\psi}' \boldsymbol{W}_t^{(1)} + \boldsymbol{\xi}' \boldsymbol{V}_{k,t}^{(1)} + \epsilon_{t+h}^{(h)}$ with $\boldsymbol{V}_{k,t}^{(1)} = (X_{k,t}, \ldots, X_{k,t-L+1})'$ and $L = 2$ for both autoregressive and distributed lag terms, and subsequently averages the predictions of these models among all $k \in \{1, \ldots, K\}$ (H. H. Kim & Swanson, 2014; Siliverstovs & Wochner, 2019).

2.6. Forecast Evaluation

We will assess forecasting performance in terms of relative Root Mean Squared Forecast Errors (RMSFE) of any two models m and b as follows,

$$\text{relative RMSFE}_{m,b}^{(h)} = \frac{\text{RMSFE}_m^{(h)}}{\text{RMSFE}_b^{(h)}}$$

where,

$$\text{RMSFE}_i^{(h)} = \left[\frac{1}{T - Q - h + 1} \sum_{t \in \mathscr{T}_f} \left(Y_{t+h}^{(h)} - \hat{Y}_{i,t+h}^{(h)} \right)^2 \right]^{1/2}$$

where $\hat{Y}_{i,t+h}^{(h)}$ denotes the h-period ahead prediction of model $i \in \{m, b\}$ (e.g. Korobilis, 2017; Siliverstovs & Wochner, 2019; Stock & Watson, 2012). Moreover, we assess superior predictive ability of our dynamic factor trees and forests against the standard dynamic factor model by means of directed Diebold Mariano (1995) tests,

$$\text{H}_0 : \text{E}\left[\left(\epsilon_{b,t+h}^{(h)} \right)^2 \right] = \text{E}\left[\left(\epsilon_{m,t+h}^{(h)} \right)^2 \right] \qquad \text{H}_A : \text{E}\left[\left(\epsilon_{b,t+h}^{(h)} \right)^2 \right] > \text{E}\left[\left(\epsilon_{m,t+h}^{(h)} \right)^2 \right]$$

and install two cautionary measures through the use of heteroscedasticity and autocorrelation robust standard errors (for $h > 1$) as well as McCracken (2007)'s critical values for nested model comparisons (cf. Siliverstovs & Wochner, 2019). Borrowing McCracken's (2007, p. 724) argument, directed testing is applied because our main models, m, (DFT, DFF) and main benchmark, b, (DFM) are nested.[4] Finally, building upon Welch and Goyal (2008), the state-dependent evaluation literature promotes the use of the cumulative sum of squared forecast error differences (CSSFED) between two models b and m,

$$\text{CSSFED}_{b,m}^{(h)}(t_0, t_1) = \sum_{t=t_0}^{t_1} \left(\epsilon_{b,t+h}^{(h)}\right)^2 - \left(\epsilon_{m,t+h}^{(h)}\right)^2$$

with continuously increasing t_0, which allows us to assess the forecast performance over time (see e.g. Siliverstovs, 2017, 2019; Siliverstovs & Wochner, 2019). While a horizontal movement of CSSFED indicates similar performance between model m and b, an upward [downward] trending series indicates persistent superiority [inferiority] of model m over b, whereas an upward [downward] jumping series indicates transient superiority [inferiority] (cf. ibid.).

3. Data

We bridge three distinct FRED datasets for our main analyses all of which are provided by the Federal Reserve Bank of St. Louis. We use 125 monthly indicators from FRED-MD and the quarterly GDP time-series from FRED-QD as dependent variable, all of which are available from 1960 until 2018 (McCracken & Ng, 2019a, 2019b; see McCracken, 2019).[5] FRED-MD and FRED-QD constitute fairly new data services ever since 2015 and 2018, respectively, that manage data revisions and real-time updates (McCracken & Ng, 2016) and find increasing use among macroeconomists (e.g. Korobilis, 2017; Medeiros et al., 2019; Siliverstovs & Wochner, 2019; Wochner, 2018). The third FRED source corresponds to the MCJH-recession probability index, which builds upon the work of Marcelle Chauvet and James Hamilton (MCJH) (2006) and provides the probability of a recession at any given quarter since October 1967 (see FRED, 2019; Hamilton, 2019).[6] Hence, a balanced panel with information from all three sources is available as of October 1967.

Similar to Siliverstovs and Wochner (2019), the following data transformations were applied: First, less than 0.35% of the datasets were classified as outliers in FRED-MD and these were substituted with the median of the previous five observations (see Stock & Watson, 2012, online appendix B) and no outliers were detected for the GDP series in FRED-QD. Second, all data were stationarity transformed as defined in McCracken and Ng (2019a; 2019b) and all the partitioning variables in use were first differenced. Following Stock and Watson (2012), the dependent variable was h-period stationarity transformed (ibid., see their Table B.2) as indicated by the super-index (h) in $Y_{t+h}^{(h)}$.

To cope with mixed frequencies, we follow Kim and Swanson (2014)'s Journal of Econometrics article who proposed to interpolate quarterly GDP values. As is well known, interpolations, however, may induce a measurement error and can affect the

dynamics of the interpolated series and its associations with other variables (Angelini, Henry, & Marcellino, 2006). To mitigate the concerns that a particular choice of interpolation method is driving our results, we will examine in total three distinct interpolation methods and will also assess the results in quarterly frequency as a robustness test (see Section 4.2). Our main specifications in Section 4.1 interpolate GDP (in levels) via Denton-Cholette (DCO) (Sax & Steiner, 2013).[7] Finally, similar to previous research (e.g. Siliverstovs & Wochner, 2019), we do not have real-time vintages of the data available, we assume the same publication structure for all variables as in the last vintage date and pursue a quasi real-time forecasting exercise.[8] For variables with ragged edges, we employ Altissimo et al. (2010)'s lagging procedure at the cost of a few observations at the beginning of the sample to ensure a balanced dataset.

4. Results

4.1. Main Results

The main results of our direct out-of-sample forecasting experiment for horizons $h = 1$ (nowcasts) and $h = 3$ (forecasts) is based on a recursively expanding scheme over a forecasting window of slightly more than 20 years (Jan. 1998 until Sep. 2018). Our proposed models can be divided into four main groups depending on the model type (DFT vs. DFF) and factor targeting (plain factors (F) vs. targeted factors (TF)). Within each of these four categories, we further distinguish between numeric (NUM) and binary measurement scales of the partitioning variable, where the latter discretize $\boldsymbol{Z}_t^{(1)}$ by assigning a binary classification to the values above and below the 40, 50, and 60th percentile of the partitioning variable, respectively (BIN40, BIN50, BIN60). Such a binary discretization (exogenously) determines "good" and "bad" states, which limits the number of possible splits and may therefore be meaningful to reduce the problem of overfitting — particularly when no cross-validation is applied (see Section 2 and 4.2; Zeileis et al., 2008; also Strobl et al., 2011).

4.1.1. Full Sample Evaluation

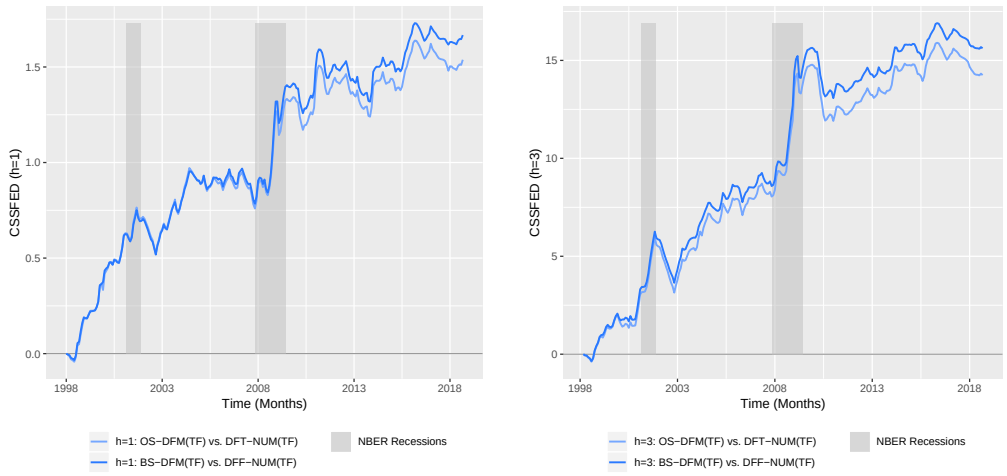
Table 1 summarizes our main results in terms of relative RMSFE against the AR2 benchmark over the full evaluation sample and highlights several interesting results: First, the dynamic factor tree and forest entries are printed in bold if they are superior in terms of RMSFE than the corresponding dynamic factor model. An inspection of Table 1 provides strong evidence in favor of the proposed models and highlights that dynamic factor trees and forests are statistically significantly superior to the standard dynamic factor model. In terms of MSFE, dynamic factor trees and forests tend to improve upon the standard dynamic factor models by almost 20%. For example, the best performing dynamic factor forest for three-month ahead predictions (DFF-NUM(TF)) achieves a relative MSFE of 0.643 whereas the standard targeted dynamic factor model from bootstrapped sample (BS-DFM(TF)) has a relative MSFE of 0.796. Second, the results also appear to be well in accord with the factor targeting literature (e.g. Bai & Ng, 2008; Stock & Watson, 2012; Boivin & Ng, 2006) in that factor

