ARTICLE IN PRESS

International Journal of Forecasting xxx (xxxx) xxx

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Reservoir computing for macroeconomic forecasting with mixed-frequency data☆

Giovanni Ballarin [a], Petros Dellaportas [b,c], Lyudmila Grigoryeva [d,e,*], Marcel Hirt [h,1], Sophie van Huellen [f,g], Juan-Pablo Ortega [h]

[a] Department of Economics, University of Mannheim, L7 3-5, Mannheim, 68131, Germany
[b] Department of Statistical Science, UCL, Gower Str., London WC1E 6BT, UK
[c] Department of Statistics, Athens University of Economics and Business, 10434 Athens, Greece
[d] Faculty of Mathematics and Statistics, University of St. Gallen, Bodanstrasse 6, CH-9000 St. Gallen, Switzerland
[e] Department of Statistics, University of Warwick, Coventry CV4 7AL, UK
[f] Global Development Institute (GDI), University of Manchester, Manchester M13 9PL, UK
[g] Department of Economics, SOAS University of London, London WC1H 0XG, UK
[h] Division of Mathematical Sciences, Nanyang Technological University, 21 Nanyang Link, 637371, Singapore

ARTICLE INFO

Keywords:
Reservoir computing
Echo state networks
Forecasting
U.S. output growth
GDP
Mixed-frequency data
Time series
Multi-Frequency Echo State Network
MIDAS
DFM

ABSTRACT

Macroeconomic forecasting has recently started embracing techniques that can deal with large-scale datasets and series with unequal release periods. Mixed-data sampling (MIDAS) and dynamic factor models (DFMs) are the two main state-of-the-art approaches to modeling series with non-homogeneous frequencies. We introduce a new framework, called the multi-frequency echo state network (MFESN), based on a relatively novel machine learning paradigm called reservoir computing. Echo state networks (ESNs) are recurrent neural networks formulated as nonlinear state-space systems with random state coefficients where only the observation map is subject to estimation. MFESNs are considerably more efficient than DFMs and can incorporate many series, as opposed to MIDAS models, which are prone to the curse of dimensionality. All methods are compared in extensive multistep forecasting exercises targeting U.S. GDP growth. We find that our MFESN models achieve superior or comparable performance over MIDAS and DFMs at a much lower computational cost.

☆ The authors acknowledge financial support from UK Research and Innovation (grant number ES/V006347/1). GB thanks the hospitality and the generosity of the University of Warwick, without which this paper would not have happened. GB is grateful to the Center for Doctoral Studies in Economics (CDSE) at the University of Mannheim for providing travel funding through the 2020 CDSE Teaching Award. The authors acknowledge support by the state of Baden-Württemberg through bwHPC. The authors acknowledge the use of the UCL Myriad High Performance Computing Facility (Myriad@UCL), and associated support services, in completing this work. MH acknowledges that the majority of his work was done when he was affiliated with UCL and hence thanks the support of UCL in this project. The authors acknowledge the support of the Swiss National Science Foundation (grant number 200021_175801/1). The authors also thank Thanos Moraitis and Alfonso Silva Ruiz for excellent research assistance and help with data collection and curation.

* Corresponding author at: Faculty of Mathematics and Statistics, University of St. Gallen, Bodanstrasse 6, CH-9000 St. Gallen, Switzerland.

E-mail addresses: Giovanni.Ballarin@gess.uni-mannheim.de (G. Ballarin), P.Dellaportas@ucl.ac.uk, Petros@aueb.gr (P. Dellaportas), Lyudmila.Grigoryeva@unisg.ch, Lyudmila.Grigoryeva@warwick.ac.uk (L. Grigoryeva), MarcelAndre.Hirt@ntu.edu.sg (M. Hirt), Sophie.vanHuellen@manchester.ac.uk, Sv8@soas.ac.uk (S. van Huellen), Juan-Pablo.Ortega@ntu.edu.sg (J.-P. Ortega).

[1] Some of the work done while at UCL.

https://doi.org/10.1016/j.ijforecast.2023.10.009

Please cite this article as: G. Ballarin, P. Dellaportas, L. Grigoryeva et al., Reservoir computing for macroeconomic forecasting with mixed-frequency data. International Journal of Forecasting (2023), https://doi.org/10.1016/j.ijforecast.2023.10.009.

1. Introduction

The availability of timely and accurate forecasts of key macroeconomic variables is of crucial importance to economic policymakers, businesses, and the banking sector alike. Fundamental macroeconomic figures, such as GDP growth, become available at low frequency with a considerable time lag and are subject to various rounds of revisions after their release. This is particularly problematic in a fast-changing and uncertain economic environment, as experienced during the Great Recession of 2007–2008 (Hindrayanto, Koopman, & de Winter, 2016) and the recent pandemic (Buell et al., 2021; Huber, Koop, Onorante, Pfarrhofer, & Schreiner, 2021). However, a large number of the potentially predictive financial market (and other macroeconomic) indicators are available at a daily or even higher frequency (Andreou, Ghysels, & Kourtellos, 2013). The desire to utilize such high-frequency data for macroeconomic forecasting has led to the exploration of techniques that can deal with large-scale datasets and series with unequal release periods (see Borio (2011, 2013), Morley (2015); we also refer the reader to Fuleky (2020) for more details regarding high-dimensional data, and to Armesto, Engemann, and Owyang (2010) and Bańbura, Giannone, Modugno, and Reichlin (2013) for a review on mixed-frequency data).

We contribute to the existing literature by proposing a new macroeconomic forecasting framework that utilizes high-dimensional and mixed-frequency input data: the multi-frequency echo state network (MFESN). The MFESN originates from a machine learning paradigm called reservoir computing (RC). RC is a family of learning models that take advantage of the information processing capabilities of complex dynamical systems (see Crutchfield, Ditto, and Sinha (2010), Legenstein and Maass (2007), Maass, Natschläger, and Markram (2002), and Lukoševičius and Jaeger (2009), Tanaka et al. (2019) for reviews). Generally speaking, RC is a versatile class of recurrent neural network (RNN) models (see Salehinejad et al. (2017) for a detailed survey). Although conventional RNNs are well suited for handling sequence data and dynamic problems, estimating their weights during the training phase is inherently difficult (Doya, 1992; Pascanu, Mikolov, & Bengio, 2013). Reservoir networks stand out due to the fact that their inner weights can be randomly generated and fixed, and only the output (readout) layer weights are subject to estimation (supervised training). An echo state network (ESN) is one of the most popular instances of RC models, with provable universality, generalization properties (see Gonon, Grigoryeva, and Ortega (2020b, 2023a), Gonon and Ortega (2021), Grigoryeva and Ortega (2018a, 2018b, 2019), and the references therein for more details), and excellent performance at forecasting, classifying, and learning dynamical systems (see Grigoryeva, Hart, and Ortega (2021), Hart, Hook, and Dawes (2021)). While conventional RNNs have been adopted for macroeconomic forecasting in a few instances (see, for example, Paranhos (2021)), to the best of our knowledge, we are the first to explore easily trainable reservoir models in this context.

Our main contribution is three-fold. First, inspired by the remarkable empirical success of ESNs in prediction tasks, we propose the multi-frequency echo state network (MFESN) framework, which allows multistep forecasting of the target variable at frequencies that are lower or the same as those of the input series. Second, we introduce two different approaches to predicting within the MFESN framework: single-reservoir MFESN (S-MFESN), and multi-reservoir MFESN (M-MFESN). S-MFESN is determined by modifying the ESN architecture to accommodate input and target variables of mixed frequencies. In M-MFESN, several ESNs are adopted to handle input time series, each ESN corresponding to a group of input variables quoted at one given frequency. Finally, we provide an extensive empirical comparative analysis of the forecasting capability of the proposed approaches in a concrete task of predicting the quarterly U.S. output growth. We inspect the forecasting capabilities of the MFESN framework compared to two well-established benchmarks widely used in the macroeconomic literature and among practitioners, and show its empirical superiority in several thoroughly conducted forecasting exercises. Moreover, as a byproduct, we propose a new data aggregation scheme that can bridge these two standard forecasting approaches, which is not available in the literature.

In our empirical study, we evaluate the multistep forecasting performance of the MFESN framework, targeting quarterly U.S. output growth (GDP growth) and utilizing a small set and a medium-sized set of monthly and daily financial and macroeconomic variables. We compare the MFESN approach against two state-of-the-art methods, mixed-data sampling (MIDAS) and the dynamic factor model (DFM), known for their ability to incorporate data of heterogeneous frequencies and utilize high-dimensional data inputs. The MIDAS model, developed in Ghysels, Santa-Clara, and Valkanov (2004), Ghysels, Sinko, and Valkanov (2007), has been widely adopted for macroeconomic forecasting with mixed-frequency data (see for instance Andreou et al. (2013), Clements and Galvão (2008, 2009), Francis, Ghysels, and Owyang (2011), Galvão (2013), Galvão and Marcellino (2010), Ghysels (2016), Ghysels and Wright (2009), Jardet and Meunier (2022), Monteforte and Moretti (2012)). However, MIDAS is prone to curse-of-dimensionality problems and performs poorly when the set of predictors is of even moderate size (Clements & Galvão, 2009; Kostrov, 2021), due to optimization-related issues. Recently, some attempts have been made in the literature to overcome these issues by employing variable selection techniques under some additional assumptions. For instance, Babii, Ghysels, and Striaukas (2022) proposes the MIDAS projection approach, which is more amenable to high-dimensional data environments under the assumption of sparsity. Even with these improvements, practical high-dimensional implementations of MIDAS remain challenging. This is in part caused by the ragged edges of the raw macroeconomic data, incomplete observations, and uneven sampling frequencies. The relative inflexibility of MIDAS regression lag specifications makes integrating daily and weekly data at true calendar frequencies (that is, without interpolation or aggregation) very complex. State-space models effectively mitigate these issues.

A strong state-of-the-art state-space competitor for our MFESN framework is the DFM, first introduced by Geweke (1977) and Sargent, Sims, et al. (1977). DFMs have become the standard workhorse for macroeconomic nowcasting and prediction (for more details, we refer the reader to Bańbura and Rünstler (2011), Chauvet, Senyuz, and Yoldas (2015), Giannone, Reichlin, and Small (2008), Hindrayanto et al. (2016), Stock and Watson (1996, 2002, 2016)). Conventional DFMs for data of multiple sampling frequencies are linear state-space models with a latent low-frequency process of interest and high-dimensional input time series. Although their linear structure lends itself to inference with likelihood-based methods and Kalman filtering, using DFMs in a high-dimensional setting is limited by the associated computational effort. For Gaussian state-space models, some of these issues can be handled with a more compact matrix representation, as in Delle Monache and Petrella (2019). Still, in the particular settings of nowcasting and forecasting GDP growth, the computational complexity is one of the main reasons why DFMs are rarely used with daily input series (see Bańbura et al. (2013)) for a detailed review and Aruoba, Diebold, and Scotti (2009) for a mixed-frequency DFM wherein the latent factor process is updated daily, with the highest input frequency being weekly. We address these numerical difficulties using novel Python libraries for auto-differentiation and using GPUs for parallel computing, such that DFMs can be estimated even in instances of high-frequency input observations. Further, to adapt the DFM to mixed-frequency tasks, we propose a new DFM aggregation scheme with an Almon polynomial structure that bridges MIDAS and the DFM for our forecasting comparison. To our knowledge, we are the first to present this aggregation scheme, which reduces the number of parameters subject to estimation. In contrast, previous DFMs, as in Bańbura and Rünstler (2011), Camacho and Pérez-Quirós (2010), Frale, Marcellino, Mazzi, and Proietti (2011), Mariano and Murasawa (2003), commonly assume a fixed aggregation scheme a priori, depending on whether the macroeconomic variable is a flow or stock variable.

To carry out a fair comparison of our MFESN framework with the state-of-the-art MIDAS and DFM models, we designed two model evaluation settings that differ regarding whether or not the financial crisis of 2007–2008 is included in the estimation period. In the first forecasting setting, all the competing models are estimated using data from January 1st, 1990, until December 31st, 2007. Their performance at forecasting into and after the financial crisis period is assessed. In the second evaluation setting, fitting is done with data largely encompassing the crisis period, again from January 1st, 1990, but now up to December 31st, 2011. In both cases, the forecasting (testing) period spans time up to the Covid-19 pandemic events, namely the fourth quarter of 2019. Along with the two state-of-the-art DFM and MIDAS models, we use the unconditional mean of the sample as a baseline benchmark against the reservoir models. We find that our ESN-inspired models attain comparable or much better performance than DFMs at a much lower computational cost, even for a relatively long forecasting horizon of four quarters. Additionally, ESNs do not suffer from curse-of-dimensionality problems, which are known to be pervasive for MIDAS models, and hence consistently outperform them in a number of forecasting exercises.

The remainder of the paper is structured as follows. Section 2 presents reservoir models and discusses their advantages, as well as estimation, hyperparameter tuning, penalization, and nonlinear multistep forecasting. In Section 3, we introduce our MFESN framework, propose the single-reservoir and multi-reservoir MFESN models, and spell out their defining features. Section 4 contains the empirical study of the comparative GDP forecasting performance of MFESNs with respect to the set of benchmark models. We assess one-step and multistep forecasting results in several setups, with a small set and a medium-sized set of regressors. We fit models with data before and after the 2007–2008 financial crisis, and with different estimation windows. Section 5 concludes and discusses future research avenues and applications. Finally, the Appendix contains information regarding data sources, forecasting figures, and formal details regarding our forecasting setups. The Supplementary Appendix, available online, provides additional figures and gives detailed information on the implementation of all models and robustness checks.

Code. Our code, the data, and all results presented in the paper are made available in the GitHub repository at github.com/RCEconModelling/Reservoir-Computing-for-Macroeconomic-Modelling.

1.1. Notation

We use the symbol $\mathbb{N}$ (respectively, $\mathbb{N}^+$) to denote the set of natural numbers with the zero element included (respectively, excluded). $\mathbb{Z}$ denotes the set of all integers. We use $\mathbb{R}$ (respectively, $\mathbb{R}_+$) to denote the set of all (respectively, positive excluding zero element) reals. We abbreviate the set $[n] = \{1, \ldots, n\}$, with $n \in \mathbb{N}^+$.

Vector notation. A column vector is denoted by a bold lowercase symbol like $\boldsymbol{r}$, and $\boldsymbol{r}^\top$ indicates its transpose. Given a vector $\boldsymbol{v} \in \mathbb{R}^n$, we denote its entries by v_i, with $i \in \{1, \ldots, n\}$; we also write $\boldsymbol{v} = (v_i)_{i\in\{1,\ldots,n\}}$. The symbols $\boldsymbol{i}_n, \boldsymbol{0}_n \in \mathbb{R}^n$ stand for vectors of length n consisting of ones and of zeros, respectively. Additionally, given $n \in \mathbb{N}^+$, $\boldsymbol{e}_n^{(i)} \in \mathbb{R}^n$, $i \in \{1, \ldots, n\}$ denotes the canonical unit vector of length n determined by $\boldsymbol{e}_n^{(i)} = (\delta_{ij})_{j\in\{1,\ldots,n\}}$. For any $\boldsymbol{v} \in \mathbb{R}^n$, $\|\boldsymbol{v}\|$ denotes its Euclidean norm.

Matrix notation. We denote by $\mathbb{M}_{n,m}$ the space of real $n \times m$ matrices with $m, n \in \mathbb{N}^+$. When $n = m$, we use the symbols $\mathbb{M}_n$ and $\mathbb{D}_n$ to refer to the space of square and diagonal matrices of order n, respectively. Given a matrix $A \in \mathbb{M}_{n,m}$, we denote its components by A_{ij} and we write $A = (A_{ij})$, with $i \in \{1, \ldots, n\}$, $j \in \{1, \ldots m\}$. The symbol $\mathbb{I}_n \in \mathbb{D}_n$ denotes the identity matrix, and the symbol $\mathbb{O}_n$ stands for the zero matrix of dimension n. For any $A \in \mathbb{M}_{n,m}$, $\|A\|_2$ denotes its matrix norm induced by the Euclidean norms in $\mathbb{R}^m$ and $\mathbb{R}^n$, and $\|A\|_2 = \sigma_{\max}(A)$, with $\sigma_{\max}(A)$ the largest singular value of A.

Input and target stochastic processes. We fix a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ on which all random variables are defined. The input and target signals are modeled by discrete-time stochastic processes $\boldsymbol{z} = (\boldsymbol{z}_t)_{t\in\mathbb{Z}}$ and $\boldsymbol{y} = (\boldsymbol{y}_t)_{t\in\mathbb{Z}}$, taking values in $\mathbb{R}^K$ and $\mathbb{R}^J$, respectively. Moreover, we write $\boldsymbol{z}(\omega) = (\boldsymbol{z}_t(\omega))_{t\in\mathbb{Z}}$ and $\boldsymbol{y}(\omega) = (\boldsymbol{y}_t(\omega))_{t\in\mathbb{Z}}$ for each outcome $\omega \in \Omega$ to denote the realizations or sample paths of $\boldsymbol{z}$ and $\boldsymbol{y}$, respectively. Since $\boldsymbol{z}$ can be seen as a random sequence in $\mathbb{R}^K$, we write $\boldsymbol{z} : \mathbb{Z}\times\Omega \longrightarrow \mathbb{R}^K$ and $\boldsymbol{z} : \Omega \longrightarrow (\mathbb{R}^K)^{\mathbb{Z}}$ interchangeably. The same applies to the analogous assignments involving $\boldsymbol{y}$.

Temporal notation. Let $(u_t)_{t\in I}$, $u_t \in \mathbb{R}$ be a (scalar) time series with I some index set (in this paper, it will always be discrete). Time series $(u_t)_{t\in I}$ will be denoted just as (u_t) when the index set I is specified by the context. We write $u_{s_1:s_2} = (u_t)_{t\in\{s_1,\dots,s_2\}}$ for integers $s_1 < s_2$ and time series (u_t). To define the concept of the sampling frequency, we must introduce an additional series, call it $(v_s)_{s\in J}$. The time index J is not the same as I. We assume that u_t is sampled at the coarsest rate; equivalently, it has the *lowest* sampling frequency, which we call the 'reference frequency' in what follows. In practice, this means that in the same window of time, u_t will be observed at most as frequently as v_s. The case when the sampling frequency of v_s is strictly higher than that of u_t is of primary interest.

We assume that all sampling happens in instants that are evenly spaced in time. Series other than the reference one and with higher sampling frequencies are given an additional time index, the tempo index, written $t, *|\kappa$, where κ is the frequency multiplier. Our tempo notation assumes that low- and high-frequency series are sampled with temporal alignment: this means that the reference time index t and the tempo index $*|\kappa$ have the following properties.

Definition 1.1. A reference time index $t \in \mathbb{N}$ and a tempo index $*|\kappa$ for a given high-frequency $\kappa \in \mathbb{N}^+$ are such that the following relations hold:

(i) $t, 0|\kappa \equiv t$

(ii) $t, \kappa|\kappa \equiv t+1$

(iii) $t, s|\kappa \equiv t + \lfloor s/\kappa \rfloor, (s \bmod \kappa)|\kappa \;\; \forall s \in \mathbb{N}$

(iv) $t, -s|\kappa \equiv (t-1) - \lfloor s/\kappa \rfloor, \kappa - (s \bmod \kappa)|\kappa \;\; \forall s \in \mathbb{N}$,

where mod is the modulo operation, and for any $x \in \mathbb{R}$, the floor operator $\lfloor x \rfloor$ outputs the greatest $z \in \mathbb{N}$ such that $z \le x$.

Since we can exchange 'frequency' and 'frequency multiplier' in the tempo notation, we make no distinction between the two terms in what follows.

Forecasting schemes. The theoretical setup and design of the forecasting exercises conducted in this paper are carefully discussed in Appendix C. There, we formally distinguish between so-called high-frequency and low-frequency forecasting in the presence of mixed-frequency data. For more details regarding time series forecasting with economic data, we refer the reader to Chen and Ghysels (2010), Clements and Galvão (2008, 2009), Jardet and Meunier (2022), and the references therein.

2. Reservoir models

In this section, we introduce reservoir computing models (Jaeger & Haas, 2004) for forecasting stochastic time series of a single frequency. We focus on a family of reservoir computing systems called echo state networks (ESNs), which have been successfully applied to forecasting deterministic dynamical systems (Arcomano et al., 2022; Pathak, Hunt, Girvan, Lu, & Ott, 2018; Pathak, Lu, Hunt, Girvan, & Ott, 2017; Wikner et al., 2021). In the following, we discuss the linear estimation of ESN model parameters, the hyperparameters tuning, the loss penalty selection, and how to carry out nonlinear forecasting.

2.1. Reservoir models

Reservoir computing (RC) models are nonlinear state-space systems that, in the forecasting setting, are defined by the following equations:

$$\boldsymbol{x}_t = F(\boldsymbol{x}_{t-1}, \boldsymbol{z}_t), \tag{2.1}$$

$$\boldsymbol{y}_{t+1} = h_{\boldsymbol{\theta}}(\boldsymbol{x}_t) + \boldsymbol{\epsilon}_t, \tag{2.2}$$

for all $t \in \mathbb{Z}$, where the state map $F : \mathbb{R}^N \times \mathbb{R}^K \to \mathbb{R}^N$, $N, K \in \mathbb{N}^+$ is called also the reservoir map, and the observation map $h_{\boldsymbol{\theta}} : \mathbb{R}^N \to \mathbb{R}^J$, $J \in \mathbb{N}^+$ is referred to as the readout layer, parameterized by $\boldsymbol{\theta} \in \Theta$. Sequences $(\boldsymbol{z}_t)_{t\in\mathbb{Z}}$, $\boldsymbol{z}_t \in \mathbb{R}^K$, and $(\boldsymbol{y}_t)_{t\in\mathbb{Z}}$, $\boldsymbol{y}_t \in \mathbb{R}^J$ stand for the input and the output (target) of the system, respectively, and $(\boldsymbol{x}_t)_{t\in\mathbb{Z}}$, $\boldsymbol{x}_t \in \mathbb{R}^N$ are the associated reservoir states. In (2.2), $(\boldsymbol{\epsilon}_t)_{t\in\mathbb{Z}}$ denotes J-dimensional independent zero-mean innovations with variance $\sigma_\epsilon^2 \mathbb{I}_J$ that are also independent of $\boldsymbol{x}_t$ across all t. Importantly, many families of RC systems have been proven to have universal approximation properties for L^p-integrable stochastic processes (Gonon & Ortega, 2020), and estimation and generalization error bounds have been established in Gonon et al. (2020b, 2023a).

