SCHOOL OF ECONOMICS AND MANAGEMENT
FACULTY OF SOCIAL SCIENCES
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric Analysis of Time Series

CREATES Research Paper 2009-18

Forecasting with Universal Approximators and a Learning Algorithm

Anders Bredahl Kock

School of Economics and Management
Aarhus University
Bartholins Allé 10, Building 1322, DK-8000 Aarhus C
Denmark

FORECASTING WITH UNIVERSAL APPROXIMATORS AND A LEARNING ALGORITHM

ANDERS BREDAHL KOCK
AARHUS UNIVERSITY AND CREATES

Abstract. This paper applies three universal approximators for forecasting. They are the Artificial Neural Networks, the Kolmogorov-Gabor polynomials, as well as the Elliptic Basis Function Networks. Even though forecast combination has a long history in econometrics focus has not been on proving loss bounds for the combination rules applied. We apply the Weighted Average Algorithm (WAA) of Kivinen and Warmuth (1999) for which such loss bounds exist. Specifically, one can bound the worst case performance of the WAA compared to the performance of the best single model in the set of models combined from. The use of universal approximators along with a combination scheme for which explicit loss bounds exist should give a solid theoretical foundation to the way the forecasts are performed. The practical performance will be investigated by considering various monthly postwar macroeconomic data sets for the G7 as well as the Scandinavian countries.

JEL classification: C22; C45; C53
Keywords: Forecasting, Universal Approximators, Elliptic Basis Function Network, Forecast Combination, Weighted Average Algorithm

1. Introduction

In this paper we examine the forecast performance of non-linear models compared to that of linear autoregressions. Linear models have the advantage that they can be understood and analyzed in great detail. However, it might be inappropriate to assume that the generating mechanism of a series is linear. Hence, non-linear models have become increasingly popular, see e.g. Granger and Teräsvirta (1993). However, the non-linear models are still restricted by the fact that modeling takes place within a prespecified family of models. Since the modeler often

Date: May 6th, 2009.

Financial support from the Center for Research in the Econometric Analysis of Time Series, *CREATES*, funded by the Danish National Research Foundation is gratefully acknowledged by the author. The author also wishes to thank Niels Haldrup, Peter Reinhard Hansen, and Timo Teräsvirta for helpful comments.

Address: Department of Economics, Aarhus University, Denmark, e-mail: akock@creates.au.dk.

has little prior knowledge regarding the functional form of the data generating process, choosing the correct family is still not an easy task. If one wants to avoid making this choice, one may apply universal approximators which are able to approximate broad classes of functions arbitrarily well in a way to be made clear in Section 2.

The universal approximators are data driven in the sense that little a priori knowledge is needed about the functional relationship between the left- and the right-hand side variables. Three of the four types of nonlinear models considered in this paper are universal approximators. Our universal approximators are the Artificial Neural Networks (ANN), the Kolmogorov-Gabor polynomials, and the Elliptic Basis Function Networks. While the ANNs have been applied frequently in forecasting studies such as Stock and Watson (1999) and Teräsvirta et al. (2005), the Kolmogorov-Gabor polynomials as well as the Elliptic Basis Function Networks have not been as frequently used.

Since forecasting is carried out using several models we obtain more than one point forecast of the same variable. But in practice one is often interested in obtaining a single forecast of a certain variable. This can be achieved by means of various forecast combination schemes. The first to study this in econometrics were Bates and Granger (1969). The literature has proliferated since then and a recent survey is given in Timmermann (2006). Two caveats apply, however, to many of these combination algorithms. First, nothing can be said a priori about the performance of the algorithm (combination rule) compared to the individual forecasts. And even if bounds are provided, they often depend on the joint distribution of the vector consisting of the forecasts made by the individual models. The *Weighted Average Algorithm* (WAA) of Kivinen and Warmuth (1999) developed in the computer science literature does not share any of these problems. First, explicit loss bounds for the worst case performance of the algorithm are available. Furthermore, these bounds do not depend on the distribution of the vector of forecasts from the individual models.

We argue that forecasting with universal approximators and combining these into a single forecast by an algorithm for which explicit bounds can be derived forms a solid theoretical foundation for combining forecasts. The empirical performance of the universal approximators as well as the WAA will be investigated by considering various monthly postwar macroeconomic data sets for the G7 and the Scandinavian countries.

As many other authors, we find that there are gains to be made by combining forecasts. However, the way one combines forecasts does not seem to be of utmost importance. In particular, the performance of the WAA is roughly the same as that of various schemes imposing equal weights. Hence, the important thing is to combine but not how to combine.

The outline of the paper is as follows. Section 2 introduces the universal approximators applied in the paper and contains a review of important theoretical results. Next, Section 3 introduces the benchmark models and Section 4 discusses forecasting with expert advice with particular emphasis on the Weighted Average Algorithm and its theoretical underpinnings. Section 5 presents the results of the application and Section 6 concludes.

2. Universal Approximators

We begin by defining precisely what we mean by universal approximators and then discuss the three types employed in this paper.
In order to define universal approximators some preliminary notation is necessary. Let X be a topological space and A a subset of X. Let $\bar{A}$ denote the closure of A. Then A is dense in X if $\bar{A} = X$. Since all topologies used in this paper will be induced by metrics we may define the closure of A as

$$\bar{A} = \{x \in X \mid \exists (x_n)_{n \geq 1} \subseteq A \text{ such that } x_n \to x\}.$$

Hence, A is dense in X if for each $x \in X$ one can choose an element $a \in A$ that is arbitrarily close (in the metric on X) to x. We are now ready to define what we mean by universal approximators.

Definition 1. *Let $\mathcal{H}$ be a subset of functions of a topological space $\mathcal{F}$. Then $\mathcal{H}$ is a universal approximator of $\mathcal{F}$ if $\bar{\mathcal{H}} = \mathcal{F}$.*

An example could be $C_C(\mathbb{R}^n)$, the compactly supported continuous functions on $\mathbb{R}^n$, being dense in $L^p(\lambda_n)$ for $1 \leq p < \infty$, where λ_n is the Lebesgue measure on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ with $\mathcal{B}(\mathbb{R}^n))$ denoting the Borel σ-field on $\mathbb{R}^n$. So for any function $f \in L^p(\lambda_n)$ one can choose a function $h \in C_C(\mathbb{R}^n)$ that is arbitrarily close to f, where closeness is expressed in terms of the metric induced by the L^p-norm. Next, we will introduce the universal approximators used in this paper.

2.1. **Artificial Neural Networks.**

Artificial Neural Networks (ANN) form a very popular family of universal approximators. They are defined in the following way:

$$\mathcal{H}_{ANN} = \left\{ h_q : \mathbb{R}^n \to \mathbb{R} \mid h_q(x) = \sum_{i=1}^{q} \beta_i G(x'\gamma_i + \delta_i), \right.$$
$$\left. \beta_i, \delta_i \in \mathbb{R}, \gamma_i \in \mathbb{R}^n,\ q \in \mathbb{N} \right\}$$

To be precise, $\mathcal{H}_{ANN}$ is the set of *single hidden layer* neural network models. Hornik et al. (1989) show that if one chooses $G : \mathbb{R} \to \mathbb{R}$ to be a nondecreasing sigmoidal function[1] (i.e. a squashing function) then $\mathcal{H}_{ANN}$ is uniformly dense on compacta in $C(\mathbb{R}^n)$. More formally, for any $f \in C(\mathbb{R}^n)$ and any compact set $K \subseteq \mathbb{R}^n$ it holds that

$$\forall \epsilon > 0 \; \exists h_q \in \mathcal{H}_{ANN} : \sup_{x \in K} |f(x) - h_q(x)| < \epsilon.$$

Furthermore, they prove that $\mathcal{H}_{ANN}$ is "dense in measure" in $M(\mathcal{B}(\mathbb{R}^n))$ which denotes the set of Borel measurable functions on $\mathbb{R}^n$. For any finite measure μ on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ they define the metric

$$\rho_\mu(f, g) = inf \{\epsilon > 0 \mid \mu(x \in \mathbb{R}^n | \; |f(x) - g(x)| > \epsilon) < \epsilon\}$$

and show that $\mathcal{H}_{ANN}$ is ρ_μ-dense in $M(\mathcal{B}(\mathbb{R}^n))$. So for any $f \in M(\mathcal{B}(\mathbb{R}^n))$ and any $\epsilon > 0$ there exists an $h \in \mathcal{H}_{ANN}$ such that $\rho_\mu(f, h) < \epsilon$. Since convergence in the metric ρ_μ is a metrization of convergence in the measure μ, this result can also be understood as establishing the existence of an $h \in \mathcal{H}_{ANN}$ for which the measure of the set $\{|f(x) - g(x)| > \epsilon\}$ can be made arbitrarily small for any $\epsilon > 0$. Hence, the term dense in measure is appropriate.
Regarding uniform denseness of $\mathcal{H}_{ANN}$ in $M(\mathcal{B}(\mathbb{R}^n))$ Hornik et al. (1989) show that for any $f \in M(\mathcal{B}(\mathbb{R}^n))$ and for any $\epsilon > 0$ there exists an $h \in \mathcal{H}_{ANN}$ and a compact set $K \subseteq \mathbb{R}^n$ such that $\mu(K^C) < \epsilon$ and $|f(x) - h(x)| < \epsilon$ for all $x \in K$.
As a final interesting result we mention that $\mathcal{H}_{ANN}$ is dense in L^p for any $p \in [1, \infty)$ if there exists a compact set $K \subseteq \mathbb{R}^n$ such that $\mu(K^C) = 0$. This is true for any finite measure μ and any squashing function G. The choice of G has not been discussed but an obvious choice is any cumulative distribution function since these are squashing functions (they are sigmoidal and non decreasing (they are even càdlàg)).

2.2. **Elliptic Basis Function Networks.** The Elliptic Basis Function Networks (EBF) introduced in Park and Sandberg (1994) have been less frequently applied in Econometrics than the more common Artificial Neural Networks. The better known Radial Basis Function Networks (RBF) may be regarded as a special case of the EBF. The

[1]The defining property of a sigmoidal function in Cybenko (1989) is $\sigma : \mathbb{R} \to \mathbb{R}$ is
$$\sigma(t) \to \begin{cases} 1 & \text{for } t \to \infty \\ 0 & \text{for } t \to -\infty \end{cases}$$

set of EBF is defined as

$$\mathcal{H}_{EBF} = \left\{ h_q : \mathbb{R}^n \to \mathbb{R} |\ h(x) = \sum_{i=1}^{q} w_i G(\frac{x_1 - c_{i1}}{\sigma_{i1}}, ..., \frac{x_n - c_{in}}{\sigma_{in}}), \right.$$
$$\left. c_i, \sigma_i \in \mathbb{R}^n, q \in \mathbb{N} \right\}.$$

The parameters c_{ij} and σ_{ij} are often referred to as the centroids and width (or smoothing) factors respectively. Though not necessary for the theorems stated below to be valid it is often assumed that $G : \mathbb{R}^n \to \mathbb{R}$ is radially symmetric. Put differently, $G(x) = G(y)$ if $||x|| = ||y||$ where $||.||$ denotes the Euclidean norm on $\mathbb{R}^n$. If G is radially symmetric and $\sigma_{ij} = \sigma_i$ for $j = 1, ..., n,\ i = 1, ..., q,\ \mathcal{H}_{EBF}$ reduces to the set of RBF Networks[2].
A frequent choice of G is the Gaussian, for which

$$G(x) = \exp\left(\frac{-\sum_{j=1}^{n} x_j^2}{2}\right)$$

For this choice the output of the i'th hidden unit is given by

$$G(\frac{x_1 - c_{i1}}{\sigma_{i1}}, ..., \frac{x_n - c_{in}}{\sigma_{in}}) = \exp\left(-\sum_{j=1}^{n} \frac{(x_j - c_{ij})^2}{2\sigma_{ij}^2}\right) \tag{2.1}$$

Formula (2.1) simply defines a rescaling of the probability density function of a multivariate Gaussian vector with a diagonal covariance matrix. From (2.1) it is seen why the c_{ij}s are called the centroids. The vector c_i determines where in $\mathbb{R}^n$ the ith hidden unit is centered. In practice one wants to center the hidden units in areas of high data intensity. Since the σ_{ij}'s are allowed to vary across j for each i the level planes of G will be elliptic (think of $n = 2$) which explains the term *Elliptic Basis Function Network*. In the RBFs the value of G only depends on the distance to the center in the sense that if $||x - c_i|| = ||y - c_i||$ the i'th hidden unit takes the same value at x and y. In other words, the level sets are circles (think of $n = 2$ again) - a special case of the ellipse.
Regarding the universal approximation ability of the EBF it follows from Park and Sandberg (1991, 1993, 1994) that if $G \in L^1(\mathbb{R}^n)$ is bounded and continuous almost everywhere and satisfies $\int_{\mathbb{R}^n} G(x)dx \neq 0$ then $\mathcal{H}_{EBF}$ is dense in $L^p(\mathbb{R}^n)$ for $1 \leq p < \infty$.
Regarding uniform approximation the following result holds. If G is continuous and satisfies the conditions above then $\mathcal{H}_{EBF}$ is uniformly dense on compacta. In other words any continuous function may be approximated arbitrarily well in the supremum norm on any compact set. For more results and details, see Park and Sandberg (1991, 1993,

[2]If $\sigma_{ij} = \sigma$ for $j = 1, ..., n,\ i = 1, ..., q$ one also calls $\mathcal{H}_{EBF}$ the set of RBF.

