FORECASTING MACROECONOMIC VARIABLES USING NEURAL NETWORK MODELS AND THREE AUTOMATED MODEL SELECTION TECHNIQUES

ANDERS BREDAHL KOCK AND TIMO TERÄSVIRTA

AARHUS UNIVERSITY AND CREATES

Abstract. When forecasting with neural network models one faces several problems, all of which influence the accuracy of the forecasts. First, neural networks are often hard to estimate due to their highly nonlinear structure. To alleviate the problem, White (2006) presented a solution (QuickNet) that converts the specification and nonlinear estimation problem into a linear model selection and estimation problem. We shall compare its performance to that of two other procedures building on the linearisation idea: the Marginal Bridge Estimator and Autometrics. Second, one must decide whether forecasting should be carried out recursively or directly. This choice is investigated in this work. The economic time series used in this study are the consumer price indices for the G7 and the Scandinavian countries. In addition, a number of simulations are carried out and results reported in the paper.

Keywords: artificial neural network, forecast comparison, model selection, nonlinear autoregressive model, nonlinear time series, root mean square forecast error, Wilcoxon's signed-rank test

JEL Classification Codes: C22, C45, C52, C53

Date: February 16, 2015.

1. Introduction

Artificial Neural Networks (ANN) have been quite popular in many areas of science for describing various phenomena and forecasting them. They have also been used in forecasting macroeconomic time series and financial series, see Kuan and Liu (1995) for a successful example on exchange rate forecasting, and Zhang et al. (1998) and Rech (2002) for more mixed results. Maasoumi et al. (1994) already modelled macroeconomic time series using ANN models, although they did not use their models for forecasting. The main argument in favour of these models is that ANNs are universal approximators, which means that they are capable of arbitrarily accurately approximating functions that satisfy only mild regularity conditions. The ANN models thus have a strong nonparametric flavour. One may therefore expect them to be a versatile tool in economic forecasting and quickly adapt to rapidly changing forecasting situations. Ahmed et al. (2010) conducted an extensive forecasting study comprising more than 1000 economic time series from the M3 competition Makridakis and Hibon (2000), and a large number of what they called machine learning tools. They concluded that the ANN model that we are going to consider, the single hidden-layer feedforward ANN model or multi-layer perceptron with one hidden layer, was one of the best or even the best performer in their study. A single hidden-layer ANN model is already a universal approximator; see Cybenko (1989) and Hornik et al. (1989).

A major problem in the application of ANN models is the specification and estimation of these models. A large number of modelling strategies have been developed for this purpose. It is possible to begin with a small model and increase its size ("specific-to-general", "bottom up", or "growing the network"). Conversely, one can specify a network with a large number of variables and hidden units or "neurons" and then reduce its size ("general-to-specific", "top down" or "pruning the network"). Since the ANN model is nonlinear in parameters, its parameters have to be estimated numerically, which may be a demanding task if the number of parameters in the model is large. White (2006) devised a clever strategy for modelling ANNs that converts the specification and ensuing nonlinear estimation problem into a linear model selection problem. This greatly simplifies the estimation stage and alleviates the computational effort. It is therefore of interest to investigate

how well this strategy, called QuickNet, performs in univariate macroeconomic forecasting. A natural benchmark in that case is a linear autoregressive model.

Quite often, application of QuickNet leads to a situation in which the number of variables in the set of candidate variables exceeds the number of observations. QuickNet handles these cases without problems, because it essentially works from specific to general and then back again. We shall also consider a one-way variant from specific to general in this study. One may want to set a maximum limit for variables to be included in the model to control its size.

There exist other modelling strategies that can also be applied to selecting the variables. In fact, White (2006) encouraged comparisons between his method and other alternatives, and here we shall follow his suggestion. In this work, we consider two additional specification techniques. One is Autometrics by Doornik (2009), see also Krolzig and Hendry (2001) and Hendry and Krolzig (2005), and the other one is the Marginal Bridge Estimator (MBE), see Huang et al. (2008). The former is designed for econometric modelling, whereas the latter one has its origins in statistics. Autometrics works from general to specific, and the same may be said about MBE. We shall compare the performance of these three methods when applying White's idea of converting the specification and estimation problem into a linear model selection problem and selecting hidden units for our ANN models. That is one of the main objectives of this paper.