In the case of an ESN model, the state and observation equations (2.1)–(2.2) are given by

$$\boldsymbol{x}_t = \alpha\boldsymbol{x}_{t-1} + (1-\alpha)\,\sigma(A\boldsymbol{x}_{t-1} + C\boldsymbol{z}_t + \boldsymbol{\zeta}), \tag{2.3}$$

$$\boldsymbol{y}_{t+1} = \boldsymbol{a} + W^\top \boldsymbol{x}_t + \boldsymbol{\epsilon}_t, \tag{2.4}$$

where $A \in \mathbb{M}_N$ is the reservoir matrix, $C \in \mathbb{M}_{N,K}$ is the input matrix, $\boldsymbol{\zeta} \in \mathbb{R}^N$ is the input shift, $\alpha \in [0, 1)$ is the leak rate, and $W \in \mathbb{M}_{N,J}$ denotes the readout coefficients. The map $\sigma : \mathbb{R} \to \mathbb{R}$ is an activation function applied elementwise, which in what follows we take to be the hyperbolic tangent. We refer to $A, C, \boldsymbol{\zeta}$ as state parameters that are randomly generated. Notice that if $A = 0$ and $\alpha = 0$, the state equation reduces to a nonlinear regression model with random coefficients (or a feedforward neural network with random weights), which is usually referred to as an extreme learning machine (Cao, Wang, Ming, & Gao, 2018; Gonon et al., 2023a).

Properties of ESN models. We focus on ESNs with the so-called echo state property (ESP), that is, when for any $\boldsymbol{z} \in (\mathbb{R}^K)^{\mathbb{Z}}$ there exists a unique $\boldsymbol{y} \in (\mathbb{R}^J)^{\mathbb{Z}}$ such that (2.3)–(2.4) hold (see Grigoryeva and Ortega (2018a, 2018b, 2019), and the references therein). One can require that

the ESP holds only on the level of the state equation. That is, for any input sequence $\boldsymbol{z} \in (\mathbb{R}^K)^{\mathbb{Z}}$, there exists a unique state sequence $\boldsymbol{x} \in (\mathbb{R}^N)^{\mathbb{Z}}$ such that (2.3) holds. The result in Corollary 3.2 in Grigoryeva and Ortega (2018a), which is also valid for the case of ESNs with the leak rate, shows that the sufficient condition of the ESP associated with (2.3) to hold is $\|A\|_2 L_\sigma < 1$, where L_σ is the Lipschitz constant of the activation function σ (in our setting, $L_{tanh} = 1$). This sufficient ESP condition has been extensively studied in the ESN literature (see Buehner and Young (2006), Jaeger (2010), Jaeger and Haas (2004), Manjunath and Jaeger (2013), Wainrib and Galtier (2016), Yildiz, Jaeger, and Kiebel (2012), Zhang, Miller, and Wang (2012)) for more details. The result in Corollary 3.2 in Grigoryeva and Ortega (2018a) also shows that this condition implies the so-called fading memory property (Boyd & Chua, 1985), which from a practical point of view means that the impact of the initial $\boldsymbol{x}_0$ is negligible for sufficiently long samples.

In the stochastic setting, part (i) of Proposition 4.2 in Grigoryeva and Ortega (2021) proves that the condition $\|A\|_2 < 1$ guarantees variance stationarity of the states associated with variance stationary inputs. Moreover, Manjunath and Ortega (2023) show that this condition implies the so-called stochastic state contractivity, ensuring a stochastic analog of the ESP. Notably, violations of $\|A\|_2 < 1$ do not have detrimental implications for the performance of ESNs in various learning tasks, as reported in multiple empirical studies.

Computational advantages of ESNs. We emphasize that the core computational advantage of ESNs is that state parameters A, C, and $\boldsymbol{\zeta}$ are randomly sampled and do not need to be estimated. Additionally, since the observation equation (2.4) is linear in $\boldsymbol{x}_t$, coefficients W can be estimated via (penalized) least squares regression, as we explain in the following subsection. The choice of the properties of state parameters determines the memory properties and forecasting performance of linear (Ballarin, Grigoryeva, & Ortega, 2023) and nonlinear ESNs (Gonon, Grigoryeva, & Ortega, 2020a), as we discuss in Section 2.2.1.

2.2. Estimation

We now discuss in detail the estimation of coefficients W in (2.4). Let a sample $(\boldsymbol{z}_t, \boldsymbol{y}_t)_{t=1}^T$ of input and target pairs be available. Given an initial state $\boldsymbol{x}_0$, the reservoir states can be computed iteratively according to state equation (2.3) as follows:

$$\boldsymbol{x}_1 = \alpha \boldsymbol{x}_0 + (1-\alpha)\sigma(A\boldsymbol{x}_0 + C\boldsymbol{z}_1 + \boldsymbol{\zeta}), \quad \ldots,$$
$$\boldsymbol{x}_T = \alpha \boldsymbol{x}_{T-1} + (1-\alpha)\sigma(A\boldsymbol{x}_{T-1} + C\boldsymbol{z}_T + \boldsymbol{\zeta}).$$

We collect the states and the targets into the state and the observation matrices, respectively, as follows:

$$X = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{T-1})^\top \in \mathbb{M}_{T-1,N},$$
$$Y = (\boldsymbol{y}_2, \boldsymbol{y}_3, \ldots, \boldsymbol{y}_T)^\top \in \mathbb{M}_{T-1,J}.$$

Consider the ridge regression estimator for W given by

$$\widehat{W}_\lambda := \underset{W \in \mathbb{R}^N}{\operatorname{argmin}} \sum_{t=1}^{T-1} \left\| \boldsymbol{y}_{t+1} - W^\top \boldsymbol{x}_t \right\|_2^2 + \lambda \|W\|_2^2$$
$$= \left(X^\top X + \lambda((T-1)\,\mathbb{I}_N)\right)^{-1} X^\top Y, \tag{2.5}$$

where $\lambda \in \mathbb{R}_+$ is the ridge penalty strength. When $\lambda \to 0$, the estimator $\widehat{W}_\lambda$ converges to the minimum-norm least squares solution (Ishwaran & Rao, 2014). In applications, ridge regression is the most commonly used estimation method applied to ESNs, as it provides a straightforward regularization scheme when both $N < T$ and $N \geq T$. This is especially important, since in practice the ESN state dimension is often chosen to be 10^3–10^4 (see for example Pathak et al., 2017). Additionally, a virtue of the ridge regression problem is the fact that the associated objective function is convex and, hence, it can be efficiently solved using stochastic gradient descent even when $\min\{N, T\}$ is large and one decides against the closed-form solution (2.5). Finally, as mentioned in the properties of reservoir systems in Section 2.1, we notice that in the presence of the fading memory property, the estimation does not depend significantly on the choice of $\boldsymbol{x}_0$ as sample size T increases.

We refer to (2.5) as the fixed-parameter estimator. In our empirical analyses, we also implement expanding and rolling window estimation strategies which update $\widehat{W}_\lambda$ as new observations become available; we refer the reader to online Supplementary Appendix D.1 for details. In the rest of the paper, for brevity, we use $\widehat{W}$ to denote the ridge estimator of coefficients W assuming that the appropriate choice of the penalty strength λ is made for each concrete situation.

2.2.1. Hyperparameter tuning

As discussed in Section 2.1, the performance of ESNs depends on the choice of randomly drawn state parameters A, C, and $\boldsymbol{\zeta}$. Much work has been put into determining optimal specifications (see for example Farkas, Bosak, & Gergel, 2016; Gonon et al., 2020a; Goudarzi et al., 2016; Grigoryeva, Henriques, Larger, & Ortega, 2015, 2016; Rodan & Tino, 2011). We construct these parameters by first sampling $\widetilde{A}$, $\widetilde{C}$, and $\widetilde{\boldsymbol{\zeta}}$ from appropriately chosen laws. Then, we normalize each element of the tuple, such that

$$\overline{A} = \widetilde{A}/\rho(\widetilde{A}), \quad \overline{C} = \widetilde{C}/\|\widetilde{C}\|, \quad \overline{\boldsymbol{\zeta}} = \widetilde{\boldsymbol{\zeta}}/\|\widetilde{\boldsymbol{\zeta}}\|, \tag{2.6}$$

where $\rho(\widetilde{A})$ denotes the spectral radius of $\widetilde{A}$. As discussed in the properties of reservoir systems in Section 2.1, the sufficient condition of the ESP is $\|A\|_2 < 1$. By this normalizing choice, we allow for some more flexibility in terms of marginal violations of the non-sharp ESP constraint. Finally, defining $A = \rho\overline{A}$, $C = \gamma\overline{C}$ and $\boldsymbol{\zeta} = \omega\overline{\boldsymbol{\zeta}}$, we can rewrite state equation (2.3) as

$$\boldsymbol{x}_t = \alpha \boldsymbol{x}_{t-1} + (1-\alpha)\sigma(\rho\overline{A}\boldsymbol{x}_{t-1} + \gamma\overline{C}\boldsymbol{z}_t + \omega\overline{\boldsymbol{\zeta}}). \tag{2.7}$$

We refer to tuple $\boldsymbol{\varphi} := (\alpha, \rho, \gamma, \omega)$ as the hyperparameters of the ESN. Specifically, $\alpha \in [0, 1)$ is the leak rate and $\rho \in \mathbb{R}_+$ is called spectral radius of the reservoir matrix. Futher, $\gamma \in \mathbb{R}_+$ is the input scaling, and $\omega \in \mathbb{R}_+$ is the shift scaling. The choice of the hyperparameters determines the properties of the state map. For simplicity, in Section 4, we choose the hyperparameters based on the empirical ESN literature. In Supplementary Appendix D.2, we also propose a general though more computationally intensive procedure to select hyperparameters in a data-driven way that could be interesting to practitioners.

2.2.2. Penalty selection

To apply ridge estimator (2.5), it is necessary to first select a penalty λ. Cross-validation (CV) is a common selection procedure for regularization strength in penalized methods such as ridge, LASSO, and elastic net. CV techniques have also been applied in the time series context (Ballarin, 2023; Kock, Medeiros, & Vasconcelos, 2020), with their validity established in Bergmeir, Hyndman, and Koo (2018).

In our empirical study, to account for temporal dependence, we use a sequential CV strategy with ten validation folds. More precisely, we reserve the last 50 observations for validation and all other previous data points for training. The first fold consists of the first five observations out of the validation set, and the model is fitted using all training data. The following validation fold comprises the next five subsequent validation observations, while the training set is expanded by five data points (from the previous fold). This procedure is repeated ten times and the CV loss is the average of the one-step-ahead forecast MSE on each fold. In expanding or rolling window setups, we rerun the CV penalty selection to ensure that estimated ESN coefficients do not induce oversmoothing. We refer the reader to Supplementary Appendix D.3 for additional details.

2.3. Relation to nonparametric regression

Together with hyperparameters and penalty strength selection, the choice of the state dimension N is a key ingredient of an ESN model. A large state space generally implies better approximation bounds (Gonon et al., 2023a; Gonon, Grigoryeva, & Ortega, 2023b). Although it is customary in the empirical literature to take N as large as possible (Lukoševičius, 2012), some recent literature discusses both the statistical risk bounds and the approximation-risk tradeoff bounds for various RC families (see Gonon et al. (2020b) and Gonon et al. (2023b) for details). Under simplified assumptions that $\alpha = 0$ and $\rho = 0$ in (2.7), ESNs have a natural connection to random-weights neural networks (Cao et al., 2018) and random projection regression (Maillard & Munos, 2012), and are thus comparable to nonparametric sieve methods. If the data were independently sampled, known results on sieve estimation would require that at most $N/T = o(1)$ up to logarithmic factors for consistency (Belloni, Chernozhukov, Chetverikov, & Kato, 2015). Chen and Christensen (2015) extended this result to β-mixing data with B-spline and wavelet sieves. Sieve rates appear to suggest that choosing $N = O(T)$ in ESNs could lead to nontrivial forecasting bias owing to poor approximation properties. Unfortunately, this comparison relies on neglecting the dynamic component of the ESN model, and as such it is only qualitative. It is, therefore, an important topic for future research.

A different but related problem is the potential degradation of forecasting performance when a model is at the interpolation threshold in the overparameterized regime, $N \geq T$. Ridge regression is also commonly applied to address generalization concerns in statistical learning (see Hastie, Tibshirani, Friedman, and Friedman (2009)). Recent work has extensively studied the link between regularization and generalization: Hastie, Montanari, Rosset, and Tibshirani (2022) show that ridgeless (that is, interpolation) solutions can be optimal in some scenarios. However, in our empirical evaluations in Section 4, cross-validation consistently selects non-zero ridge penalties, confirming that ridge penalization plays an important role in ESN forecasting performance.

2.4. ESN forecasting

We are primarily interested in using ESN models to construct conditional forecasts of target variables. Given that the conditional mean is the best mean square error estimator for the h-step-ahead target $\boldsymbol{y}_{t+h}$, $h \geq 1$, our main focus is approximating

$$\widehat{\boldsymbol{y}}_{t+h|t} \coloneqq \mathbb{E}\left[\boldsymbol{y}_{t+h}|\boldsymbol{x}_{0:t}, \boldsymbol{z}_{0:t}\right].$$

The case $h = 1$ is trivial, since the ESN model is estimated by regressing $\boldsymbol{y}_{t+1}$ on state $\boldsymbol{x}_t$, and thus we can set $\widetilde{\boldsymbol{y}}_{t+1|t} = \widehat{W}^{\top}\boldsymbol{x}_t$. However, when $h > 1$, the nonlinear state dynamics preclude a direct computation of the conditional mean. This is in contrast to linear models like VARMAs or DFMs, where the assumption of linearity implies that conditional expectations reduce to simple matrix–vector operations. In particular, linear models are such that the variance (and any other higher-order moments) of the noise term does not impact the conditional mean forecast.

Let $p_\theta(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{z}_t)$ and $g_\theta(\boldsymbol{y}_{t+1}|\boldsymbol{x}_t)$ be the state transition and observation densities, respectively. Then, for $h > 1$,

$$\widehat{\boldsymbol{y}}_{t+h|t} = \int \boldsymbol{y}_{t+h}\, g_\theta(\boldsymbol{y}_{t+h}|\boldsymbol{x}_{t+h-1}) \times \prod_{j=1}^{h-1} p_\theta(\boldsymbol{x}_{t+j}|\boldsymbol{x}_{t+j-1}, \boldsymbol{z}_{t+j})\nu(\boldsymbol{z}_{t+j}|\boldsymbol{x}_{t+j-1})\mathrm{d}\boldsymbol{z}_{t+j}\mathrm{d}\boldsymbol{x}_{t+j}\mathrm{d}\boldsymbol{y}_{t+h}, \tag{2.8}$$

where $\nu(\boldsymbol{z}_{t+j}|\boldsymbol{x}_{t+j-1})$ is the conditional density of inputs. Here, we introduce the additional assumption that $\boldsymbol{x}_{t+j-1}$ is sufficient to condition on past states and inputs. That is,

$$\nu(\boldsymbol{z}_{t+j}|\boldsymbol{x}_{t+j-1}) \equiv \nu(\boldsymbol{z}_{t+j}|\boldsymbol{x}_{0:t+j-1}, \boldsymbol{z}_{0:t+j-1}). \tag{2.9}$$

Some elements in the expectation integral are not directly available. Specifically, while an ESN explicitly models both $p_\theta(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{z}_t)$ and $g_\theta(\boldsymbol{y}_{t+1}|\boldsymbol{x}_t)$, the density $\nu(\boldsymbol{z}_{t+j}|\boldsymbol{x}_{t+j-1})$ is unavailable.

In the remaining part of this subsection, we present a novel ESN-based approach to forecasting the target variable. Our idea is to enrich the ESN model with an auxiliary observation equation for the input covariates. As we demonstrate in Section 4, our proposed method shows superior performance with respect to the standard state-of-the-art benchmarks.

2.4.1. Multi-step forecasting of targets via iterative forecasting of inputs

In general, we are interested in constructing forecasts of target variables that are not the same as the model inputs. To do so, we resolve the issue of the intractability of (2.8) while simultaneously capitalizing on the available results using ESNs to forecast dynamical systems. More explicitly, we add to the ESN specification (2.3)–(2.4) an equation that allows sidestepping modeling of the density ν directly, thus making the computation of $\widehat{\boldsymbol{y}}_{t+h|t}$ feasible even when $h > 1$.

Consider the ESN where the reservoir states $(\boldsymbol{x}_t)_{t\in\mathbb{Z}}$ follow (2.3), while the target sequence is the same as the input sequence $(\boldsymbol{z}_t)_{t\in\mathbb{Z}}$:

$$\boldsymbol{x}_t = \alpha\boldsymbol{x}_{t-1} + (1-\alpha)\sigma(A\boldsymbol{x}_{t-1} + C\boldsymbol{z}_t + \boldsymbol{\zeta}) \tag{2.10}$$

$$\boldsymbol{z}_{t+1} = \mathcal{W}^\top\boldsymbol{x}_t + \boldsymbol{u}_{t+1}. \tag{2.11}$$

Here, we use symbol $\mathcal{W}$ for the output coefficients to separate this case from the general ESN equations (2.3)–(2.4). In (2.11), $(\boldsymbol{u}_t)_{t\in\mathbb{Z}}$ denotes K-dimensional independent zero-mean innovations with variance $\sigma_u^2\mathbb{I}_K$ that are also independent of $\boldsymbol{x}_t$ across all t.

In this case, the reservoir map $F(\boldsymbol{x}_{t-1}, \boldsymbol{z}_t)$ in (2.1) is determined by (2.10), and it is possible to re-feed the forecasted variables back into the state equation as inputs. This yields the following state recursion:

$$\boldsymbol{x}_t = F(\boldsymbol{x}_{t-1}, \mathcal{W}^\top\boldsymbol{x}_{t-1} + \boldsymbol{u}_t) =: G_\theta(\boldsymbol{x}_{t-1}, \boldsymbol{u}_t),$$

where the subscript θ denotes the dependence on the model coefficients. In the reservoir computing literature, regimes where the ESN state equation is iteratively fed with the model outputs are called autonomous (Gonon et al., 2020a). They are widely and successfully utilized to predict deterministic dynamical systems. Indeed, in those instances, provided that the ridge estimate $\widehat{\mathcal{W}}$ is available from data according to Section 2.2, the $h >$ 1-step autonomous state iteration is given by

$$F_\theta^*(\boldsymbol{x}_t) := \alpha\boldsymbol{x}_t + (1-\alpha)\sigma((A + C\widehat{\mathcal{W}}^\top)\boldsymbol{x}_t + \boldsymbol{\zeta})$$

and

$$\boldsymbol{x}_{t+h} = \underbrace{F_\theta^* \circ F_\theta^* \circ \cdots \circ F_\theta^*}_{h \text{ times}}(\boldsymbol{x}_t).$$

Hence one can directly obtain the h-step-ahead predictions of the input time series as $\boldsymbol{z}_{t+h} = \widehat{\mathcal{W}}^\top\boldsymbol{x}_{t+h-1}$.

In the case of stochastic target variables, assuming (2.9), we notice that for the conditional forecast of the states, it holds that

$$\begin{aligned}\widehat{\boldsymbol{x}}_{t+1|t} &= \mathbb{E}\left[\boldsymbol{x}_{t+1}|\boldsymbol{x}_{0:t}, \boldsymbol{z}_{0:t}\right] \\ &= \int \boldsymbol{x}_t\, p_\theta(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{z}_t)\nu(\boldsymbol{z}_t|\boldsymbol{x}_{t-1})\mathrm{d}\boldsymbol{z}_t \\ &= \int G_\theta(\boldsymbol{x}_{t-1}, \boldsymbol{u}_t)\phi(\boldsymbol{u}_t)\mathrm{d}\boldsymbol{u}_t,\end{aligned} \tag{2.12}$$

where the density ϕ of $\boldsymbol{u}_t$ is, again, unavailable. Note that even under the assumption $\boldsymbol{u}_t \sim \mathcal{N}(\boldsymbol{0}, \Sigma_{\boldsymbol{u}})$, which is standard in the filtering literature, the presence of a nonlinear map G_θ makes the computation of the forecasts of $\boldsymbol{z}_{t+h}$ a non-straightforward exercise. Nevertheless, this forecast construction can be readily used when one is interested exclusively in predicting the time series $\boldsymbol{z}_t$.

Whenever the final goal of the exercise is forecasting h steps ahead some other explained variable $\boldsymbol{y}_{t+h}$, additional issues arise. In this case, one needs to compute the conditional expectation in (2.8), which is intractable even under Gaussian assumptions on the innovations. One option is to apply particle filtering techniques such as bootstrap sampling or sequential importance sampling (SIS) to evaluate the expectation (Doucet, de Freitas, & Gordon, 2001). We emphasize that the state dimension is usually chosen to be large, and hence implementing filtering techniques requires some care.

Our approach is to avoid dealing with the nonlinear densities involved in (2.8) with the help of (2.12) and, instead, to reduce the computation of the conditional expectation $\widehat{\boldsymbol{y}}_{t+h|t}$ to a composition of functions. By the linearity of the observation equation (2.4) and the assumption of independence in the zero-mean noise $\boldsymbol{\epsilon}_{t+h}$, we write

$$\begin{aligned}\widehat{\boldsymbol{y}}_{t+h|t} = W^\top\widehat{\boldsymbol{x}}_{t+h-1|t} = \int W^\top\boldsymbol{x}_{t+h-1} \\ \times \prod_{j=1}^{h-1} p_\theta(\boldsymbol{x}_{t+j}|\boldsymbol{x}_{t+j-1}, \boldsymbol{z}_{t+j})\nu(\boldsymbol{z}_{t+j}|\boldsymbol{x}_{t+j-1})\mathrm{d}\boldsymbol{x}_{t+j}\mathrm{d}\boldsymbol{z}_{t+j}\end{aligned}$$

and use the approximation

$$\widehat{\boldsymbol{y}}_{t+h|t} \approx \widetilde{\boldsymbol{y}}_{t+h} = W^\top \underbrace{F_\theta^* \circ F_\theta^* \circ \cdots \circ F_\theta^*}_{h-1 \text{ times}}(\boldsymbol{x}_t), \tag{2.13}$$

which originates from

$$\begin{aligned}\widehat{\boldsymbol{x}}_{t|t-1} &= \int G_\theta(\boldsymbol{x}_{t-1}, \boldsymbol{u}_t)\phi(\boldsymbol{u}_t)\mathrm{d}\boldsymbol{u}_t \\ &\approx G_\theta(\boldsymbol{x}_{t-1}, \mathbb{E}[\boldsymbol{u}_t]) = F(\boldsymbol{x}_{t-1}, \mathcal{W}^\top\boldsymbol{x}_{t-1}) \equiv F_\theta^*(\boldsymbol{x}_{t-1}),\end{aligned} \tag{2.14}$$

where $\boldsymbol{u}_t$ is assumed to be zero-mean. The validity of (2.14) itself requires implicit assumptions on the nature of the distribution of $\boldsymbol{u}_t$, but here we want to keep the analysis of $\widehat{\boldsymbol{y}}_{t+h|t}$ to a minimum and just use the insights from the dynamical systems ESN literature. We are hence not delving deeper into alternative approaches to estimate forecasts or, more generally, to compute conditional expectations of ESN models with stochastic inputs.

3. Multi-frequency echo state models

In this subsection, we construct a broad class of ESN models that can accommodate input and target time series sampled at distinct sampling frequencies. We call this family of reservoir models multi-frequency echo state networks (MFESNs). The state-space structure of MFESNs is naturally amenable to the setting of time series with mixed frequencies. Additionally, the prediction strategy discussed in Section 2.4 is straightforward to extend to MFESNs.

We present two groups of MFESN architectures. The first is based on a single echo state network architecture and we call these models single-reservoir multi-frequency

echo state networks (S-MFESNs). The second group, referred to as multi-reservoir multi-frequency echo state networks (M-MFESNs), allows for as many state equations as the number of distinct sampling frequencies present in the input data.