1994).
Finally, we notice that a probability density function is an obvious choice for G in light of the above conditions on G. This is in contrast with the sigmoidal hidden units in the case of ANNs. A cumulative distribution function was an obvious choice in that case. This illustrates an interesting difference between the EBF (RBF) networks and the ANN. The former may be seen as local approaches since probability distribution functions tend to zero as the distance from the centroids goes to infinity. So the hidden units are only active close (locally) to their centroids. The latter may be seen as a global approach since the hidden units take values close to one for sufficiently large $x'\gamma + \delta$. Put differently we do not have that a cumulative distribution function tends to 0 as the norm of the input vector goes to ∞. Global effects can, however, cancel out and become local.

2.3. **Kolmogorov-Gabor polynomials.** The set of Kolmogorov-Gabor polynomials is defined as follows:

$$\mathcal{H}_{KG} = \Bigg\{ h_q : \mathbb{R}^n \to \mathbb{R} \mid h_q(x) = \phi + \sum_{i_1=1}^{n} \phi_{i_1} x_{i_1} + \sum_{i_1=1}^{n} \sum_{i_2=i_1}^{n} \phi_{i_1 i_2} x_{i_1} x_{i_2} + \dots$$
$$+ \sum_{i_1=1}^{n} \sum_{i_2=i_1}^{n} \dots \sum_{i_q=i_{q-1}}^{n} \phi_{i_1 i_2 \dots i_q} x_{i_1} x_{i_2} \dots x_{i_q},\ \phi, \phi_{i_1 \dots i_j} \in \mathbb{R}, 1 \le j \le q,\ q \in \mathbb{N} \Bigg\}.$$

The Kolmogorov-Gabor polynomials are qth degree polynomials with all possible cross-products included.
By the Stone-Weierstrass Theorem it follows that $\mathcal{H}_{KG}$ is uniformly dense on compacta in $C(\mathbb{R}^n)$. $\mathcal{H}_{KG}$ is clearly an algebra of (real) functions on $K \subseteq \mathbb{R}^n$ for any compact set K. It vanishes at no point since $\mathcal{H}_{KG} \ni h(x) = 1 + \sum_{i=1}^{n} x_i^2 > 0$ for all $x \in K$. $\mathcal{H}_{KG}$ also separates points in K. To see this, let $x, y \in \mathbb{R}^n$ and assume $x \neq y$. So for at least one $1 \le i \le n$ it holds that $x_i \neq y_i$. Since $h(x) = x_i$ belongs to $\mathcal{H}_{KG}$, the uniform closure of $\mathcal{H}_{KG}$ consists of all continuous functions on K.

3. Benchmark Models

3.1. **Smooth Transition Models.** The Smooth Transition regression model is not a universal approximator. The reason for including it in this work is that it is a benchmark nonlinear model. A standard Smooth Transition regression model is given by

$$y_t = \phi' x_t + \theta' x_t G(s_t) + \varepsilon_t \tag{3.1}$$

where G is the transition function. As is usual in the literature, we choose G to be the logistic function, i.e.

$$G(s_t) = \frac{1}{1 + \exp(-\gamma(s_t - c))}$$

where s_t is the transition variable. Examples include $s_t = y_{t-d}$ for some $d \geq 1$ or $s_t = t$. The parameter γ controls the speed of transition and c indicates the position of the transition function.

3.2. **Autoregressions.** Finally, the pth order linear autoregression

$$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + u_t \tag{3.2}$$

for some $p \in \mathbb{N}$ is employed as a benchmark in order to check whether or not the universal approximators are able to outperform it when it comes to forecasting.

4. Forecasting with Experts

This section introduces forecasting with experts with particular emphasis on the Weighted Average Algorithm (WAA) of Kivinen and Warmuth (1999). To establish the notation consider a setting with n experts (models) and let $\mathcal{E}_i$ denote expert i, $i = 1, 2, ..., n$, and l the number of trials, i.e., the number of forecasts made with each model. Now consider a sequence $S_l = \{(\mathbf{x}_{t+\tau,t}, y_{t+\tau})\}_{t=1}^{l}$ where $(\mathbf{x}_{t+\tau,t}, y_{t+\tau}) \in [0,1]^{n+1}$ for $t \in \{1, 2, ..., l\}$; for every t we have access to a vector of forecasts $\mathbf{x}_{t+\tau,t} = (x_{t+\tau,t,1}, ..., x_{t+\tau,t,n}) \in [0,1]^n$ whose elements are the τ period ahead forecasts made by each expert at time t. $y_{t+\tau} \in [0,1]$ denotes the actual outcome of the variable to be forecast. By doing so, expert i incurs a loss $L(y_{t+\tau}, x_{t+\tau,t,i})$. A frequently applied loss function to be used in this paper is the quadratic loss, i.e. $L(y_{t+\tau}, x_{t+\tau,t,i}) = (y_{t+\tau} - x_{t+\tau,t,i})^2$. The total loss of expert i given the sequence S_l is defined as $L_{\mathcal{E}_i}(S_l) = \sum_{t=1}^{l} L(y_{t+\tau}, x_{t+\tau,t,i})$. Similarly, the total loss of an algorithm A that gives a sequence of forecast $\{\hat{y}_{t+\tau,t}\}_{t=1}^{l}$ is $L_A(S_l) = \sum_{t=1}^{l} L(y_{t+\tau}, \hat{y}_{t+\tau,t})$.
In economic applications one cannot assume that $(\mathbf{x}_{t+\tau,t}, y_{t+\tau}) \in [0,1]^{n+1}$. This assumption can, however, be relaxed to $(\mathbf{x}_{t+\tau,t}, y_{t+\tau}) \in [a,b]^{n+1}$. Thus, by choosing $[a,b]$ sufficiently wide one may circumvent this problem. The exact choice of $[a,b]$ depends on the problem at hand and the procedure used in this paper to determine it will be described in Section 5.

The Weighted Average Algorithm. The Weighted Average Algorithm (WAA) of Kivinen and Warmuth (1999) provides a way of combining the forecasts $\mathbf{x}_{t+\tau,t}$ of the experts at trial t into a single forecast. As mentioned in the introduction, an attractive feature of the WAA is that as opposed to many other forecast combination schemes applied in econometrics (see e.g. Timmermann (2006)) the WAA does not make any assumptions regarding the joint distribution of the forecasts made by the experts. The loss bounds presented below are purely arithmetic results that hold for any distribution of $\mathbf{x}_{t+\tau,t}$. Letting $\mathbf{v}_t$ denote a probability vector of weights, i.e.,

$$\mathbf{v}_t \in \left\{ \mathbf{s} \in \mathbb{R}^n | \sum_{i=1}^n s_i = 1,\ s_i \geq 0,\ i = 1, 2, ..., n \right\}$$

the forecasts of the WAA are given by $\hat{y}_{t+\tau,t} = \mathbf{v}_t \mathbf{x}_{t+\tau,t},\ t = 1, ..., l$. This way of forecasting explains the terminology *Weighted Average Algorithm* since the forecasts made by the algorithm are weighted averages of the forecasts made by the experts. The weights in the WAA are constructed in the following way:

(1) Initialize the algorithm by choosing $\mathbf{v}_1$. If no prior knowledge regarding the performance of the experts is available, an obvious choice is to give equal weights to all of them.
(2) For $t = 1, ..., l$
   (a) Observe vector of expert forecasts $\mathbf{x}_{t+\tau,t}$.
   (b) Calculate the forecast of the algorithm, $\hat{y}_{t+\tau,t} = \mathbf{v}_t' \mathbf{x}_{t+\tau,t}$.
   (c) Observe the actual value of y_t.
   (d) Update the weights according to

$$v_{t+1,i} = \frac{v_{t,i} \exp(-L(y_t, x_{t,t-\tau,i})/c)}{\sum_{i=1}^n v_{t,i} \exp(-L(y_t, x_{t,t-\tau,i})/c)}$$

   where the denominator ensures that the weights sum to 1 and c is a positive constant further to be defined below.

Notice that if two experts, $\mathcal{E}_1$ and $\mathcal{E}_2$ have $v_{t,1}/v_{t,2} \neq 1$ due to differences in their previous performance but perform equally well in all future periods we will not have $v_{t,1}/v_{t,2} \to 1$ as $t \to \infty$. Their ratio will stay unchanged unless they actually incur different losses.
What makes the WAA attractive from a theoretical point of view is the following result in Kivinen and Warmuth (1999):

Theorem 1. *Let $L(y, x)$ be a convex twice differentiable loss function of x for every y. Assume $L_2'(y, y) = 0$. Letting WAA denote the Weighted Average Algorithm with uniform initial weights, i.e. $v_{1,i} = 1/n$, and $S_l = \{(\mathbf{x}_{t+\tau,t}, y_{t+\tau})\}_{t=1}^l$ an arbitrary input sequence, it holds that*

$$L_{WAA}(S_l) \leq \left(\min_{1 \leq i \leq n} L_{\mathcal{E}_i}(S_l) \right) + c \ln(n) \tag{4.1}$$

where c is a constant that depends on the loss function.

In particular, Kivinen and Warmuth (1999) show that it is enough that

$$c \geq \sup_{0 \leq x, y \leq 1} \frac{\left(L_2'(y,x)\right)^2}{L_{22}''(y,x)} \tag{4.2}$$

in order for the inequality (4.1) to be valid. For a quadratic loss function this implies $c \geq 2$. As mentioned above, one cannot in general know in advance that $(\mathbf{x}_{t+\tau,t}, y_{t+\tau}) \in [0,1]^{n+1}$. But if there exists an interval $[a,b]$ such that $(\mathbf{x}_{t+\tau,t}, y_{t+\tau}) \in [a,b]^{n+1}$ then inequality (4.1) is still valid if one chooses

$$c \geq \sup_{a \leq x, y \leq b} \frac{\left(\frac{d}{dx}(y-x)^2\right)^2}{\frac{d^2}{dx^2}(y-x)^2} = \sup_{a \leq x, y \leq b} \frac{\left(-2(y-x)\right)^2}{2} = 2(b-a)^2.$$

Regarding the conditions on c for other loss functions we refer to Kivinen and Warmuth (1999).
The inequality (4.1) is the theoretical foundation of the Weighted Average Algorithm since it gives an explicit bound on the loss of the algorithm as compared to the best expert in the set of experts. In particular, the WAA will perform no worse than the best expert plus some constant independent of the number of trials. This implies

$$\limsup_{l \to \infty} \frac{L_{WAA}(S_l)}{l} \leq \limsup_{l \to \infty} \frac{\left(\min_{1 \leq i \leq n} L_{\mathcal{E}_i}(S_l)\right)}{l}.$$

Thus, the average loss of the WAA will be no larger than the average loss incurred by the best expert as the number of trials (forecasts) approaches infinity.