The focus in this study is on multiperiod forecasting. There are two ways of generating multiperiod forecasts. One consists of building a single model and generating the forecasts for more than one period ahead recursively. The other one, called direct forecasting, implies that a separate model is built for each forecasting horizon, and no recursions are involved. For discussions, see for example Teräsvirta (2006), Teräsvirta et al. (2010, Chapter 14), or Kock and Teräsvirta (2011a). In nonlinear forecasting, the latter method appears to be more common, see for example Stock and Watson (1999) and Marcellino (2002), whereas Teräsvirta et al. (2005) constitutes an example of the former alternative. A systematic comparison of the performance of the two methods exists, see Marcellino et al. (2006), but it is restricted to linear autoregressive models. Our aim is to extend these comparisons to nonlinear ANN models.

Nonlinear models can sometimes generate obviously insane forecasts. One way of alleviating this problem is to use insanity filters as in Swanson and White (1995, 1997a,b) who discuss this issue. We will compare two filters to the unfiltered forecasts and see how they impact on the forecasting performance of the neural networks.

In this work the ANN models are augmented by including lags of the variable to be forecast linearly in them. As a result, the augmented models nest a linear autoregressive model. It is well known that if the data-generating process is linear, the augmented ANN model is not even locally identified; see for example Lee et al. (1993), Teräsvirta et al. (1993) or Teräsvirta et al. (2010, Chapter 5) for discussion. A general discussion of identification problems in ANN models can be found in Hwang and Ding (1997). It may then be advisable to first test linearity of each series under consideration before applying any ANN modelling strategy to it. But then, it may also be argued that linearity tests are unnecessary, because the set of candidate variables can be (and in our case is) defined to include both linear lags and hidden units. Our empirical results (not reported in detail in this paper) suggest that testing linearity before building ANN models is of little help in forecasting with these models.

The main criterion of comparing forecasts is the Root Mean Square Forecast Error (RMSFE), which implies a quadratic loss function. Other alternatives are possible, but RMSFE is commonly used and thus applied here. We rank the methods, which makes some comparisons possible.

It might be desirable to compare White's method with modelling strategies which are not based on linearising the problem but in which statistical methods such as hypothesis testing and nonlinear maximum likelihood estimation are applied. Examples of these include Swanson and White (1995, 1997a,b), Anders and Korn (1999) and Medeiros et al. (2006). These approaches do, however, require plenty of human resources, unless the number of time series under consideration and forecasts generated from them are small. This is because nonlinear iterative estimation is hard to automate. Each estimation needs a nonnegligible amount of attention from the user, and when the number of time series to be considered is large, ANN model building and forecasting tend to require a substantial amount of resources.

In this paper we investigate the forecasting performance of the aforementioned techniques. We first conduct a set of small simulations to see how well these techniques perform when the data are generated by a known nonlinear model or by a stationary linear autoregressive model. The economic data sets consist of the monthly consumer price index series from the 1960's until the end of 2009 for the G7 as well as the four Scandinavian countries. We have also implemented the techniques on unemployment series for the same countries. The results show very little variation and are omitted[1].

The plan of the paper is as follows. The neural network model is presented in Section 2 and estimation techniques in Section 3. The recursive and direct forecasting methods are discussed in Section 4 and the results are summarized in Section 5. Section 6 concludes.