3.1. Single-reservoir MFESN

Recall that in the temporal notation of Definition 1.1, we reserve t to be the reference time index, which is also used for the target variable, and all other frequencies are measured with respect to the reference frequency.

Consider L collections of different time series. We assume that the lth collection, $l \in [L]$, consists of n_l time series that are sampled at a common frequency κ_l and contain observations $(\boldsymbol{z}^{(l)}_{t,s|\kappa_l})_{t,s}$ with $\boldsymbol{z}^{(l)}_{t,s|\kappa_l} \in \mathbb{R}^{n_l}$ for all $t \in \mathbb{Z}$ and $s \in \{0,\dots,\kappa_l - 1\}$. Let $\kappa_{\max} = \max_l \kappa_l$ be the highest sampling frequency among the L time series groups, and let $q_l := \kappa_{\max}/\kappa_l$ indicate how low each κ_l sampling frequency is with respect to $\kappa_{\max}$. We can now stack together and repeat the observations in a way that is consistent with the high-frequency index by defining

$$\boldsymbol{z}_{t,s|\kappa_{\max}} := \left(\boldsymbol{z}^{(1)\top}_{t,\lfloor s/q_1\rfloor|\kappa_1}, \boldsymbol{z}^{(2)\top}_{t,\lfloor s/q_2\rfloor|\kappa_2}, \dots, \boldsymbol{z}^{(L)\top}_{t,\lfloor s/q_L\rfloor|\kappa_L}\right)^\top \in \mathbb{R}^{\sum_{l=1}^L n_l},$$
$$s \in \{0,\dots,\kappa_{\max}-1\},$$

where for all $l \in [L]$, $\boldsymbol{z}^{(l)}_{0,0|\kappa_l} = \boldsymbol{0}_{n_l}$. Thus, it is possible to write a single high-frequency ESN as

$$\boldsymbol{x}_{t,s|\kappa_{\max}} = \alpha \boldsymbol{x}_{t,s-1|\kappa_{\max}} + (1-\alpha)\sigma(A\boldsymbol{x}_{t,s-1|\kappa_{\max}} + C\boldsymbol{z}_{t,s|\kappa_{\max}} + \boldsymbol{\zeta}), \tag{3.1}$$

$$\boldsymbol{z}_{t,s+1|\kappa_{\max}} = \mathcal{W}^\top \boldsymbol{x}_{t,s|\kappa_{\max}} + \boldsymbol{u}_{t,s+1|\kappa_{\max}}, \tag{3.2}$$

where $\mathcal{W} \in \mathbb{M}_{N,\sum_{l=1}^L n_l}$ and $s > 0$. We term this class of MFESN models single-reservoir multi-frequency ESNs (S-MFESNs).

Notice that equations (3.1)–(3.2) of the S-MFESN model prescribe the dynamics at the highest frequency, $\kappa_{\max}$. In order to forecast a lower-frequency target, we map high-frequency states $\boldsymbol{x}_{t,s|\kappa_{\max}}$ to low-frequency targets $\boldsymbol{y}_{t+1} \in \mathbb{R}^J$ by introducing a state alignment scheme. An aligned S-MFESN uses the most recent state with respect to the reference time index t to construct the forecast. More precisely, the state equation of an S-MFESN is iterated $\kappa_{\max}$ times until the state $\boldsymbol{x}_{t-1,\kappa_{\max}|\kappa_{\max}} = \boldsymbol{x}_{t,0|\kappa_{\max}}$ is obtained, and then target $\boldsymbol{y}_{t+1}$ is forecast with the following observation equation:

$$\boldsymbol{y}_{t+1} = W^\top \boldsymbol{x}_{t,0|\kappa_{\max}} + \boldsymbol{\epsilon}_{t+1}, \quad W \in \mathbb{M}_{N,J}. \tag{3.3}$$

Estimation of aligned S-MFESN. Both coefficient matrices W and $\mathcal{W}$ can be estimated as explained in Section 2.2 under appropriate choices of corresponding penalty strengths. In particular, in order to obtain $\widehat{\mathcal{W}}$, the state and the observation matrices in (2.5) are given by

$$X_{\kappa_{\max}} = (\boldsymbol{x}_{1,0|\kappa_{\max}}, \dots, \boldsymbol{x}_{1,\kappa_{\max}-1|\kappa_{\max}}, \dots, \boldsymbol{x}_{T-1,0|\kappa_{\max}}, \dots, \boldsymbol{x}_{T-1,\kappa_{\max}-1|\kappa_{\max}})^\top \in \mathbb{M}_{(T-1)\kappa_{\max}-1,N},$$
$$Y_{\kappa_{\max}} = (\boldsymbol{z}_{1,1|\kappa_{\max}}, \dots, \boldsymbol{z}_{1,\kappa_{\max}|\kappa_{\max}}, \dots, \boldsymbol{z}_{T-1,1|\kappa_{\max}}, \dots, \boldsymbol{z}_{T-1,\kappa_{\max}|\kappa_{\max}})^\top \in \mathbb{M}_{(T-1)\kappa_{\max}-1,\sum_{l=1}^L n_l},$$

while

$$X = \left(\boldsymbol{x}_{1,0|\kappa_{\max}}, \boldsymbol{x}_{2,0|\kappa_{\max}}, \dots, \boldsymbol{x}_{T-1,0|\kappa_{\max}}\right)^\top \in \mathbb{M}_{T-1,N},$$
$$Y = (\boldsymbol{y}_2, \dots, \boldsymbol{y}_T)^\top \in \mathbb{M}_{T-1,J},$$

are used for the estimation of $\widehat{W}$. We note that the state equation (3.1) of S-MFESN can be initialized by $\boldsymbol{x}_{0,0|\kappa_{\max}}$, which under the fading memory property is inconsequential for sufficiently long samples (see the discussion in Section 2).

Forecasting with aligned S-MFESN. Let $\widehat{W}$ and $\widehat{\mathcal{W}}$ be the sample estimates of the readout matrices as explained above. The fitted high-frequency autonomous state transition map associated with (3.1) is given by

$$F_{\kappa_{\max}}(\boldsymbol{x}_{t,s-1|\kappa_{\max}}) := \alpha \boldsymbol{x}_{t,s-1|\kappa_{\max}} + (1-\alpha)\,\sigma\left((A + C\widehat{\mathcal{W}}^\top)\boldsymbol{x}_{t,s-1|\kappa_{\max}} + \boldsymbol{\zeta}\right), \tag{3.4}$$

which, composed with itself exactly $\kappa_{\max}$ times, yields the target-frequency-aligned autonomous state transition map:

$$F(\boldsymbol{x}_{t,0|\kappa_{\max}}) := \underbrace{F_{\kappa_{\max}} \circ F_{\kappa_{\max}} \circ \cdots \circ F_{\kappa_{\max}}}_{\kappa_{\max}\ \text{times}}(\boldsymbol{x}_{t,0|\kappa_{\max}}). \tag{3.5}$$

Finally, from (2.13) the h-step-ahead low-frequency forecasts, $h \in \mathbb{N}$, can be computed as

$$\widetilde{y}_{T+h|T} = \widehat{W}^\top\left(\underbrace{F \circ F \circ \cdots \circ F}_{h-1\ \text{times}}(\boldsymbol{x}_{T,0|\kappa_{\max}})\right). \tag{3.6}$$

Fig. 1 gives a graphical diagram of the one-step forecasting procedure for an S-MFESN. Additionally, Fig. 12 in Supplementary Appendix K provides a similar diagram for the case of multistep forecasts.

The following example illustrates this proposed forecasting strategy for the case of quarterly GDP forecasting using monthly and daily series inputs.

Example 3.1. Suppose that we wish to use an aligned S-MFESN model to forecast a quarterly one-dimensional target (y_t) using $n_{(\mathrm{m})}$ monthly and $n_{(\mathrm{d})}$ daily series, $(\boldsymbol{z}^{(\mathrm{m})}_{t,s|\kappa_1})$ and $(\boldsymbol{z}^{(\mathrm{d})}_{t,s|\kappa_2})$, respectively. We adopt the assumption that daily data are released 24 days over each calendar month and hence $\kappa_1 = 3$, $\kappa_2 = 72$, and $\kappa_{\max} = 72$, while $q_1 = 24$ and $q_2 = 1$. Let $t, *|72$ be the temporal index with a quarterly reference frequency. The input vector for the S-MFESN state equation consistent with the daily frequency is given by

$$\boldsymbol{z}^{(\mathrm{m},\mathrm{d})}_{t,s|72} := (\boldsymbol{z}^{(\mathrm{m})\top}_{t,\lfloor s/24\rfloor|3}, \boldsymbol{z}^{(\mathrm{d})\top}_{t,s|72})^\top \in \mathbb{R}^{n_{(\mathrm{m})}+n_{(\mathrm{d})}} \text{ with}$$
$$\boldsymbol{z}^{(\mathrm{d})}_{0,0|3} = \boldsymbol{0}_{n_{(\mathrm{d})}} \text{ and } \boldsymbol{z}^{(\mathrm{m})}_{0,0|24} = \boldsymbol{0}_{n_{(\mathrm{m})}}.$$

The complete S-MFESN model with the state-space dimension N can be written as

$$\boldsymbol{x}^{(\mathrm{m},\mathrm{d})}_{t,s|72} = \alpha \boldsymbol{x}^{(\mathrm{m},\mathrm{d})}_{t,s-1|72} + (1-\alpha)\sigma(A\boldsymbol{x}^{(\mathrm{m},\mathrm{d})}_{t,s-1|72} + C\boldsymbol{z}^{(\mathrm{m},\mathrm{d})}_{t,s|72} + \boldsymbol{\zeta}), \tag{3.7}$$

$$\boldsymbol{z}^{(\mathrm{m},\mathrm{d})}_{t,s+1|72} = \mathcal{W}^\top \boldsymbol{x}^{(\mathrm{m},\mathrm{d})}_{t,s|72} + \boldsymbol{u}_{t,s+1|72}, \tag{3.8}$$

$$y_{t+1} = W^\top \boldsymbol{x}^{(\mathrm{m},\mathrm{d})}_{t,0|\kappa_{\max}} + \boldsymbol{\epsilon}_{t+1}, \tag{3.9}$$

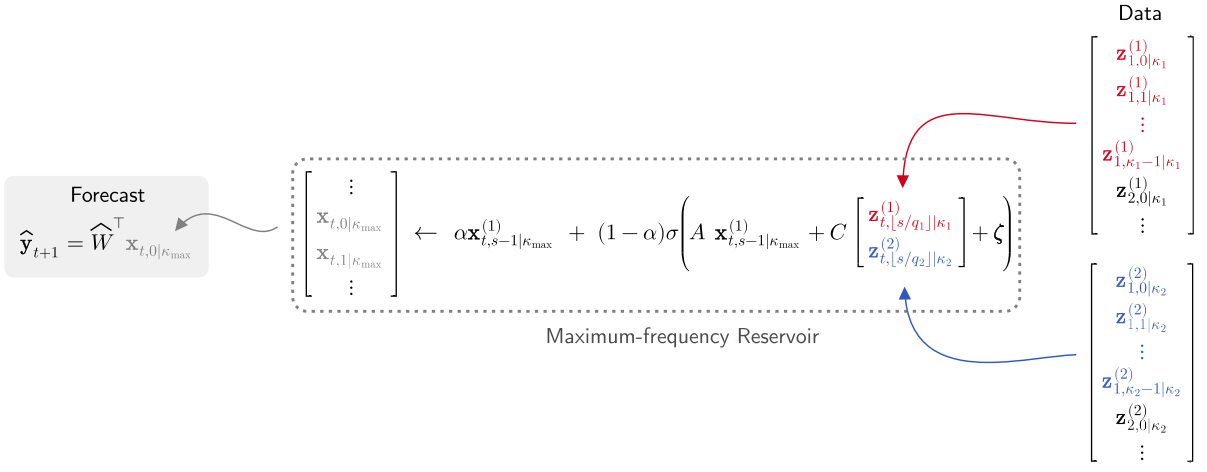


Fig. 1. Scheme of a single-reservoir MFESN (S-MFESN) model combining input data sampled at two frequencies with state alignment and estimation for one-step-ahead forecasting of the target series.

where the state equations (3.7)–(3.8) are run in their own maximum frequency temporal index $s > 0$, and only the states $\boldsymbol{x}_{t-1,\kappa_{\max}|\kappa_{\max}} = \boldsymbol{x}_{t,0|\kappa_{\max}}$ are used in the observation equation (3.9). Provided the input-target pairs sample of length T, the coefficient matrices $\mathcal{W} \in \mathbb{M}_{N,n_{(\mathtt{m})}+n_{(\mathtt{d})}}$ in (3.8) and $W \in \mathbb{R}^N$ in (3.9) can be estimated via ridge regression, as explained above.

From (3.4), the high-frequency autonomous state transition map is given by

$$F_{72}^{(\mathtt{m},\mathtt{d})}(\boldsymbol{x}_{t,s-1|72}^{(\mathtt{m},\mathtt{d})}) := \alpha \boldsymbol{x}_{t,s-1|72}^{(\mathtt{m},\mathtt{d})} + (1-\alpha)\,\sigma\left((A + C\widehat{\mathcal{W}}^\top)\boldsymbol{x}_{t,s-1|72}^{(\mathtt{m},\mathtt{d})} + \boldsymbol{\zeta}\right),$$

which, composed with itself exactly 72 times, by (3.5) yields the target-frequency-aligned autonomous state transition map:

$$F^{(\mathtt{m},\mathtt{d})}(\boldsymbol{x}_{t,0|72}^{(\mathtt{m},\mathtt{d})}) := \underbrace{F_{72}^{(\mathtt{m},\mathtt{d})} \circ F_{72}^{(\mathtt{m},\mathtt{d})} \cdots \circ F_{72}^{(\mathtt{m},\mathtt{d})}}_{72 \text{ times}}(\boldsymbol{x}_{t,0|72}^{(\mathtt{m},\mathtt{d})}).$$

By applying $F^{(\mathtt{m},\mathtt{d})}$ to state $\boldsymbol{x}_{t,0|72}^{(\mathtt{m},\mathtt{d})}$ we iterate the S-MFESN forward in time to provide an estimate for $\boldsymbol{x}_{t+1,0|72}^{(\mathtt{m},\mathtt{d})}$, which can then be linearly projected using $\widehat{W}$ to yield a forecast for y_{t+2}. For the target variable, as well as forecasts, we do not use our temporal notation for the sake of compactness and clarity of exposition. Finally, the quarterly forecasts for $h \in \mathbb{N}$ can be computed using (3.6) as follows:

$$\widetilde{y}_{T+h|T} = \widehat{W}^\top \Big(\underbrace{F^{(\mathtt{m},\mathtt{d})} \circ F^{(\mathtt{m},\mathtt{d})} \circ \cdots \circ F^{(\mathtt{m},\mathtt{d})}}_{h-1 \text{ times}}(\boldsymbol{x}_{T,0|72}^{(\mathtt{m},\mathtt{d})})\Big).$$

3.2. Multi-reservoir MFESN

Constructing an MFESN with a single reservoir is not necessarily the most effective modeling strategy. Having more than one reservoir allows for more flexible modeling of state dynamics for different subsets of input variables sampled at common frequencies. For example, suppose quarterly and monthly data are used as regressors. Our presentation is general enough to accommodate other types of partitioning of series into the corresponding reservoir models. We leave it to future research to test other approaches based, for instance, on markets or data types, as done by van Huellen et al. (2020).

Assume again L groups of series with input observations $(\boldsymbol{z}_{t,s|\kappa_l}^{(l)})_{t,s}$ with $\boldsymbol{z}_{t,s|\kappa_l}^{(l)} \in \mathbb{R}^{n_l}$, $l \in [L]$, for all $t \in \mathbb{Z}$ and $s \in \{0, \ldots, \kappa_l - 1\}$ sampled at common frequencies $\{\kappa_1, \ldots, \kappa_L\}$, respectively. For each of the L groups of input series, we define the corresponding ESN model as

$$\boldsymbol{x}_{t,s|\kappa_l}^{(l)} = \alpha_l \boldsymbol{x}_{t,s-1|\kappa_l}^{(l)} + (1-\alpha_l)\,\sigma(A_l \boldsymbol{x}_{t,s-1|\kappa_l}^{(l)} + C_l \boldsymbol{z}_{t,s|\kappa_l}^{(l)} + \boldsymbol{\zeta}_l), \tag{3.10}$$

$$\boldsymbol{z}_{t,s+1|\kappa_l}^{(l)} = \mathcal{W}_l^\top \boldsymbol{x}_{t,s|\kappa_l}^{(l)} + \boldsymbol{u}_{t,s+1|\kappa_l}^{(l)}, \quad l \in [L], \tag{3.11}$$

with $s > 0$, $\mathcal{W}_l \in \mathbb{M}_{N_l,n_l}$ with N_l the dimension of the state space. Notice that the time index s is different for each l according to our temporal notation introduced in Definition 1.1, and each state equation runs at its own frequency κ_l. The dimensions $\{N_1, N_2, \ldots, N_L\}$ of the state spaces can be chosen for the L reservoir models individually. Additionally, multiple reservoirs have the associated hyperparameter tuples $\{\boldsymbol{\varphi}_1, \ldots, \boldsymbol{\varphi}_L\}$ to be tuned. This requires some care whenever one wants to optimize all hyperparameters jointly. Since there are L reservoir state equations, we call this class of MFESN models multi-reservoir multi-frequency ESNs (M-MFESNs).

Similar to the S-MFESN, all L state equations are each iterated κ_l times until the states $\boldsymbol{x}_{t-1,\kappa_l|\kappa_l}^{(l)} = \boldsymbol{x}_{t,0|\kappa_l}^{(l)}$ are obtained. The aligned M-MFESN observation equation is given by

$$\boldsymbol{y}_{t+1} = W^\top \boldsymbol{x}_{t,L} + \boldsymbol{\epsilon}_{t+1}, \quad \text{with}$$
$$\boldsymbol{x}_{t,L} = \begin{pmatrix} \boldsymbol{x}_{t,0|\kappa_1}^{(1)} \\ \vdots \\ \boldsymbol{x}_{t,0|\kappa_L}^{(L)} \end{pmatrix} \in \mathbb{R}^{\sum_{l=1}^L N_l}, \quad W \in \mathbb{M}_{\sum_{l=1}^L N_l J}. \tag{3.12}$$

Estimation of aligned M-MFESN. The coefficient matrices W_l, $l \in [L]$, and $\mathcal{W}$ can be estimated similarly to the case of S-MFESN. The state and observation matrices for the estimation of $\widehat{\mathcal{W}}_l$, $l \in [L]$, in (2.5) are constructed as follows:

$$X^{(l)} = (\boldsymbol{x}^{(l)}_{1,0|\kappa_l}, \ldots, \boldsymbol{x}^{(l)}_{1,\kappa_l-1|\kappa_l}, \ldots, \boldsymbol{x}^{(l)}_{T-1,0|\kappa_l}, \ldots, \boldsymbol{x}^{(l)}_{T-1,\kappa_l-1|\kappa_l})^\top \in \mathbb{M}_{(T-1)\kappa_l-1,N_l},$$
$$Y^{(l)} = (\boldsymbol{z}^{(l)}_{1,1|\kappa_l}, \ldots, \boldsymbol{z}^{(l)}_{1,\kappa_l|\kappa_l}, \ldots, \boldsymbol{z}^{(l)}_{T-1,1|\kappa_l}, \ldots, \boldsymbol{z}^{(l)}_{T-1,\kappa_l|\kappa_l})^\top \in \mathbb{M}_{(T-1)\kappa_l-1,n_l},$$

while, with the notation as in (3.12),

$$X = \left(\boldsymbol{x}_{1,L}, \boldsymbol{x}_{2,L}, \ldots, \boldsymbol{x}_{T-1,L}\right)^\top \in \mathbb{M}_{T-1,\sum_{l=1}^{L} N_l},$$
$$Y = (\boldsymbol{y}_2, \ldots, \boldsymbol{y}_T)^\top \in \mathbb{M}_{T-1,J},$$

are used for the estimation of $\widehat{W}$. Again, the state equation (3.10) of M-MFESN can be started with $\boldsymbol{x}^{(l)}_{0,0|\kappa_l} = \boldsymbol{0}_{N_l}$; see Section 2 for more details.

Forecasting with aligned M-MFESN. Let $\widehat{W}$ and $\widehat{\mathcal{W}}_l$, $l \in [L]$ be the sample estimates of the readout matrices. For any $l \in [L]$, the κ_l-frequency autonomous state transition map is given by

$$F^{(l)}_{\kappa_l}(\boldsymbol{x}^{(l)}_{t,s-1|\kappa_l}) := \alpha_l \boldsymbol{x}^{(l)}_{t,s-1|\kappa_l} + (1-\alpha_l)\,\sigma\left((A_l + C_l \widehat{\mathcal{W}}_l^\top)\boldsymbol{x}^{(l)}_{t,s-1|\kappa_l} + \boldsymbol{\zeta}_l\right). \tag{3.13}$$

The target-frequency-aligned autonomous state transition map associated with each frequency l is hence defined as

$$F^{(l)}(\boldsymbol{x}_{t,0|\kappa_l}) := \underbrace{F^{(l)}_{\kappa_l} \circ F^{(l)}_{\kappa_l} \circ \cdots \circ F^{(l)}_{\kappa_l}}_{\kappa_l \text{ times}}(\boldsymbol{x}^{(l)}_{t,0|\kappa_l}). \tag{3.14}$$

Finally, from (2.13), the h-step-ahead forecasts can be computed as

$$\widetilde{y}_{T+h|T} = \widehat{W}^\top \begin{pmatrix} \underbrace{F^{(1)} \circ F^{(1)} \circ \cdots \circ F^{(1)}}_{h-1 \text{ times}}(\boldsymbol{x}^{(1)}_{T,0|\kappa_1}) \\ \vdots \\ \underbrace{F^{(L)} \circ F^{(L)} \circ \cdots \circ F^{(L)}}_{h-1 \text{ times}}(\boldsymbol{x}^{(L)}_{T,0|\kappa_L}) \end{pmatrix}. \tag{3.15}$$

In Fig. 2 we provide a diagram for the case of one-step-ahead forecasting with an aligned M-MFESN involving regressors of only two frequencies. Fig. 13 in Supplementary Appendix K provides a similar diagram for the case of multistep forecasting.

Example 3.2. Similar to Example 3.1, we aim to forecast a quarterly target with monthly and daily series, but this time we use an M-MFESN model. We have to define two independent state equations, one for monthly and one for daily series; in the observation equations, two states must be aligned temporally and stacked to form the full set of regressors. The data again consist of quarterly (y_t), $n_{(\mathrm{m})}$ monthly series ($\boldsymbol{z}^{(\mathrm{m})}_{t,s|3}$), and $n_{(\mathrm{d})}$ daily series ($\boldsymbol{z}^{(\mathrm{d})}_{t,s|72}$).