Bounds with respect to random selection of experts. Kivinen and Warmuth (1999) also prove a result regarding the performance of the Weighted Average Algorithm compared to the expected loss of using a random selection of experts. *Random selection* means that at each trial one forecasts according to expert $\mathcal{E}_i$ with probability u_i. The expected total loss using the probability vector $\mathbf{u}$ is then given by $L_{\mathbf{u}}^{avg}(S_l) = \sum_{i=1}^{n} u_i L_{\mathcal{E}_i}(S_l) = \sum_{t=1}^{l} \sum_{i=1}^{n} u_i L(y_{t+\tau}, x_{t+\tau,t,i})$. We also need to define the relative entropy (also referred to as the *Kullback-Leibler divergence*[3]) between two discrete distributions in order to introduce

[3]Though one might conjecture that the Kullback-Leibler divergence defines a metric on the space of (discrete) probability distributions this is not the case. In particular, it is not symmetric, nor does it satisfy the triangle inequality. It is,

the desired bound. Let $\mathbf{u}$ and $\mathbf{v}$ be probability vectors in $\mathbb{R}^n$. Then the Kullback-Leibler divergence is defined as $d_{re}(\mathbf{u},\mathbf{v}) = \sum_{i=1}^{n} u_i \ln\left(\frac{u_i}{v_i}\right)$. The following result holds:

Theorem 2. *Let $L(y,x)$ be a convex twice differentiable loss function of x for every y. Assume $L_2'(y,y) = 0$. Let $\mathbf{v}_1$ be any vector of initial weights for the WAA and let $S_l = \{(\mathbf{x}_{t+\tau,t}, y_{t+\tau})\}_{t=1}^{l}$ be any input sequence and $\mathbf{u}$ an arbitrary probability vector. Then it holds that*

$$L_{WAA}(S_l) \leq L_{\mathbf{u}}^{avg}(S_l) + c d_{re}(\mathbf{u},\mathbf{v}_1) \tag{4.3}$$

This inequality states that no matter how one chooses $\mathbf{u}$ the Weighted Average Algorithm will not do much worse than forecasting by choosing experts at random according to $\mathbf{u}$.
If one chooses equal initial weights, i.e. $v_{1,i} = 1/n$ for $1 \leq i \leq n$, and chooses $u_k = 1$ and $u_j = 0$ for $1 \leq k, j \leq n,\ j \neq k$ then one obtains $L_{\mathbf{u}}^{avg}(S_l) = L_{\mathcal{E}_k}(S_l)$ and $c d_{re}(\mathbf{u},\mathbf{v}_1) = c\ln(n)$. Hence, (4.3) becomes $L_{WAA}(S_l) \leq L_{\mathcal{E}_k}(S_l) + c\ln(n)$, for all $1 \leq k \leq n$, and just as in (4.1), $L_{WAA}(S_l) \leq \left(\min_{1\leq i\leq n} L_{\mathcal{E}_i}(S_l)\right) + c\ln(n)$. This means that the bound in (4.1) is obtained again.

The results of this section give the theoretical motivation for applying the WAA in econometric forecasting. At this point it should be noticed, however, that the WAA is only one of many learning algorithms developed in computer science that might be of interest for econometricians. More elaborate algorithms that give tighter loss bounds should also be confronted with economic data. A drawback of the WAA is that due to its non-differentiability it does not apply to the absolute loss function $L(y_{t+\tau}, x_{t+\tau,t,i}) = |y_{t+\tau} - x_{t+\tau,t,i}|$. Other algorithms such as the *Hedge-β* algorithm by Freund and Schapire (1997) are available in this case.

5. Application

In order to investigate the performance of the models introduced in Sections 2 and 3 as well as the Weighted Average Algorithm we consider monthly postwar macroeconomic data sets for the G7 countries as well as the Scandinavian countries including Finland. Five different macroeconomic series were considered for each country: Inflation (CPI), Industrial Production (IP), long term Interest Rates (I), narrow Money Supply (M), and Unemployment (U). For some countries certain series were missing, and in total 47 series were analyzed. The series have been obtained from the OECD Main Economic Indicators

however, nonnegative; this can be established using Gibbs' inequality which in turn follows from $\ln(x) \leq x - 1$.

database and the IMF database. The starting date of the majority of the series is 1960 and most series are available until 2008. The series were seasonally adjusted except for CPI and I. For CPI, IP, and M the models are specified for yearly growth rates.

5.1. **Estimation and Forecasting Methodology.** In this section we describe the details of the estimation procedure for the various models as well as the details of the forecasting procedure. For all series forecasts were made 1, 3, 6, 12 and 24 months ahead. 72 forecasts were made at each horizon for each series. For all series, specification, estimation, and forecasting were carried out using an expanding window (a recursive scheme) with the last window closing 24 months prior to the last observation. All models were respecified and reestimated each time the window was expanded by one observation. All models were univariate and nested the linear autoregression. The details of the individual models will follow.

Autoregressions. For each window, autoregressions with up to five lags were estimated. The one with the lowest value of our Choice Criterion $CC = \log(MSE) + \delta k/T$, where k denotes the number of parameters, $\delta = 1$ and T is the number of observations in the window, was chosen for forecasting. Compared to the Akaike Information Criterion (AIC), in which $\delta = 2$, CC is a rather liberal criterion. Since autoregressions are affine, $E(y_{t+\tau}|\mathcal{F}_t)$ equals the skeleton forecast τ periods made at t (i.e. the recursive forecast ignoring the noise) where $\mathcal{F}_t$ is the σ-algebra generated by $\{y_s\}_{s=1}^{t}$. Even though preliminary experiments indicated that an insanity filter as introduced in Swanson and White (1995) was not necessary for the linear autoregression, we adopted the following rule in order to safeguard ourselves against too extreme forecasts. If a forecast did not belong to the interval given by the last observation of the estimation window plus/minus three times the standard deviation of the 120 most recent observations in the window it was replaced by the last observation of the window. In the words of Swanson and White (1995), craziness was replaced by ignorance. The purpose of this insanity filter is of course to weed out unreasonable forecasts and thereby more closely mimic the behavior of a real forecaster. The reason for only calculating the standard deviation s based on the last 120 observations of the window is that for many data sets the standard deviation of all observations in the window is often very high due to large historic fluctuations. As a result, basing s on all observations in the window would lead to occasionally accepting of wild forecasts.

No Change Forecasts. In order to investigate whether any of the estimated models was able to beat naive No Change (NC) forecasts these were also included. Inability to beat the NC forecasts can be seen as an indication of a martingale (e.g. a random walk) like behavior of the series considered.

Smooth Transition Autoregressions. For each window a search over lag orders up to five was performed. The transition variables searched over were 1, 2, 6, and 12 lags of the left-hand side variable. The model with the lowest CC value was chosen for forecasting. This model could be, and was indeed quite often, a linear one. In order to avoid biased forecasts the forecasts were generated numerically. This was done the bootstrap approach as in Teräsvirta et al. (2005). It works in the following way:

Let $y_t = f(y_{t-1}, ..., y_{t-p}; \theta) + \varepsilon_t$ for some parameter vector θ. Letting h denote the maximal forecast horizon and N_B the number of bootstrap replications we resampled $h-1$ errors N_B times. Put differently, we created $(\hat{\varepsilon}^i_{t+1,t}, ..., \hat{\varepsilon}^i_{t+h-1,t})$ for $i = 1, ..., N_B$ and generated the τ-step ahead forecast in the following way

$$\hat{y}_{t+\tau,t} = \frac{1}{N_B} \sum_{i=1}^{N_B} f(\hat{y}^i_{t+\tau-1,t} + \hat{\varepsilon}^i_{t+\tau-1,t}, ..., \hat{y}^i_{t+\tau-p,t} + \hat{\varepsilon}^i_{t+\tau-p,t}) \tag{5.1}$$

with $\hat{y}^i_{t+\tau-j,t} + \hat{\varepsilon}^i_{t+\tau-j,t}$ replaced by $y_{t+\tau-j}$ for $j \geq \tau,\ j = 1, ..., p$. In this paper $N_B = 1000$ was used. Furthermore, an insanity filter was applied at the level of the individual bootstrap replications. Specifically, if *any* forecast of a bootstrap sample path did not belong to the interval consisting of the last observation of the given window plus minus 3 times the standard deviation of the 120 most recent observations of the window, then the whole sample path was discarded.

Kolmogorov-Gabor polynomials. For each window we searched over models with a maximum number of five lags and the maximal degree of the polynomial being five. Since the number of parameters increases rapidly in the number of lags as well as the degree of the polynomial we implemented a parameter cap of 50 such that specifications containing more than 50 parameters were ignored. Among the remaining models the one with the lowest CC value was chosen. Due to the non-affinity of the Kolmogorov-Gabor polynomials the forecasts were implemented using the bootstrap technique outlined above. An insanity filter of the same kind as previously explained was used. This turned out to be a useful strategy since the Kolmogorov-Gabor polynomials sometimes generate explosive forecasts.

Artificial Neural Networks. In each window *single hidden layer feedforward* networks with a maximum number of five lags and five hidden units were estimated. Model specification and estimation were carried out using the *QuickNet* algorithm of White (2006). The model with the lowest CC value was chosen for forecasting and as for all non-affine models the forecasts were generated using a bootstrap approach combined with the insanity filter.

Elliptic Basis Function Networks. Models with at most five lags and no more than five hidden units were estimated in each window. G was chosen as in (2.1). In other words the models considered were of the form

$$y_t = \alpha_q + \beta_q' x_t + \sum_{i=1}^{q} w_{i,q} \exp\left(\sum_{j=1}^{n} \frac{(x_j - c_{ij})^2}{2(n\sigma_{ij})^2}\right) \tag{5.2}$$

where $x_t = (y_{t-1}, ..., y_{t-n}),\ n = 1, .., 5$. The multiplication of σ_{ij} by n was done for practical reasons. Initial experiments showed that the hidden units had a very small radius of activity – in particular if many explanatory variables were included. This made the EBF less interesting and also resulted in large values of $w_{i,q}$ giving many unreasonable forecasts. By scaling the width parameters proportionally to the number of explanatory variables this numerical problem was alleviated.

There are many ways to estimate EBFs. We settled for a procedure which learns the centroids and smoothing factors unsupervised. After having determined these the problem is linear and the w_i's can be found by linear regression. This somehow resembles the structure of the *QuickNet* since the problem is split into two parts. First, the centroids and smoothing parameters are found by a grid search. Having done this, the problem becomes linear and one can determine the w's by linear regression.

Several ways exist to determine the centroids and the width parameters. We adopted a grid search over a grid constructed in the following way. For each n $y_{t-i},\ i = 1, ..., n$, was divided into five clusters of equally many observation, i.e. the splits were made at the quintiles[4]. Within each cluster we calculated the mean as well as the standard deviation and regarded these as pairs. This yielded a grid of cardinality 5^n.

We are now in a position to describe the details of the algorithm. Let x_t be given, i.e. consider a fixed vector of explanatory variables.

(1) Determine $\hat{\alpha}_0$ and $\hat{\beta}_0$ by regressing y_t on a constant and x_t. Also calculate the value of the choice criterion CC.
(2) For $q = 1, ..., 5$, add hidden units one by one in the following way. Evaluate the q'th hidden unit at each grid point (which has not been chosen previously) and determine $\hat{\alpha}_q$ $\hat{\beta}_q$ and $\hat{w}_{i,q},\ i = 1, ..., q$ by OLS. Notice that the weights of previously added hidden units as well as $\hat{\alpha}_q$ and $\hat{\beta}_q$ are allowed to change as further hidden units are added whereas the centroids and smoothing parameters remain fixed once they have been

[4]One could also split each explanatory variable into more clusters than five and thereby obtain an even finer grid. Here five clusters were chosen for each series since the maximum number of hidden units was five

determined. This resembles the *QuickNet* algorithm of White (2006). Calculate the value of the choice criterion.

(3) Repeat (1) and (2) for all choices of explanatory variables which in our case corresponds to $x_t = (y_{t-1}, ..., y_{t-n})$, $n = 1, .., 5$. Choose $\hat{n} \in \{1, .., 5\}$ and $\hat{q} \in \{0, ..., 5\}$ such that they minimize the choice criterion and forecast with the parameter estimates corresponding to these values.

The forecasts of the Elliptic Basis Function Network were produced by the same bootstrap procedure as for the aforementioned nonlinear models. Unreasonable forecasts were weeded out by the insanity filter.

5.2. **Results.** Tables 2 through 7 in the Appendix report the relative Root Mean Square Forecast Errors (RMSFE) of each point forecast relative to the RMSFE of the linear autoregressive specification. The numbers in brackets are the RMSFE of the linear autoregressive specification. Empty sections in the tables identify series for which data was unavailable. In addition to these ratios for the individual models described above we also performed three forecast combinations with equal weights: (i) Equal weighting of all models (EQ(All)), (ii) Equal weighting of all nonlinear models (EQ(NL)), and (iii) Equal weighting of all universal approximators (EQ(UA)). The No Change forecasts were regarded as linear models in this context.