	Benchmarks		**Dynamic Factor Trees**		**Dynamic Factor Forests**	
Horizon	Results		Models	rRMSFE	Models	rRMSFE
h=1	HMN	Factors	OS-DFM(F)	*0.841*	BS-DFM(F)	*0.839*
	1.001		OS-DFM-RP(F)	0.773***	BS-DFM-RP(F)	0.772***
			DFT-NUM(F)	**0.762*****	DFF-NUM(F)	**0.754*****
	AR4		DFT-BIN40(F)	**0.766*****	DFF-BIN40(F)	**0.765*****
	0.924		DFT-BIN50(F)	**0.763*****	DFF-BIN50(F)	**0.758*****
			DFT-BIN60(F)	**0.767*****	DFF-BIN60(F)	**0.766*****
	ARL	Targeted Factors	OS-DFM(TF)	*0.802*	BS-DFM(TF)	*0.800*
	0.894		OS-DFM-RP(TF)	0.755***	BS-DFM-RP(TF)	0.754***
			DFT-NUM(TF)	**0.732*****	DFF-NUM(TF)	**0.724*****
	CADL		DFT-BIN40(TF)	**0.739*****	DFF-BIN40(TF)	**0.738*****
	0.922		DFT-BIN50(TF)	**0.739*****	DFF-BIN50(TF)	**0.736*****
			DFT-BIN60(TF)	**0.743*****	DFF-BIN60(TF)	**0.739*****
h=3	HMN	Factors	OS-DFM(F)	*0.914*	BS-DFM(F)	*0.909*
	1.014		OS-DFM-RP(F)	0.781***	BS-DFM-RP(F)	0.780***
			DFT-NUM(F)	**0.831*****	DFF-NUM(F)	**0.814*****
	AR4		DFT-BIN40(F)	**0.820*****	DFF-BIN40(F)	**0.814*****
	0.970		DFT-BIN50(F)	**0.827*****	DFF-BIN50(F)	**0.821*****
			DFT-BIN60(F)	**0.829*****	DFF-BIN60(F)	**0.822*****
	ARL	Targeted Factors	OS-DFM(TF)	*0.894*	BS-DFM(TF)	*0.892*
	0.975		OS-DFM-RP(TF)	0.775***	BS-DFM-RP(TF)	0.774***
			DFT-NUM(TF)	**0.812*****	DFF-NUM(TF)	**0.802*****
	CADL		DFT-BIN40(TF)	**0.813*****	DFF-BIN40(TF)	**0.808*****
	0.943		DFT-BIN50(TF)	**0.819*****	DFF-BIN50(TF)	**0.811*****
			DFT-BIN60(TF)	**0.820*****	DFF-BIN60(TF)	**0.811*****

Notes: The table entries show the relative root mean squared forecast error (RMSFE) of a particular model against the AR2 benchmark (settings: recursive scheme; DCO interpolation; Jan. 1998 first vintage; MCJH-REC-IDX partitioning variable). DFT and DFF models employ model splitting at the conventional 5% significance level. The star symbols indicate the level of statistical significance from a directed Diebold Mariano (1995) test that assesses superiority in predictive performance of dynamic factor trees and forests (namely, DFT(F), DFT(TF), DFF(F) and DFF(TF)) or recession probability augmented dynamic factor models (namely, OS-DFM-RP(F), OS-DFM-RP(TF), BS-DFM-RP(F), BS-DFM-RP(TF)) against the corresponding (italicized) standard dynamic factor model (namely, OS-DFM(F), OS-DFM(TF), BS-DFM(F), and BS-DFM(TF)). For example, superior predictive ability of the DFT(TF) model is assessed against the OS-DFM(TF) model. The HMN, AR4, ARL and CADL benchmarks are compared against DFM(F). For nested model comparisons, McCracken (2007)'s critical values are used. The entries for DFTs and DFFs are bold if they have equal or lower MSFE than the corresponding DFM. The best model of each of the four groups per horizon is underlined. Factors are targeted based on a hard-threshold in the spirit of Bai and Ng (2008). The symbols $^{+}$, *, **, *** indicate significance at the 15%, 10%, 5% and 1% level, respectively. For more details, see Section 4.1.

Table 1: Main Results (relative RMSFE)

targeting consistently helps to improve predictive performance. Moreover, while numeric and binary partitioning variables perform comparably well, numeric partitions are able to meaningfully exploit the richer information and perform slightly superior. Third, consistent with the tree-based ensemble and forecast combination literature (e.g. Medeiros et al., 2019; Breiman, 2001a; James et al., 2013; also Elliott & Timmermann, 2016), dynamic factor forests systematically improve upon dynamic factor trees in all cases. Fourth, dynamic factor trees and forests perform about equally well as recession probability augmented dynamic factor models (DFM-RP) — which is broadly in line with Chauvet and Potter (2013) in that that DFM-RP models are closely related to their Markov-switching dynamic factor model, which also achieved strong predictive performance in their setting. In fact, while DFTs and DFFs tend to be first-best for nowcasts ($h = 1$), the DFM-RP models tend yield the first-best forecasts ($h = 3$).



2.a: Nowcast ($h = 1$) **2.b:** Forecast ($h = 3$)

Notes: This figure shows the CSSFED of the conventional dynamic factor model against the top-performing dynamic factor trees and forests in Table 1 (settings: recursive scheme; DCO interpolation; Jan. 1998 first vintage; MCJH-REC-IDX partitioning variable). An upward [downward] movement of CSSFED indicates superior [inferior] performance of DFT or DFF over DFM (Siliverstovs, 2017, 2019; Siliverstovs & Wochner, 2019). While the absolute levels of two CSSFED in Figure 2.a and 2.b may not be compared quantitatively because the CSSFED are not measured on similar scales (non-standardized SFED) (cf. Siliverstovs & Wochner, 2019), they reveal qualitatively similar dynamics.

Figure 2: DFTs and DFFs Relative Performance against DFMs over Time (CSSFED)

4.1.2. Sub-sample Evaluation

To better understand the evolution of these forecasting improvements, we follow the burgeoning state-dependent forecast evaluation literature and examine (in addition to full sample forecast evaluations above) also those for the sub-samples in boom and bust periods according to NBER (e.g. Chauvet & Potter, 2013; Fossati, 2018; Siliverstovs, 2017; Siliverstovs & Wochner, 2019). Table 2 reveals that the DFT and DFF models perform better than the DFM in both recessionary and expansionary subsamples. Specifically, during recessions they typically achieve seizable and statistically significant improvements over the DFM, whereas in expansions the improvements are still present but more moderate in terms of size and significance. Especially at higher forecasting horizons, the performance during expansions is similar to those of the AR2 benchmark, which is again in accord with the state-dependent forecast evaluation literature (e.g. Siliverstovs & Wochner, 2019).

Building upon Siliverstovs (2017; 2019) and Siliverstovs and Wochner (2019), Figure 2 provides the CSSFED of the conventional dynamic factor model with targeted factors from bootstrapped samples (b = BS-DFM(TF)) against the targeted dynamic factor tree (m_1 = DFT-NUM(TF)) as well as against the targeted dynamic factor forest (m_2 = DFF-NUM(TF)). As explained in Section 2.6, the figure visualizes how the relative performance of these models evolves over the evaluation window: On the one hand, the CSSFED series show a fairly consistent upward trend (with a slightly diminishing slope) and have, on the other, larger upward jumps during recessions (and slight to moderate deteriorations in the aftermath of recessions). This indicates that targeted dynamic factor trees and forests tend to perform persistently superior to dynamic factor models in general and especially during recessions.

Horizon		BM Results		Dynamic Factor Trees Models	Dynamic Factor Trees rRMSFE	Dynamic Factor Forests Models	Dynamic Factor Forests rRMSFE
h=1	Expansion	HMN 0.991	Factors	OS-DFM(F) OS-DFM-RP(F)	*0.979* 0.912***	BS-DFM(F) BS-DFM-RP(F)	*0.976* 0.910***
		AR4 0.966		DFT-NUM(F) DFT-BIN50(F)	**0.921*** **0.922***	DFF-NUM(F) DFF-BIN50(F)	**0.911*** **0.915***
		ARL 0.936+	Fact.	OS-DFM(TF) OS-DFM-RP(TF)	*0.941* 0.881***	BS-DFM(TF) BS-DFM-RP(TF)	*0.939* 0.879***
		CADL 0.956	Trgt.	DFT-NUM(TF) DFT-BIN50(TF)	**0.875*** **0.884***	DFF-NUM(TF) DFF-BIN50(TF)	**0.866*** **0.880***
	Recession	HMN 1.012	Factors	OS-DFM(F) OS-DFM-RP(F)	*0.655* 0.584**	BS-DFM(F) BS-DFM-RP(F)	*0.654* 0.584**
		AR4 0.877		DFT-NUM(F) DFT-BIN50(F)	**0.533*** **0.535***	DFF-NUM(F) DFF-BIN50(F)	**0.530*** **0.535***
		ARL 0.847	Fact.	OS-DFM(TF) OS-DFM-RP(TF)	*0.613* 0.587	BS-DFM(TF) BS-DFM-RP(TF)	*0.612* 0.586
		CADL 0.883	Trgt.	DFT-NUM(TF) DFT-BIN50(TF)	**0.531*** **0.535***	DFF-NUM(TF) DFF-BIN50(TF)	**0.523*** **0.534***
h=3	Expansion	HMN 1.003*	Factors	OS-DFM(F) OS-DFM-RP(F)	*1.067* 0.982***	BS-DFM(F) BS-DFM-RP(F)	*1.064* 0.982***
		AR4 0.999		DFT-NUM(F) DFT-BIN50(F)	**1.035*** **1.028****	DFF-NUM(F) DFF-BIN50(F)	**1.015**** **1.022****
		ARL 1.000	Fact.	OS-DFM(TF) OS-DFM-RP(TF)	*1.058* 0.975***	BS-DFM(TF) BS-DFM-RP(TF)	*1.059* 0.976***
		CADL 0.971+	Trgt.	DFT-NUM(TF) DFT-BIN50(TF)	**0.997**** **1.004****	DFF-NUM(TF) DFF-BIN50(TF)	**0.986*** **0.999****
	Recession	HMN 1.023	Factors	OS-DFM(F) OS-DFM-RP(F)	*0.769* 0.567***	BS-DFM(F) BS-DFM-RP(F)	*0.761* 0.565***
		AR4 0.946		DFT-NUM(F) DFT-BIN50(F)	**0.618*** **0.616***	DFF-NUM(F) DFF-BIN50(F)	**0.601*** **0.610***
		ARL 0.954	Fact.	OS-DFM(TF) OS-DFM-RP(TF)	*0.735* 0.561***	BS-DFM(TF) BS-DFM-RP(TF)	*0.730* 0.559***
		CADL 0.919	Trgt.	DFT-NUM(TF) DFT-BIN50(TF)	**0.620*** **0.627***	DFF-NUM(TF) DFF-BIN50(TF)	**0.610*** **0.615***

Notes: Building upon Chauvet and Potter (2013) and Siliverstovs and Wochner (2019), the table entries show the relative RMSFE of a particular model against the AR2 benchmark for expansionary and recessionary sub-samples separately (settings: recursive scheme; DCO interpolation; Jan. 1998 first vintage; MCJH-REC-IDX partitioning variable). Recessions and expansions are determined according to NBER. For $h = 1$ and $h = 3$, the forecast evaluation window contains 249 and 247 monthly time periods, respectively; the expansionary sub-sample contains 221 and 219 observations and the remaining observations belong to the recessionary sub-sample. BM stands for benchmarks. For more details about table entries, see notes in Table 1.