The aligned M-MFESN model with two reservoirs of dimensions $N_{(\mathrm{m})}$ and $N_{(\mathrm{d})}$ is respectively given by

$$\boldsymbol{x}^{(\mathrm{m})}_{t,s|3} = \alpha_1 \boldsymbol{x}^{(\mathrm{m})}_{t,s-1|3} + (1-\alpha_1)\,\sigma(A_1 \boldsymbol{x}^{(\mathrm{m})}_{t,s-1|3} + C_1 \boldsymbol{z}^{(\mathrm{m})}_{t,s|3} + \boldsymbol{\zeta}_1), \tag{3.16}$$

$$\boldsymbol{z}^{(\mathrm{m})}_{t,s+1|3} = \mathcal{W}^\top_{(\mathrm{m})} \boldsymbol{x}^{(\mathrm{m})}_{t,s|3} + \boldsymbol{u}^{(\mathrm{m})}_{t,s+1|3}, \tag{3.17}$$

$$\boldsymbol{x}^{(\mathrm{d})}_{t,s|72} = \alpha_2 \boldsymbol{x}^{(\mathrm{d})}_{t,s-1|72} + (1-\alpha_2)\,\sigma(A_2 \boldsymbol{x}^{(\mathrm{d})}_{t,s-1|72} + C_2 \boldsymbol{z}^{(\mathrm{d})}_{t,s|72} + \boldsymbol{\zeta}_2), \tag{3.18}$$

$$\boldsymbol{z}^{(\mathrm{d})}_{t,s+1|72} = \mathcal{W}^\top_{(\mathrm{d})} \boldsymbol{x}^{(\mathrm{d})}_{t,s|72} + \boldsymbol{u}^{(\mathrm{d})}_{t,s+1|72}, \tag{3.19}$$

$$y_{t+1} = W^\top \begin{pmatrix} \boldsymbol{x}^{(\mathrm{m})}_{t,0|3} \\ \boldsymbol{x}^{(\mathrm{d})}_{t,0|72} \end{pmatrix} + \epsilon_{t+1}, \tag{3.20}$$

where $s > 0$, $\mathcal{W}_{(\mathrm{m})} \in \mathbb{M}_{N_{(\mathrm{m})},n_{(\mathrm{m})}}$, $\mathcal{W}_{(\mathrm{d})} \in \mathbb{M}_{N_{(\mathrm{d})},n_{(\mathrm{d})}}$ and $W \in \mathbb{R}^{N_{(\mathrm{m})}+N_{(\mathrm{d})}}$. Here, the monthly reservoir ($\boldsymbol{x}^{(\mathrm{m})}_{t,s|3}$) has a temporal index of frequency 3, while that of the daily reservoir ($\boldsymbol{x}^{(\mathrm{d})}_{t,s|72}$) is 72; the high-frequency index s is different for the two models. Notice that in an M-MFESN model it is necessary to introduce two additional observation equations for the states, that is, (3.17) and (3.19). Notice that the state equations are each iterated κ_l times to collect the states to be aligned in the observation equation (3.20). Again, the sample-based estimates of coefficient matrices $\widehat{\mathcal{W}}_{(\mathrm{m})}$, $\widehat{\mathcal{W}}_{(\mathrm{d})}$ and $\widehat{W}$ in (3.17), (3.18), and (3.20), respectively, can be obtained via the ridge regression, as discussed above.

Exactly as in Example 3.1, using (3.13) we can introduce high-frequency autonomous state maps $F^{(\mathrm{m})}_3$ and $F^{(\mathrm{d})}_{72}$:

$$F^{(\mathrm{m})}_3(\boldsymbol{x}^{(\mathrm{m})}_{t,s-1|3}) := \alpha_1 \boldsymbol{x}^{(\mathrm{m})}_{t,s-1|3} + (1-\alpha_1)\,\sigma\left((A_1 + C_1 \widehat{\mathcal{W}}^\top_{(\mathrm{m})})\boldsymbol{x}^{(\mathrm{m})}_{t,s-1|3} + \boldsymbol{\zeta}_1\right),$$
$$F^{(\mathrm{d})}_{72}(\boldsymbol{x}^{(\mathrm{d})}_{t,s-1|72}) := \alpha_2 \boldsymbol{x}^{(\mathrm{d})}_{t,s-1|72} + (1-\alpha_2)\,\sigma\left((A_2 + C_2 \widehat{\mathcal{W}}^\top_{(\mathrm{d})})\boldsymbol{x}^{(\mathrm{d})}_{t,s-1|72} + \boldsymbol{\zeta}_2\right),$$

as well as their target-frequency aligned counterparts $F^{(\mathrm{m})}$ and $F^{(\mathrm{d})}$, by (3.14):

$$F^{(\mathrm{m})}(\boldsymbol{x}^{(\mathrm{m})}_{t,0|3}) := \underbrace{F^{(\mathrm{m})}_3 \circ F^{(\mathrm{m})}_3 \circ F^{(\mathrm{m})}_3}_{3 \text{ times}}(\boldsymbol{x}^{(\mathrm{m})}_{t,0|3}),$$
$$F^{(\mathrm{d})}(\boldsymbol{x}^{(\mathrm{d})}_{t,0|72}) := \underbrace{F^{(\mathrm{d})}_{72} \circ F^{(\mathrm{d})}_{72} \circ \cdots \circ F^{(\mathrm{d})}_{72}}_{72 \text{ times}}(\boldsymbol{x}^{(\mathrm{d})}_{t,0|72}).$$

The h-step-ahead forecasts can be computed using the approximation in (3.15):

$$\widetilde{y}_{T+h|T} = \widehat{W}^\top \begin{pmatrix} \underbrace{F^{(\mathrm{m})} \circ F^{(\mathrm{m})} \circ \cdots \circ F^{(\mathrm{m})}}_{h-1 \text{ times}}(\boldsymbol{x}^{(\mathrm{m})}_{T,0|3}) \\ \underbrace{F^{(\mathrm{d})} \circ F^{(\mathrm{d})} \circ \cdots \circ F^{(\mathrm{d})}}_{h-1 \text{ times}}(\boldsymbol{x}^{(\mathrm{d})}_{T,0|72}) \end{pmatrix}.$$

In this case, it is important to note that while both $F^{(\mathrm{m})}$ and $F^{(\mathrm{d})}$ are composed $h-1$ times at step h, the underlying number of autonomous reservoir iterations is different for the monthly and daily reservoirs, namely 3 and 72, and depends on their own frequencies. This also suggests that one should take into account the different time dynamics

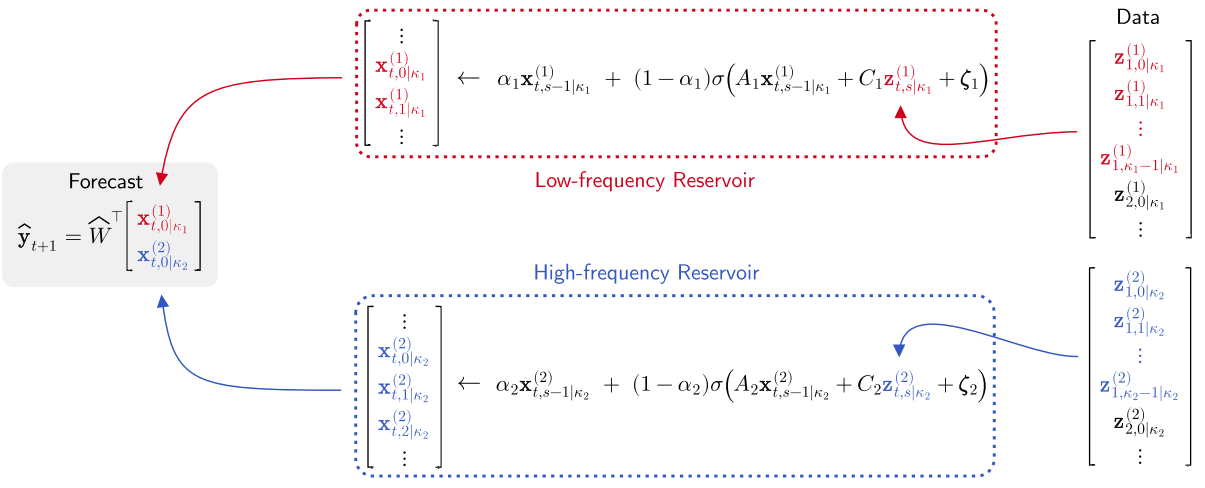


Fig. 2. Scheme of a multi-reservoir MFESN (M-MFESN) model combining input data sampled at two frequencies with state alignment and estimation for one-step-ahead forecasting of the target series.

when, for example, tuning M-MFESN hyperparameters $\boldsymbol{\varphi}^{(\mathtt{m})}$ and $\boldsymbol{\varphi}^{(\mathtt{d})}$, as proposed in Subsection D.2.

4. Empirical study

In this section, we compare the forecasting performance of our proposed MFESN to state-of-the-art benchmarks. We use a combination of macroeconomic and financial data sampled at low- and high-frequency intervals, respectively. Our empirical exercises encompass several setups, with a small and a medium-sized set of regressors, fitting models with data before and after the 2007–2008 crisis, and with fixed, rolling, and expanding estimation windows.

4.1. Data

Two sets of predictors of different sizes are compiled: small-MD with nine predictors and medium-MD with 33 predictors at monthly and daily frequencies. The reference frequency is quarterly: this is the frequency at which the target variable, U.S. GDP growth, is available. Seasonally adjusted quarterly and monthly data are obtained from the Federal Reserve Bank of St. Louis Monthly (FRED-MD) and Quarterly (FRED-QD) Databases for Macroeconomic Research (see McCracken and Ng (2016, 2020) for details). Daily data are obtained from Refinitiv Datastream, a subscription-based data service. All data are the last revised vintage data. The macroeconomic target and predictors, their transformations, and availability are provided in full detail in Table A.9 in Appendix A.

The selection of predictors follows the seminal work by Stock and Watson (1996, 2006), in which the FRED-MD and FRED-QD data are proposed. Variations of their dataset have been used profusely in the literature (for example, see Boivin and Ng (2005), Hatzius, Hooper, Mishkin, Schoenholtz, and Watson (2010), Marcellino, Stock, and Watson (2006)). Indicators from ten macroeconomic and financial categories are considered: (1) output and income, (2) labor market, (3) housing, (4) orders and inventories, (5) price indices, (6) money and credit, (7) interest rates, (8) exchange rates, (9) equity, and (10) derivatives. The latter five categories represent financial market conditions and are sourced at daily frequency. The exception is interest rates, which move relatively slowly and enter as monthly aggregates, available in the FRED-MD data. We refer to this dataset as medium-MD. A subset of predictors is selected for the small-MD dataset by choosing variables that have been identified as leading indicators in the empirical literature (Andreou et al., 2013; Carriero, Galvão, & Kapetanios, 2019; Clements & Galvão, 2008; Ferrara, Marsilli, & Ortega, 2014; Ingenito & Trehan, 1996; Jardet & Meunier, 2022; Marsilli, 2014). Data availability is an additional criterion, and predictors unavailable before 1990 are not considered. This excludes the VIX volatility index, which has been identified as a leading indicator in some studies, for example in Andreou et al. (2013), Jardet and Meunier (2022).

We follow instructions by McCracken and Ng (2016, 2020) on pre-processing macroeconomic predictors before they are used as input for forecasting. These are mainly differenced for detrending. We further transform financial predictors to capture market disequilibrium and volatility. Disequilibrium indicators, such as interest rate spreads, have been found to be more relevant for macroeconomic prediction than routine changes captured by differencing (see Borio and Lowe (2002), Gramlich, Miller, Oet, and Ong (2010), Qin, van Huellen, Wang, and Moraitis

(2022)). In addition to disequilibrium indicators, realized stock market volatility has been found to improve macroeconomic predictions (Chauvet et al., 2015). In the absence of intraday trading data from the 1990s onward, which prevents us from utilizing conventional daily realized volatility indicators, we extract volatility indicators from daily price series by fitting a GARCH(1,1) by Bollerslev (1986).[2] In addition to volatility of stock and commodity prices, term structure indicators are used. The term structure is forward-looking, capturing information about future demand and supply, and has been found to be a leading predictor of GDP growth (see for example Hong and Yogo (2012), Kang and Kwon (2020)).

The data span the period from January 1st, 1990, to December 31st, 2019.[3] We are interested in evaluating model performance under two stylized settings. First, a researcher fits all models up until the Great Recession, including data from 1990Q1 to 2007Q4. Second, fitting is done with data largely encompassing the crisis period, again from 1990Q1 but now up to 2011Q4. In both cases, the testing sample ranges from the next GDP growth observation after fitting up to 2019Q4. All exercises exclude the global Covid-19 economic depression, as we consider it an extreme, unpredictable event that induced significant structural changes in the underlying macroeconomic dynamics.[4]

To avoid having to handle the many edge cases that daily data in their raw calendar releases involve, we use an interpolation approach. We set ex ante the number of working days in any month to be exactly 24: given that in forecasting the most recent information sets are more relevant, when interpolating daily data over months with fewer than 24 calendar observations, we linearly interpolate the missing data starting from a month's beginning (using the previous month's last observation). The choice of 24 as a daily frequency is transparent by noting that this is the closest number to actual commonly observed data releases, whilst also being a multiple of both four (the approximate number of weeks per month) and six (the upper bound on the number of working days per week).

[2] We include a control `scale = 1` to ensure convergence of the optimization algorithm, and we only include a constant mean term in the return process for simplicity.

[3] In the small-MD dataset experiments we make a small variation and instead include data starting from January 1st, 1975, but *only* for the initial CV selection of ridge penalties for MFESN models. Our aim is to make sure that at least for the fixed window estimation strategy – where λ is cross-validated once, and only one $\widehat{W}$ is estimated – the ridge estimator is robust. In practice, when we compare to expanding and rolling window estimators, where λ is re-selected at each window, we find that extending the initial CV data window has little impact on out-of-sample performance.

[4] In the macroeconomic literature this falls under the category of 'natural disaster' events, and should not be naïvely modeled together with previous observations. In this section, we therefore avoid dealing with post-Covid-19 macroeconomic data altogether.

4.2. Models

In this section, we present the set of models that we use throughout our empirical exercises. For a general overview, Table 4.1 summarizes all models, including hyperparameters. In our analysis, we compare the competing models based on several performance measures, which we introduce in Supplementary Appendix E.

4.2.1. Benchmarks

Unconditional mean. We use the unconditional mean of the sample used for fitting as a baseline benchmark. For GDP growth forecasting, there is evidence that the unconditional mean produces forecasts that are competitive with linear models such as VARs in terms of mean square forecasting errors (MSFEs), even at relatively short horizons (Arora, Little, & McSharry, 2013). It is therefore an important reference for the performance of all other models, and we report relative MSFEs with respect to the unconditional mean in the tables below.

AR(1) model. A simple autoregressive process of order one on the target variable is included as a benchmark model.[5] This is also a common benchmark in the literature, as AR(1) models are often able to capture key dynamics and produce meaningful forecasts for macroeconomic variables (Bai & Ng, 2008; Stock & Watson, 2002). We emphasize that since AR(1) model is fit to the series of quarterly GDP targets and does not use any additional information, its forecasts are identical for both the small-MD and medium-MD samples.

Mixed-data sampling (MIDAS). The first mixed-frequency model benchmark is given by a MIDAS model (Ghysels et al., 2004, 2007). Our dynamic MIDAS specification includes autoregressive lags of the target series and uses an Almon weighting scheme. As shown in Bai, Ghysels, and Wright (2013), exponential Almon MIDAS regressions are related to dynamic factor models, which we also consider as benchmarks. The MIDAS model includes three lags of quarterly GDP target variable, and 30 daily and nine monthly lags for all daily and monthly series, respectively. This model prescription allows for some parsimony, as the Almon polynomial weighing reduces the number of daily and monthly lag coefficients.

A thorough description of our MIDAS implementation can be found in Supplementary Appendix G. To make optimization more efficient, we use explicit expressions for MIDAS loss gradients, as in Kostrov (2021). The MIDAS estimation can be hard to perform in practice, due to the complexity of nonlinear optimization. First, exponential weighting schemes might require computing floating-point numbers that exceed numerical precision. Therefore, it is a better choice to start the gradient descent close to the origin of the parameter space. Second, even with this choice of starting points, one may encounter issues with the optimization results, since the Almon-scheme MIDAS loss can have a large number of distinct local minima. In Supplementary Appendix J.1 we

[5] Suggested by an anonymous referee.

Table 4.1
Table of models used in applied forecasting exercises. MFESN hyperparameters are defined with respect to normalized state parameters; cf. (2.6).

Model name	Description	Specification
Mean	Unconditional mean of target series over estimation sample.	None
AR(1)	Autoregressive model of target series estimated using OLS.	None
MIDAS	Almon-weighted MIDAS regression, linear (unconstrained) autoregressive component.	Autoregressive lags: 3 Monthly freq. lags: 9 Daily freq. lags: 30
DFM [A]	Stock aggregation, VAR(1) factor process.	Factors:5 for Small-MD 10 for Medium-MD
DFM [B]	Almon aggregation, VAR(1) factor process	Factors:5 for Small-MD 10 for Medium-MD
singleESN [A]	S-MFESN model: Sparse-normal $\widetilde{A}$, sparse-uniform $\widetilde{C}$, $\widetilde{\zeta} = \mathbf{0}$. Isotropic ridge regression fit.	Reservoir dim: 30 Sparsity: 33.3% $\rho = 0.5$, $\gamma = 1$, $\alpha = 0.1$
singleESN [B]	S-MFESN model: Sparse-normal $\widetilde{A}$, sparse-uniform $\widetilde{C}$, $\widetilde{\zeta} = \mathbf{0}$. Isotropic ridge regression fit.	Reservoir dim: 120 Sparsity: 8.3% $\rho = 0.5$, $\gamma = 1$, $\alpha = 0.1$
multiESN [A]	M-MFESN model: Monthly and daily frequency reservoirs. Sparse-normal $\widetilde{A}_1$, $\widetilde{A}_2$, sparse-uniform $\widetilde{C}_1$, $\widetilde{C}_2$, $\widetilde{\zeta}_1 = \mathbf{0}$, $\widetilde{\zeta}_2 = \mathbf{0}$. Isotropic ridge regression fit	Reservoir dims: M=100, D=20 Sparsity: M=10%, D=50% M: $\rho = 0.5$, $\gamma = 1.5$, $\alpha = 0$ D: $\rho = 0.5$, $\gamma = 0.5$, $\alpha = 0.1$
multiESN [B]	M-MFESN model: Monthly and daily frequency reservoirs. Sparse-normal $\widetilde{A}_1$, $\widetilde{A}_2$, sparse-uniform $\widetilde{C}_1$, $\widetilde{C}_2$, $\widetilde{\zeta}_1 = \mathbf{0}$, $\widetilde{\zeta}_2 = \mathbf{0}$. Isotropic ridge regression fit.	Reservoir dims: M=100, D=20 Sparsity: M=10%, D=50% M: $\rho = 0.08$, $\gamma = 0.25$, $\alpha = 0.3$ D: $\rho = 0.01$, $\gamma = 0.01$, $\alpha = 0.99$

document, using a simple replication experiment, that even small changes in the initial conditions can result in different local minima picked by the numerical optimization algorithm.[6] These important robustness issues are present even when using closed-form gradients and multi-start optimization routines for the MIDAS models. The computational issues become more pronounced as the number of MIDAS parameters increase, unless a careful model/variable selection step is performed. Therefore, we do not include any MIDAS model specifications in the medium-MD setup.

Dynamic factor model (DFM). The DFM framework has been extensively applied in macroeconometrics, starting with Geweke (1977) and Sargent et al. (1977). A DFM specification assumes that predictable dynamics of a large set of time series can be explained by a small number of factors with an autoregressive dependence (see for example Doz, Giannone, and Reichlin (2011), Forni, Hallin, Lippi, and Reichlin (2005), Stock and Watson (2016)). We generalize the standard two-frequency DFM modeling setup (Bańbura & Modugno, 2014; Mariano & Murasawa, 2003) to a flexible mixed-frequency DFM that encompasses any number of data frequencies. Moreover, we derive a novel weighting scheme that effectively links the MIDAS and DFM approaches. For a detailed discussion of our factor model setup, we refer the reader to Supplementary Appendix H. Two distinct DFM specifications are used. The first one, termed DFM [A], uses the standard linear aggregation scheme, as provided in Example H.1, while the second is a variation that implements an Almon weighting scheme, as presented in Example H.2 (we name it DFM [B]). The latter is similar to a MIDAS-type aggregation scheme (Marcellino & Schumacher, 2010): the factor structure effectively mitigates the parameter proliferation.

A key choice for a DFM model is the dimension of the factor process. While a number of methods have been developed over the years to systematically derive the number of factors (see for example the review by Stock and Watson (2016)), commonly used macroeconomic panels feature a number of challenges, such as weak factors (Onatski, 2012). Moreover, as mentioned in Supplementary Appendix H.1, factor number selection in the mixed-frequency setting has not been sufficiently addressed in the literature. To sidestep these issues, we construct both DFM models with five unobserved factors

[6] We set the initial coefficient values to zero in all empirical exercises.

Table 4.2
Execution time in seconds for model estimation measured over a single run on a quad-core computer. MFESN model timing includes ridge penalty cross-validation. MIDAS estimation time refers to optimization from a single initial value. DFM models were estimated on a single-core server and times are adjusted by a factor of 1/4 for comparison.

Execution time (in seconds) for model estimation									
Dataset	Mean	AR(1)	MIDAS	DFM		singleESN		multiESN	
				[A]	[B]	[A]	[B]	[A]	[B]
Small-MD	0.1	0.7	1.3	40.5	85.5	2.6	4.5	15.3	14.6
Medium-MD	0.1	0.8	–	48.0	226.5	2.5	5.7	17.7	14.7

for small-MD and 10 for medium-MD, respectively, and assume that they follow a VAR(1) process.

One extant issue with integrating daily data is their very high release frequency compared to monthly and especially quarterly releases: computationally, this can be extremely taxing, which might be one of the reasons why to our knowledge we are the first to provide DFM forecasts that include daily data. Our solution is to reduce aggregate daily data every six days by averaging, thus leaving four observations per month. This considerably eases the computational burden of estimating coefficients and latent states (12 versus 72 daily observations per quarter).

4.2.2. Multi-frequency ESNs

The first set of ESNs we propose is given by two S-MFESN models, based on Example 3.1. One model uses a reservoir of 30 neurons (we call it singleESN [A]). The other has a larger reservoir of dimension 120 (named singleESN [B]). The sparsity degree of state parameters for both models is set to be $10/N$, where N is the reservoir size. Both MFESNs share the same hyperparameters: $\rho = 0.5$, $\gamma = 1$, and $\alpha = 0.1$ (see (2.7)). These values have not been tuned but are presumed credible given other ESN implementations in the literature. To make a fair comparison with DFMs, we fit the S-MFESN models using six-day-averaged daily data. Note here that for MFESN models the computational gains of averaging are negligible, and are most apparent when tuning the ridge penalty via cross-validation.