As mentioned in Section 4 one must assume the existence of an interval $[a, b]$ such that $(\mathbf{x}_{t+\tau,t}, y_{t+\tau}) \in [a, b]^{n+1}$ in order to give explicit loss bounds for the WAA. Of course one can not know in advance that the forecasts as well as the realizations will belong to an interval $[a, b]$. One can circumvent this problem by choosing $[a, b]$ to be a wide interval. Our solution to this problem was the following. Let Y denote the last observation in the first estimation window and s the standard deviation of last 120 observations of the first estimation window. Then we chose $[a, b] = [Y - 3s, Y + 3s]$ which in the vast majority of the cases was more than wide enough to contain all forecasts and realizations. The corresponding value of c was denoted c_B. The reason for only calculating s on the basis of the last 120 observations was the same as explained in the treatment of the insanity filter.

The results for the WAA with $c = c_B$ were called $WAA(c_B)$. In order to investigate the performance of the WAA for smaller values of c, i.e. a faster adjustment of the weights towards the models that have performed well in the more recent past we calculated the forecasts of the WAA with $c = c_L = \frac{c_B}{100}$. These forecasts were called $WAA(c_L)$. Furthermore, the performance of the WAA was investigated with a constant low value $c = 1$. These forecasts were called $WAA(1)$. All WAA type forecasts were applied separately to each horizon. This allowed the WAA to attach different weights to each model for different horizons. This is sensible since models performing well in short term forecasting need not perform well in long term forecasting and vice versa. All remaining schemes were applied separately to each horizon. Furthermore, we performed forecasts using Inverse MSE weights, i.e.

$$w_{i,t}^{\tau} = \frac{1/MSE_{i,t-1}^{\tau}}{\sum\limits_{i=1}^{n} 1/MSE_{i,t-1}^{\tau}} \tag{5.3}$$

with $MSE_{i,t-1}^{\tau} = \frac{1}{\tau} \sum\limits_{j=t-\tau}^{t-1} (x_{j,i} - y_j)^2$ and τ indicating the window length used in the construction of the MSE. An expanding window that used all previous forecast errors in the construction of $w_{i,t}^{\tau}$, $MSE(All)$, as well as a rolling window, $MSE(Short)$, that included the past 6 forecast errors, were applied.

Finally, we generated forecasts by choosing the forecast at a given horizon to be equal to the forecast from the individual model that had the smallest most recent realized forecast error at that horizon. These forecasts are called *Last* in Tables 2 through 7.

Inspection of Tables 2 through 7 in the Appendix reveals that no single model systematically outperforms the others. This is in line with Teräsvirta et al. (2005). One notices that for some data sets there are models which have a RMSFE of 1 at all horizons. This indicates that the linear specification was chosen in each window for this model class. In general, it is seen that larger gains are to be made from the nonlinear models at longer forecast horizons. This is where one finds most of the low RMSFE ($<< 1$). The Money Supply series seem to be odd in the sense that this conclusion does not hold for them.

Table 1 summarizes Tables 2 through 7. It shows the RMSFE ratios as well as the rank of each model across all data sets and all horizons (Overall), all horizons (Horizon), and all data sets (Data). Since the loss of a τ period ahead forecast will not be observed until τ periods have elapsed one would have to initialize the weights of the loss based algorithms (WAA and MSE) in some fashion until the first loss of the individual models are realized or simply not start the comparison until the first losses at the relevant horizon are realized. We chose the latter. Consequently, for τ period ahead forecasts the actual number of evaluation periods was 72-τ.

Table 1 confirms that the gains from the non linear models are larger at the longer forecast horizons.

The No Change forecasts and the Elliptic Basis Function Networks are the individual model classes that perform best overall. The performance of the No Change forecasts is rather volatile. It is the best model for the inflation and interest rate series but the worst one for the industrial production and money supply series. The performance of the EBF is more stable.

One also notices that the individual models in general occupy the high ranks. This confirms that there are gains to be made from forecast combination. Overall weighting by the inverse MSE using all historic losses is the best procedure. However, the RMSFE ratios of all weighting schemes are very close to each other. Initially one might find it disappointing that "intelligent" schemes based on historic losses are unable to outperform naive schemes such as equal weighting. This is not surprising, however. Consider a 24-month forecast. The deviation from the truth (the loss) will not be revealed until 24 month have elapsed. The forecaster cannot use this loss in the loss-based algorithms prior to its revelation. This implies that the losses used in the loss-based algorithms stem from the performance of the individual models at least 24 month ago and this may not reflect the recent performance of the individual models which is what one is really interested in when making the combinations. This is also a possible explanation to the poor results obtained from simply forecasting with the model that had the lowest most recently realized loss at the relevant horizon. In conclusion, there are gains to be made from combining forecasts as compared to using only a single model. This is in line with Stock and Watson (1999) and Teräsvirta et al. (2005). The way one combines is not so important in the sense that RMSFE ratios are roughly equal for all combination schemes.

Finally, Figures 1 and 2 in the Appendix show the development of the weights in the WAA for $c = c_B$ and $c = c_L$ for the industrial production series for the UK at all horizons. This series was chosen since it illustrates some interesting features of the WAA. When inspecting Table 6 it is not surprising that as the forecast horizon increases more weight is given to the Kolmogorov-Gabor polynomials since these did well at the long horizons. By the same token it is sensible that the linear autoregression receives less weight as the forecast horizon becomes longer. When inspecting the development of the weights at the 24-month horizon in Figure 2 it is interesting to see how the weights of the WAA can adapt over time as the relative performance of the individual models changes. In particular, the relative weights assigned to the Elliptic Basis Function Networks and the Kolmogorov-Gabor polynomials changes around period 35.

		Horizon					Data				
	Overall	1	3	6	12	24	CPI	IP	I	M	U
AR	1.000(14)	1.000(10)	1.000(11)	1.000(15)	1.000(11)	1.000(15)	1.000(15)	1.000(10)	1.000(12)	1.000(1)	1.000(10)
NC	0.980(9)	0.999(8)	0.989(10)	0.977(9)	1.008(14)	0.926(10)	0.852(1)	1.118(15)	0.868(1)	1.154(15)	0.990(9)
STR	0.990(11)	1.015(12)	1.006(12)	0.992(12)	1.014(15)	0.923(8)	0.882(12)	1.064(14)	0.962(9)	1.048(10)	1.028(14)
KG	1.005(15)	1.058(15)	1.009(13)	0.991(11)	0.994(10)	0.972(14)	0.873(10)	1.008(12)	1.075(14)	1.082(13)	1.030(15)
ANN	0.998(13)	1.026(14)	1.011(14)	0.998(14)	1.003(12)	0.951(12)	0.877(11)	0.995(9)	1.080(15)	1.049(11)	1.020(13)
EBF	0.980(10)	0.999(9)	0.988(9)	0.981(10)	0.979(9)	0.954(13)	0.927(14)	0.981(7)	0.991(11)	1.017(2)	1.004(11)
EQ(All)	0.948(4)	0.984(4)	0.965(2)	0.950(4)	0.950(5)	0.893(4)	0.860(4)	0.972(4)	0.955(8)	1.027(6)	0.968(2)
EQ(NL)	0.958(7)	0.993(6)	0.973(6)	0.958(7)	0.961(7)	0.905(7)	0.857(2)	0.980(6)	0.984(10)	1.023(3)	0.982(6)
EQ(UA)	0.966(8)	1.001(11)	0.977(8)	0.965(8)	0.965(8)	0.923(9)	0.864(7)	0.978(5)	1.011(13)	1.028(8)	0.986(8)
WAA(c_B)	0.948(3)	0.984(3)	0.965(3)	0.949(3)	0.946(3)	0.894(6)	0.863(6)	0.969(2)	0.951(7)	1.027(7)	0.968(3)
WAA(c_L)	0.950(5)	0.988(5)	0.971(5)	0.954(6)	0.948(4)	0.889(3)	0.869(9)	0.992(8)	0.913(3)	1.039(9)	0.980(5)
WAA(1)	0.954(6)	0.994(7)	0.977(7)	0.954(5)	0.950(6)	0.893(5)	0.868(8)	1.003(11)	0.905(2)	1.061(12)	0.983(7)
MSE(All)	0.939(1)	0.983(1)	0.964(1)	0.942(1)	0.934(1)	0.872(1)	0.857(3)	0.965(1)	0.922(4)	1.026(4)	0.969(4)
MSE(6)	0.941(2)	0.984(2)	0.966(4)	0.945(2)	0.937(2)	0.875(2)	0.863(5)	0.970(3)	0.925(5)	1.026(5)	0.965(1)
Last	0.993(12)	1.021(13)	1.016(15)	0.995(13)	1.006(13)	0.927(11)	0.909(13)	1.057(13)	0.949(6)	1.086(14)	1.009(12)

TABLE 1. The Table shows RMSFE ratios as well as the corresponding ranks (in ascending order) for the overall, the horizonwise, and datawise performance of each forecast procedure. The ratios were calculated by taking the average over the relevant ratios from Tables 2 through 7. For example, the horizonwise performance was calculated by taking the average of all ratios at a fixed forecast horizon across all data sets.

6. Conclusions

In this paper we consider the forecasting performance of non-linear models relative to linear autoregressions. Three of the model classes employed are universal approximators - the Kolmogorov-Gabor polynomials, the Artificial Neural Networks, and the Elliptic Basis Function Networks.

The gains from using non-linear models seem to be larger at the longer forecast horizons than at the short horizons. Regarding forecast combinations it is found that there are gains to be made by combining forecasts; in particular the performance of various weighting schemes is more stable than the one of individual models. However, the particular way of combining the forecasts turns out to be less important. Naive equal weighting schemes perform as well as more elaborate loss-based algorithms.

The fact that there are gains to made by combining forecasts agrees with the findings of Stock and Watson (1999) and Teräsvirta et al. (2005). It should be mentioned, however, that the former authors produce their forecasts directly while we, in agreement with the latter, produce them by iterating forward.

Since the only learning algorithm discussed in this paper is the WAA, a possible next step could be to investigate the performance of other algorithms for which explicit loss bounds exist. These algorithms are interesting since the weights in them have a theoretical foundation as opposed to e.g. equal weighting.

7. Appendix

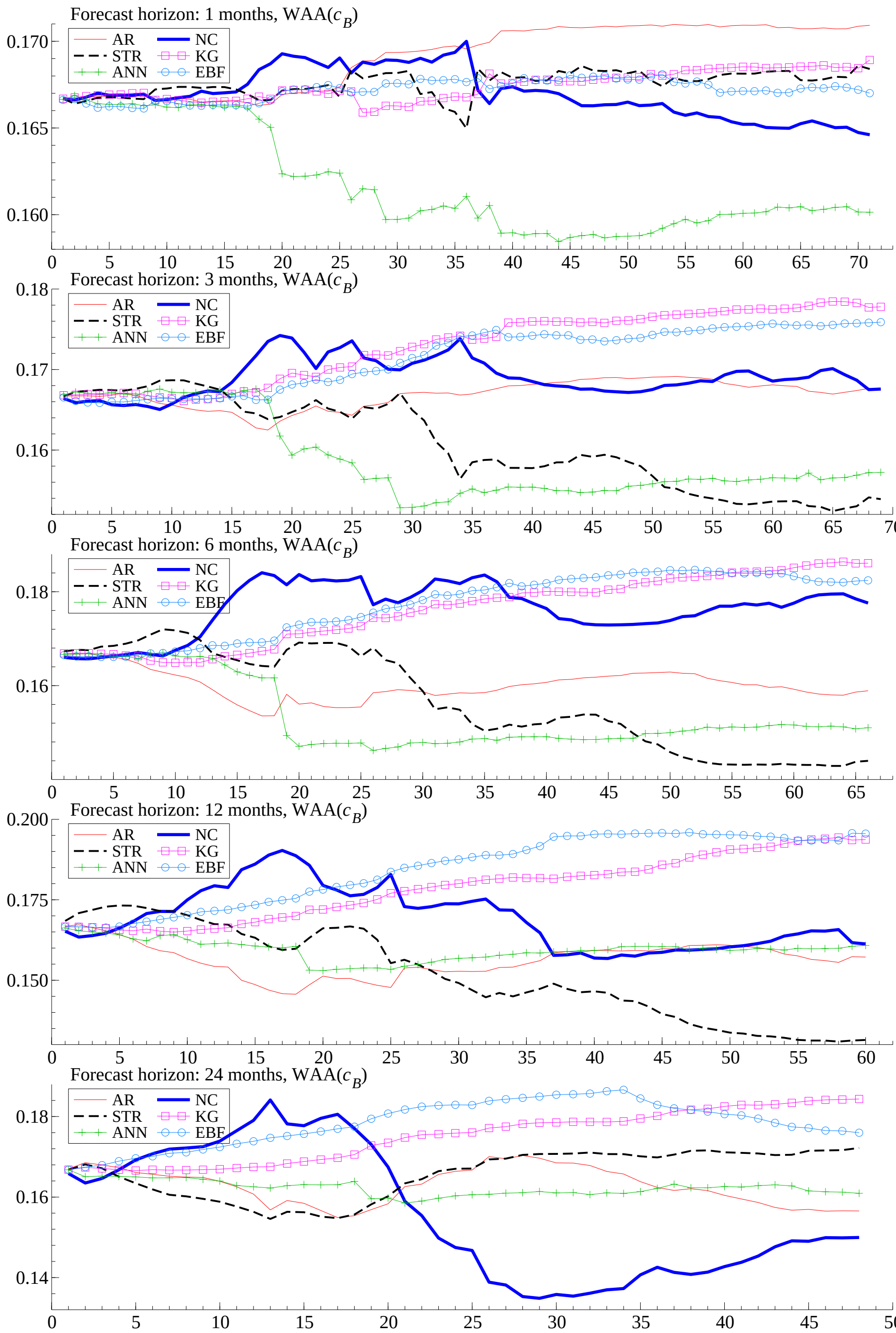

FIGURE 1. Development of the weights of the WAA applied to the Industrial Production series for the UK with $c = c_B$.