Table 2: Main Results for Sub-samples (relative RMSFE)

4.2. Robustness Results

This section examines how sensitive our estimation results are with respect to several modelling choices and assumptions. We will hereafter assess alternatives typically encountered in the relevant literature (e.g. H. H. Kim & Swanson, 2014; Siliverstovs & Wochner, 2019; Stock & Watson, 2012).

4.2.1. Alternative Interpolations

To mitigate the concerns that our results are driven by a particular choice of interpolation method, we shall examine three distinct ones: Our main specifications interpolate GDP (in levels) via Denton-Cholette (DCO), which seeks to preserve the movement at a higher frequency (Sax & Steiner, 2013). In a notable contribution to the Journal Econometrics, Kim and Swanson (2014) propose to interpolate GDP via Chow-Lin, which can either be applied to stationary or co-integrated series (Sax & Steiner, 2013, p. 80ff.). In the former case, GDP (in growth rates) is interpolated from a dynamic factor model with monthly factors (CLU).[9] In the latter, GDP (in levels) is interpolated from three monthly co-integrated series (CL3).[10] All three interpolations (DCO, CL3, CLU) qualify as equally appropriate as there has not yet emerged a general consensus on a first best interpolation method (Guérin & Marcellino, 2013).

Table A.1 shows the result for the CLU-based interpolations of GDP growth. As can be seen the results are fairly similar to DCO-interpolations both in terms of size and significance. Still all DFTs and DFFs outperform the DFMs and tend to be substantially better than conventional benchmarks (e.g. HMN, ARL). Likewise, Table A.2 shows that the main results can be qualitatively maintained for CL3 interpolations but are quantitatively slightly weaker in terms of size and significance.

4.2.2. Quarterly Frequency

Table A.3 provides the estimation results for quarterly series by quarterly aggregation of the monthly FRED-MD series (cf. Foroni & Marcellino, 2013). As expected, the substantial reduction of sample size by two thirds results in weaker model performance because the power of the parameter instability tests is decisively weakened.[11] For example, while the monthly nowcasts ($h_m = 1$) for the DFT-NUM(TF) find parameter instabilities in virtually every time-period and split the model accordingly (see Figure 2); the equivalent quarterly nowcasts ($h_q = 0$) of the DFT-NUM(TF) split only in very few time-periods. The similar performance between DFT-NUM(TF) and OS-DFM(TF) models can therefore be attributed to the fact that there are only a few observations responsible for performance differences because the two models are formally equivalent in the absence of any splits (see Section 2). Nevertheless, aggregating a large number of trees proves helpful and dynamic factor forests show a fairly robust performance at quarterly frequencies: Almost all DFFs still outperform DFMs — and do so often significantly.

4.2.3. Rolling Windows

While our main results are based on recursively expanding windows, Table A.4 examines the effect of using rolling schemes with window lengths of 300 months (see e.g. H. H. Kim & Swanson, 2014; Stock & Watson, 2012). The table shows that our main findings are robust to this change. In fact, the results strengthen our main findings in that the improvements of DFTs and DFFs over DFMs tend to be slightly more accentuated under rolling than under recursive windows.

4.2.4. Extended Forecasting Window

While the forecast evaluation window for the main specification in Section 4.1 includes over 20 years (1998-2018) with almost 250 monthly predictions, we examine an extension of the evaluation window to 1985-2018 with over 400 monthly observations (Siliverstovs & Wochner, 2019; Wochner, 2018). Table A.5 shows that the main results still persist but are weaker in terms of size (especially for nowcasts), which indicates that the nowcasting over the post-millennial period appears to be better than over the pre-millennial period (also see Wochner, 2018). As the sources of these differences cannot be attributed to insufficient power due to shorter sample sizes (because DFTs and DFFs tend to split over the pre-millennial era), they appear to be, at least in part, structural in the sense that the proposed methodology appears to work worse over the pre-millennial period (1985-2000). Moreover, Table A.5 makes evident that the DFM-RP models are the first best models in each group at all horizons.

4.2.5. Alternative Recession Probability Indices

While our main specifications make use of the MCJH-index (see Chauvet & Hamilton, 2006; Hamilton, 2019), the Federal Reserve Bank of Philadelphia conducts the Survey of Professional Forecasters (SPF) (2018; 2019) and provides the mean survey response concerning the probability of a recession in the current (SPF-REC1-IDX) and next quarter (SPF-REC2-IDX) ever since 1968. These survey-based indicators shall be used as alternative partitioning variables to the data-based version of the MCJH-index. In analogy to our main analyses, the quarterly SPF-REC1 and SPF-REC2 indices are interpolated to monthly frequencies via Denton-Cholette (Sax & Steiner, 2013). The results in Table A.6 and A.7 provide empirical support for the proposed dynamic factor trees and forests for these alternative recession probability indices: At all horizons the results are only slightly weaker and still outperform the standard DFM and rank in about half of all cases as the first best model among all those considered. This suggests that the proposed methodology is not dependent on the MCJH-index and generalizes to alternative recession probability indicators.

4.2.6. Absence of Hyper-Parameter Tuning

As outlined in Section 2, the hyper-parameters of the DFTs and DFFs, such as minimum node size, were defined via cross-validation. In the absence of such parameter tuning, we would expect the performance to be less pronounced (cf. James et al., 2013; Zeileis & Hothorn, 2015) and this is indeed what Table A.8 shows: The numeric DFTs and DFFs tend to perform slightly worse in the absence of cross-validation. As expected, binary models tend to perform often equally well with or without parameter tuning because they can only have a single split and are consequentially less sensible to parameter tuning.

4.2.7. Single Factor and Ten Factors

While the main results are all based on five dynamic factors, Table A.9 and A.10 examine the model performance with only a single factor and ten factors (cf. Siliver-

stovs & Wochner, 2019). The ten [one] factor models tend to perform slightly superior [inferior] than the five factor models, but the main results stay qualitatively the same in that DFTs and DFFs remain superior to the standard DFMs.

4.2.8. Publication Lag

The MCJH-Index has a fairly consistent publication structure and is released with a publication lag of 4 to 6 months.[12] While the main results assumed the MCJH-index to be released together with all monthly indicators, i.e. having a one month lag, we may alternatively seek to predict the missing values arising from lagged publications (cf. Bulligan et al., 2015). Inspired by Kim and Swanson (2016), we infer the missing values on the current edge, $Z^{(1)}_{t+h}$, with a simple regime-switching model that alternates between predictions from a distributed lag model and a beta regression model (cf. Caribari-Neto & Zeileis, 2019; Hill, Griffiths, Lim, & Lim, 2011).[13] Table A.11 provides the estimation results and indicates that the timely release of the MCJH-index contains indeed relevant information — particularly for the 3-months ahead forecasts. While the nowcasting results ($h = 1$) remain robust to publication lags, the forecasting results ($h = 3$) are weaker but the DFFs still tend to outperform DFMs slightly.

These findings suggest that the employed MCJH-index entails both advantages and disadvantages: On the one hand, it appears to be the one that yields the best performance in our modelling framework among all indices considered. On the other, the index is unfortunately not (yet) available in a monthly frequency and publication lags appear to weaken model performance to some extent. In light of these results, a high-frequency real-time release of the MCJH-index in the spirit of Aruoba, Diebold and Scotti (2009)'s daily business cycle indicator appears to be desirable. In any case, the analyses have made clear that weaker performance due to publication lags can only hardly be considered as a weakness of the proposed dynamic factor trees and forests but constitute rather a limitation of the data.

5. Conclusion

This study proposed dynamic factor trees and forests based on Zeileis et al. (2008)'s model-based recursive partitioning algorithm, which allows to embed theory-led dynamic factor models within tree-based machine learning ensembles conditional on the state of the business cycle (see Section 1 and 2). In our out-of-sample forecasting exercise for short-term GDP growth predictions, we find strong and statistically significant empirical evidence that they systematically outperform standard dynamic factor models in both expansive and contractive periods. These observations corroborate the idea that the proposed state-dependent models are indeed able to beneficially exploit the time-varying dynamics in good and bad times — much like sailors who benefit from maneuvering their ships differently in stormy and calm seas (cf. K. Kim & Swanson, 2016; Diebold & Rudebusch, 1996).

We see several possible avenues for future research: First, the proposed modelling framework will have to be examined for alternative dependent variables. It is, however,

not clear à-priori whether it will work as well for other variables because these series may have structurally distinct characteristics. For instance, Corradi and Swanson (2014)'s test for structural instability in dynamic factor models rejects the null of stability for GDP but fails to reject the null for the S&P500, producer price index or the 10-year Treasury-bond rate. Second, despite the fact that DFTs and DFFs tend to perform overall generally superior to DFMs, they tend to perform slightly worse than DFMs in the aftermath of recessions. Hence, the hyper-parameter α could be endogenized (α_t) so as to allow splitting ($\alpha_t > 0$) in some time periods t and disallow it in others ($\alpha_t = 0$) (cf. e.g. K. Kim & Swanson, 2016). Third, the present work could be readily extended to a two-stage machine learning framework where the recession probability index itself is also derived via recent advances in (statistical) machine learning (e.g. Döpke, Fritsche, & Pierdzioch, 2015). Finally, while we used only a single partitioning variable in DFTs and DFFs, the framework can also be readily generalized to multiple partitioning variables, which may reveal further relevant sub-states (cf. Zeileis et al., 2008). For example, when inflation is used as dependent variable, one could think of the different Federal Reserve Bank regimes (e.g. Greenspan, Bernanke, Yellen) as potentially relevant categorical partitioning variables (cf. Fernandez-Villaverde & Rubio-Ramirez, 2013). Zeileis et al. (2008)'s framework is general enough to encompass exciting future research along these lines.