Our second set of proposed models consists of two M-MFESNs according to Example 3.2. Both models have two reservoirs, one for each data frequency – monthly and daily – with 100 and 20 neurons, respectively. Sparsity degrees are again adjusted to be $10/N$, where N is the reservoir state dimension. The first M-MFESN has hyperparameters that are hand-selected among reasonable values: we note that the monthly-frequency reservoir has no state leak and a larger input scaling, while the daily-frequency reservoir features smaller scaling than usual (to avoid compressing high-volatility events with the activation function) and the same leak rate as in the S-MFESN models (we call this specification multiESN [A]). For the second M-MFESN, we change the hyperparameters more radically: we aim to set up a model that has a very high input memory (Ballarin et al., 2023), and that features long-term smoothing of states. Note that here, input scaling values are small, spectral radii are an order of magnitude smaller than in previous models, and leak rates are large (we term this model multiESN [B]).

4.3. Results

We start by commenting on the computational efficiency of competing models and report execution times (in seconds) in Table 4.2. Firstly, DFM models appear to be the most computationally effortful models among all specifications. For the small-MD dataset, the simplest MFESN models, that is, singleESN [A] and [B], have execution times which are at most 3.5 times higher than the MIDAS model, while still being at least 15.6 times computationally cheaper than any of the DFM models. The more resource-demanding MFESN models, multiESN [A] and [B], are nevertheless at least 2.6 times faster to run than the best DFM model (DFM [A]). When moving to the medium-MD dataset – where the MIDAS model is not a feasible choice, as explained above – the most inefficient MFESN model (singleESN [B]) still outperforms the best DFM model, DFM [A], by 8.4 times, while the same holds for the multiESN [A] model versus DFM [A] model by 2.7 times. We can conclude that our proposed MFESN architectures provide an attractive and computationally efficient framework for GDP forecasting in the multifrequency framework, which is feasible for computations on low-cost machine configurations available to practitioners.

Competing forecasts are compared using the model confidence set (MCS) test derived by Hansen, Huang, and Shek (2011). One should note that due to the intrinsic nature of data availability of macroeconomic time series and panels, our sample sizes are modest. This implies that the small-sample sensitivities of the MCS test need to be taken into account when evaluating our comparisons. Recent analyses of the finite sample properties of the MCS methodology have shown that it requires signal-to-noise ratios which are unattainable in most empirical settings, an issue that undermines its applicability (Aparicio & de Prado, 2018). Given this fact, we also conduct pairwise model comparison tests with the modified Diebold–Mariano (MDM) test for predictive accuracy (Diebold & Mariano, 2002; Harvey, Leybourne, & Newbold, 1997).

Table 4.3
Relative MSFEs and model confidence set (MCS) comparisons between models in one-step-ahead forecasting exercises. The unconditional mean MSFE is used as a reference. MCS columns show inclusion among best models.

One-Step-ahead GDP forecasting – small-MD dataset												
Model	Fixed parameters				Expanding window				Rolling window			
	2007		2011		2007		2011		2007		2011	
	MSFE	MCS	MSFE	MCS	MSFE	MCS	MSFE	MCS	MSFE	MCS	MSFE	MCS
Mean	1.000	*	1.000	**	1.000	**	1.000	**	1.000	**	1.000	**
AR(1)	0.758	*	1.230	**	0.789	**	1.226	**	0.775	**	1.209	**
MIDAS	**0.533**	**	1.300		0.596	**	1.129	*	0.709	**	1.170	*
DFM [A]	0.799	*	1.337		0.980	*	1.320		0.919	*	1.226	
DFM [B]	0.885		1.221	**	0.982	*	1.022	**	0.948		1.028	**
singleESN [A]	0.721	**	1.015	**	0.597	**	0.867	**	**0.529**	**	0.863	**
singleESN [B]	0.758	*	**0.921**	**	0.602	**	**0.844**	**	0.561	**	0.930	**
multiESN [A]	0.802	*	1.250		0.635	**	0.874	**	0.621	**	**0.859**	**
multiESN [B]	0.590	**	0.969	**	**0.552**	**	0.895	**	0.530	**	0.921	**

* Indicates inclusion at 90% confidence.
** Indicates inclusion at 75% confidence.

As we also provide multiple-step-ahead forecasts, we test for the best subset of models uniformly across all horizons using the uniform multi-horizon MCS (uMCS) test proposed by Quaedvlieg (2021). Since there is relatively little systematic knowledge regarding the power properties of the uMCS test in small samples, our inclusion of this procedure is meant as a statistical counterpoint to simple relative forecasting error comparisons, which provide limited information about the significance of performance differences. We provide more details on our implementation of the test in Supplementary Appendix F. Finally, we do not report uMCS test outcomes for the expanding window setup, as Quaedvlieg (2021) argues that in such contexts the test is invalid.

4.3.1. Small dataset

We begin our discussion of the small-MD forecasting results by reviewing Table 4.3. For both sample setups (2007 and 2011) and all three estimation strategies (fixed, expanding, and rolling windows) we provide relative MSFE metrics, with the unconditional mean being used as a reference. Plots of each of the model's forecasts are given in Figs. 3 and 4 in Appendix B; additional plots for cumulative SFE, cumulative RMSFEs, and other metrics can be found in Supplementary Appendix K.

The overall finding is that MFESN models perform excellent, and, when we exclude the 2007 fixed parameters setup, they perform the best. It is easy to see from Fig. 3(a) why the 2007 fixed window estimation case is different from the other cases: the 2008 financial crisis induced a deep drop in quarter-to-quarter GDP growth that was in stark contrast with previous business-cycle fluctuations. By keeping the model parameters fixed and using only information from 1990 to 2007 – periods where systematic fluctuations were small – the DFM and MFESN models are fit to produce smooth, low-volatility forecasts. MIDAS, on the other hand, yields an exponential smoothing which can be more responsive to changes in monthly and daily series. From Figs. 3(b) and (c), it is possible to see that expanding and rolling window estimation resolves this weakness of state-space models. At the same time, the AR(1) model outperforms the unconditional mean only in the 2007 sample with fixed parameters, losing to the MIDAS model in all but one scenario.

Table 4.3 shows that MFESN models always perform better than the mean in terms of the MSFE, something which no other model class achieves across all setups. In both expanding and rolling window setups they also always outperform the AR(1) model. Furthermore, at least one MFESN model for each subclass (single or multi-reservoir) is always included in the model confidence set at the highest confidence level. Again, recall that the MCS test of Hansen et al. (2011) might be distorted due to the modest sample sizes considered, and even more so in the 2011 test sample. To complement the MCS, we provide graphical tables for pairwise modified Diebold–Mariano (MDM) tests, with 10% level rejections highlighted in Fig. 14, Supplementary Appendix K. The MDM tests broadly agree with the results of Table 4.3, although they do not account for multiple testing and therefore cannot be interpreted as yielding subsets of the most accurate forecasting models in a statistical sense.

For multiple-step-ahead forecasts, the relative RMSFE and uMCS are reported in Tables 4.4 and 4.5: we constrain our exercise to $h \in \{1, \ldots, 8\}$ steps, since we are interested in GDP growth forecasts within two years. Note that for $h = 1$ our results are similar, but do not reduce to the one-step-ahead results. To make correct multistep RMSFE evaluations and execute the uMCS procedure, one must select h different vectors of residuals of the same length: this implies that residuals at the end of the forecasting sample must be trimmed off to compute short-term multistep RMSFEs that are comparable to the long-term ones. Generally, we notice that the MIDAS models, as well as the S-MFESNs, provide the worst-performing multistep

Table 4.4
Relative RMSFEs and uniform multi-horizon model confidence set (uMCS) comparisons between models in multiple-step-ahead forecasting exercises. The unconditional mean RMSFE used as reference. FIX: fixed parameters, EW: expanding window, and RW: rolling window. The uMCS columns show inclusion among best models.

Multistep-ahead GDP forecasting – small-MD dataset – 2007 sample										
Setup	Model	Horizon								uMCS
		1	2	3	4	5	6	7	8	
FIX	Mean	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
FIX	AR(1)	0.870	**0.950**	**0.982**	0.991	0.992	0.991	0.992	0.992	**
FIX	MIDAS	0.823	1.672	2.737	1.816	2.213	2.791	1.888	1.921	
FIX	DFM [A]	0.890	0.969	1.014	1.077	1.341	1.701	2.001	2.180	*
FIX	DFM [B]	0.937	1.069	1.202	1.344	1.799	2.310	2.638	2.801	
FIX	singleESN [A]	0.852	0.994	0.995	0.995	0.993	0.991	0.991	0.991	*
FIX	singleESN [B]	0.871	0.986	0.989	**0.989**	**0.985**	**0.981**	**0.981**	**0.981**	**
FIX	multiESN [A]	0.898	0.980	0.990	0.991	0.988	0.985	0.985	0.985	**
FIX	multiESN [B]	**0.767**	0.954	0.983	0.991	0.991	0.990	0.991	0.991	**
EW	Mean	1.000	1.000	1.000	1.000	1.000	1.000	**1.000**	**1.000**	–
EW	AR(1)	0.887	0.922	0.951	0.962	0.957	**0.981**	1.001	1.008	–
EW	MIDAS	0.814	1.283	1.518	1.596	1.697	1.391	1.951	1.800	–
EW	DFM [A]	0.985	1.109	1.123	1.114	1.217	1.226	1.241	1.539	–
EW	DFM [B]	0.989	1.082	1.149	1.199	1.315	1.412	1.373	1.425	–
EW	singleESN [A]	0.771	1.260	1.485	1.564	2.070	2.728	2.550	2.834	–
EW	singleESN [B]	0.772	1.031	1.135	1.319	1.831	2.279	2.449	2.556	–
EW	multiESN [A]	0.792	0.897	0.941	0.976	1.015	1.240	1.377	1.227	–
EW	multiESN [B]	**0.740**	**0.853**	**0.894**	**0.911**	**0.873**	0.993	1.020	1.020	–
RW	Mean	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	*
RW	AR(1)	0.898	0.943	0.968	0.974	0.963	**0.968**	0.970	**0.962**	**
RW	MIDAS	0.933	1.438	1.642	1.993	1.794	1.661	1.816	1.973	*
RW	DFM [A]	0.931	1.017	1.033	1.020	1.024	1.003	**0.918**	1.062	*
RW	DFM [B]	0.942	0.973	0.970	1.045	1.059	1.203	1.225	1.263	*
RW	singleESN [A]	**0.714**	1.320	1.693	1.972	2.733	3.669	3.391	3.719	*
RW	singleESN [B]	0.737	1.100	1.248	1.667	2.327	2.765	2.842	2.792	*
RW	multiESN [A]	0.773	0.972	1.053	1.111	1.187	1.293	1.505	1.131	*
RW	multiESN [B]	0.716	**0.895**	**0.916**	**0.926**	**0.890**	1.041	1.102	1.105	**

* Indicates inclusion at 90% confidence.
** Indicates inclusion at 75% confidence.

forecasts, with RMSFEs considerably exceeding the unconditional mean baseline after horizon 1. Figs. 5 and 6 in Appendix B reproduce the RMSFE numbers of the aforementioned tables graphically.

For MIDAS, we already discussed how the existence of multiple loss minima can generate numerical instabilities. Model re-fitting at each horizon can amplify this problem, as the loss landscape itself changes as new observations are added to the fitting sample. We provide more discussion in Supplementary Appendix J.1. In the case of S-MFESN models, the reason is structural: we discussed how in our framework multistep MFESN forecasting entails iterating the state map, which can have multiple attraction (stable) points. If the hyperparameters and estimated full model $\widehat{W}$s jointly do not define a contraction, the limit of the multistep forecast does not have to be the estimated MFSEN model intercept. However, Figs. 5 and 6 show that our M-MFESN models, multiESN [A] and multiESN [B], both perform on par or better than the DFM models, even after horizon $h = 4$. For example, in the 2007 expanding and rolling window experiments, multiESN [B] is able to outperform both DFMs and an unconditional mean forecast by meaningful margins for forecasts up to a year into the future.

4.3.2. Medium dataset

We now present the results for the medium-MD dataset, which includes more than 30 regressors and many high-frequency daily series. The same metrics as in the previous subsection are used for this dataset to evaluate the relative performance of different methods.

The main difference in our empirical exercises is that now we a priori exclude MIDAS from the set of forecasting methods, as explained in detail in Section 4.2.1. Table 4.6 showcases the relative performance of the DFM and MFESN models in the medium-MD forecast setup. We

Table 4.5
Relative RMSFEs and uniform multi-horizon model confidence set (uMCS) comparisons between models in multiple-step-ahead forecasting exercises. The unconditional mean RMSFE used as reference. FIX: Fixed parameters, EW: expanding window, and RW: rolling window. The uMCS columns show inclusion among best models.

Multistep-ahead GDP forecasting – small-MD dataset – 2011 sample										
Setup	Model	Horizon								uMCS
		1	2	3	4	5	6	7	8	
FIX	Mean	1.000	1.000	1.000	1.000	**1.000**	1.000	1.000	1.000	**
FIX	AR(1)	1.119	1.031	1.008	1.001	1.001	0.999	**0.999**	**0.998**	*
FIX	MIDAS	1.090	1.721	1.793	2.203	2.363	1.997	2.846	2.328	
FIX	DFM [A]	1.112	1.051	0.999	1.079	1.084	1.025	1.020	1.061	*
FIX	DFM [B]	1.058	**0.945**	**0.916**	1.003	1.012	**0.970**	1.038	1.033	**
FIX	singleESN [A]	0.978	1.705	2.561	2.704	3.314	3.151	2.999	3.316	
FIX	singleESN [B]	**0.930**	1.095	1.885	2.356	2.650	2.704	2.880	2.844	**
FIX	multiESN [A]	1.059	1.148	1.262	1.312	1.339	1.409	1.424	1.162	
FIX	multiESN [B]	0.981	1.007	0.985	**0.994**	1.008	0.999	**0.999**	**0.998**	**
EW	Mean	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	–
EW	AR(1)	1.117	1.033	1.011	1.002	1.007	1.003	1.004	1.003	–
EW	MIDAS	1.005	1.382	1.339	1.354	1.609	1.444	1.803	1.263	–
EW	DFM [A]	1.144	1.132	1.057	1.093	1.076	1.067	1.038	1.016	–
EW	DFM [B]	0.985	**0.940**	**0.918**	0.995	1.010	**0.980**	1.050	**0.971**	–
EW	singleESN [A]	0.935	1.645	2.184	1.929	2.388	1.959	1.810	2.266	–
EW	singleESN [B]	**0.911**	1.092	1.101	1.529	2.195	1.843	1.847	2.060	–
EW	multiESN [A]	0.922	0.965	1.089	0.978	**0.977**	1.043	1.278	0.995	–
EW	multiESN [B]	0.944	0.992	0.978	**0.977**	0.991	0.985	**0.990**	0.996	–
RW	Mean	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
RW	AR(1)	1.080	1.000	0.984	**0.989**	0.982	0.976	**0.963**	0.968	
RW	MIDAS	1.051	1.303	1.310	1.674	1.762	1.467	1.643	1.463	
RW	DFM [A]	1.061	1.033	1.012	1.088	1.077	1.015	1.040	1.069	
RW	DFM [B]	0.947	**0.893**	**0.901**	1.009	1.040	**0.966**	1.030	**0.949**	**
RW	singleESN [A]	0.919	1.788	2.359	2.483	2.981	2.401	2.234	2.690	
RW	singleESN [B]	0.944	1.132	1.214	1.762	2.608	2.552	2.517	2.541	
RW	multiESN [A]	**0.896**	1.047	1.222	1.124	1.122	1.410	1.666	1.316	
RW	multiESN [B]	0.940	1.003	0.969	**0.989**	**0.979**	0.972	0.967	0.961	**

* Indicates inclusion at 90% confidence.
** Indicates inclusion at 75% confidence.

Table 4.6
Relative MSFEs and model confidence set (MCS) comparisons between models in one-step-ahead forecasting exercises. The unconditional mean MSFE is used as a reference. The MCS columns show inclusion among best models.

One-step-ahead GDP forecasting – medium-MD dataset												
Model	Fixed parameters				Expanding window				Rolling window			
	2007		2011		2007		2011		2007		2011	
	MSFE	MCS	MSFE	MCS	MSFE	MCS	MSFE	MCS	MSFE	MCS	MSFE	MCS
Mean	1.000		1.000	**	1.000	**	1.000	**	1.000	**	1.000	**
AR(1)	0.758	*	1.230	**	0.789	**	1.226	**	0.775	*	1.209	**
DFM [A]	0.841	*	1.325	*	0.682	**	1.272	**	0.747	*	1.517	**
DFM [B]	1.118	*	1.408	**	0.821	*	1.117	**	0.926		1.186	**
singleESN [A]	0.967	*	1.717	*	0.775	**	1.072	**	0.791	*	1.493	*
singleESN [B]	0.826	*	1.278	**	0.655	**	1.028	**	0.561	**	0.944	**
multiESN [A]	0.901	*	1.080	**	0.618	**	0.913	**	0.556	**	0.884	**
multiESN [B]	**0.682**	**	**0.748**	**	**0.587**	**	**0.774**	**	**0.547**	**	**0.728**	**

* Indicates inclusion at 90% confidence.
** Indicates inclusion at 75% confidence.

Table 4.7
Relative RMSFEs and uniform multi-horizon model confidence set (uMCS) comparisons between models in multiple-steps-ahead forecasting exercises. The unconditional mean RMSFE used as reference. FIX: fixed parameters, EW: expanding window, and RW: rolling window. The uMCS columns show inclusion among best models.

Multistep-ahead GDP forecasting – medium-MD dataset – 2007 sample.										
Setup	Model	Horizon								uMCS
		1	2	3	4	5	6	7	8	
FIX	Mean	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	*
FIX	AR(1)	0.870	0.950	0.982	0.991	0.992	0.991	0.992	0.992	
FIX	DFM [A]	0.914	**0.947**	**0.955**	0.988	1.015	1.027	1.034	0.995	**
FIX	DFM [B]	1.046	1.204	1.293	1.341	1.649	1.984	2.101	2.070	*
FIX	singleESN [A]	0.985	0.995	0.995	0.995	0.994	0.992	0.992	0.992	*
FIX	singleESN [B]	0.912	0.985	0.985	**0.985**	**0.980**	**0.976**	**0.976**	**0.976**	*
FIX	multiESN [A]	0.950	0.993	0.994	0.994	0.992	0.990	0.990	0.990	*
FIX	multiESN [B]	**0.826**	0.972	0.988	0.990	0.989	0.986	0.985	0.985	*
EW	Mean	1.000	1.000	1.000	1.000	1.000	1.000	**1.000**	**1.000**	–
EW	AR(1)	0.887	0.922	0.951	0.962	0.957	0.981	1.001	1.008	–
EW	DFM [A]	0.805	0.916	0.978	1.038	1.077	1.126	1.077	1.073	–
EW	DFM [B]	0.893	1.134	1.418	1.567	2.238	2.964	3.375	3.629	–
EW	singleESN [A]	0.879	1.125	1.305	1.442	1.860	2.166	2.361	2.443	–
EW	singleESN [B]	0.802	1.174	1.439	1.744	2.305	2.869	2.935	3.167	–
EW	multiESN [A]	0.780	0.935	1.012	1.005	1.093	1.337	1.328	1.313	–
EW	multiESN [B]	**0.760**	**0.874**	**0.911**	**0.891**	**0.863**	**0.971**	1.030	1.051	–
RW	Mean	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
RW	AR(1)	0.898	0.943	0.968	0.974	0.963	**0.968**	**0.970**	**0.962**	
RW	DFM [A]	0.837	0.913	0.924	0.954	1.012	0.997	1.018	1.005	
RW	DFM [B]	0.932	1.116	1.232	1.414	1.952	2.704	3.183	3.294	
RW	singleESN [A]	0.873	1.274	1.530	1.652	2.095	2.575	2.786	3.014	
RW	singleESN [B]	0.732	1.190	1.490	1.712	2.218	2.861	2.967	3.094	
RW	multiESN [A]	0.732	0.914	0.960	1.011	1.202	1.618	1.683	1.572	
RW	multiESN [B]	**0.731**	**0.871**	**0.875**	**0.844**	**0.771**	0.971	1.014	1.014	**

* Indicates inclusion at 90% confidence.
** Indicates inclusion at 75% confidence.

find that the MFESN model multiESN [B] performs best in all setups, particularly under fixed parameters, where MCS testing reveals that it is the only model included at a 75% confidence level. Of course, for the MCS results we must again take into account the relatively small sample size, which could distort the selection of best model subsets. MDM tests of Fig. 16 in Supplementary Appendix K largely agree with the MCS results: in the fixed-parameter setup, any pairwise comparison of an alternative model against MFESN multiESN [B] is rejected in favor of the latter. A visual inspection of one-step-ahead forecasts in Figs. 7 and 8 in Appendix B also shows that DFM models estimated over the medium-MD datasets produce forecasts with larger variability than MFESN methods, which is likely the key driver of the difference in performance.

The multistep-ahead experiments are run as for the small-MD dataset, with a maximum horizon of eight quarters. Tables 4.7 and 4.8 present the relative RMSFE performance of multistep forecasts for all models, and we use Figs. 9 and 10 of RMSFEs as references for our discussion, available in Appendix B. What can be seen visually – and what is also reproduced in the tables – is that the multi-reservoir MFESN models and DFM model [A] have the best performance up to four quarters ahead; overall, taking into account also the longer term, expanding or rolling window estimation of model multiESN [B] yields the best forecasting results in the 2007 sample setup. The post-crisis 2011 sample setup makes the comparison harder, as the DFM and M-MFESN models largely produce results in line with the unconditional sample mean. This evaluation is confirmed by uMCS tests, consistently with the multistep results obtained with the small-MD dataset.

5. Conclusions

Macroeconomic forecasting – especially long-term forecasting of macroeconomic aggregates – is a topic of crucial importance for institutional policymakers, private companies, and economic researchers. Given the modern-day availability of big-data resources, methods capable of integrating heterogeneous data sources are increasingly sought to provide more precise and robust forecasts.

This paper presented a new methodological framework inspired by the reservoir computing literature to deal with data sampled at multiple frequencies and with multiple-step-ahead forecasts. We then took echo state networks – a type of RC models – and formally extended them to model data with multiple release frequencies. Our discussion encompassed model fitting, hyperparameter tuning, and forecast computation. As a result, we

Table 4.8
Relative RMSFEs and uniform multi-horizon model confidence set (uMCS) comparisons between models in multiple-step-ahead forecasting exercises. The unconditional mean RMSFE used as reference. FIX: Fixed parameters, EW: expanding window, and RW: rolling window. The uMCS columns show inclusion among best models.