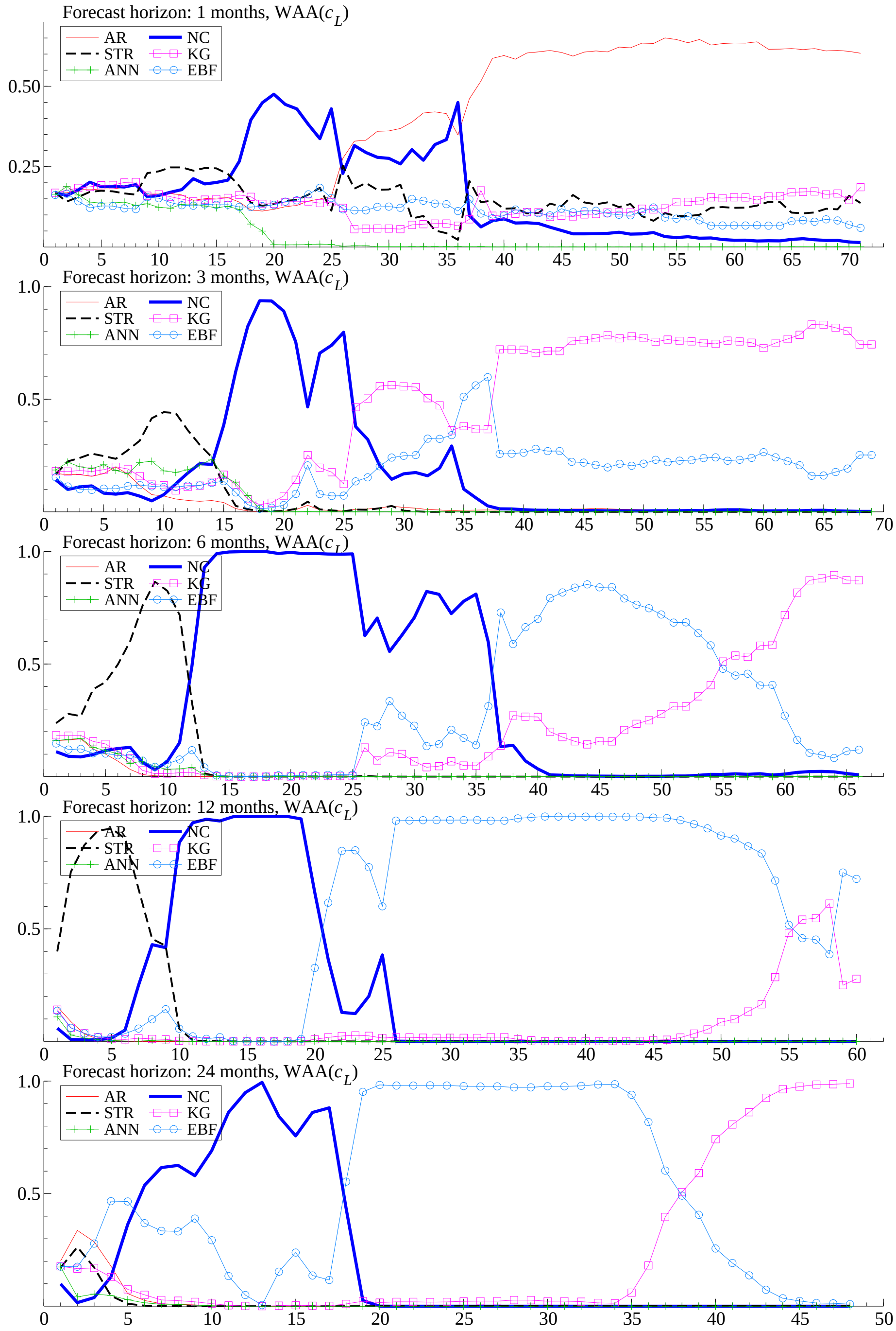

FIGURE 2. Development of the weights of the WAA applied to the Industrial Production series for the UK with $c = c_L$.

	Inflation					Industrial Production					Interest Rate					Money Supply					Unemployment				
Canada	1	3	6	12	24	1	3	6	12	24	1	3	6	12	24	1	3	6	12	24	1	3	6	12	24
AR	[0.538]	[0.940]	[1.295]	[1.644]	[1.284]	[0.519]	[0.769]	[1.165]	[1.253]	[1.377]	[0.172]	[0.274]	[0.399]	[0.672]	[1.144]	[0.923]	[1.674]	[2.263]	[2.510]	[1.823]	[0.149]	[0.248]	[0.363]	[0.475]	[0.813]
NC	0.972	0.983	0.977	0.966	0.794	1.085	1.089	1.062	1.135	0.826	0.949	0.896	0.804	0.666	0.583	1.033	1.086	1.204	1.580	1.186	0.993	0.973	0.966	0.922	0.942
STR	0.999	0.990	0.998	0.991	0.901	0.992	0.968	0.970	1.017	0.730	0.982	0.940	0.867	0.745	0.691	0.996	1.007	1.030	1.198	0.923	1.016	1.033	1.053	1.116	1.132
KG	0.985	0.962	0.940	0.885	0.659	1.044	1.180	1.271	1.308	0.988	1.010	1.025	1.070	1.112	1.078	1.027	1.031	1.106	1.370	1.074	1.030	1.059	1.079	1.080	1.079
ANN	1.013	1.000	0.993	0.966	0.843	1.021	1.111	1.125	1.116	0.881	0.993	1.000	0.995	0.921	0.816	1.021	1.010	1.034	1.189	1.020	1.000	1.000	1.000	1.000	1.000
EBF	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.038	1.067	1.075	1.189	0.931	1.000	1.000	1.000	1.000	1.000
EQ(All)	0.992	0.985	0.980	0.961	0.838	1.005	1.029	1.032	1.010	0.826	0.981	0.964	0.939	0.896	0.849	1.006	1.019	1.062	1.249	0.979	1.001	1.001	1.007	1.010	1.022
EQ(NL)	0.996	0.984	0.978	0.954	0.830	0.999	1.039	1.058	1.043	0.840	0.988	0.979	0.968	0.934	0.882	1.007	1.018	1.055	1.233	0.977	1.007	1.019	1.028	1.042	1.049
EQ(UA)	0.996	0.983	0.973	0.944	0.814	1.011	1.079	1.106	1.084	0.911	0.992	0.999	1.013	1.004	0.951	1.014	1.024	1.064	1.246	1.001	1.006	1.015	1.022	1.020	1.022
WAA(c_B)	0.992	0.985	0.980	0.960	0.834	1.005	1.029	1.031	1.005	0.825	0.981	0.964	0.939	0.894	0.842	1.006	1.019	1.062	1.241	0.977	1.001	1.001	1.007	1.010	1.023
WAA(c_L)	0.990	0.985	0.974	0.910	0.746	1.005	1.016	1.002	0.914	0.718	0.981	0.960	0.899	0.763	0.629	1.008	1.020	1.078	1.297	1.002	1.001	1.001	1.006	1.006	1.046
WAA(1)	0.989	0.984	0.977	0.906	0.743	1.004	1.007	0.997	1.019	0.725	0.981	0.956	0.869	0.732	0.606	1.012	1.041	1.128	1.334	1.000	1.001	1.000	1.006	1.004	1.052
MSE(All)	0.991	0.986	0.982	0.960	0.819	1.001	1.004	1.013	0.980	0.786	0.980	0.961	0.921	0.816	0.723	1.006	1.021	1.075	1.274	0.965	1.001	1.001	1.006	1.010	1.030
MSE(6)	0.992	0.987	0.982	0.961	0.832	1.002	1.016	1.044	1.013	0.830	0.980	0.962	0.918	0.814	0.732	1.008	1.019	1.074	1.266	0.944	1.002	1.000	1.009	1.003	1.031
Last	0.975	1.000	0.985	0.977	0.930	1.060	1.063	1.019	1.158	0.882	0.977	0.950	0.916	0.730	0.588	1.022	1.079	1.119	1.371	1.018	1.002	1.001	1.037	1.046	1.072

	Inflation					Industrial Production					Interest Rate					Money Supply					Unemployment				
Denmark	1	3	6	12	24	1	3	6	12	24	1	3	6	12	24	1	3	6	12	24	1	3	6	12	24
AR	[0.256]	[0.416]	[0.634]	[1.048]	[1.472]						[0.197]	[0.394]	[0.573]	[0.920]	[1.272]	[1.307]	[2.705]	[3.655]	[5.303]	[6.379]	[0.098]	[0.197]	[0.359]	[0.774]	[1.944]
NC	0.937	0.857	0.782	0.715	0.609						0.834	0.850	0.824	0.731	0.628	1.121	1.104	1.210	1.372	1.469	1.059	1.109	1.117	1.042	0.861
STR	1.000	0.902	0.752	0.706	0.629						1.121	1.050	1.036	1.027	1.016	0.993	0.991	1.048	1.121	1.086	1.016	1.055	1.107	1.101	1.106
KG	0.993	0.826	0.752	0.632	0.534						1.006	1.076	1.434	1.727	1.911	0.964	0.914	0.958	1.218	1.202	1.024	1.042	1.020	1.020	0.938
ANN	0.920	0.858	0.734	0.675	0.610						1.022	1.317	1.646	1.701	1.629	1.015	1.045	1.054	1.155	1.149	1.004	1.016	1.003	1.001	0.978
EBF	0.977	0.905	0.847	0.771	0.681						0.929	0.900	0.881	0.805	0.723	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
EQ(All)	0.952	0.864	0.774	0.707	0.640						0.900	0.969	1.061	1.118	1.111	0.975	0.974	1.006	1.118	1.125	0.996	1.004	1.008	1.003	0.969
EQ(NL)	0.950	0.851	0.747	0.676	0.603						0.906	1.010	1.170	1.272	1.296	0.964	0.968	0.993	1.108	1.096	1.001	1.015	1.016	1.016	0.999
EQ(UA)	0.942	0.841	0.750	0.669	0.596						0.876	1.011	1.225	1.359	1.393	0.967	0.969	0.983	1.108	1.100	1.000	1.009	0.997	0.999	0.969
WAA(c_B)	0.952	0.863	0.770	0.697	0.622						0.900	0.969	1.051	1.070	1.002	0.975	0.974	1.005	1.111	1.116	0.996	1.004	1.008	1.003	0.969
WAA(c_L)	0.942	0.843	0.785	0.698	0.605						0.893	0.931	0.868	0.763	0.669	0.981	0.969	1.013	1.016	1.046	0.996	1.006	1.016	1.042	0.965
WAA(1)	0.950	0.849	0.758	0.682	0.596						0.887	0.909	0.852	0.746	0.646	1.040	1.023	1.048	1.038	1.033	0.996	1.009	1.035	1.095	0.977
MSE(All)	0.947	0.857	0.759	0.679	0.599						0.886	0.932	0.933	0.853	0.749	0.975	0.982	1.012	1.107	1.091	0.997	1.007	1.019	1.032	0.980
MSE(6)	0.951	0.869	0.781	0.723	0.590						0.896	0.947	0.951	0.843	0.754	0.980	0.986	1.023	1.129	1.117	0.995	0.993	0.999	1.026	1.015
Last	0.956	0.889	0.828	0.829	0.601						0.952	1.076	0.918	0.806	0.660	1.096	1.021	1.027	1.274	1.168	0.994	1.011	1.020	1.041	1.023

Table 2. Root Mean Square Forecast Error ratios of each model with the linear autoregression being the benchmark. The first row in each table holds the Root Mean Square Forecast Error of the linear autoregression in sharp parentheses.