Notes

1. Theoretical equilibrium models often assume that macroeconomic dynamics are driven by a few shocks, such as technology or monetary policy shocks (Bai & Ng, 2007; Giannone, Reichlin, & Sala, 2006). Dynamic factor models are consistent with these assumptions and rest on the idea that the co-movements among many economic variables can be decomposed into two (orthogonal) sources: A "common component" (common shocks or common factors) that captures a few unobserved factors that govern the dynamics of many variables plus a "idiosyncratic component", which captures residual peculiarities of each individual series (e.g. D'Agostino, Giannone, Lenza, & Modugno, 2016; Diebold & Rudebusch, 1996; Forni et al., 2009; Giannone et al., 2006; Stock & Watson, 2006, 2016, 2017). Furthermore, theoretical business cycle equilibrium models can be shown to take a factor-like structure in case of measurement errors (cf. e.g. Giannone et al., 2006; Diebold & Rudebusch, 1996).

2. As indicated in the text, parts of this section follow closely Zeileis et al. (2008) as well as Zeileis and Hothorn (2015) (also see Kopf et al., 2013, for a thorough review).

3. Equation (2) is, for instance, closely related to standard regime switching models, such as threshold autoregressive (TAR) models: Specifically, if there were no factors present and if the partitioning variable corresponded to lagged values of the dependent variable, then equation (2) would come close to a TAR-like structure (cf. e.g. Kock & Teräsvirta, 2011, p. 62ff.).

4. The determination of McCracken's (2007) critical values for two nested models requires the difference in the number of model parameters, k_2 (excess parameters). For example, comparing DFM-RP versus DFM, we simply have $k_2 = 1$. As seen in Section 2.3.3, dynamic factor trees with a binary dependent variable are mathematically equivalent to fully-interacted dynamic factor models. More generally, for λ_t many terminal nodes in the dynamic factor tree in period t, there are $(\lambda_t - 1) \times (1 + L + R)$ many excess parameters relative to the dynamic factor model (i.e. one additional intercept, L additional autoregressive lags and R additional factors per additional terminal node). To determine k_2 for trees [and forests], we use the average number of terminal nodes, $\bar{\lambda} = 1/|\mathscr{S}_f| \sum_{t=1}^{T} \lambda_t$ [and $\bar{\lambda} = 1/|\mathscr{S}_f| \sum_{t=1}^{T} 1/B \sum_{b=1}^{B} \lambda_{b,t}$]. Moreover, since the maximum k_2 provided by McCracken (2007) is limited to 10, we use $k_2 = \min(\bar{\lambda}, 10)$.

5. The raw datasets for FRED-MD actually provide information for 128 monthly and 248 quarterly macroeconomic indicators (see McCracken & Ng, 2019a, 2019b). From FRED-MD, we dropped monthly and quarterly datasets which are not available at the beginning of 1960 (cf. Wochner, 2018) and retained quarterly GDP from FRED-QD. Factors are subsequently extracted from the 375 contemporaneous, first and second

order lags of the monthly indicators (cf. H. H. Kim & Swanson, 2014). Moreover, to retain the `UMCSENTx` index (FRED-MD), which is only available in quarterly frequency until Q4-1977, the initial values were interpolated via Denton-Cholette (Sax & Steiner, 2013).

6. The FRED mnemonic of the index is `JHGDPBRINDX`. We will cope with the mixed-frequency problem by interpolating quarterly recession probability indices via Denton-Cholette (Sax & Steiner, 2013) and will examine the quarterly frequency equivalent as a robustness test.

7. Flow variables are typically interpolated such that the sum or mean of the monthly interpolated values correspond to the quarterly observed value (Chipman & Lapham, 1995, p. 89) and the sum is typically chosen for flow variables, such as GDP or national income (ibid. p. 92). However whether the mean or the sum is chosen for the interpolation of GDP levels is mathematically irrelevant for the subsequent derivation of the logarithmic growth rate, which is given as $Y_t^{(h)} = \ln(y_t) - \ln(y_{t-h})$ with y_t denoting the level of GDP in period t (cf. e.g. Siliverstovs & Wochner, 2019; Stock & Watson, 2012). To see this, let $\dot{y}_t$ denote the interpolated level of GDP (meeting the sum constraint) and $\ddot{y}_t$ denote the interpolated level of GDP (meeting the mean constraint). As each quarter has $\nu = 3$ months, it follows that each $\dot{y}_t$ is a third of $\ddot{y}_t$, i.e. $\nu \times \dot{y}_t = \ddot{y}_t$, and thus the logarithmic growth rate for both series is mathematically equivalent: $\ddot{Y}_t^{(h)} = \ln(\ddot{y}_t) - \ln(\ddot{y}_{t-h}) = \ln(\nu \times \dot{y}_t) - \ln(\nu \times \dot{y}_{t-h}) = \ln(\dot{y}_t) - \ln(\dot{y}_{t-h}) = \dot{Y}_t^{(h)}$.

8. Notice that we assume the recession probability index to be released together with FRED-MD variables. The robustness of this assumption will be assessed in Section 4.2.8.

9. As previously explained, we use h-period differenced growth rates, $Y_t^{(h)} = \ln(y_t) - \ln(y_{t-h})$ as dependent variables, where y_t is the level of GDP in period t (cf. Stock & Watson, 2012; Siliverstovs & Wochner, 2019; also see McCracken & Ng, 2019a). While the quarterly growth rates, $Y_t^{(3)}$, can be simply derived from the interpolated GDP levels via DCO and CL3, this task is more complex for CLU as it does not impute the levels, y_t, but monthly growth rates, $Y_t^{(1)}$. Hence, for $h = 3$, we will reconstruct the quarterly growth rates, $Y_t^{(3)}$, from the monthly growth rates, $Y_t^{(1)}$, by summing over the three adjacent values of $Y_t^{(1)}$, because

$$\begin{aligned} Y_t^{(3)} &= Y_t^{(1)} + Y_{t-1}^{(1)} + Y_{t-2}^{(1)} \\ &= \big(\ln(y_t) - \ln(y_{t-1})\big) + \big(\ln(y_{t-1}) - \ln(y_{t-2})\big) + \big(\ln(y_{t-2}) - \ln(y_{t-3})\big) \\ &= \ln(y_t) - \ln(y_{t-3}). \end{aligned}$$

10. The Chow-Lin interpolation "CL3" is obtained via the three co-integrated series with FRED-mnemonics `DPCERA3M086SBEA`, `SRVPRD` and `CE16OV`. Stationarity of the errors is warranted in that the (augmented) Dickey-Fuller test (Trapletti, Hornik, & LeBaron, 2019) for the residuals of the regression in levels rejects the null of non-stationarity at the 1% significance level.

11. While this argument relates to the estimation sample, a similar argument applies to the subsequent tests on superior predictive performance for the evaluation sample.

12. For example, for vintage dates in Jan, Feb and May 2018 (Q1 2018), the index is available until Sep 2017 (Q3 2017). Likewise, for vintage dates in April, May and June 2018 (Q2 2018), the index is available until Dec 2017 (Q4 2017) (see Hamilton, 2019).

13. Formally, the regime-switching model is given as (see K. Kim & Swanson, 2016, for a related specification),

$$\hat{Z}_{t+h}^{(1)} = \mathbb{1}(\hat{z}_{t+h} < \varsigma)\hat{Z}_{\text{DL},t+h}^{(1)} + \big(1 - \mathbb{1}(\hat{z}_{t+h} < \varsigma)\big)\hat{Z}_{\text{BETA},t+h}^{(1)},$$

where $\hat{Z}_{\text{DL},t+h}^{(1)}$ designates the prediction from a distributed lag model with BIC chosen number of $Y_{t-i_Y}^{(1)}$ lags with $i_Y \in \{0, 1, \dots, 4\}$ (cf. Hill et al., 2011) and $\hat{Z}_{\text{BETA},t+h}^{(1)}$ is the predicted value of a beta-regression with a logit-link function for the mean and precision equation (see Caribari-Neto & Zeileis, 2019). Both mean and precision equations are BIC optimized over $Y_{t-i_Y}^{(1)}$ lags with $i_Y \in \{0, 1, \dots, 4\}$, the first i_F linear factors $F_{i_F,t}^{(1)}$ with $i_F \in \{1, \dots, 10\}$ and the first i_Q quadratic factors $\big(F_{i_Q,t}^{(1)}\big)^2$ with $i_Q \in \{0, \dots, 10\}$ (cf. Bai & Ng, 2008). Finally, $\varsigma = 0.2$ and $\hat{z}_{t+h}$ corresponds to the predicted raw value of the index from the beta regression.

Acknowledgements

We gratefully acknowledge valuable inputs and comments from Jan-Egbert Sturm, Philipp Baumann, Boriss Siliverstovs, Samad Sarferaz, Alexander Rathke, Petra Ehmann and Oliver Müller. The paper also benefitted from discussions with participants at the SkiLLS Seminar in Engelberg (Switzerland). As indicated in the text, parts of this work build upon and follow earlier work of the author: see Wochner (2018) and Siliverstovs and Wochner (2019).

References

Altissimo, F., Cristadoro, R., Forni, M., Lippi, M., & Veronese, G. (2010). New Eurocoin: Tracking Economic Growth in Real Time. *The Review of Economics and Statistics*, *92*(4), 1024–1034.

Andrews, D. W. (1993). Tests for Parameter Instability and Structural Change with Unknown Change Point. *Econometrica*, 821–856.

Angelini, E., Henry, J., & Marcellino, M. (2006). Interpolation and Backdating with a Large Information Set. *Journal of Economic Dynamics and Control*, *30*(12), 2693–2724.

Appenzeller, T. (2017). The Scientists' Apprentice. *Science*, *357*(6346), 16–17.

Aruoba, S. B., Diebold, F. X., & Scotti, C. (2009). Real-time Measurement of Business Conditions. *Journal of Business & Economic Statistics*, *27*(4), 417–427.

Athey, S., & Imbens, G. W. (2017). The State of Applied Econometrics: Causality and Policy Evaluation. *Journal of Economic Perspectives*, *31*(2), 3–32.

Athey, S., & Imbens, G. W. (2019). Machine Learning Methods that Economists Should Know About. *Annual Review of Economics*, *11*.

Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized Random Forests. *The Annals of Statistics*, *47*(2), 1148–1178.