Multistep-ahead GDP forecasting – medium-MD dataset – 2011 sample										
Setup	Model	Horizon								uMCS
		1	2	3	4	5	6	7	8	
FIX	Mean	1.000	1.000	1.000	1.000	**1.000**	1.000	1.000	1.000	*
FIX	AR(1)	1.119	1.031	1.008	1.001	1.001	0.999	**0.999**	**0.998**	**
FIX	DFM [A]	1.126	0.987	0.962	1.054	1.031	**0.988**	1.001	1.002	**
FIX	DFM [B]	1.149	0.987	**0.885**	1.064	1.142	1.134	1.273	1.296	
FIX	singleESN [A]	1.283	1.921	2.527	3.038	3.285	3.154	3.193	3.655	
FIX	singleESN [B]	1.059	1.523	1.918	2.417	2.812	2.683	2.703	2.970	
FIX	multiESN [A]	1.011	1.061	1.434	1.477	1.748	2.030	2.023	1.994	
FIX	multiESN [B]	**0.841**	**0.945**	0.997	**0.978**	1.004	1.015	1.013	1.014	**
EW	Mean	1.000	1.000	1.000	1.000	**1.000**	1.000	1.000	1.000	-
EW	AR(1)	1.117	1.033	1.011	1.002	1.007	1.003	1.004	1.003	-
EW	DFM [A]	1.092	0.942	**0.944**	1.049	1.026	**0.994**	**0.996**	**0.999**	-
EW	DFM [B]	0.971	1.046	1.031	1.114	1.238	1.116	1.223	1.310	-
EW	singleESN [A]	1.039	1.451	1.980	2.385	2.699	2.353	2.506	2.608	-
EW	singleESN [B]	0.992	1.828	2.465	3.072	3.547	3.357	3.368	3.610	-
EW	multiESN [A]	0.934	1.014	1.391	1.252	1.371	1.369	1.228	1.279	-
EW	multiESN [B]	**0.857**	**0.931**	1.003	**0.973**	1.002	1.009	1.025	1.029	-
RW	Mean	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	**
RW	AR(1)	1.080	1.000	0.984	**0.989**	**0.982**	**0.976**	**0.963**	**0.968**	**
RW	DFM [A]	1.113	0.982	**0.927**	1.038	1.030	0.997	1.016	1.028	*
RW	DFM [B]	0.881	0.996	1.021	1.098	1.150	1.114	1.114	1.212	**
RW	singleESN [A]	1.193	2.267	3.265	3.580	4.090	3.790	4.015	4.562	
RW	singleESN [B]	0.927	1.933	2.612	3.265	3.753	3.567	3.556	3.792	
RW	multiESN [A]	0.900	1.049	1.500	1.465	1.789	1.707	1.505	1.462	
RW	multiESN [B]	**0.816**	**0.916**	0.977	1.009	**0.982**	0.988	0.974	0.981	**

* Indicates inclusion at 90% confidence.
** Indicates inclusion at 75% confidence.

provided two classes of models, single- and multiple-reservoir multi-frequency ESNs, that can be effectively applied to our empirical setup: forecasting U.S. GDP growth using monthly and daily data series. Along with the unconditional mean and AR(1) model, we considered two well-known methods, MIDAS and DFMs, as the current benchmarks available in the literature. In our applications, we found that MFESN models were computationally more efficient and easier to implement than DFMs and MIDAS, respectively, and performed better than or as well as the benchmarks in terms of the MSFE. These improvements were statistically significant in a number of setups, as shown by our MCS and MDM tests. Thus, we argue that our machine learning-based methodology can be a useful addition to the toolbox of contemporary macroeconomic forecasters.

Lastly, we wish to highlight the many potential areas of research that we believe would be interesting to explore in the future. We did not discuss the role of the distribution from which we sample the entries of the reservoir matrices. While it is known that these can have significant effects on the forecasting capacity of an ESN model, the literature lacks definitive theoretical results (even for dynamical systems applications) or systematic studies with stochastic inputs and targets. The hyperparameter tuning routine we developed cannot separate individual hyperparameters or tackle the identification problem. Moreover, we assumed that the ridge regression penalty strength, λ, is tuned ex ante: it would be interesting and desirable to understand if it is possible to jointly tune λ and $\boldsymbol{\varphi}$, or rather if one can fully separate their selection. In our preliminary experiments, we noticed that the roles of the ridge penalty and the input scaling, for example, cannot be trivially disentangled, thus prompting the ψ-form normalization. Model selection for the dimension of MFESN models is another question that would be key to exploring and designing more efficient and effective ESN models, especially when dealing with multiple frequencies and reservoirs. Finally, practitioners may be interested in identifying the combination of frequencies in the regressor series that would lead to the most accurate GDP forecasts produced by MFESN models.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Juan-Pablo Ortega is an Associate Editor of IJF.

Appendix A. Data table

See Table A.9.

Table A.9
Variables, frequencies, and transformations for small and medium-sized datasets.

SM	Start Date	T	Code	Name	Description
Quarterly					
XX	31/03/1959	5	GDPC1	Y	Real Gross Domestic Product
Monthly					
XX	30/01/1959	5	INDPRO	XM1	Industrial Production Index
XX	30/01/1959	5	PAYEMS	XM4	Payroll All Employees: Total nonfarm
XX	30/01/1959	4	HOUST	XM5	Housing Starts: Total New Privately Owned
XX	30/01/1959	5	RETAILx	XM7	Retail and Food Services Sales
XX	31/01/1973	5	TWEXMMTH	XM11	Nominal effective exchange rate US
XX	30/01/1959	2	FEDFUNDS	XM12	Effective Federal Funds Rate
XX	30/01/1959	1	BAAFFM	XM14	Moody's Baa Corporate Bond Minus FEDFUNDS
XX	30/01/1959	1	COMPAPFFx	XM15	3-Month Commercial Paper Minus FEDFUNDS
X	30/01/1959	2	CUMFNS	XM2	Capacity Utilization: Manufacturing
X	30/01/1959	2	UNRATE	XM3	Civilian Unemployment Rate
X	30/01/1959	5	DPCERA3M086SBEA	XM6	Real personal consumption expenditures
X	30/01/1959	5	AMDMNOx	XM8	New Orders for Durable Goods
X	31/01/1978	2	UMCSENTx	XM9	Consumer Sentiment Index
X	30/01/1959	6	WPSFD49207	XM10	PPI: Finished Goods
X	30/01/1959	1	AAAFFM	XM13	Moody's Aaa Corporate Bond Minus FEDFUNDS
X	30/01/1959	1	TB3SMFFM	XM16	3-Month Treasury C Minus FEDFUNDS
X	30/01/1959	1	T10YFFM	XM17	10-Year Treasury C Minus FEDFUNDS
X	30/01/1959	2	GS1	XM18	1-Year Treasury Rate
X	30/01/1959	2	GS10	XM19	10-Year Treasury Rate
X	30/01/1959	1	GS10-TB3MS	XM20	10-Year Treasury Rate - 3-Month Treasury Bill
Daily					
XX	30/01/1959	8	DJINDUS	XD3	DJ Industrial price index
X	31/12/1963	8	S&PCOMP	XD1	S&P500 price index
X	01/05/1982	1	ISPCS00-S&PCOMP[a]	XD2	S&P500 basis spread
X	11/09/1989	8	SP5EIND	XD4	S&P Industrial price index
X	31/12/1969	8	GSCITOT	XD5	Spot commodity price index
X	10/01/1983	8	CRUDOIL	XD6	Spot price oil
X	02/01/1979	8	GOLDHAR	XD7	Spot price gold
X	30/03/1982	8	WHEATSF	XD8	Spot price wheat
X	01/11/1983	8	COCOAIC,COCINUS[b]	XD9	Spot price cocoa
X	30/03/1983	1	NCLC.03-NCLC.01	XD10	Futures price oil term structure
X	30/10/1978	1	NGCC.03-NGCC.01	XD11	Futures price gold term structure
X	02/01/1975	1	CWFC.03-CWFC.01	XD12	Futures price wheat term structure
X	02/01/1973	1	NCCC.03-NCCC.01	XD13	Futures price cocoa term structure

Notes: S and M stand for small and medium datasets, respectively. An 'X' indicates selection into the dataset. 'Start Date' is the date for which the series is first available (before data transformations). Following (McCracken & Ng, 2016, 2020), the transformation codes in column 'T' indicate with D for difference and log for natural logarithm 1: none, 2: D, 3: DD, 4: Log, 5: Dlog, 6: DDlog, 7: percentage change, 8: GARCH volatility. 'Codes' are the codes in the FRED-QD and FRED-MD datasets for quarterly and monthly data and Datastream mnemonic for the remaining frequencies. Missing values due to public holidays are interpolated by averaging over the previous five observations.
[a] Available until 20/09/2021.
[b] Average before 29/12/2017, COCINUS mean adjusted thereafter.

Appendix B. Forecasting figures

See Figs. 3–10.

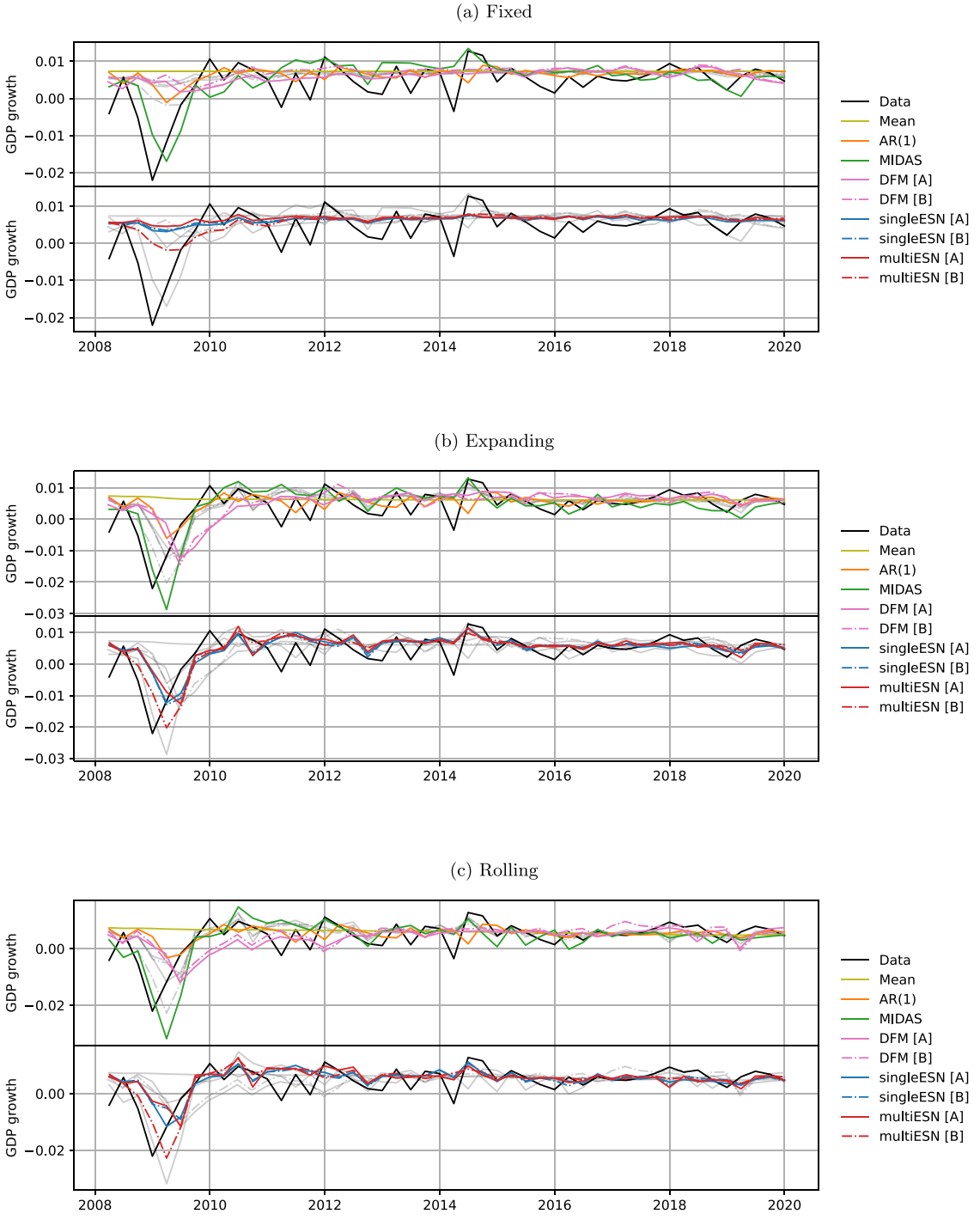
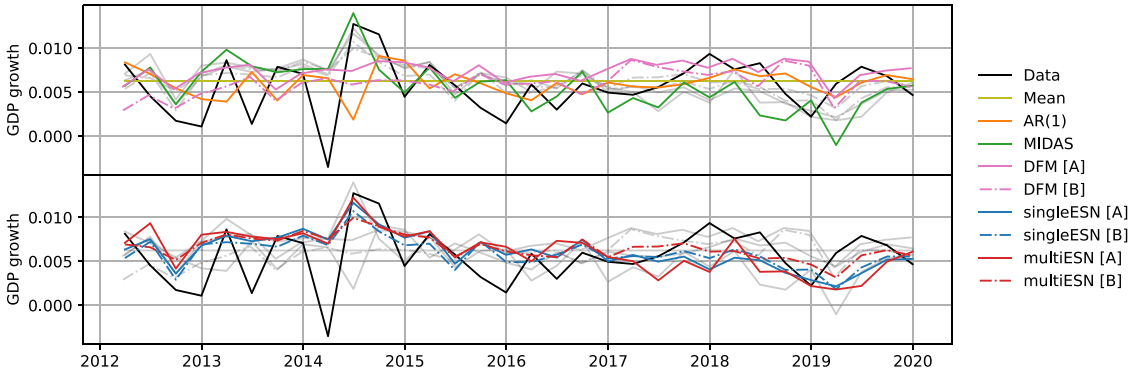
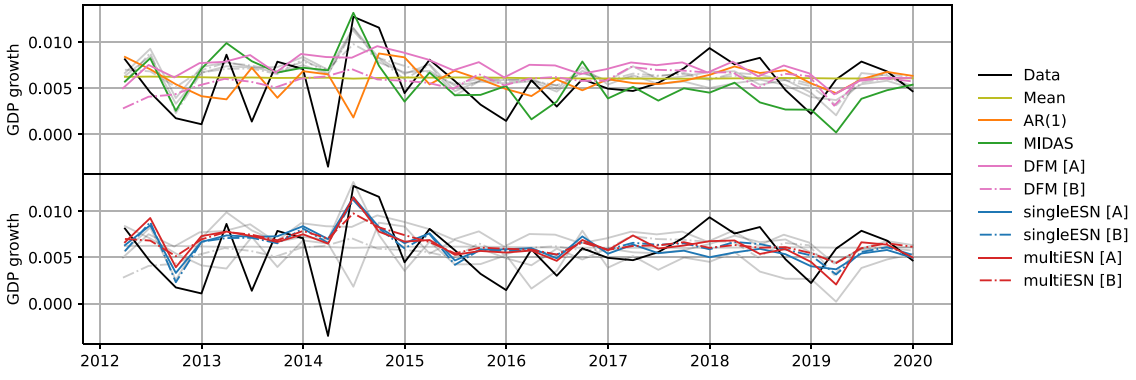


Fig. 3. One-step-ahead GDP forecasting – 2007 sample – small-MD dataset.

(a) Fixed



(b) Expanding



(c) Rolling

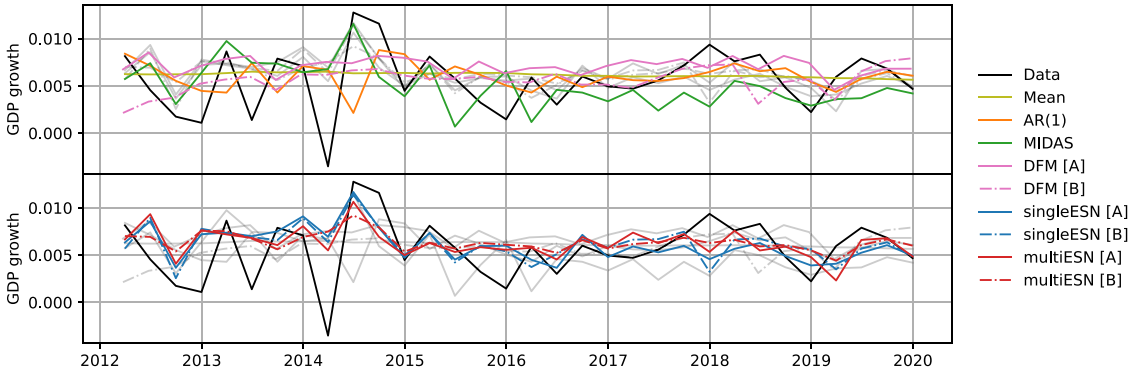
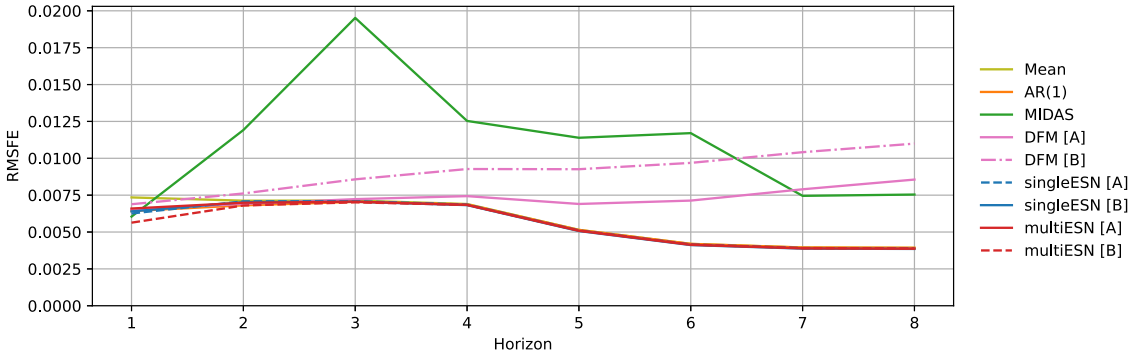
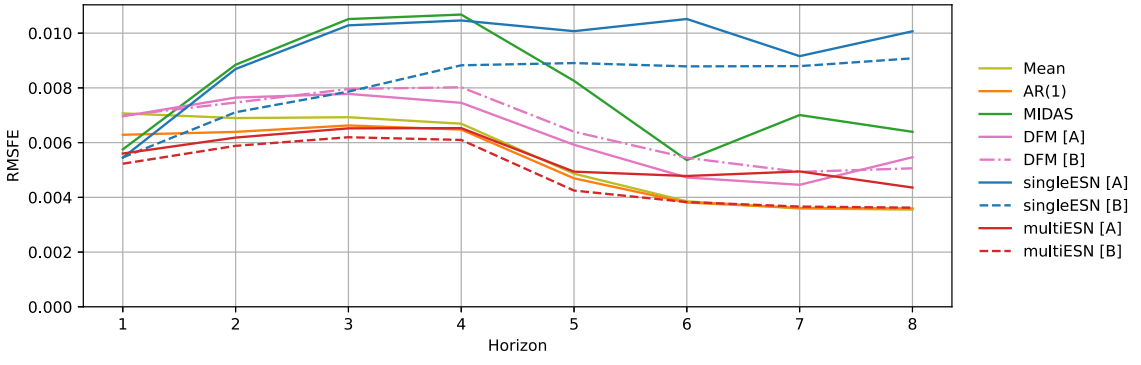


Fig. 4. One-step-ahead GDP forecasting – 2011 Sample – small-MD dataset.

(a) Fixed



(b) Expanding



(c) Rolling

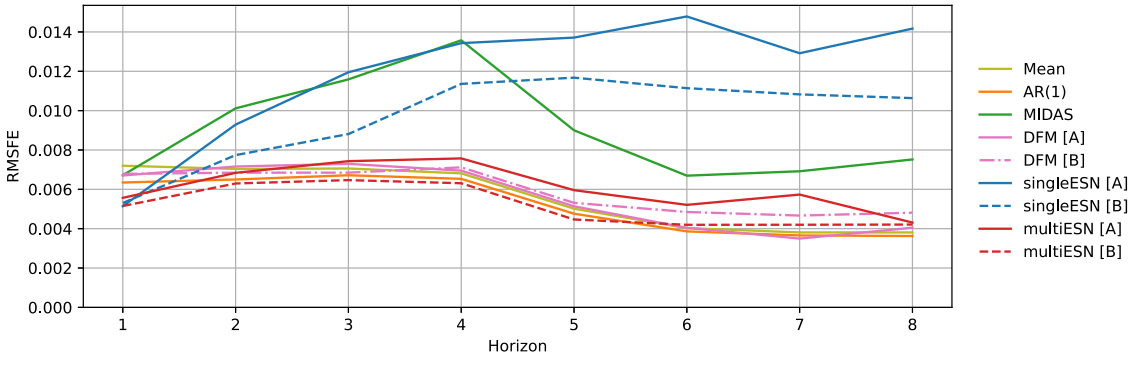


Fig. 5. Multistep-ahead GDP forecasting, RMSFE – 2007 sample – small-MD dataset.