Finland	Inflation					Industrial Production					Interest Rate					Money Supply					Unemployment				
	1	3	6	12	24	1	3	6	12	24	1	3	6	12	24	1	3	6	12	24	1	3	6	12	24
AR	[0.320]	[0.544]	[0.794]	[1.277]	[1.329]	[3.489]	[4.751]	[5.154]	[5.176]	[6.058]											[0.100]	[0.224]	[0.344]	[0.446]	[0.945]
NC	0.945	0.869	0.792	0.748	1.324	0.964	1.032	1.262	1.535	1.278											0.894	0.911	0.940	1.272	1.309
STR	0.998	0.919	0.799	0.733	1.114	0.987	0.986	1.081	1.152	0.978											0.826	0.984	1.033	1.102	1.381
KG	1.014	0.946	0.815	0.734	1.095	1.001	0.991	1.015	1.045	1.115											0.901	0.900	0.865	0.899	0.846
ANN	1.000	1.022	1.016	1.036	1.300	0.956	0.966	1.026	1.095	1.015											1.004	1.064	1.187	1.444	1.494
EBF	1.000	1.000	1.000	1.000	1.000	0.950	0.939	0.976	1.027	1.002											1.000	1.000	1.000	1.000	1.000
EQ(All)	0.982	0.943	0.865	0.794	0.924	0.939	0.953	1.031	1.122	1.033											0.885	0.931	0.945	1.018	1.118
EQ(NL)	0.997	0.962	0.878	0.807	0.952	0.944	0.947	1.004	1.067	1.019											0.879	0.955	0.977	1.015	1.119
EQ(UA)	0.997	0.980	0.917	0.864	0.965	0.955	0.957	0.998	1.047	1.037											0.957	0.973	0.988	1.041	1.063
WAA(c_B)	0.982	0.942	0.861	0.795	1.161	0.939	0.954	1.029	1.102	1.033											0.885	0.931	0.944	1.018	1.117
WAA(c_L)	0.979	0.914	0.802	0.751	1.328	0.968	0.999	1.063	1.042	1.003											0.885	0.930	0.937	1.003	1.067
WAA(1)	0.979	0.914	0.802	0.750	1.328	1.003	0.993	1.068	1.045	0.997											0.881	0.923	0.943	1.019	1.137
MSE(All)	0.984	0.935	0.833	0.753	1.038	0.938	0.956	1.030	1.088	1.032											0.876	0.921	0.901	0.988	1.102
MSE(6)	0.985	0.942	0.837	0.729	1.061	0.938	0.975	1.030	1.103	1.044											0.875	0.929	0.905	0.942	1.048
Last	1.015	0.944	0.887	0.707	1.138	0.948	1.024	1.208	1.096	1.185											0.880	1.029	0.887	0.928	1.067

France	Inflation					Industrial Production					Interest Rate					Money Supply					Unemployment				
	1	3	6	12	24	1	3	6	12	24	1	3	6	12	24	1	3	6	12	24	1	3	6	12	24
AR	[0.319]	[0.534]	[0.621]	[0.774]	[1.103]	[1.426]	[1.811]	[2.246]	[2.396]	[2.008]	[0.154]	[0.327]	[0.485]	[0.764]	[0.976]						[0.061]	[0.114]	[0.222]	[0.480]	[1.018]
NC	0.939	0.821	0.707	0.632	0.591	1.044	1.011	0.971	1.041	1.187	0.992	0.935	0.876	0.797	0.733						1.061	1.194	1.163	1.063	0.927
STR	0.930	0.810	0.693	0.663	0.654	1.010	0.987	0.986	0.959	0.887	0.993	0.994	0.986	0.965	0.937						1.034	1.035	1.009	0.924	0.901
KG	0.997	0.821	0.689	0.622	0.610	0.998	0.933	0.819	0.795	0.806	0.989	0.957	0.968	0.958	0.979						1.497	1.251	1.176	1.223	1.377
ANN	0.916	0.798	0.705	0.637	0.591	0.959	0.880	0.778	0.756	0.791	1.000	1.000	1.000	1.000	1.000						1.003	1.004	1.006	1.003	1.001
EBF	0.926	0.805	0.694	0.697	0.639	0.972	0.949	0.877	0.896	0.924	1.000	1.000	1.000	1.000	1.000						1.000	1.000	1.000	1.000	1.000
EQ(All)	0.936	0.825	0.726	0.685	0.656	0.967	0.907	0.840	0.816	0.809	0.991	0.975	0.965	0.945	0.921						1.014	1.008	1.007	1.000	1.018
EQ(NL)	0.929	0.799	0.683	0.646	0.620	0.959	0.894	0.814	0.790	0.784	0.992	0.983	0.984	0.976	0.969						1.040	1.023	1.020	1.015	1.056
EQ(UA)	0.930	0.797	0.683	0.643	0.610	0.958	0.901	0.805	0.794	0.811	0.993	0.980	0.984	0.981	0.982						1.063	1.033	1.036	1.057	1.113
WAA(c_B)	0.936	0.824	0.722	0.678	0.646	0.967	0.907	0.840	0.815	0.805	0.991	0.975	0.964	0.944	0.919						1.014	1.008	1.007	1.000	1.018
WAA(c_L)	0.932	0.813	0.689	0.659	0.618	0.974	0.920	0.841	0.789	0.759	0.991	0.973	0.947	0.876	0.779						1.012	1.008	1.018	1.016	1.036
WAA(1)	0.933	0.811	0.690	0.654	0.616	0.992	0.947	0.850	0.768	0.789	0.991	0.972	0.926	0.842	0.762						1.013	1.008	1.016	1.014	1.029
MSE(All)	0.936	0.826	0.708	0.683	0.645	0.968	0.916	0.844	0.816	0.791	0.991	0.972	0.955	0.917	0.854						1.011	1.007	1.012	1.008	1.040
MSE(6)	0.938	0.829	0.705	0.679	0.646	0.971	0.911	0.836	0.849	0.815	0.991	0.971	0.950	0.923	0.859						1.012	1.014	1.021	0.998	1.006
Last	0.953	0.893	0.778	0.795	0.729	1.021	0.985	0.846	0.958	0.946	0.982	0.984	0.938	0.873	0.733						1.106	1.123	1.134	1.105	1.041

TABLE 3. Root Mean Square Forecast Error ratios of each model with the linear autoregression being the benchmark. The first row in each table holds the Root Mean Square Forecast Error of the linear autoregression in sharp parentheses.

	Inflation					Industrial Production					Interest Rate					Money Supply					Unemployment				
Germany	1	3	6	12	24	1	3	6	12	24	1	3	6	12	24	1	3	6	12	24	1	3	6	12	24
AR	[0.316]	[0.426]	[0.536]	[0.714]	[0.534]	[1.464]	[1.976]	[2.837]	[2.512]	[2.836]	[0.155]	[0.342]	[0.543]	[0.913]	[1.292]	[2.255]	[3.323]	[2.544]	[3.160]	[3.141]	[0.122]	[0.301]	[0.521]	[0.967]	[1.896]
NC	0.976	0.980	0.950	0.912	1.406	1.086	1.069	1.126	1.312	1.241	0.988	0.892	0.785	0.664	0.539	1.039	1.119	1.331	1.539	1.499	1.025	1.017	0.992	0.987	0.965
STR	1.011	1.021	1.015	0.987	1.187	1.015	1.043	1.055	1.379	1.118	0.982	0.936	0.861	0.755	0.638	1.035	1.100	1.113	1.141	1.001	0.943	0.997	1.015	1.022	0.982
KG	1.068	1.046	1.044	1.010	1.168	1.044	1.006	1.004	1.041	0.930	0.985	0.897	0.848	0.790	0.763	1.031	1.074	1.073	1.052	0.966	1.173	1.003	0.937	0.946	0.861
ANN	1.001	1.001	1.001	1.000	1.000	1.030	1.019	1.017	1.016	0.974	1.000	1.000	1.000	1.000	1.000	1.043	1.098	1.080	1.064	1.045	0.995	0.971	0.955	0.962	0.953
EBF	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.975	1.011	1.047	1.033	1.033
EQ(All)	1.005	1.002	0.991	0.965	1.035	0.997	0.994	1.001	1.063	1.008	0.968	0.933	0.889	0.831	0.736	1.010	1.037	1.048	1.118	1.052	0.978	0.959	0.956	0.969	0.954
EQ(NL)	1.016	1.013	1.007	0.984	1.020	0.997	0.986	0.985	1.063	0.986	0.977	0.938	0.901	0.849	0.752	1.012	1.040	1.016	1.056	0.999	0.984	0.957	0.955	0.968	0.946
EQ(UA)	1.019	1.011	1.007	0.988	0.990	1.013	0.999	1.000	1.010	0.961	0.981	0.950	0.934	0.910	0.860	1.012	1.033	1.008	1.032	0.999	1.006	0.950	0.941	0.954	0.935
WAA(c_B)	1.005	1.002	0.991	0.966	1.037	0.997	0.994	1.001	1.065	1.009	0.968	0.933	0.888	0.827	0.730	1.010	1.038	1.037	1.076	1.032	0.978	0.959	0.956	0.969	0.957
WAA(c_L)	1.003	1.001	0.997	1.010	1.334	1.001	1.035	1.077	1.354	1.001	0.968	0.926	0.829	0.697	0.564	1.015	1.068	1.000	1.000	1.000	0.979	0.963	0.963	0.985	0.988
WAA(1)	1.002	1.001	1.002	1.025	1.360	1.013	1.046	1.015	1.377	1.007	0.968	0.923	0.816	0.682	0.557	1.026	1.060	1.000	1.000	1.000	0.980	0.967	0.972	0.975	1.005
MSE(All)	1.005	1.001	0.992	0.974	1.103	0.996	0.991	0.992	1.073	0.987	0.963	0.925	0.865	0.767	0.626	1.012	1.026	1.030	1.065	1.008	0.981	0.963	0.959	0.972	0.964
MSE(6)	1.005	1.002	0.991	0.986	1.078	1.001	0.991	0.983	1.094	1.022	0.965	0.921	0.865	0.766	0.648	1.012	1.027	1.045	1.094	1.043	0.985	0.964	0.951	0.967	0.972
Last	1.008	1.049	0.999	1.024	1.347	1.029	1.038	0.999	1.194	1.181	1.017	0.982	0.880	0.761	0.572	1.057	1.077	1.155	1.185	1.142	0.946	1.088	1.039	0.946	1.042

	Inflation					Industrial Production					Interest Rate					Money Supply					Unemployment				
Italy	1	3	6	12	24	1	3	6	12	24	1	3	6	12	24	1	3	6	12	24	1	3	6	12	24
AR	[0.160]	[0.281]	[0.545]	[1.010]	[1.874]	[1.261]	[2.118]	[3.026]	[2.817]	[3.6039	[0.147]	[0.321]	[0.502]	[0.806]	[0.992]										
NC	0.810	0.725	0.544	0.406	0.286	0.923	0.943	0.947	1.054	0.792	1.014	0.945	0.874	0.792	0.721										
STR	0.994	0.822	0.627	0.425	0.297	1.164	1.217	1.222	1.555	0.791	0.997	0.998	0.986	0.945	0.846										
KG	0.944	0.792	0.642	0.508	0.497	1.058	0.977	0.953	0.918	0.877	1.025	1.017	0.970	0.954	0.982										
ANN	0.949	0.863	0.693	0.491	0.421	1.026	0.950	0.956	0.921	0.831	0.996	0.993	0.993	1.007	0.997										
EBF	0.985	0.827	0.630	0.474	0.411	1.057	1.028	0.988	0.906	0.877	1.000	1.000	1.000	1.000	1.000										
EQ(All)	0.930	0.791	0.642	0.507	0.459	0.988	0.963	0.946	0.930	0.791	0.989	0.980	0.956	0.930	0.862										
EQ(NL)	0.957	0.795	0.617	0.448	0.394	1.030	0.993	0.980	1.010	0.830	0.999	0.995	0.976	0.957	0.888										
EQ(UA)	0.948	0.797	0.628	0.471	0.435	1.033	0.977	0.958	0.896	0.848	1.001	0.996	0.976	0.963	0.905										
WAA(c_B)	0.930	0.791	0.641	0.499	0.437	0.988	0.962	0.947	0.939	0.788	0.989	0.980	0.956	0.930	0.862										
WAA(c_L)	0.927	0.780	0.608	0.440	0.301	0.964	0.996	1.154	1.417	0.768	0.989	0.980	0.952	0.914	0.808										
WAA(1)	0.924	0.772	0.601	0.437	0.291	0.950	1.071	1.175	1.432	0.814	0.989	0.977	0.917	0.829	0.742										
MSE(All)	0.926	0.785	0.620	0.455	0.319	0.976	0.964	0.953	0.949	0.784	0.985	0.975	0.942	0.896	0.811										
MSE(6)	0.926	0.790	0.624	0.468	0.326	0.964	0.954	0.944	0.915	0.759	0.986	0.975	0.944	0.904	0.877										
Last	0.881	0.873	0.810	0.488	0.356	0.983	0.978	1.015	1.094	0.774	1.035	1.016	0.949	0.903	0.873										

TABLE 4. Root Mean Square Forecast Error ratios of each model with the linear autoregression being the benchmark. The first row in each table holds the Root Mean Square Forecast Error of the linear autoregression in sharp parentheses.