Bai, J., & Han, X. (2016). Structural Changes in High Dimensional Factor Models. *Frontiers of Economics in China*, *11*(1), 9.

Bai, J., & Ng, S. (2007). Determining the Number of Primitive Shocks in Factor Models. *Journal of Business & Economic Statistics*, *25*(1), 52–60.

Bai, J., & Ng, S. (2008). Forecasting Economic Time Series using Targeted Predictors. *Journal of Econometrics*, *146*(2), 304–317.

Bai, J., & Wang, P. (2016). Econometric Analysis of Large Factor Models. *Annual Review of Economics*, *8*, 53–80.

Banerjee, A., Marcellino, M., & Masten, I. (2008). Forecasting Macroeconomic Variables Using Diffusion Indexes in Short Samples with Structural Change. In *Forecasting in the Presence of Structural breaks and Model Uncertainty* (pp. 149–194). Emerald Group Publishing Limited.

Bates, B. J., Plagborg-Møller, M., Stock, J. H., & Watson, M. W. (2013). Consistent Factor Estimation in Dynamic Factor Models with Structural Instability. *Journal of Econometrics*, *177*(2), 289–304.

Boivin, J., & Ng, S. (2006). Are More Data Always Better for Factor Analysis? *Journal of Econometrics*, *132*(1), 169–194.

Breiman, L. (1996). Bagging Predictors. *Machine Learning*, *24*(2), 123–140.

Breiman, L. (2001a). Random Forests. *Machine Learning*, *45*(1), 5–32.

Breiman, L. (2001b). Statistical Modeling: The Two Cultures. *Statistical Science*, *16*(3), 199–231.

Breitung, J., & Eickmeier, S. (2011). Testing for Structural Breaks in Dynamic Factor Models. *Journal of Econometrics*, *163*(1), 71–84.

Bulligan, G., Marcellino, M., & Venditti, F. (2015). Forecasting Economic Activity with Targeted Predictors. *International Journal of Forecasting*, *31*(1), 188–206.

Canty, A. J. (2002). Resampling Methods in R: The boot Package. *R News*, *2*(3),

2–7.

Caribari-Neto, F., & Zeileis, A. (2019). *Beta Regression in R* (Tech. Rep.). (R Vignette)

Chauvet, M., & Hamilton, J. D. (2006). Dating Business Cycle Turning Points. *Contributions to Economic Analysis*, *276*, 1–54.

Chauvet, M., & Potter, S. (2013). Forecasting Output. In *Handbook of Economic Forecasting* (Vol. 2, p. 141-194). Elsevier.

Chen, B., & Hong, Y. (2012). Testing for Smooth Structural Changes in Time Series Models via Nonparametric Regression. *Econometrica*, *80*(3), 1157–1183.

Chen, J. C., Dunn, A., Hood, K. K., Driessen, A., & Batch, A. (2019). Off to the Races: A Comparison of Machine Learning and Alternative Data for Predicting Economic Indicators. In *Big Data for 21st Century Economic Statistics.* University of Chicago Press.

Chen, X., & Ishwaran, H. (2012). Random Forests for Genomic Data Analysis. *Genomics*, *99*(6), 323–329.

Chipman, J. S., & Lapham, B. J. (1995). Interpolation of Economic Time Series, with Application to German and Swedish Data. In *Econometrics of Short and Unreliable Time Series* (pp. 89–139). Springer.

Corradi, V., & Swanson, N. R. (2014). Testing for Structural Instability of Factor Augmented Forecasting Models. *Journal of Econometrics*, *182*, 100-118.

Coulombe, P. G., Leroux, M., Stevanovic, D., & Surprenant, S. (2019). *How is Machine Learning Useful for Macroeconomic Forecasting?* (Tech. Rep.). Working Paper.

D'Agostino, A., Giannone, D., Lenza, M., & Modugno, M. (2016). Nowcasting Business Cycles: A Bayesian Approach to Dynamic Heterogeneous Factor Models. In *Dynamic Factor Models* (pp. 569–594). Emerald Group Publishing Limited.

Del Negro, M., Hasegawa, R. B., & Schorfheide, F. (2016). Dynamic Prediction Pools: An Investigation of Financial Frictions and Forecasting Performance. *Journal of Econometrics*, *192*(2), 391–405.

Diebold, F. X. (2003). Big Data Dynamic Factor Models for Macroeconomic Measurement and Forecasting. In *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress of the Econometric Society* (pp. 115–122).

Diebold, F. X., & Mariano, R. S. (1995). Comparing Predictive Accuracy. *Journal of Business & Economic statistics*, *20*(1), 134–144.

Diebold, F. X., & Rudebusch, G. D. (1996). Measuring Business Cycles: A Modern Perspective. *The Review of Economics and Statistics*, *78*(1), 67–77.

Döpke, J., Fritsche, U., & Pierdzioch, C. (2015). *Predicting Recessions in Germany with Boosted Regression Trees* (Tech. Rep.). DEP (Socioeconomics) Discussion Papers, Macroeconomics and Finance Series.

Doz, C., & Fuleky, P. (2019). Dynamic Factor Models.

Elliott, G., & Timmermann, A. (2016). *Economic Forecasting.* Princeton University Press.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level Classification of Skin Cancer with Deep Neural Networks. *Nature*, *542*(7639), 115.

Fernandez-Villaverde, I., & Rubio-Ramirez, I. F. (2013). Macroeconomics and Volatility: Data, Models, and Estimation. In *Advances in Economics and Econometrics: Tenth World Congress* (Vol. 3, p. 137).

Forni, M., Giannone, D., Lippi, M., & Reichlin, L. (2009). Opening the Black Box: Structural Factor Models with Large Cross Sections. *Econometric Theory*, *25*(5), 1319–1347.

Foroni, C., & Marcellino, M. (2013). A Survey of Econometric Methods for Mixed-Frequency Data.

Fossati, S. (2018). *A Test for State-Dependent Predictive Ability based on a Markov-Switching Framework.* University of Alberta.

FRED. (2019). GDP-Based Recession Indicator Index. Retrieved 24-05-2018, from `https://fred.stlouisfed.org/series/JHGDPBRINDX` (FRED Mnemonic: JHGDPBRINDX)

Garcia, M. G., Medeiros, M. C., & Vasconcelos, G. F. (2017). Real-Time Inflation Forecasting with High-Dimensional Models: The Case of Brazil. *International Journal of Forecasting*, *33*(3), 679–693.

Garge, N. R., Bobashev, G., & Eggleston, B. (2013). Random Forest Methodology for Model-based Recursive Partitioning: the mobForest Package for R. *BMC Bioinformatics*, *14*(1), 125.

Giannone, D., Reichlin, L., & Sala, L. (2006). VARs, Common Factors and the Empirical Validation of Equilibrium Business Cycle Models. *Journal of Econometrics*, *132*(1), 257–279.

Guérin, P., & Marcellino, M. (2013). Markov-Switching MIDAS Models. *Journal of Business & Economic Statistics*, *31*(1), 45–56.

Hamilton, J. D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica*, *57*(2), 357-384.

Hamilton, J. D. (2019). *The Econbrowser Recession Indicator Index.* Retrieved 10-12-2018, from `http://econbrowser.com/recession-index`

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning* (Vol. 2). Springer, New York.

Hill, R. C., Griffiths, W. E., Lim, G. C., & Lim, M. A. (2011). *Principles of Econometrics* (Fourth Edition ed.). Wiley Hoboken, NJ.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning* (Vol. 112). Springer.

Khaidem, L., Saha, S., & Dey, S. R. (2016). Predicting the Direction of Stock Market Prices Using Random Forest. *Applied Mathematical Finance*, *1*, 1–20.

Kim, H. H., & Swanson, N. R. (2014). Forecasting Financial and Macroeconomic Variables Using Data Reduction Methods: New Empirical Evidence. *Journal of Econometrics*, *178*, 352–367.

Kim, H. H., & Swanson, N. R. (2018). Mining Big Data using Parsimonious Factor, Machine Learning, Variable Selection and Shrinkage Methods. *International Journal of Forecasting*, *34*(2), 339–354.

Kim, K., & Swanson, N. R. (2016). Mixing Mixed Frequency and Diffusion Indices in Good Times and in Bad.

Kock, A. B., & Teräsvirta, T. (2011). Forecasting with Nonlinear Time Series Models.

Oxford Handbook of Economic Forecasting, 61–87.

Kock, A. B., & Teräsvirta, T. (2016). Forecasting Macroeconomic Variables using Neural Network Models and Three Automated Model Selection Techniques. *Econometric Reviews*, *35*(8-10), 1753–1779.

Kopf, J., Augustin, T., & Strobl, C. (2013). The Potential of Model-Based Recursive Partitioning in the Social Sciences: Revisiting Ockham's Razor. In *Contemporary Issues in Exploratory Data Mining in the Behavioral Sciences* (pp. 97–117). Routledge.

Korobilis, D. (2017). *Forecasting with many Predictors using Message Passing Algorithms.* Retrieved 15-12-2018, from `https://ssrn.com/abstract=2977838`

Kuan, C.-M. (2002). *Lecture on the Markov Switching Model* (Tech. Rep.).

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep Learning. *Nature*, *521*(7553), 436.

McAfee, A., & Brynjolfsson, E. (2017). *Machine, Platform, Crowd: Harnessing our Digital Future.* WW Norton & Company.

McCracken, M. W. (2007). Asymptotics for Out of Sample Tests of Granger Causality. *Journal of Econometrics*, *140*(2), 719–752.

McCracken, M. W. (2019). *FRED-MD: A Monthly Database for Macroeconomic Research.* Federal Reserve Bank of St. Louis. Retrieved 12-03-2019, from `https://research.stlouisfed.org/econ/mccracken/fred-databases/`

McCracken, M. W., & Ng, S. (2016). FRED-MD: A Monthly Database for Macroeconomic Research. *Journal of Business & Economic Statistics*, *34*(4), 574–589.