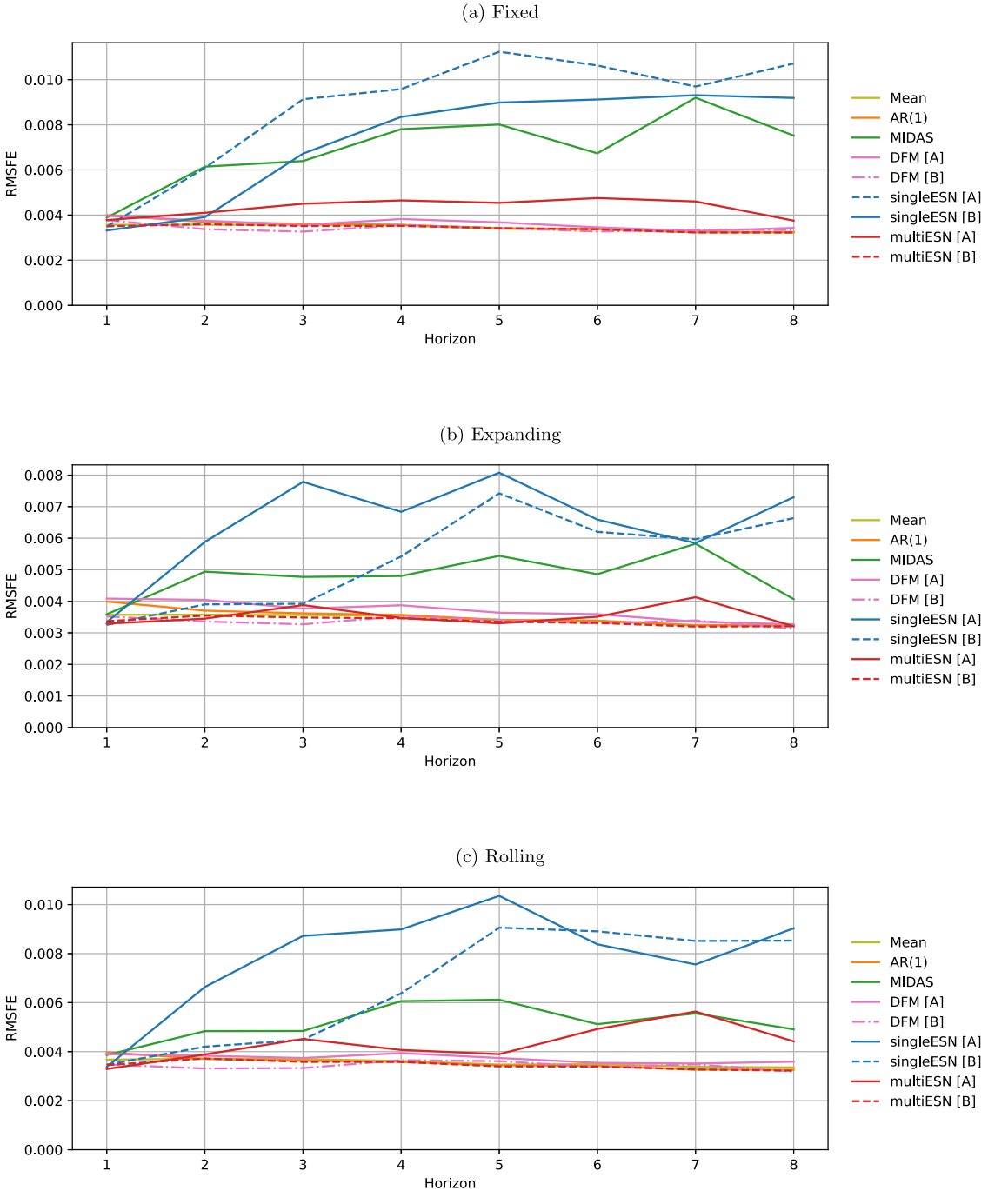
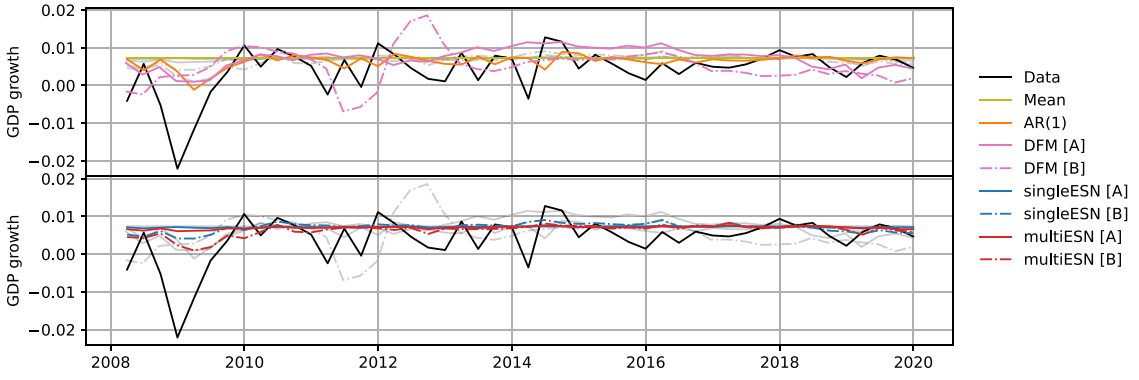
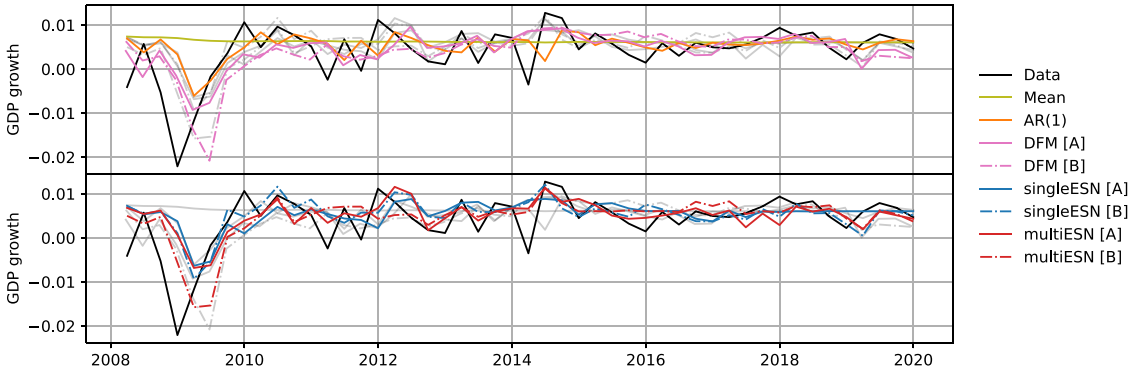


Fig. 6. Multistep-ahead GDP forecasting, RMSFE – 2011 sample – small-MD dataset.

(a) Fixed



(b) Expanding



(c) Rolling

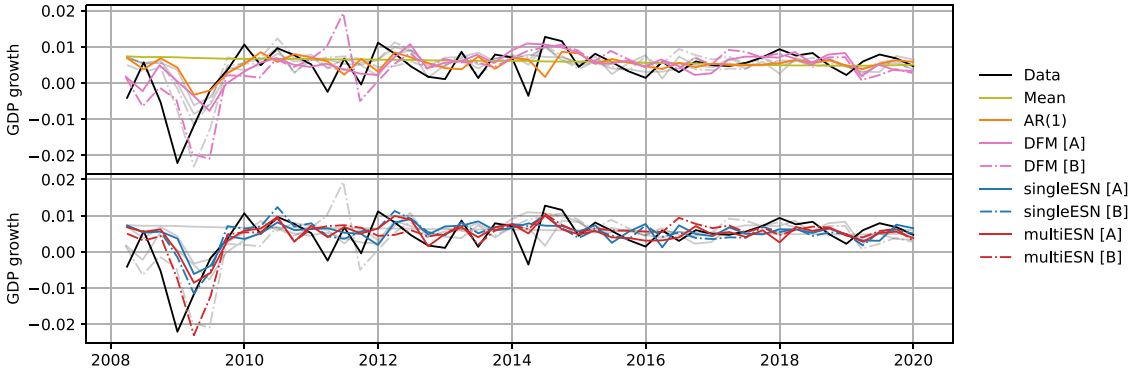
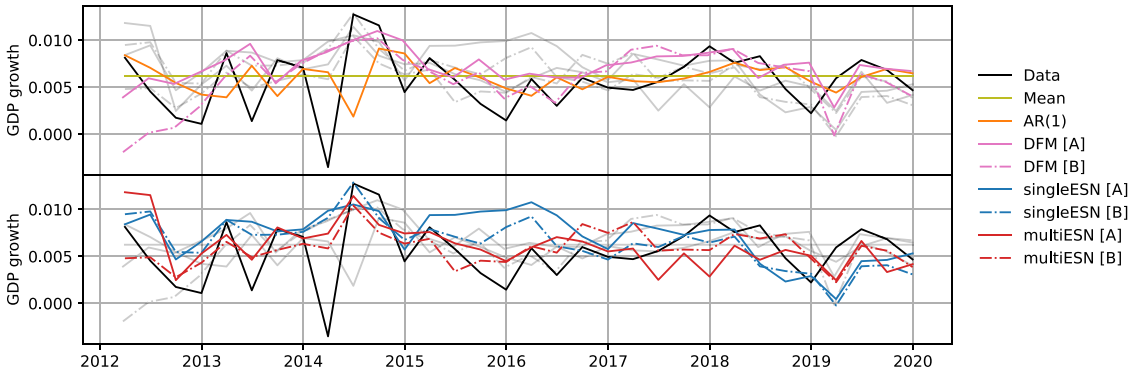
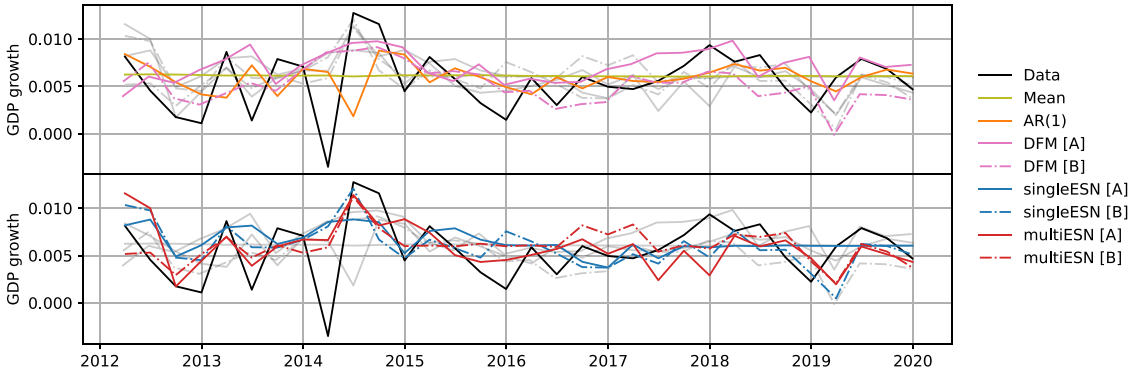


Fig. 7. One-step-ahead GDP forecasting – 2007 sample – medium-MD dataset.

(a) Fixed



(b) Expanding



(c) Rolling

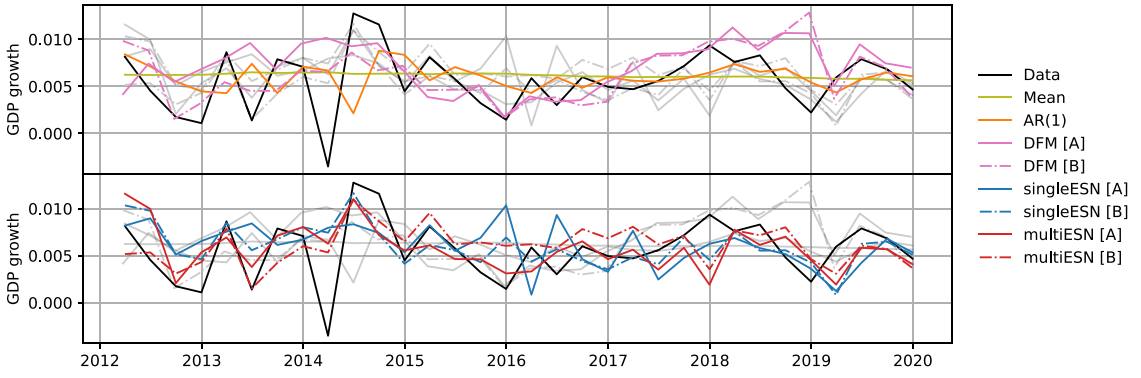


Fig. 8. One-step-ahead GDP forecasting – 2011 sample – medium-MD dataset.

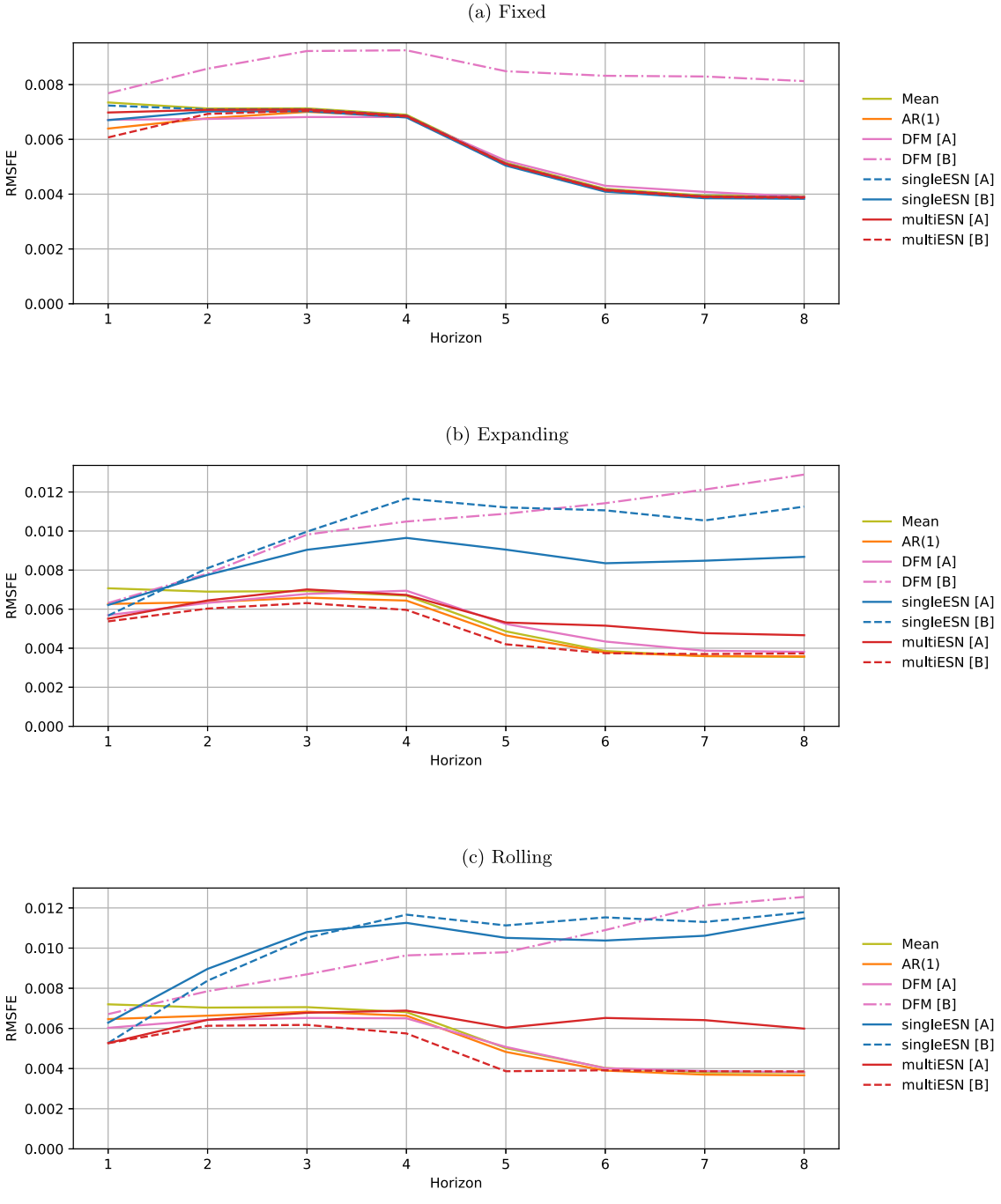
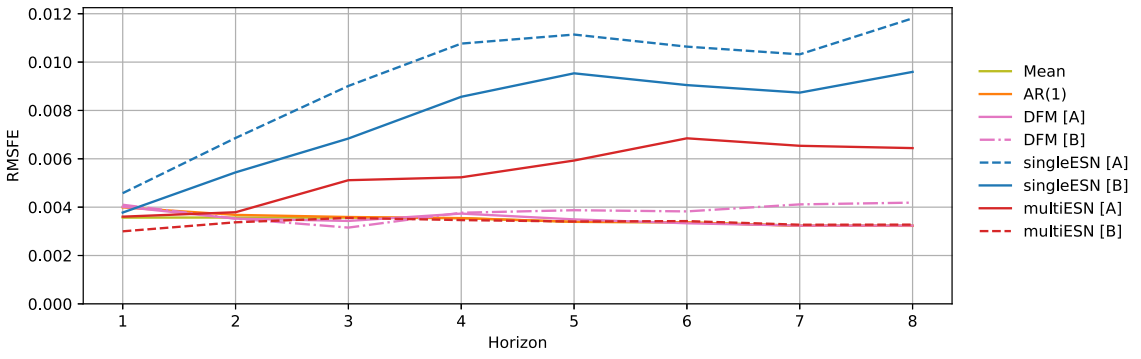
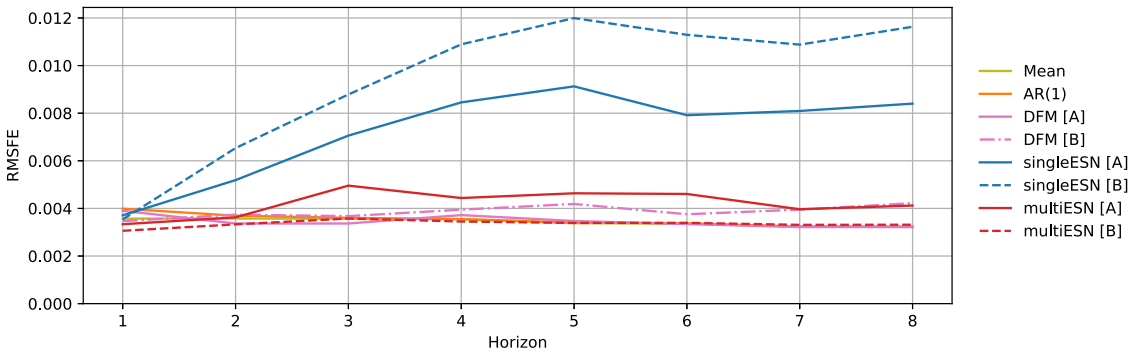


Fig. 9. Multistep-ahead GDP forecasting, RMSFE – 2007 sample – medium-MD dataset.

(a) Fixed



(b) Expanding



(c) Rolling

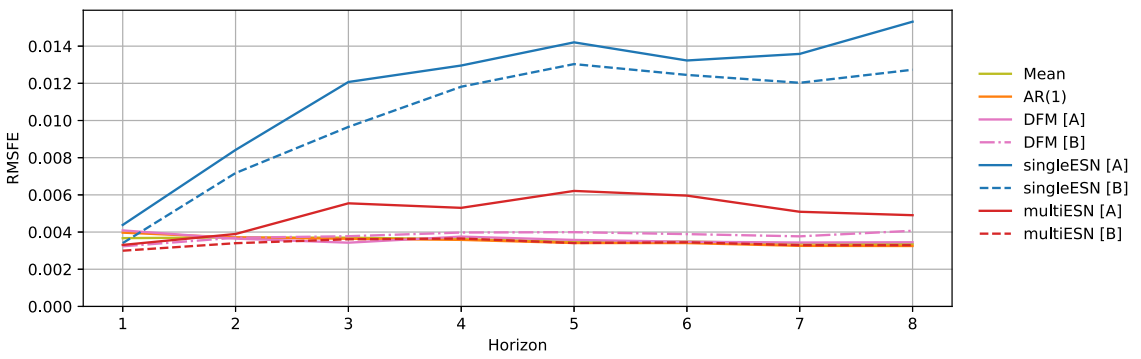


Fig. 10. Multistep-ahead GDP forecasting, RMSFE – 2011 sample – medium-MD dataset.

Appendix C. Forecasting schemes

To clarify the design of the forecasting experiments conducted in this paper, we present two different types of prediction illustrated in Fig. 11.

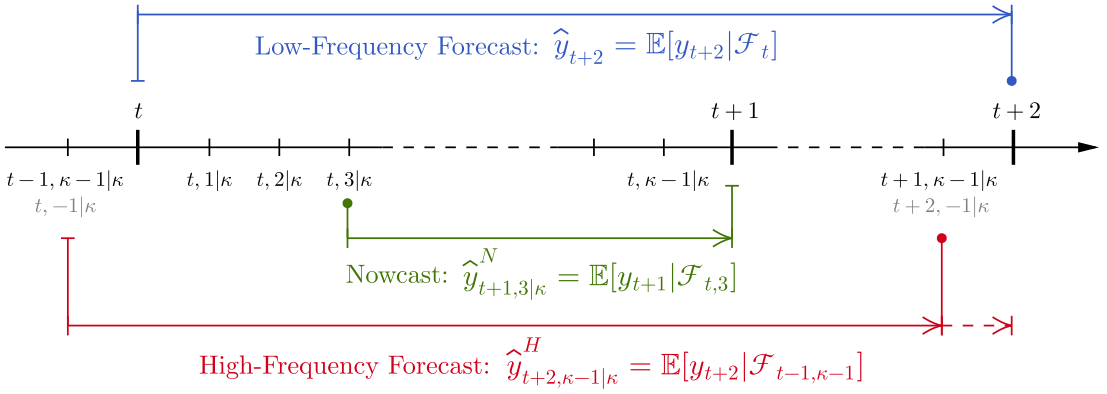


Fig. 11. Diagram of the low-/high-frequency forecasting and nowcasting schemes in tempo notation. Arrows point to time indices of the forecast target, and solid dots indicate the high-frequency time placeholder for the constructed high-frequency forecasts.

Let t denote time in the reference frequency of the target series (y_t), and suppose a regressor (z_r) of frequency κ is included in the forecasting model. The notation can be readily extended to include multiple regressors. Let $h \geq 0$ be a low-frequency prediction horizon counted from the last available observation of (y_t). Let $l \geq 0$ be a high-frequency horizon with respect to frequency κ.

Low-frequency forecasting. We call an h-step-ahead forecast 'low-frequency' when predictions for the target variable are constructed only at the end of the low-frequency periods. The information set which is used at the time of h-step-ahead low-frequency forecasting at t is the σ-algebra defined as

$$\mathcal{F}_t = \sigma\left(\{y_t, y_{t-1}, y_{t-2}, \ldots, z_{t,0|\kappa}, z_{t,-1|\kappa}, z_{t,-2|\kappa}, \ldots\}\right) \tag{C.1}$$

and, when using the mean square error as a loss, the optimal forecast is given by

$$\widehat{y}_{t+h} = \mathbb{E}\left[y_{t+h}|\mathcal{F}_t\right]. \tag{C.2}$$

High-frequency forecasting. In this forecasting scheme, one may also use high-frequency regressors to produce additional high-frequency forecasts of the low-frequency target variable. For example, in the case of a target released at the end of each year and having monthly quoted covariates, the low-frequency forecasting scheme corresponds to constructing forecasts always at the end of the last month of the year (December). At the same time, with all the information collected up to the end of December, there are other possibilities to construct forecasts. In particular, the forecaster could consider placing herself at the end of any other month of the year instead and construct predictions for the monthly proxy of the yearly variable for the next hth year.

In this scheme, one often artificially reduces the information set. Although not all the available information is exploited, this procedure has its benefits: first, it renders high-frequency forecast instances; second, it takes into account misspecification due to a seasonal response of (y_t) to (z_r). This is especially important whenever multiple time series with different sampling frequencies are combined in one model and seasonality effects are either difficult to detect or impossible to avoid. In the context of macroeconomic forecasting, we again refer the reader to Chen and Ghysels (2010), Clements and Galvão (2008, 2009) and Jardet and Meunier (2022), where these questions are carefully discussed.