	Inflation					Industrial Production					Interest Rate					Money Supply					Unemployment				
Japan	1	3	6	12	24	1	3	6	12	24	1	3	6	12	24	1	3	6	12	24	1	3	6	12	24
AR	[0.257]	[0.443]	[0.601]	[0.794]	[1.161]	[1.615]	[2.870]	[4.807]	[4.804]	[3.7869	[0.132]	[0.243]	[0.337]	[0.466]	[0.3309	[2.057]	[4.723]	[7.800]	[8.905]	[7.355]	[0.129]	[0.197]	[0.295]	[0.530]	[1.083]
NC	0.969	0.912	0.837	0.812	0.551	1.064	1.180	1.186	1.422	0.808	0.985	0.991	0.982	0.973	1.117	0.996	1.032	1.064	1.207	0.692	1.009	0.928	0.873	0.706	0.629
STR	1.003	0.937	0.823	0.830	0.642	1.134	1.224	1.208	1.301	0.630	1.002	0.991	1.010	1.005	1.415	1.173	0.912	0.887	1.077	0.937	1.127	1.117	1.104	0.987	0.966
KG	1.026	0.963	1.057	0.869	0.543	1.011	1.006	1.004	1.004	0.966	1.118	1.161	0.998	0.961	0.597	1.438	1.126	0.954	0.941	1.244	1.112	1.110	1.063	0.871	0.777
ANN	1.051	1.026	0.938	0.734	0.532	1.030	1.019	1.038	1.191	0.670	1.462	1.306	1.133	1.173	1.000	1.013	1.011	1.007	1.054	1.000	0.998	1.002	0.983	0.928	0.900
EBF	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.007	1.020	1.036	1.123	0.938	1.000	1.002	1.004	1.006	1.011	1.009	1.004	0.999	0.980	0.955
EQ(All)	0.971	0.910	0.853	0.724	0.478	1.012	1.032	1.034	1.075	0.750	1.014	0.939	0.955	0.960	0.733	0.987	0.946	0.957	1.025	0.956	0.995	0.989	0.961	0.875	0.856
EQ(NL)	0.973	0.910	0.863	0.717	0.462	1.030	1.044	1.044	1.083	0.776	1.044	0.941	0.953	0.970	0.634	1.015	0.931	0.930	1.001	1.044	1.008	1.016	0.991	0.902	0.881
EQ(UA)	0.976	0.949	0.945	0.779	0.549	1.010	1.004	1.007	1.028	0.839	1.083	0.969	0.971	1.017	0.608	1.045	0.977	0.965	0.988	1.080	0.998	1.001	0.974	0.888	0.861
WAA(c_B)	0.971	0.910	0.853	0.719	0.465	1.012	1.033	1.038	1.093	0.725	1.014	0.939	0.956	0.961	0.742	0.986	0.953	0.964	1.039	1.039	0.995	0.989	0.961	0.877	0.858
WAA(c_L)	0.976	0.924	0.864	0.826	0.568	1.014	1.028	1.044	1.359	0.666	1.011	0.962	0.986	1.017	1.146	1.143	1.101	1.083	1.061	1.056	0.994	0.982	0.958	0.918	1.011
WAA(1)	0.977	0.926	0.860	0.826	0.567	1.011	1.039	1.036	1.370	0.676	1.008	0.988	1.009	1.019	1.274	1.147	1.111	1.096	1.062	1.087	0.994	0.985	0.962	0.913	0.983
MSE(All)	0.979	0.919	0.806	0.748	0.512	1.012	1.031	1.035	1.090	0.716	1.005	0.971	0.968	0.971	0.996	0.985	0.952	0.960	1.026	0.995	0.995	0.988	0.954	0.904	0.941
MSE(6)	0.985	0.904	0.843	0.813	0.599	1.010	1.021	1.033	1.083	0.703	1.000	0.994	0.969	0.974	0.941	0.995	0.970	0.957	0.996	0.850	1.000	0.992	0.922	0.855	0.941
Last	1.032	0.994	0.894	0.989	0.683	1.054	1.042	1.061	1.370	0.752	1.106	1.114	1.009	1.040	1.268	1.083	0.980	0.995	1.108	0.774	1.051	1.037	0.948	0.829	0.983

	Inflation					Industrial Production					Interest Rate					Money Supply					Unemployment				
Norway	1	3	6	12	24	1	3	6	12	24	1	3	6	12	24	1	3	6	12	24	1	3	6	12	24
AR	[0.621]	[1.283]	[1.646]	[2.132]	[1.813]	[4.569]	[5.003]	[5.962]	[6.788]	[6.501]	[0.248]	[0.491]	[0.684]	[1.165]	[1.076]						[0.109]	[0.231]	[0.408]	[0.758]	[1.283]
NC	1.000	0.978	0.960	0.939	0.934	1.073	1.136	1.075	1.233	0.860	0.967	0.950	0.973	0.957	1.036						1.005	1.006	1.002	1.000	0.997
STR	1.006	1.011	0.983	0.957	0.949	1.023	1.170	1.256	1.397	0.881	1.044	1.069	1.068	1.056	1.275						0.988	0.986	0.980	0.976	0.964
KG	1.034	1.016	1.004	0.993	1.002	1.051	1.009	1.012	1.038	0.956	1.161	1.079	1.120	1.074	1.427						0.985	0.994	0.998	0.975	0.830
ANN	1.022	1.004	0.972	0.896	0.928	1.192	1.018	0.991	1.051	0.881	1.360	1.161	1.185	1.065	1.097						0.992	0.985	0.972	0.974	0.948
EBF	1.000	1.000	1.000	1.000	1.000	0.996	0.971	0.972	0.992	0.912	1.028	1.038	1.045	1.032	1.068						1.012	1.019	1.018	1.008	1.003
EQ(All)	1.001	0.993	0.973	0.942	0.882	0.971	0.978	0.979	1.058	0.817	1.025	1.009	1.030	1.008	1.103						0.989	0.989	0.986	0.981	0.952
EQ(NL)	1.008	1.000	0.977	0.942	0.899	1.002	0.996	1.018	1.097	0.873	1.059	1.035	1.060	1.028	1.162						0.984	0.985	0.981	0.974	0.931
EQ(UA)	1.012	1.001	0.986	0.955	0.944	1.051	0.985	0.983	1.020	0.904	1.088	1.044	1.075	1.033	1.162						0.984	0.987	0.983	0.975	0.921
WAA(c_B)	1.001	0.993	0.976	0.930	0.909	0.970	0.977	0.978	1.034	0.789	1.025	1.009	1.030	1.008	1.110						0.989	0.989	0.986	0.982	0.955
WAA(c_L)	1.006	1.000	0.995	0.903	0.967	1.014	1.011	1.031	1.000	0.860	1.021	1.001	1.032	1.049	1.240						0.989	0.992	0.997	1.004	1.001
WAA(1)	1.002	1.001	0.992	0.911	0.965	1.038	1.013	1.041	0.996	0.860	1.018	0.995	1.031	1.061	1.265						0.989	0.993	0.998	1.005	1.002
MSE(All)	1.001	0.994	0.975	0.940	0.891	0.969	0.979	0.985	1.035	0.769	1.022	1.007	1.029	1.015	1.114						0.989	0.990	0.986	0.984	0.967
MSE(6)	1.001	0.995	0.978	0.957	0.877	0.974	0.973	1.000	1.042	0.780	1.023	1.003	1.027	1.006	1.136						0.988	0.984	0.982	0.981	0.969
Last	0.985	1.037	0.980	1.017	0.863	1.069	1.057	1.225	1.112	0.872	1.034	1.080	0.997	1.048	1.268						0.975	0.949	0.924	0.955	0.966

TABLE 5. Root Mean Square Forecast Error ratios of each model with the linear autoregression being the benchmark. The first row in each table holds the Root Mean Square Forecast Error of the linear autoregression in sharp parentheses.

	Inflation					Industrial Production					Interest Rate					Money Supply					Unemployment				
Sweden	1	3	6	12	24	1	3	6	12	24	1	3	6	12	24	1	3	6	12	24	1	3	6	12	24
AR	[0.356]	[0.617]	[0.887]	[1.369]	[1.656]	[2.588]	[2.435]	[2.437]	[2.778]	[4.086]	[0.165]	[0.351]	[0.560]	[0.9449	[1.394]	[1.818]	[2.379]	[2.998]	[3.650]	[4.009]	[0.215]	[0.349]	[0.541]	[0.895]	[1.323]
NC	0.973	0.904	0.816	0.797	1.048	1.189	1.196	1.325	1.525	1.270	1.006	0.919	0.866	0.770	0.660	1.028	1.062	1.063	1.252	1.307	1.034	0.989	1.004	0.992	0.992
STR	1.007	0.963	0.869	0.842	1.080	1.003	1.005	1.050	1.114	1.071	0.996	0.989	0.971	0.952	0.896	1.022	1.053	1.084	1.299	1.047	1.067	1.167	1.310	1.408	1.093
KG	0.992	0.923	0.848	0.811	0.941	1.037	1.037	1.007	1.018	1.082	1.039	1.024	0.984	0.969	1.042	1.056	1.008	0.988	1.045	1.090	1.206	1.263	1.175	1.251	1.057
ANN	0.995	0.940	0.852	0.826	1.015	1.047	1.038	0.971	1.045	1.085	1.000	1.000	1.000	1.000	1.000	1.042	1.113	1.000	1.016	1.017	1.002	1.004	1.000	1.004	0.985
EBF	1.000	1.000	1.000	1.000	1.000	1.027	1.021	1.049	1.128	1.085	0.984	0.972	0.977	0.966	0.973	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
EQ(All)	0.990	0.943	0.874	0.840	0.934	1.032	1.022	1.033	1.120	1.083	0.989	0.967	0.949	0.926	0.894	1.011	1.017	0.990	1.059	1.031	0.999	0.995	0.997	0.990	0.924
EQ(NL)	0.994	0.946	0.873	0.838	0.947	1.019	1.012	1.001	1.063	1.076	0.991	0.981	0.967	0.958	0.945	1.015	1.022	0.990	1.060	1.024	1.016	1.040	1.045	1.039	0.914
EQ(UA)	0.993	0.945	0.881	0.848	0.927	1.028	1.023	0.994	1.054	1.082	0.991	0.982	0.971	0.964	0.963	1.016	1.021	0.977	1.003	1.023	1.018	1.035	1.025	1.033	0.928
WAA(c_B)	0.990	0.943	0.874	0.841	0.954	1.032	1.022	1.032	1.111	1.080	0.989	0.967	0.949	0.926	0.891	1.011	1.017	0.990	1.059	1.032	0.999	0.995	0.997	0.986	0.924
WAA(c_L)	0.990	0.944	0.872	0.824	1.110	1.018	1.013	1.004	1.022	1.052	0.989	0.967	0.940	0.886	0.707	1.011	1.017	0.996	1.072	1.060	0.999	0.987	0.982	0.989	1.090
WAA(1)	0.990	0.948	0.871	0.810	1.112	1.019	1.040	1.025	1.022	1.031	0.989	0.965	0.913	0.853	0.674	1.016	1.033	1.007	1.123	1.148	0.999	0.984	0.991	1.027	1.074
MSE(All)	0.988	0.944	0.874	0.820	0.994	1.030	1.020	1.026	1.076	1.067	0.990	0.967	0.941	0.923	0.773	1.011	1.018	0.993	1.060	1.068	1.001	1.000	1.007	1.002	0.887
MSE(6)	0.987	0.942	0.872	0.797	1.006	1.030	1.021	1.029	1.088	1.095	0.990	0.968	0.935	0.904	0.780	1.013	1.021	1.003	1.090	1.071	1.001	1.007	1.015	1.012	0.911
Last	0.991	0.943	0.872	0.802	1.058	1.148	1.128	1.229	1.132	1.207	1.073	0.991	0.930	0.870	0.660	1.013	1.108	1.107	1.252	1.193	1.092	1.089	1.224	1.233	0.991