2. The model

We begin by briefly introducing the Artificial Neural Network (ANN) model and reviewing some of its properties. The techniques for specifying the structure of the model and estimating the parameters will be considered in the next section. Our model is the so-called single-hidden-layer feedforward autoregressive neural network model or single-hidden-layer perceptron

$$y_t = \beta_0'\mathbf{z}_t + \sum_{j=1}^{q} \beta_j (1 + \exp\{\gamma_j'\mathbf{z}_t\})^{-1} + \varepsilon_t \tag{1}$$

where $\mathbf{z}_t = (1, y_{t-1}, ..., y_{t-p})'$, $\beta_0 = (\beta_{00}, \beta_{01}, ..., \beta_{0p})'$, $\gamma_j = (\gamma_{j0}, \gamma_{j1}, ..., \gamma_{jp})'$ and $\varepsilon_t \sim \text{iid}\mathcal{N}(0, \sigma^2)$. The weak stationarity condition of (1) is the same as that of the corresponding linear AR(p) model, namely, all roots of $1 - \sum_{i=1}^{p} \beta_{0i} z^i$ should lie outside the unit circle.

The ANN model is a so-called universal approximator in the following sense. Suppose there is a functional relationship between y and $\mathbf{z}$: $y = H(\mathbf{z})$. Then under appropriate regularity conditions for any $\delta > 0$ there exists a positive integer $q < \infty$ such that $\left\| H(\mathbf{z}) - \sum_{j=1}^{q} \beta_j (1 + \exp\{\gamma_j'\mathbf{z}\})^{-1} \right\| < \delta$ where $\|\cdot\|$ is an appropriate norm on the function space H is an element of. This indicates that (1) is a very flexible functional form and thus in principle capable of satisfactorily approximating various nonlinear processes.

[1]The results for the unemployment series can be found at `sites.google.com/site/andersbkock`

Before forecasting with the model (1), the number of logistic functions or hidden units, q, has to be specified and its parameters estimated. Various specification techniques have been proposed in the literature. One possibility is to begin with a large model (large q) and reduce the size of the model. One can also begin with a small model and add hidden units. Either way, one has to estimate the parameters of the model which, given that the model is heavily nonlinear, may be numerically demanding, in particular when q is large. For discussions, see for example Fine (1999, Chapter 6), Goffe et al. (1994), or Simon (1999).

Nevertheless, if the parameter vectors γ_j, $j = 1, ..., q$, are known, the model is linear in parameters. This opens up the possibility of combining specification and estimation into a single linear model selection problem. As already mentioned, White (2006) suggested this technique for specifying and estimating artificial neural network models. The linear model selection problem encountered is the one of choosing a subset of variables from the set

$$S = \{y_{t-i}, i = 1, ..., p; (1 + \exp\{\gamma_j' \mathbf{z}_t\})^{-1}, j = 1, ..., M\} \tag{2}$$

of $M + p$ elements where M is large. How well the data-generating process can be approximated by an ANN model depends on the size of S. Furthermore, in a typical macroeconomic application the size of S is likely to exceed the number of observations. Model selection techniques that can handle such a situation are discussed in the next section.

The neural network model (1) is not the only possible universal approximator for this application. White (2006) mentioned ridgelets, Candès (1998, 2003), as an alternative. Polynomials would probably in this context not be the best possible class of universal approximators. The fit of the estimated polynomials often deteriorates at both ends of the series they describe, which is not a desirable feature in forecasting economic variables such as growth rates. Another universal approximator, the Fourier Flexible Form (FFF), was discussed in Gallant (1984). In applying the FFF, the problem of constructing the variables would have two aspects. One would have to choose the linear combinations $\gamma_j' \mathbf{z}_t$, but also determine the number of frequencies in the sum of trigonometric components. We settle for the ANN model, because it is, alongside the polynomials,

probably the most commonly used universal approximator, and because QuickNet was originally designed to solve the specification and estimation problem for this model.

3. Modelling with three automated model selection algorithms

We consider three model selection algorithms that apply to our modelling problem, in which the number of variables exceeds the number of observations. They are Autometrics, see Doornik (2009), Marginal Bridge Estimator (MBE), see Huang et al. (2008), and QuickNet. Autometrics is built on the principle of moving from general to specific, which means beginning with a large model and gradually reducing its size. QuickNet may be characterised as a specific-to-general-to specific procedure, although we shall also report results on a simplified specific-to-general version. The starting-point of MBE also involves all variables, but the process of selecting the final model is very different from Autometrics. We shall now describe these three techniques in more detail, beginning with Autometrics.