Let the forecaster place herself at time t: she wishes to construct a high-frequency forecast for some $t, l|\kappa$ with $l \in \mathbb{N}$. The maximal information set available at t is $\mathcal{F}_t$, as in (C.1). However, if she uses $\mathcal{F}_t$, then the forecast for $t, l|\kappa$ coincides with the low-frequency forecast and is given by (C.2) for any l. Notice that the forecasts can be constructed using the reduced information sets instead. Let $h = \lceil l/\kappa \rceil$, $\ell = l \bmod \kappa$, and $m = h - \lfloor l/\kappa \rfloor$, and define

$$\begin{aligned}\mathcal{F}_{t-m,\ell} &= \sigma\left(\{y_{t-m}, y_{t-1-m}, \ldots, z_{t-m,\ell|\kappa},\ z_{t-m,(\ell-1)|\kappa},\ z_{t-m,(\ell-2)|\kappa}, \ldots\}\right) \\ &= \sigma\left(\{y_{t-m}, y_{t-1-m}, \ldots, z_{t+1-m,-(\kappa-\ell)|\kappa},\ z_{t+1-m,-(\kappa-\ell)+1|\kappa}, z_{t+1-m,-(\kappa-\ell+2)|\kappa}, \ldots\}\right).\end{aligned}$$

The high-frequency forecast information sets nest the low-frequency forecasting setup, since $\mathcal{F}_{t-m,\ell} \equiv \mathcal{F}_t$ if $l = \kappa h$ for $h \in \mathbb{N}$, and the forecast for the high-frequency proxy constructed for the moments $t, l|\kappa$ for the low-frequency variable is provided by the conditional expectation

$$\widehat{y}^{H}_{t+h,\ell|\kappa} = \mathbb{E}\left[y_{t+h}|\mathcal{F}_{t-m,\ell}\right].$$

It is easy to see that if the forecaster is interested in nowcasting, it can be readily obtained by taking $m = 0$ and writing for all $0 < \ell \leq \kappa - 1$:

$$\widehat{y}^{N}_{t+1,\ell|\kappa} = \mathbb{E}\left[y_{t+1}|\mathcal{F}_{t,\ell}\right].$$

Nowcasting. We call 'nowcasting' the setup in which one constructs a high-frequency proxy for a yet-unobserved target which will be available at the end of the current low-frequency period. As such, we construct a nowcast only for horizons $0 < l \leq \kappa - 1$; notice that $l = \kappa$ yields a contemporaneous regression at $t + 1$, while $l = 0$ falls into the category of low-frequency forecasting, hence both these cases are excluded. The σ-algebras that are used in order to construct nowcasts $\widehat{y}_{t+1,\ell|\kappa}$ are given by

$$\begin{aligned}\mathcal{F}_{t,\ell|\kappa} &= \sigma\left(\left\{y_t, y_{t-1}, \ldots, z_{t,\ell|\kappa},\ z_{t,(\ell-1)|\kappa},\ z_{t,(\ell-2)|\kappa}, \ldots\right\}\right)\\ &= \sigma\left(\left\{y_t, y_{t-1}, \ldots, z_{t+1,-(\kappa-\ell)|\kappa},\ z_{t+1,-(\kappa-\ell)+1|\kappa}, z_{t+1,-(\kappa-\ell+2)|\kappa}, \ldots\right\}\right).\end{aligned}$$

The l-step nowcast for the high-frequency proxy constructed at moments $t, \ell|\kappa$ of the current period for the low-frequency variable, which becomes available at $t + 1, 0|\kappa \equiv t + 1$, is provided by the conditional expectation

$$\widehat{y}^{N}_{t+1,\ell|\kappa} = \mathbb{E}\left[y_{t+1}|\mathcal{F}_{t,\ell}\right].$$

Multicasting. One always aims to construct one-step and multistep forecasts by using all the available information at a given point in time. It is, therefore, natural to compare models by constructing high-frequency nowcasts for the target variable to be released at the end of the current period and its high-frequency proxy forecasts for the next periods. To avoid confusion, we refer to this situation as 'multicasting'. More explicitly, provided that the forecaster finds herself at time index $t, s|\kappa$ and is interested in all the forecasts up to some maximal low-frequency horizon $H \geq 1$, for each $1 \leq l \leq H\kappa$, the multicasting scheme yields the following combination:

(a) *Nowcasting* when $0 < l \leq \kappa - 1$ and $\ell = l$: $\widehat{y}^{N}_{t+1,\ell|\kappa} = \mathbb{E}\left[y_{t+1}|\mathcal{F}_{t,\ell}\right]$

(b) Forecasting when $l > \kappa - 1$:

- *Low-frequency forecasting* if l satisfies $l \bmod \kappa = 0$: $\widehat{y}_{t+h} = \mathbb{E}\left[y_{t+h}|\mathcal{F}_t\right]$
- *High-frequency forecasting* if $l \bmod \kappa \neq 0$: $\mathcal{F}_{t,\ell}$: $\widehat{y}^{H}_{t+h,\ell|\kappa} = \mathbb{E}\left[y_{t+h}|\mathcal{F}_{t,\ell}\right]$.

Appendix D. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ijforecast.2023.10.009.

References

Andreou, E., Ghysels, E., & Kourtellos, A. (2013). Should macroeconomic forecasters use daily financial data and how? *Journal of Business & Economic Statistics*, *31*(2), 240–251.

Aparicio, D., & de Prado, M. L. (2018). How hard is it to pick the right model? MCS and backtest overfitting. *Algorithmic Finance*, *7*(1–2), 53–61.

Arcomano, T., Szunyogh, I., Wikner, A., Pathak, J., Hunt, B. R., & Ott, E. (2022). A hybrid approach to atmospheric modeling that combines machine learning with a physics-based numerical model. *Journal of Advances in Modeling Earth Systems*, *14*(3), e2021MS002712.

Armesto, M. T., Engemann, K. M., & Owyang, M. T. (2010). Forecasting with mixed frequencies. *Federal Reserve Bank of St. Louis Review*, *92*(6), 521–536.

Arora, S., Little, M. A., & McSharry, P. E. (2013). Nonlinear and nonparametric modeling approaches for probabilistic forecasting of the US gross national product. *Studies in Nonlinear Dynamics & Econometrics*, *17*(4), 395–420.

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

Babii, A., Ghysels, E., & Striaukas, J. (2022). Machine learning time series regressions with an application to nowcasting. *Journal of Business & Economic Statistics*, *40*(3), 1094–1106.

Bai, J., Ghysels, E., & Wright, J. H. (2013). State space models and MIDAS regressions. *Econometric Reviews*, *32*(7), 779–813.

Bai, J., & Ng, S. (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics*, *146*(2), 304–317.

Ballarin, G. (2023). Ridge regularized estimation of VAR models for inference. Preprint.

Ballarin, G., Grigoryeva, L., & Ortega, J.-P. (2023). Memory of recurrent networks: Do we compute it right?. Preprint.

Bańbura, M., Giannone, D., Modugno, M., & Reichlin, L. (2013). Nowcasting and the real-time data flow. In *Handbook of economic forecasting* (pp. 195–237). Elsevier.

Bańbura, M., & Modugno, M. (2014). Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data. *Journal of Applied Econometrics*, *29*(1), 133–160.

Bańbura, M., & Rünstler, G. (2011). A look into the factor model black box: Publication lags and the role of hard and soft data in forecasting GDP. *International Journal of Forecasting*, *27*(2), 333–346.

Belloni, A., Chernozhukov, V., Chetverikov, D., & Kato, K. (2015). Some new asymptotic theory for least squares series: Pointwise and uniform results. *Journal of Econometrics*, *186*(2), 345–366.

Bergmeir, C., Hyndman, R. J., & Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis*, *120*, 70–83.

Boivin, J., & Ng, S. (2005). *Understanding and comparing factor-based forecasts*: *Technical report*, National Bureau of Economic Research.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327.

Borio, C. (2011). Rediscovering the macroeconomic roots of financial stability policy: Journey, challenges, and a way forward. *Annual Review of Financial Economics*, *3*(1), 87–117.

Borio, C. (2013). The great financial crisis: Setting priorities for new statistics. *Journal of Banking Regulation*, *14*(3–4), 306–317.

Borio, C., & Lowe, P. W. (2002). Asset prices, financial and monetary stability: Exploring the nexus. *SSRN Electronic Journal*.

Boyd, S., & Chua, L. (1985). Fading memory and the problem of approximating nonlinear operators with Volterra series. *IEEE Transactions on Circuits and Systems*, *32*(11), 1150–1161.

Buehner, M., & Young, P. (2006). A tighter bound for the echo state property. *IEEE Transactions on Neural Networks*, *17*(3), 820–824.

Buell, B., Cherif, R., Chen, C., Hyeon, Tang, J., & Wendt, N. (2021). *Impact of COVID-19: Nowcasting and big data to track economic activity in Sub-Saharan Africa. Vol. 124*: *IMF working paper*, (pp. 1–61). IMF.

Camacho, M., & Pérez-Quirós, G. (2010). Introducing the euro-sting: Short-term indicator of euro area growth. *Journal of Applied Economics*, *25*, 663–694.
Cao, W., Wang, X., Ming, Z., & Gao, J. (2018). A review on neural networks with random weights. *Neurocomputing*, *275*, 278–287.
Carriero, A., Galvão, A. B., & Kapetanios, G. (2019). A comprehensive evaluation of macroeconomic forecasting methods. *International Journal of Forecasting*, *35*(4), 1226–1239.
Chauvet, M., Senyuz, Z., & Yoldas, E. (2015). What does financial volatility tell us about macroeconomic fluctuations? In *Finance and Economics Discussion*: *Journal of Economic Dynamics & Control*, In *Finance and Economics Discussion*: *52*.340–360,
Chen, X., & Christensen, T. M. (2015). Optimal uniform convergence rates and asymptotic normality for series estimators under weak dependence and weak conditions. *Journal of Econometrics*, *188*(2), 447–465.
Chen, X., & Ghysels, E. (2010). News - good or bad - and its impact on volatility predictions over multiple horizons. *The Review of Financial Studies*, *24*(1), 46–81.
Clements, M., & Galvão, A. (2008). Macroeconomic forecasting with mixed-frequency data: Forecasting output growth in the United States. *Journal of Business & Economic Statistics*, *26*, 546–554.
Clements, M. P., & Galvão, A. (2009). Forecasting US output growth using leading indicators: An appraisal using MIDAS models. *Journal of Applied Econometrics*, *7*(7), 1187–1206.
Crutchfield, J. P., Ditto, W. L., & Sinha, S. (2010). Introduction to focus issue: Intrinsic and designed computation: Information processing in dynamical systems - beyond the digital hegemony. *Chaos*, *20*(3), Article 037101.
Delle Monache, D., & Petrella, I. (2019). Efficient matrix approach for classical inference in state space models. *Economics Letters*, *181*, 22–27.
Diebold, F. X., & Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *20*(1), 134–144.
Doucet, A., de Freitas, N., & Gordon, N. J. (Eds.), (2001). *Statistics for engineering and information science*, *Sequential monte carlo methods in practice*. Springer.
Doya, K. (1992). Bifurcations in the learning of recurrent neural networks. In *Proceedings of IEEE international symposium on circuits and systems. Vol. 6* (pp. 2777–2780).
Doz, C., Giannone, D., & Reichlin, L. (2011). A two-step estimator for large approximate dynamic factor models based on Kalman filtering. *Journal of Econometrics*, *164*(1), 188–205.
Farkas, I., Bosak, R., & Gergel, P. (2016). Computational analysis of memory capacity in echo state networks. *Neural Networks*, *83*, 109–120.
Ferrara, L., Marsilli, C., & Ortega, J.-P. (2014). Forecasting growth during the great recession: Is financial volatility the missing ingredient? *Economic Modelling*, *36*, 44–50.
Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2005). The generalized dynamic factor model: One-sided estimation and forecasting. *Journal of the American Statistical Association*, *100*(471), 830–840.
Frale, C., Marcellino, M., Mazzi, G. L., & Proietti, T. (2011). EUROMIND: A monthly indicator of the euro area economic conditions. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *174*, 439–470.
Francis, N., Ghysels, E., & Owyang, M. T. (2011). *The low-frequency impact of daily monetary policy shocks*: *Technical report*, Federal Reserve Bank of St. Louis.
Fuleky, P. (Ed.), (2020). *Macroeconomic forecasting in the era of big data*. Springer International Publishing.
Galvão, A. B. (2013). Changes in predictive ability with mixed frequency data. *International Journal of Forecasting*, *29*(3), 395–410.
Galvão, A. B., & Marcellino, M. (2010). *Endogenous monetary policy regimes and the great moderation*: *Technical report*, EUI.
Geweke, J. (1977). The dynamic factor analysis of economic time series. In *Latent variables in socio-economic models*. North-Holland.
Ghysels, E. (2016). Macroeconomics and the reality of mixed frequency data. *Journal of Econometrics*, *193*(2), 294–314.
Ghysels, E., Santa-Clara, P., & Valkanov, R. (2004). *The MIDAS touch: Mixed data sampling regression models*: *Technical report 919*, UCLA: Finance.
Ghysels, E., Sinko, A., & Valkanov, R. (2007). MIDAS regressions: Further results and new directions. *Econometric Reviews*, *26*(1), 53–90.
Ghysels, E., & Wright, J. H. (2009). Forecasting professional forecasters. *Journal of Business & Economic Statistics*, *27*(4), 504–516.
Giannone, D., Reichlin, L., & Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, *55*(4), 665–676.
Gonon, L., Grigoryeva, L., & Ortega, J.-P. (2020a). Memory and forecasting capacities of nonlinear recurrent networks. *Physica D*, *414*(132721), 1–13..
Gonon, L., Grigoryeva, L., & Ortega, J.-P. (2020b). Risk bounds for reservoir computing. *Journal of Machine Learning Research*, *21*(240), 1–61.
Gonon, L., Grigoryeva, L., & Ortega, J.-P. (2023a). Approximation error estimates for random neural networks and reservoir systems. *Annals of Applied Probability*, *33*(1), 28–69.
Gonon, L., Grigoryeva, L., & Ortega, J. P. (2023b). Infinite-dimensional reservoir computing. Arxiv preprint.
Gonon, L., & Ortega, J.-P. (2020). Reservoir computing universality with stochastic inputs. *IEEE Transactions on Neural Networks and Learning Systems*, *31*(1), 100–112.
Gonon, L., & Ortega, J.-P. (2021). Fading memory echo state networks are universal. *Neural Networks*, *138*, 10–13.
Goudarzi, A., Marzen, S., Banda, P., Feldman, G., Lakin, M. R., Teuscher, C., et al. (2016). *Memory and information processing in recurrent neural networks*: *Technical report*, Portland State University.
Gramlich, D., Miller, G. L., Oet, M. V., & Ong, S. J. (2010). Early warning systems for systemic banking risk: Critical review and modeling implications. *Banks and Bank Systems*, *5*(2), 199–211.
Grigoryeva, L., Hart, A., & Ortega, J.-P. (2021). Learning strange attractors with reservoir systems. *Nonlinearity*, *36*(9), 4674.
Grigoryeva, L., Henriques, J., Larger, L., & Ortega, J.-P. (2015). Optimal nonlinear information processing capacity in delay-based reservoir computers. *Scientific Reports*, *5*(12858), 1–11.
Grigoryeva, L., Henriques, J., Larger, L., & Ortega, J.-P. (2016). Nonlinear memory capacity of parallel time-delay reservoir computers in the processing of multidimensional signals. *Neural Computation*, *28*, 1411–1451.
Grigoryeva, L., & Ortega, J.-P. (2018a). Echo state networks are universal. *Neural Networks*, *108*, 495–508.
Grigoryeva, L., & Ortega, J.-P. (2018b). Universal discrete-time reservoir computers with stochastic inputs and linear readouts using non-homogeneous state-affine systems. *Journal of Machine Learning Research*, *19*(24), 1–40.
Grigoryeva, L., & Ortega, J.-P. (2019). Differentiable reservoir computing. *Journal of Machine Learning Research*, *20*(179), 1–62.
Grigoryeva, L., & Ortega, J.-P. (2021). Dimension reduction in recurrent networks by canonicalization. *Journal of Geometric Mechanics*, *13*(4), 647–677.
Hansen, P. R., Huang, Z., & Shek, H. H. (2011). Realized GARCH: A joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*, *27*(6), 877–906.
Hart, A. G., Hook, J. L., & Dawes, J. H. P. (2021). Echo state networks trained by Tikhonov least squares are L2(μ) approximators of ergodic dynamical systems. *Physica D: Nonlinear Phenomena*, *421*, Article 132882.
Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, *13*(2), 281–291.
Hastie, T., Montanari, A., Rosset, S., & Tibshirani, R. J. (2022). Surprises in high-dimensional ridgeless least squares interpolation. *The Annals of Statistics*, *50*(2), 949–986.
Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (2nd ed.). Springer.
Hatzius, J., Hooper, P., Mishkin, F., Schoenholtz, K., & Watson, M. (2010). *Financial conditions indexes: A fresh look after the financial crisis*: *Technical report*, National Bureau of Economic Research.
Hindrayanto, I., Koopman, S. J., & de Winter, J. (2016). Forecasting and nowcasting economic growth in the euro area using factor models. *International Journal of Forecasting*, *32*(4), 1284–1305.
Hong, H., & Yogo, M. (2012). What does futures market interest tell us about the macroeconomy and asset prices? *Journal of Financial Economics*, *105*(3), 473–490.
Huber, F., Koop, G., Onorante, L., Pfarrhofer, M., & Schreiner, J. (2021). *Nowcasting in a pandemic using non-parametric mixed frequency VARs. Vol. 2510*: *ECB working paper series*, (pp. 1–40). ECB.

Ingenito, R., & Trehan, B. (1996). Using monthly data to predict quarterly output. *Econometric Reviews*, 3–11.
Ishwaran, H., & Rao, J. (2014). Geometry and properties of generalized ridge regression in high dimensions. In S. Ahmed (Ed.), *Contemporary mathematics. Vol. 622* (pp. 81–93). Providence, Rhode Island: American Mathematical Society.
Jaeger, H. (2010). *The 'echo state' approach to analysing and training recurrent neural networks with an erratum note*: *Technical report*, German National Research Center for Information Technology.
Jaeger, H., & Haas, H. (2004). Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science*, *304*(5667), 78–80.
Jardet, C., & Meunier, B. (2022). Nowcasting world GDP growth with high-frequency data. *Journal of Forecasting*, *41*(6), 1181–1200.
Kang, J., & Kwon, K. Y. (2020). Can commodity futures risk factors predict economic growth? *Journal of Futures Markets*, *40*(12), 1825–1860.
Kock, A. B., Medeiros, M., & Vasconcelos, G. (2020). Penalized time series regression. In P. Fuleky (Ed.), *Advanced studies in theoretical and applied econometrics*, *Macroeconomic forecasting in the era of big data: theory and practice* (pp. 193–228). Cham: Springer International Publishing.
Kostrov, A. (2021). *Essays on the use of MIDAS regressions in banking and finance* (Ph.D. thesis), Universität St. Gallen.
Legenstein, R., & Maass, W. (2007). What makes a dynamical system computationally powerful? In S. Haykin (Ed.), *New directions in statistical signal processing: from systems to brain*. Cambridge, MA: MIT Press.
Lukoševičius, M. (2012). A practical guide to applying echo state networks. In G. Montavon, G. B. Orr, & K.-R. Müller (Eds.), *Lecture notes in computer science*, *Neural networks: tricks of the trade: second edition* (pp. 659–686). Berlin, Heidelberg: Springer.
Lukoševičius, M., & Jaeger, H. (2009). Reservoir computing approaches to recurrent neural network training. *Computer Science Review*, *3*(3), 127–149.
Maass, W., Natschläger, T., & Markram, H. (2002). Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural Computation*, *14*, 2531–2560.
Maillard, O.-A., & Munos, R. (2012). Linear regression with random projections. *Journal of Machine Learning Research*, *13*(89), 2735–2772.
Manjunath, G., & Jaeger, H. (2013). Echo state property linked to an input: Exploring a fundamental characteristic of recurrent neural networks. *Neural Computation*, *25*(3), 671–696.
Manjunath, G., & Ortega, J.-P. (2023). Transport in reservoir computing. *Physica D: Nonlinear Phenomena*, *449*, Article 133744.
Marcellino, M., & Schumacher, C. (2010). Factor MIDAS for nowcasting and forecasting with ragged-edge data: A model comparison for German GDP. *Oxford Bulletin of Economics and Statistics*, *72*(4), 518–550.
Marcellino, M., Stock, J. H., & Watson, M. W. (2006). A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics*, *135*(1–2), 499–526.
Mariano, R. S., & Murasawa, Y. (2003). A new coincident index of business cycles based on monthly and quarterly series. *Journal of Applied Econometrics*, *18*(4), 427–443.
Marsilli, C. (2014). *Mixed-frequency modeling and economic forecasting* (Ph.D. thesis), (p. 140). Université de Franche-Comté.
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. (2020). *FRED-QD: A quarterly database for macroeconomic research*: *Technical report*, Federal Reserve Bank of St. Louis.
Monteforte, L., & Moretti, G. (2012). Real-time forecasts of inflation: The role of financial variables. *Journal of Forecasting*, *32*(1), 51–61.
Morley, J. (2015). Macro-finance linkages. *Journal of Economic Surveys*, *30*(4), 698–711.
Onatski, A. (2012). Asymptotics of the principal components estimator of large factor models with weakly influential factors. *Journal of Econometrics*, *168*(2), 244–258.
Paranhos, L. (2021). Predicting inflation with neural networks.
Pascanu, R., Mikolov, T., & Bengio, Y. (2013). On the difficulty of training recurrent neural networks. In *International conference on machine learning. Vol. 28, no. 3* (pp. 1310–1318). PMLR.
Pathak, J., Hunt, B., Girvan, M., Lu, Z., & Ott, E. (2018). Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach. *Physical Review Letters*, *120*(2), 24102.
Pathak, J., Lu, Z., Hunt, B. R., Girvan, M., & Ott, E. (2017). Using machine learning to replicate chaotic attractors and calculate Lyapunov exponents from data. *Chaos*, *27*(12).
Qin, D., van Huellen, S., Wang, Q. C., & Moraitis, T. (2022). Algorithmic modelling of financial conditions for macro predictive purposes: Pilot application to USA data. *Econometrics*, *10*(2), 22.
Quaedvlieg, R. (2021). Multi-horizon forecast comparison. *Journal of Business & Economic Statistics*, *39*(1), 40–53.
Rodan, A., & Tino, P. (2011). Minimum complexity echo state network. *IEEE Transactions on Neural Networks*, *22*(1), 131–144.
Salehinejad, H., Baarbe, J., Sankar, S., Barfett, J., Colak, E., & Valaee, S. (2017). Recent advances in recurrent neural networks.
Sargent, T. J., Sims, C. A., et al. (1977). Business cycle modeling without pretending to have too much a priori economic theory. *New Methods in Business Cycle Research*, *1*, 145–168.
Stock, J. H., & Watson, M. W. (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business & Economic Statistics*, *14*(1), 11–30.
Stock, J. H., & Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, *20*(2), 147–162.
Stock, J. H., & Watson, M. W. (2006). Forecasting with many predictors. In G. Elliot, C. W. Granger, & A. Timmermann (Eds.), *Handbook of economic forecasting. Vol. 1, no. 05*. Elsevier.
Stock, J. H., & Watson, M. W. (2016). Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. In *Handbook of macroeconomics. Vol. 2* (pp. 415–525). Elsevier.
Tanaka, G., Yamane, T., Héroux, J. B., Nakane, R., Kanazawa, N., Takeda, S., et al. (2019). Recent advances in physical reservoir computing: A review. *Neural Networks*, *115*, 100–123.
van Huellen, S., Qin, D., Lu, S., Wang, H., Wang, Q. C., & Moraitis, T. (2020). Modelling opportunity cost effects in money demand due to openness. *International Journal of Finance & Economics*, *27*(1), 697–744.
Wainrib, G., & Galtier, M. N. (2016). A local echo state property through the largest Lyapunov exponent. *Neural Networks*, *76*, 39–45.
Wikner, A., Pathak, J., Hunt, B. R., Szunyogh, I., Girvan, M., & Ott, E. (2021). Using data assimilation to train a hybrid forecast system that combines machine-learning and knowledge-based components. *Chaos. An Interdisciplinary Journal of Nonlinear Science*, *31*(5), 53114.
Yildiz, I. B., Jaeger, H., & Kiebel, S. J. (2012). Re-visiting the echo state property. *Neural Networks*, *35*, 1–9.
Zhang, B., Miller, D. J., & Wang, Y. (2012). Nonlinear system modeling with random matrices: Echo state networks revisited. *IEEE Transactions on Neural Networks and Learning Systems*, *23*(1), 175–182.