	Inflation					Industrial Production					Interest Rate					Money Supply					Unemployment				
UK	1	3	6	12	24	1	3	6	12	24	1	3	6	12	24	1	3	6	12	24	1	3	6	12	24
AR	[0.264]	[0.502]	[0.694]	[1.060]	[1.347]	[1.290]	[1.576]	[1.953]	[2.004]	[1.806]	[0.147]	[0.329]	[0.450]	[0.706]	[0.841]						[0.045]	[0.072]	[0.121]	[0.263]	[0.656]
NC	0.892	0.779	0.655	0.522	0.571	1.063	0.995	0.934	1.159	0.964	1.021	0.935	0.863	0.772	0.617						1.011	1.052	1.016	0.749	0.464
STR	0.980	0.904	0.776	0.535	0.575	1.025	1.104	1.089	1.219	0.797	0.987	0.959	0.914	0.856	0.713						0.973	0.938	0.900	0.700	0.423
KG	0.955	0.823	0.703	0.593	0.579	1.020	0.926	0.879	0.896	0.777	1.054	1.122	1.226	1.286	1.489						1.062	1.029	0.995	0.804	0.783
ANN	0.974	0.799	0.731	0.604	0.537	1.106	1.078	1.043	1.026	0.854	1.000	0.998	0.997	1.002	1.042						1.015	1.097	1.189	1.081	0.902
EBF	0.995	0.992	1.005	1.034	0.887	1.038	0.932	0.891	0.917	0.950	1.014	1.020	1.034	1.021	1.012						1.051	1.111	1.141	0.991	0.782
EQ(All)	0.941	0.843	0.730	0.584	0.604	1.007	0.948	0.913	0.948	0.788	0.993	0.988	0.975	0.947	0.905						0.988	0.940	0.892	0.741	0.637
EQ(NL)	0.951	0.844	0.735	0.596	0.586	1.010	0.955	0.936	0.972	0.827	1.001	1.009	1.017	1.007	1.016						1.007	0.987	0.951	0.766	0.626
EQ(UA)	0.949	0.837	0.745	0.642	0.603	1.027	0.947	0.921	0.929	0.843	1.009	1.035	1.070	1.083	1.158						1.022	1.015	0.994	0.828	0.719
WAA(c_B)	0.941	0.843	0.729	0.584	0.590	1.008	0.948	0.912	0.947	0.785	0.993	0.988	0.974	0.945	0.888						0.988	0.940	0.892	0.741	0.635
WAA(c_L)	0.941	0.831	0.699	0.620	0.567	1.021	0.971	0.999	1.038	0.772	0.993	0.982	0.933	0.862	0.625						0.988	0.941	0.894	0.743	0.564
WAA(1)	0.938	0.820	0.673	0.644	0.570	1.036	1.000	1.041	1.029	0.776	0.993	0.978	0.904	0.819	0.626						0.988	0.941	0.899	0.748	0.532
MSE(All)	0.942	0.839	0.706	0.628	0.568	1.009	0.948	0.913	0.951	0.776	0.993	0.983	0.949	0.864	0.701						0.989	0.948	0.909	0.732	0.515
MSE(6)	0.944	0.845	0.718	0.624	0.564	1.014	0.948	0.919	0.948	0.767	0.997	0.993	0.953	0.898	0.691						0.991	0.943	0.886	0.737	0.527
Last	1.026	0.919	0.771	0.768	0.600	1.084	1.051	1.036	1.203	0.898	1.062	1.082	1.038	0.932	0.653						1.054	1.017	0.968	0.899	0.644

TABLE 6. Root Mean Square Forecast Error ratios of each model with the linear autoregression being the benchmark. The first row in each table holds the Root Mean Square Forecast Error of the linear autoregression in sharp parentheses.

	Inflation					Industrial Production					Interest Rate					Money Supply					Unemployment				
US	1	3	6	12	24	1	3	6	12	24	1	3	6	12	24	1	3	6	12	24	1	3	6	12	24
AR	[0.406]	[0.709]	[0.912]	[1.209]	[1.030]	[0.729]	[1.439]	[2.294]	[1.722]	[1.417]	[0.215]	[0.425]	[0.524]	[0.727]	[0.612]	[1.348]	[2.031]	[2.544]	[3.179]	[4.996]	[0.117]	[0.218]	[0.346]	[0.500]	[0.846]
NC	0.923	0.931	0.981	1.007	1.223	1.026	1.077	1.056	1.782	1.300	1.029	0.979	0.942	0.898	0.867	0.948	1.006	1.009	0.984	1.081	1.064	1.144	1.224	1.037	0.855
STR	1.002	0.976	1.003	1.030	1.247	0.999	0.940	0.863	1.315	1.114	1.004	0.992	0.962	0.901	0.787	1.065	1.039	1.025	1.016	1.035	1.019	1.069	1.149	1.139	0.921
KG	1.076	0.995	0.969	0.971	1.120	0.968	0.968	0.948	1.520	1.131	1.136	1.033	1.042	1.046	1.038	1.372	1.042	1.041	0.997	1.057	0.999	1.060	1.109	0.941	0.840
ANN	0.996	0.965	0.958	0.948	0.950	0.964	0.929	0.895	1.365	0.984	1.004	1.001	1.000	1.000	0.994	1.062	1.057	1.033	1.005	1.023	1.000	1.000	1.000	1.000	1.000
EBF	0.973	0.929	0.980	0.931	1.000	0.985	0.923	0.890	1.060	0.949	1.000	1.000	1.000	1.000	1.000	1.035	1.034	1.029	1.008	1.078	1.004	1.009	1.011	1.000	1.000
EQ(All)	0.986	0.958	0.974	0.971	1.062	0.970	0.933	0.885	1.195	1.002	0.999	0.993	0.977	0.953	0.866	1.020	1.007	1.006	0.986	1.035	0.990	1.006	1.036	0.964	0.930
EQ(NL)	1.002	0.958	0.971	0.964	1.064	0.972	0.922	0.878	1.298	1.019	1.011	1.001	0.991	0.973	0.927	1.057	1.016	1.015	0.995	1.045	0.991	1.015	1.042	0.986	0.936
EQ(UA)	1.004	0.955	0.964	0.945	1.012	0.966	0.922	0.887	1.295	1.001	1.017	1.006	1.008	1.009	1.007	1.097	1.027	1.019	0.993	1.050	0.989	1.007	1.025	0.967	0.943
WAA(c_B)	0.985	0.958	0.974	0.972	1.069	0.970	0.934	0.892	1.172	1.004	0.999	0.993	0.977	0.952	0.847	1.019	1.007	1.006	0.987	1.035	0.991	1.006	1.036	0.967	0.938
WAA(c_L)	0.963	0.945	0.981	0.983	1.124	0.975	0.956	0.995	1.044	0.917	0.999	0.992	0.967	0.924	0.882	1.011	1.016	1.010	1.013	1.011	0.992	1.024	1.074	1.022	1.000
WAA(1)	0.967	0.947	0.981	0.984	1.125	0.976	0.975	1.001	1.048	0.924	0.999	0.993	0.972	0.936	0.884	1.141	1.047	1.001	1.038	1.000	0.992	1.023	1.074	1.022	1.000
MSE(All)	0.982	0.957	0.973	0.973	1.079	0.971	0.931	0.902	1.167	0.987	0.993	0.993	0.975	0.945	0.835	1.017	1.006	1.006	0.987	1.035	0.994	1.021	1.060	0.963	0.958
MSE(6)	0.983	0.957	0.974	0.972	1.106	0.967	0.935	0.932	1.193	0.970	0.996	0.994	0.979	0.950	0.829	1.023	1.008	1.002	0.986	1.036	0.991	1.009	1.055	0.947	0.922
Last	0.988	0.948	1.009	1.019	1.188	0.978	0.895	1.017	1.557	1.001	1.152	1.044	0.999	0.973	0.949	1.013	1.032	1.028	0.988	1.092	1.029	1.054	1.091	0.917	0.924

TABLE 7. Root Mean Square Forecast Error ratios of each model with the linear autoregression being the benchmark. The first row in each table holds the Root Mean Square Forecast Error of the linear autoregression in sharp parentheses.

References

Bates, J. and C. Granger (1969). Combination of forecasts . *Operations Research Quarterly 20*(4), 451–468.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems (MCSS) 2*(4), 303–314.

Freund, Y. and R. Schapire (1997). A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences 55*(1), 119–139.

Granger, C. and T. Teräsvirta (1993). *Modelling Nonlinear Economic Relationships*. Oxford University Press, USA.

Hornik, K., M. Stinchcombe, and H. White (1989). Multilayer feedforward networks are universal approximators. *Neural Networks 2*(5), 359–366.

Kivinen, J. and M. Warmuth (1999). Averaging Expert Predictions. *In Paul Fischer and Hans Ulrich Simon, editors, Proceedings of the 4th European Conference on Computational Learning Theory EuroCOLT '99*, 153–167.

Park, J. and I. Sandberg (1991). Universal approximation using radial-basis-function networks. *Neural computation 3*(2), 246–257.

Park, J. and I. Sandberg (1993). Approximation and radial-basis-function networks. *Neural Computation 5*(2), 305–316.

Park, J. and I. Sandberg (1994). Nonlinear approximations using elliptic basis function networks. *Circuits, Systems, and Signal Processing 13*(1), 99–113.

Stock, J. and M. Watson (1999). A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series, in R.F. Engle and H. White (eds). *Cointegration, Causality and Forecasting: A Festschrift in honour of Clive WJ Granger*, 1–44.

Swanson, N. and H. White (1995). A model-selection approach to assessing the information in the term structure using linear models and artificial neural networks. *Journal of Business & Economic Statistics*, 265–275.

Teräsvirta, T., D. van Dijk, and M. Medeiros (2005). Linear models, smooth transition autoregressions, and neural networks for forecasting macroeconomic time series: A re-examination. *International Journal of Forecasting 21*(4), 755–774.

Timmermann, A. (2006). Forecast Combination, in G. Elliott, C.W.J. Granger, A Timmermann (eds). *Handbook of Economic Forecasting 1*, 135–196.

White, H. (2006). Approximate nonlinear forecasting methods, in G. Elliott, C.W.J. Granger, A Timmermann (eds). *Handbook of Economic Forecasting 1*, 459–512.

Research Papers 2009

CREATES

Center for Research in Econometric Analysis of Time Series

2009-03: Andrés González, Kirstin Hubrich and Timo Teräsvirta: Forecasting inflation with gradual regime shifts and exogenous information

2009-4: Theis Lange: First and second order non-linear cointegration models

2009-5: Tim Bollerslev, Natalia Sizova and George Tauchen: Volatility in Equilibrium: Asymmetries and Dynamic Dependencies

2009-6: Anders Tolver Jensen and Theis Lange: On IGARCH and convergence of the QMLE for misspecified GARCH models

2009-7: Jeroen V.K. Rombouts and Lars Stentoft: Bayesian Option Pricing Using Mixed Normal Heteroskedasticity Models

2009-8: Torben B. Rasmussen: Jump Testing and the Speed of Market Adjustment

2009-9: Dennis Kristensen and Andrew Ang: Testing Conditional Factor Models

2009-10: José Fajardo and Ernesto Mordecki: Skewness Premium with Lévy Processes

2009-11: Lasse Bork: Estimating US Monetary Policy Shocks Using a Factor-Augmented Vector Autoregression: An EM Algorithm Approach

2009-12: Konstantinos Fokianos, Anders Rahbek and Dag Tjøstheim: Poisson Autoregression

2009-13: Peter Reinhard Hansen and Guillaume Horel: Quadratic Variation by Markov Chains

2009-14: Dennis Kristensen and Antonio Mele: Adding and Subtracting Black-Scholes: A New Approach to Approximating Derivative Prices in Continuous Time Models

2009-15: Charlotte Christiansen, Angelo Ranaldo and Paul Söderllind: The Time-Varying Systematic Risk of Carry Trade Strategies

2009-16: Ingmar Nolte and Valeri Voev: Least Squares Inference on Integrated Volatility and the Relationship between Efficient Prices and Noise

2009-17: Tom Engsted: Statistical vs. Economic Significance in Economics and Econometrics: Further comments on McCloskey & Ziliak

2009-18: Anders Bredahl Kock: Forecasting with Universal Approximators and a Learning Algorithm