McCracken, M. W., & Ng, S. (2019a). *FRED-MD: A Monthly Database for Macroeconomic Research.* Federal Reserve Bank of St. Louis. Retrieved 12-03-2019, from `https://s3.amazonaws.com/files.fred.stlouisfed.org/fred-md/Appendix_Tables_Update.pdf` (Online Appendix for FRED-MD)

McCracken, M. W., & Ng, S. (2019b). *FRED-QD: A Quarterly Database for Macroeconomic Research.* Federal Reserve Bank of St. Louis. Retrieved 12-03-2019, from `https://s3.amazonaws.com/files.fred.stlouisfed.org/fred-md/FRED-QD_appendix.pdf` (Online Appendix for FRED-QD)

Medeiros, M. C., Vasconcelos, G. F., Veiga, Á., & Zilberman, E. (2019). Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods. *Journal of Business & Economic Statistics*, 1–45.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The Ethics of Algorithms: Mapping the Debate. *Big Data & Society*, *3*(2), 1–21.

Mittelstadt, B. D., & Floridi, L. (2016). The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts. *Science and Engineering Ethics*, *22*(2), 303–341.

Mukherjee, S. (2017). AI versus MD - What Happens when Diagnosis is Automated? *The New Yorker*, *3*. (Annals of Medicine)

Mullainathan, S., & Spiess, J. (2017). Machine Learning: An Applied Econometric Approach. *Journal of Economic Perspectives*, *31*(2), 87–106.

Nyblom, J. (1989). Testing for the Constancy of Parameters over Time. *Journal of the American Statistical Association*, *84*(405), 223–230.

Pesaran, M. H., & Timmermann, A. (2007). Selection of Estimation Window in the Presence of Breaks. *Journal of Econometrics*, *137*(1), 134–161.

Politis, D. N., & Romano, J. P. (1994). The Stationary Bootstrap. *Journal of the American Statistical Association*, *89*(428), 1303–1313.

Racine, J. (2000). Consistent Cross-Validatory Model-Selection for Dependent Data: hv-Block Cross-Validation. *Journal of econometrics*, *99*(1), 39–61.

Rossi, B. (2013). Advances in Forecasting under Instability. In *Handbook of economic forecasting* (Vol. 2, pp. 1203–1324). Elsevier.

Sax, C., & Steiner, P. (2013). Temporal Disaggregation of Time Series. *The R Journal*, *5*(2).

Seibold, H., Zeileis, A., & Hothorn, T. (2016). Model-based Recursive Partitioning for Subgroup Analyses. *The International Journal of Biostatistics*, *12*(1), 45–63.

Siliverstovs, B. (2017). Dissecting Models' Forecasting Performance. *Economic Modelling*, *67*, 294–299.

Siliverstovs, B. (2019). Assessing Nowcast Accuracy of US GDP Growth in Real Time: The Role of Booms and Busts. *Empirical Economics*. (Forthcoming)

Siliverstovs, B., & Wochner, D. (2018). Google Trends and Reality: Do the Proportions Match? Appraising the Informational Value of Online Search Behavior: Evidence from Swiss Tourism Regions. *Journal of Economic Behavior & Organization*, *145*, 1–23.

Siliverstovs, B., & Wochner, D. (2019). Recessions as Breadwinner for Forecasters — State-Dependent Evaluation of Predictive Ability: Evidence from Big Macroeconomic US Data. *KOF Working Paper*(463).

SPF. (2018). *Survey of professional forecasters: Documentation* (Tech. Rep.). Retrieved 04-11-2019, from `https://www.philadelphiafed.org/-/media/research-and-data/real-time-center/survey-of-professional-forecasters/spf-documentation.pdf?la=en`

SPF. (2019). *Probability variables: Survey of professional forecasters.* Retrieved 04-11-2019, from `https://www.philadelphiafed.org/research-and-data/real-time-center/survey-of-professional-forecasters/historical-data/probability-variables`

Stock, J. H., & Watson, M. (2009). Forecasting in Dynamic Factor Models Subject to Structural Instability. *The Methodology and Practice of Econometrics. A Festschrift in Honour of David F. Hendry*, *173*, 205.

Stock, J. H., & Watson, M. W. (2006). Forecasting with Many Predictors. *Handbook of Economic Forecasting*, *1*, 515–554.

Stock, J. H., & Watson, M. W. (2011). Dynamic Factor Models. In *The Oxford Handbook of Economic Forecasting* (pp. 35–59). Oxford University Press.

Stock, J. H., & Watson, M. W. (2012). Generalized Shrinkage Methods for Forecasting using Many Predictors. *Journal of Business & Economic Statistics*, *30*(4), 481–493.

Stock, J. H., & Watson, M. W. (2016). Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics. *Handbook of Macroeconomics*, *2*, 415–525.

Stock, J. H., & Watson, M. W. (2017). Twenty Years of Time Series Econometrics in Ten Pictures. *Journal of Economic Perspectives*, *31*(2), 59–86.

Strobl, C., Wickelmaier, F., & Zeileis, A. (2011). Accounting for Individual Differ-

ences in Bradley-Terry Models by Means of Recursive Partitioning. *Journal of Educational and Behavioral Statistics*, *36*(2), 135–153.

Trapletti, A., Hornik, K., & LeBaron, B. (2019). tseries: Time Series Analysis and Computational Finance. Retrieved 2019-10-28, from `https://cran.r-project.org/web/packages/tseries/index.html`

Varian, H. R. (2014). Big Data: New Tricks for Econometrics. *The Journal of Economic Perspectives*, *28*(2), 3–27.

Varian, H. R. (2016). Causal Inference in Economics and Marketing. *Proceedings of the National Academy of Sciences*, *113*(27), 7310–7315.

Veltri, G. A. (2017). Big Data is Not Only About Data: The Two Cultures of Modelling. *Big Data & Society*, *4*(1), 1–6.

Welch, I., & Goyal, A. (2008). A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. *The Review of Financial Studies*, *21*(4), 1455–1508.

Wochner, D. (2018). Reducing Dimensionality for Economic Time Series Forecasting: A Double Predictor Targeting Strategy.

Yousuf, K. (2019). *Essays in High Dimensional Time Series Analysis* (Unpublished doctoral dissertation). Columbia University.

Zeileis, A. (2005). A Unified Approach to Structural Change Tests based on ML Scores, F Statistics, and OLS Residuals. *Econometric Reviews*, *24*(4), 445–466.

Zeileis, A., & Hornik, K. (2007). Generalized M-Fluctuation Tests for Parameter Instability. *Statistica Neerlandica*, *61*(4), 488–508.

Zeileis, A., & Hothorn, T. (2015). Parties, Models, Mobsters: A New Implementation of Model-Based Recursive Partitioning in R. Retrieved 2019-10-15, from `https://cran.rstudio.org/web/packages/partykit/vignettes/mob.pdf` (R Package vignette)

Zeileis, A., Hothorn, T., & Hornik, K. (2008). Model-based Recursive Partitioning. *Journal of Computational and Graphical Statistics*, *17*(2), 492–514.

A. Appendix

	Benchmarks		**Dynamic Factor Trees**		**Dynamic Factor Forests**	
Horizon	Results		Models	rRMSFE	Models	rRMSFE
h=1	HMN	Factors	OS-DFM(F)	*0.892*	BS-DFM(F)	*0.890*
	1.063		OS-DFM-RP(F)	0.864***	BS-DFM-RP(F)	0.863***
	AR4		DFT-NUM(F)	**0.837****	DFF-NUM(F)	**0.834****
	0.983		DFT-BIN50(F)	**0.840****	DFF-BIN50(F)	**0.837****
	ARL	Trgt. Fact.	OS-DFM(TF)	*0.867*	BS-DFM(TF)	*0.866*
	1.003		OS-DFM-RP(TF)	0.852*	BS-DFM-RP(TF)	0.851*
	CADL		DFT-NUM(TF)	**0.823****	DFF-NUM(TF)	**0.818****
	0.931		DFT-BIN50(TF)	**0.826****	DFF-BIN50(TF)	**0.824****
h=3	HMN	Factors	OS-DFM(F)	*0.883*	BS-DFM(F)	*0.880*
	1.078		OS-DFM-RP(F)	0.773***	BS-DFM-RP(F)	0.772***
	AR4		DFT-NUM(F)	**0.792****	DFF-NUM(F)	**0.779****
	0.989		DFT-BIN50(F)	**0.785****	DFF-BIN50(F)	**0.782****
	ARL	Trgt. Fact.	OS-DFM(TF)	*0.864*	BS-DFM(TF)	*0.862*
	1.010		OS-DFM-RP(TF)	0.766***	BS-DFM-RP(TF)	0.765***
	CADL		DFT-NUM(TF)	**0.783****	DFF-NUM(TF)	**0.767****
	0.932		DFT-BIN50(TF)	**0.780****	DFF-BIN50(TF)	**0.778****

Notes: The table entries show the relative RMSFE of a particular model against the AR2 benchmark (settings: recursive scheme; CLU interpolation; Jan. 1998 first vintage; MCJH-REC-IDX partitioning variable). For more details, see notes in Table 1.

Table A.1: Robustness Results: Chow-Lin (CLU) Interpolation

	Benchmarks		**Dynamic Factor Trees**		**Dynamic Factor Forests**	
Horizon	Results		Models	rRMSFE	Models	rRMSFE
h=1	HMN	Factors	OS-DFM(F)	*0.943*	BS-DFM(F)	*0.942*
	1.004		OS-DFM-RP(F)	0.918**	BS-DFM-RP(F)	0.918**
	AR4		DFT-NUM(F)	**0.918****	DFF-NUM(F)	**0.908****
	0.990		DFT-BIN50(F)	**0.928****	DFF-BIN50(F)	**0.914****
	ARL	Trgt. Fact.	OS-DFM(TF)	*0.927*	BS-DFM(TF)	*0.926*
	0.998		OS-DFM-RP(TF)	0.905**	BS-DFM-RP(TF)	0.904**
	CADL		DFT-NUM(TF)	**0.900****	DFF-NUM(TF)	**0.893****
	0.971		DFT-BIN50(TF)	**0.919***	DFF-BIN50(TF)	**0.898****
h=3	HMN	Factors	OS-DFM(F)	*0.930*	BS-DFM(F)	*0.928*
	1.014		OS-DFM-RP(F)	0.835***	BS-DFM-RP(F)	0.835***
	AR4		DFT-NUM(F)	**0.865****	DFF-NUM(F)	**0.849****
	0.980		DFT-BIN50(F)	**0.865****	DFF-BIN50(F)	**0.861****
	ARL	Trgt. Fact.	OS-DFM(TF)	*0.931*	BS-DFM(TF)	*0.930*
	1.012		OS-DFM-RP(TF)	0.843***	BS-DFM-RP(TF)	0.844***
	CADL		DFT-NUM(TF)	**0.904****	DFF-NUM(TF)	**0.867****
	0.956		DFT-BIN50(TF)	**0.873****	DFF-BIN50(TF)	**0.871****

Notes: The table entries show the relative RMSFE of a particular model against the AR2 benchmark (settings: recursive scheme; CL3 interpolation; Jan. 1998 first vintage; MCJH-REC-IDX partitioning variable). For more details, see notes in Table 1.

Table A.2: Robustness Results: Chow-Lin (CL3) Interpolation

	Benchmarks		**Dynamic Factor Trees**		**Dynamic Factor Forests**	
Horizon	Results		Models	rRMSFE	Models	rRMSFE
h=0	HMN	Factors	OS-DFM(F)	*0.739*	BS-DFM(F)	*0.744*
	1.115		OS-DFM-RP(F)	0.721	BS-DFM-RP(F)	0.723*
	AR4		DFT-NUM(F)	**0.738**	DFF-NUM(F)	**0.690*****
	1.004		DFT-BIN50(F)	**0.738**	DFF-BIN50(F)	**0.709*****
	ARL	Trgt. Fact.	OS-DFM(TF)	*0.754*	BS-DFM(TF)	*0.762*
	1.025		OS-DFM-RP(TF)	0.727*	BS-DFM-RP(TF)	0.732*
	CADL		DFT-NUM(TF)	0.755	DFF-NUM(TF)	**0.708*****
	0.869		DFT-BIN50(TF)	0.764	DFF-BIN50(TF)	**0.725*****
h=1	HMN	Factors	OS-DFM(F)	*0.973*	BS-DFM(F)	*0.977*
	1.060		OS-DFM-RP(F)	0.855***	BS-DFM-RP(F)	0.861***
	AR4		DFT-NUM(F)	**0.916*****	DFF-NUM(F)	**0.924*****
	1.006		DFT-BIN50(F)	**0.922*****	DFF-BIN50(F)	**0.929*****
	ARL	Trgt. Fact.	OS-DFM(TF)	*0.866*	BS-DFM(TF)	*0.874*
	1.002		OS-DFM-RP(TF)	0.792***	BS-DFM-RP(TF)	0.799***
	CADL		DFT-NUM(TF)	**0.826****	DFF-NUM(TF)	0.886
	0.945		DFT-BIN50(TF)	**0.866**	DFF-BIN50(TF)	**0.863****

Notes: The table entries show the relative RMSFE of a particular model against the AR2 benchmark (settings: recursive scheme; quarterly frequency; Q1 Jan. 1998 first vintage; MCJH-REC-IDX partitioning variable). The results for the predictions in the 3rd month of every quarter are shown. The DM-tests require $h = 1$ for both forecasting horizons and were derived accordingly. For more details, see notes in Table 1.

Table A.3: Robustness Results: Quarterly Frequency

	Benchmarks		**Dynamic Factor Trees**		**Dynamic Factor Forests**	
Horizon	Results		Models	rRMSFE	Models	rRMSFE
h=1	HMN	Factors	OS-DFM(F)	*0.885*	BS-DFM(F)	*0.881*
	0.998		OS-DFM-RP(F)	0.800***	BS-DFM-RP(F)	0.797***
	AR4		DFT-NUM(F)	**0.780*****	DFF-NUM(F)	**0.772*****
	0.935		DFT-BIN50(F)	**0.789*****	DFF-BIN50(F)	**0.792*****
	ARL	Trgt. Fact.	OS-DFM(TF)	*0.814*	BS-DFM(TF)	*0.813*
	0.909		OS-DFM-RP(TF)	0.750***	BS-DFM-RP(TF)	0.747***
	CADL		DFT-NUM(TF)	**0.730*****	DFF-NUM(TF)	**0.719*****
	0.938		DFT-BIN50(TF)	**0.736*****	DFF-BIN50(TF)	**0.739*****
h=3	HMN	Factors	OS-DFM(F)	*0.964*	BS-DFM(F)	*0.956*
	1.020		OS-DFM-RP(F)	0.806***	BS-DFM-RP(F)	0.799***
	AR4		DFT-NUM(F)	**0.860*****	DFF-NUM(F)	**0.824*****
	0.973		DFT-BIN50(F)	**0.844*****	DFF-BIN50(F)	**0.828*****
	ARL	Trgt. Fact.	OS-DFM(TF)	*0.940*	BS-DFM(TF)	*0.933*
	0.980		OS-DFM-RP(TF)	0.767***	BS-DFM-RP(TF)	0.764***
	CADL		DFT-NUM(TF)	**0.752*****	DFF-NUM(TF)	**0.748*****
	0.960		DFT-BIN50(TF)	**0.819*****	DFF-BIN50(TF)	**0.823*****

Notes: The table entries show the relative RMSFE of a particular model against the AR2 benchmark (settings: rolling scheme; DCO interpolation; Jan. 1998 first vintage; MCJH-REC-IDX partitioning variable). The rolling window has a length of 300 observations. For more details, see notes in Table 1.

Table A.4: Robustness Results: Rolling Windows

	Benchmarks		**Dynamic Factor Trees**		**Dynamic Factor Forests**	
Horizon	Results		Models	rRMSFE	Models	rRMSFE
h=1	HMN	Factors	OS-DFM(F)	*0.865*	BS-DFM(F)	*0.864*
	0.998		OS-DFM-RP(F)	0.803***	BS-DFM-RP(F)	0.803***
	AR4		DFT-NUM(F)	**0.830****	DFF-NUM(F)	**0.820****
	0.927		DFT-BIN50(F)	**0.828****	DFF-BIN50(F)	**0.819****
	ARL	Trgt. Fact.	OS-DFM(TF)	*0.831*	BS-DFM(TF)	*0.830*
	0.894		OS-DFM-RP(TF)	0.780***	BS-DFM-RP(TF)	0.781***
	CADL		DFT-NUM(TF)	**0.806****	DFF-NUM(TF)	**0.794****
	0.924		DFT-BIN50(TF)	**0.810****	DFF-BIN50(TF)	**0.797****
h=3	HMN	Factors	OS-DFM(F)	*0.947*	BS-DFM(F)	*0.943*
	1.015		OS-DFM-RP(F)	0.818***	BS-DFM-RP(F)	0.818***
	AR4		DFT-NUM(F)	**0.889****	DFF-NUM(F)	**0.873****
	0.975		DFT-BIN50(F)	**0.888****	DFF-BIN50(F)	**0.882****
	ARL	Trgt. Fact.	OS-DFM(TF)	*0.931*	BS-DFM(TF)	*0.926*
	0.987		OS-DFM-RP(TF)	0.807***	BS-DFM-RP(TF)	0.806***
	CADL		DFT-NUM(TF)	**0.880****	DFF-NUM(TF)	**0.867****
	0.944		DFT-BIN50(TF)	**0.883****	DFF-BIN50(TF)	**0.871****

Notes: The table entries show the relative RMSFE of a particular model against the AR2 benchmark (settings: recursive scheme; DCO interpolation; Jan. 1985 first vintage; MCJH-REC-IDX partitioning variable). For more details, see notes in Table 1.

Table A.5: Robustness Results: Extended Forecasting Window (1985-2018)

	Benchmarks		**Dynamic Factor Trees**		**Dynamic Factor Forests**	
Horizon	Results		Models	rRMSFE	Models	rRMSFE
h=1	HMN	Factors	OS-DFM(F)	*0.841*	BS-DFM(F)	*0.839*
	1.001		OS-DFM-RP(F)	0.813***	BS-DFM-RP(F)	0.813***
	AR4		DFT-NUM(F)	**0.835**	DFF-NUM(F)	**0.832**
	0.924		DFT-BIN50(F)	**0.834**	DFF-BIN50(F)	**0.826****
	ARL	Trgt. Fact.	OS-DFM(TF)	*0.802*	BS-DFM(TF)	*0.800*
	0.894		OS-DFM-RP(TF)	0.779***	BS-DFM-RP(TF)	0.779***
	CADL		DFT-NUM(TF)	**0.777****	DFF-NUM(TF)	**0.776****
	0.922		DFT-BIN50(TF)	**0.776****	DFF-BIN50(TF)	**0.772****
h=3	HMN	Factors	OS-DFM(F)	*0.914*	BS-DFM(F)	*0.909*
	1.014		OS-DFM-RP(F)	0.908**	BS-DFM-RP(F)	0.909
	AR4		DFT-NUM(F)	**0.910**	DFF-NUM(F)	**0.906**
	0.970		DFT-BIN50(F)	**0.912**	DFF-BIN50(F)	**0.907**
	ARL	Trgt. Fact.	OS-DFM(TF)	*0.894*	BS-DFM(TF)	*0.892*
	0.975		OS-DFM-RP(TF)	0.886**	BS-DFM-RP(TF)	0.887**
	CADL		DFT-NUM(TF)	**0.881****	DFF-NUM(TF)	**0.884****
	0.943		DFT-BIN50(TF)	**0.884****	DFF-BIN50(TF)	**0.885****

Notes: The table entries show the relative RMSFE of a particular model against the AR2 benchmark (settings: recursive scheme; DCO interpolation; Jan. 1998 first vintage; SPF-REC1-IDX partitioning variable). For more details, see notes in Table 1.

Table A.6: Robustness Results: Recession Probability Index (SPF-REC1-IDX)

	Benchmarks		**Dynamic Factor Trees**		**Dynamic Factor Forests**	
Horizon	Results		Models	rRMSFE	Models	rRMSFE
h=1	HMN	Factors	OS-DFM(F)	*0.841*	BS-DFM(F)	*0.839*
	1.001		OS-DFM-RP(F)	0.806***	BS-DFM-RP(F)	0.809***
	AR4		DFT-NUM(F)	**0.797****	DFF-NUM(F)	**0.794****
	0.924		DFT-BIN50(F)	**0.798****	DFF-BIN50(F)	**0.803****
	ARL	Trgt. Fact.	OS-DFM(TF)	*0.802*	BS-DFM(TF)	*0.800*
	0.894		OS-DFM-RP(TF)	0.767***	BS-DFM-RP(TF)	0.770***
	CADL		DFT-NUM(TF)	**0.773****	DFF-NUM(TF)	**0.758****
	0.922		DFT-BIN50(TF)	**0.759****	DFF-BIN50(TF)	**0.761****
h=3	HMN	Factors	OS-DFM(F)	*0.914*	BS-DFM(F)	*0.909*
	1.014		OS-DFM-RP(F)	0.868**	BS-DFM-RP(F)	0.870**
	AR4		DFT-NUM(F)	**0.863****	DFF-NUM(F)	**0.871****
	0.970		DFT-BIN50(F)	**0.886***	DFF-BIN50(F)	**0.885***
	ARL	Trgt. Fact.	OS-DFM(TF)	*0.894*	BS-DFM(TF)	*0.892*
	0.975		OS-DFM-RP(TF)	0.849**	BS-DFM-RP(TF)	0.851**
	CADL		DFT-NUM(TF)	**0.882***	DFF-NUM(TF)	**0.876***
	0.943		DFT-BIN50(TF)	**0.868****	DFF-BIN50(TF)	**0.869***

Notes: The table entries show the relative RMSFE of a particular model against the AR2 benchmark (settings: recursive scheme; DCO interpolation; Jan. 1998 first vintage; SPF-REC2-IDX partitioning variable). For more details, see notes in Table 1.

Table A.7: Robustness Results: Recession Probability Index (SPF-REC2-IDX)

	Benchmarks		**Dynamic Factor Trees**		**Dynamic Factor Forests**	
Horizon	Results		Models	rRMSFE	Models	rRMSFE
h=1	HMN	Factors	OS-DFM(F)	*0.841*	BS-DFM(F)	*0.839*
	1.001		OS-DFM-RP(F)	0.773***	BS-DFM-RP(F)	0.772***
	AR4		DFT-NUM(F)	**0.777****	DFF-NUM(F)	**0.752****
	0.924		DFT-BIN50(F)	**0.763****	DFF-BIN50(F)	**0.759****
	ARL	Trgt. Fact.	OS-DFM(TF)	*0.802*	BS-DFM(TF)	*0.800*
	0.894		OS-DFM-RP(TF)	0.755***	BS-DFM-RP(TF)	0.754***
	CADL		DFT-NUM(TF)	**0.756****	DFF-NUM(TF)	**0.729****
	0.922		DFT-BIN50(TF)	**0.739****	DFF-BIN50(TF)	**0.734****
h=3	HMN	Factors	OS-DFM(F)	*0.914*	BS-DFM(F)	*0.909*
	1.014		OS-DFM-RP(F)	0.781***	BS-DFM-RP(F)	0.780***
	AR4		DFT-NUM(F)	**0.844****	DFF-NUM(F)	**0.816****
	0.970		DFT-BIN50(F)	**0.827****	DFF-BIN50(F)	**0.821****
	ARL	Trgt. Fact.	OS-DFM(TF)	*0.894*	BS-DFM(TF)	*0.892*
	0.975		OS-DFM-RP(TF)	0.775***	BS-DFM-RP(TF)	0.774***
	CADL		DFT-NUM(TF)	**0.853****	DFF-NUM(TF)	**0.803****
	0.943		DFT-BIN50(TF)	**0.819****	DFF-BIN50(TF)	**0.812****

Notes: The table entries show the relative RMSFE of a particular model against the AR2 benchmark (settings: recursive scheme; DCO interpolation; Jan. 1998 first vintage; MCJH-REC-IDX partitioning variable). The hyper-parameter for minsize was not determined via cross-validation. For more details, see notes in Table 1.

Table A.8: Robustness Results: Non-cross-validated Minsize

Horizon	Benchmarks		Dynamic Factor Trees		Dynamic Factor Forests	
	Results		Models	rRMSFE	Models	rRMSFE
h=1	HMN	Factors	OS-DFM(F)	*0.873*	BS-DFM(F)	*0.872*
	1.001		OS-DFM-RP(F)	0.806***	BS-DFM-RP(F)	0.806***
	AR4		DFT-NUM(F)	**0.784****	DFF-NUM(F)	**0.772****
	0.924		DFT-BIN50(F)	**0.791****	DFF-BIN50(F)	**0.785****
	ARL	Trgt. Fact.	OS-DFM(TF)	*0.895*	BS-DFM(TF)	*0.896*
	0.894		OS-DFM-RP(TF)	0.827***	BS-DFM-RP(TF)	0.828***
	CADL		DFT-NUM(TF)	**0.798****	DFF-NUM(TF)	**0.796****
	0.922		DFT-BIN50(TF)	**0.808****	DFF-BIN50(TF)	**0.803****
h=3	HMN	Factors	OS-DFM(F)	*0.897*	BS-DFM(F)	*0.879*
	1.014		OS-DFM-RP(F)	0.753***	BS-DFM-RP(F)	0.753***
	AR4		DFT-NUM(F)	**0.808****	DFF-NUM(F)	**0.781****
	0.970		DFT-BIN50(F)	**0.796****	DFF-BIN50(F)	**0.791****
	ARL	Trgt. Fact.	OS-DFM(TF)	*0.915*	BS-DFM(TF)	*0.908*
	0.975		OS-DFM-RP(TF)	0.777***	BS-DFM-RP(TF)	0.779***
	CADL		DFT-NUM(TF)	**0.817****	DFF-NUM(TF)	**0.803****
	0.943		DFT-BIN50(TF)	**0.811****	DFF-BIN50(TF)	**0.810****

Notes: The table entries show the relative RMSFE of a particular model against the AR2 benchmark (settings: recursive scheme; DCO interpolation; Jan. 1998 first vintage; MCJH-REC-IDX partitioning variable). The DFMs, DFTs and DFFs were estimated with 1 factor. For more details, see notes in Table 1.

Table A.9: Robustness Results: One Factor

Horizon	Benchmarks		Dynamic Factor Trees		Dynamic Factor Forests	
	Results		Models	rRMSFE	Models	rRMSFE
h=1	HMN	Factors	OS-DFM(F)	*0.767*	BS-DFM(F)	*0.765*
	1.001		OS-DFM-RP(F)	0.738***	BS-DFM-RP(F)	0.735***
	AR4		DFT-NUM(F)	**0.720****	DFF-NUM(F)	**0.710****
	0.924		DFT-BIN50(F)	**0.727****	DFF-BIN50(F)	**0.719****
	ARL	Trgt. Fact.	OS-DFM(TF)	*0.772*	BS-DFM(TF)	*0.773*
	0.894		OS-DFM-RP(TF)	0.748**	BS-DFM-RP(TF)	0.748**
	CADL		DFT-NUM(TF)	**0.715****	DFF-NUM(TF)	**0.708****
	0.922		DFT-BIN50(TF)	**0.723****	DFF-BIN50(TF)	**0.719****
h=3	HMN	Factors	OS-DFM(F)	*0.838*	BS-DFM(F)	*0.837*
	1.014		OS-DFM-RP(F)	0.746***	BS-DFM-RP(F)	0.746***
	AR4		DFT-NUM(F)	**0.763****	DFF-NUM(F)	**0.763****
	0.970		DFT-BIN50(F)	**0.778****	DFF-BIN50(F)	**0.773****
	ARL	Trgt. Fact.	OS-DFM(TF)	*0.843*	BS-DFM(TF)	*0.843*
	0.975		OS-DFM-RP(TF)	0.749***	BS-DFM-RP(TF)	0.749***
	CADL		DFT-NUM(TF)	**0.772****	DFF-NUM(TF)	**0.757****
	0.943		DFT-BIN50(TF)	**0.776****	DFF-BIN50(TF)	**0.771****

Notes: The table entries show the relative RMSFE of a particular model against the AR2 benchmark (settings: recursive scheme; DCO interpolation; Jan. 1998 first vintage; MCJH-REC-IDX partitioning variable). The DFMs, DFTs and DFFs were estimated with 10 factors. For more details, see notes in Table 1.

Table A.10: Robustness Results: Ten Factors

	Benchmarks		**Dynamic Factor Trees**		**Dynamic Factor Forests**	
Horizon	Results		Models	rRMSFE	Models	rRMSFE
h=1	HMN	Factors	OS-DFM(F)	*0.841*	BS-DFM(F)	*0.839*
	1.001		OS-DFM-RP(F)	0.806***	BS-DFM-RP(F)	0.804***
	AR4		DFT-NUM(F)	**0.818****	DFF-NUM(F)	**0.802*****
	0.924		DFT-BIN50(F)	**0.813****	DFF-BIN50(F)	**0.808*****
	ARL	Trgt. Fact.	OS-DFM(TF)	*0.802*	BS-DFM(TF)	*0.800*
	0.894		OS-DFM-RP(TF)	0.784*	BS-DFM-RP(TF)	0.782*
	CADL		DFT-NUM(TF)	**0.794**	DFF-NUM(TF)	**0.780****
	0.922		DFT-BIN50(TF)	**0.789***	DFF-BIN50(TF)	**0.783****
h=3	HMN	Factors	OS-DFM(F)	*0.914*	BS-DFM(F)	*0.910*
	1.014		OS-DFM-RP(F)	0.898	BS-DFM-RP(F)	0.895
	AR4		DFT-NUM(F)	0.927	DFF-NUM(F)	**0.898***
	0.970		DFT-BIN50(F)	**0.912**	DFF-BIN50(F)	**0.907**
	ARL	Trgt. Fact.	OS-DFM(TF)	*0.894*	BS-DFM(TF)	*0.892*
	0.975		OS-DFM-RP(TF)	0.886	BS-DFM-RP(TF)	0.885
	CADL		DFT-NUM(TF)	0.903	DFF-NUM(TF)	**0.892**
	0.943		DFT-BIN50(TF)	0.902	DFF-BIN50(TF)	0.897

Notes: The table entries show the relative RMSFE of a particular model against the AR2 benchmark (settings: recursive scheme; DCO interpolation; Jan. 1998 first vintage; MCJH-REC-IDX partitioning variable). Assuming the absence of a publication lag of the MCJH-REC-IDX is relaxed and missing values are inferred from a regime-switching model (see Section 4.2.8). For more details, see notes in Table 1.

Table A.11: Robustness Results: Publication Lag