3.1. **Autometrics.** Modelling begins with a linear model called the General Unrestricted Model (GUM). When the number of variables is less than the number of observations the GUM contains all candidate variables. The model is subjected to significance tests. If all variables have statistically significant coefficient estimates, the GUM is the final model. Otherwise, because there is no unique way of going from general to specific, the algorithm searches simpler models using different search paths. It does this by removing variables with insignificant coefficients. When the model cannot be reduced any more, it is subjected to diagnostic tests. If it passes the tests, it is called a terminal model. If not, the algorithm backtracks. Since there are many search paths, there will in general be several terminal models as well[2,3].

After reaching this stage, Autometrics forms the union of the terminal models and tests the terminal models against it. The union of the models that pass the tests forms a new GUM. The general-to-specific testing procedure is then repeated and a new set of terminal models obtained. If all models in this set are rejected against the new union model, the union will be the final

[2]Following the advice of Jurgen Doornik (personal communication) we also ran Autometrics without diagnostic tests, but this had little effect on the results.
[3]The version of Autometrics used in this paper is the one to be found in PcGive 13.0.

model. Otherwise, modelling restarts with yet another GUM and continues until a final model has been reached.

When the number of variables exceeds the number of observations. Hendry and Krolzig (2005) suggested dividing the variables into subsets, each of which contains fewer variables than observations. Autometrics does this automatically. This implies that at the outset there exists more than one GUM. Each of these GUMs now forms a starting-point for Autometrics and the algorithm yields a set of terminal models for each GUM. The terminal models derived from all subsets of variables or all GUMs are merged to form a single union model. If the number of variables in this model is less than the number of observations model selection proceeds from this union model as described above.

Autometrics is partly a black box. The user can, however, affect the outcomes by selecting a number of settings, such as the significance level of the tests the algorithm relies on.

3.2. **Marginal Bridge Estimator.** MBE is designed for situations often occurring in statistical and genomic applications in which there is a large number of candidate variables but only a small subset of these may belong to the model. Following Huang et al. (2008), consider first the Bridge Estimator (BE). This is a shrinkage estimator for a linear regression model

$$y_i = \alpha + \beta' \mathbf{x}_i + \varepsilon_i, \ i = 1, ..., n \tag{3}$$

where $\mathbf{x}_i = (x_{i1}, ..., x_{ip_n})'$ is a $p_n \times 1$ observation vector (p_n may increase in n but $p_n < n$) and $\alpha = 0$ without loss of generality. Furthermore, $\varepsilon_i \sim \text{iid}(0, \sigma^2)$. BE is a result of minimizing

$$L(\beta) = \sum_{i=1}^{n} \left(y_i - \beta' \mathbf{x}_i\right)^2 + \lambda_n \sum_{k=1}^{p_n} |\beta_k|^\gamma \tag{4}$$

where $\gamma > 0$, and $\lambda_n > 0$ determines the size of the penalty. Let the true parameter vector be $\beta_0 = (\beta_{10}', \beta_{20}')'$ with β_{10} having nonzero entries, $\beta_{20} = \mathbf{0}$, and let $\hat{\beta}_n = (\hat{\beta}_{1n}', \hat{\beta}_{2n}')'$ be the corresponding estimator from (4). BE minimizes the OLS objective function plus a penalty for parameters different from zero. Hence, it shrinks estimates towards zero. Huang et al. (2008)

showed that under regularity conditions parameters are i) estimated consistently ($\hat{\beta}_n \to \beta_n$ in probability), ii) the truly zero parameters are set to zero $\left(P(\hat{\beta}_{2n} = 0) \to 1\right)$ and iii) the asymptotic distribution of the estimators of nonzero parameters is the same as if only these had been included in the model. This means that the parameters of the nonzero coefficients are estimated (asymptotically) as efficiently as if only the relevant variables had been included in the original model.

For BE to possess this property one needs $p_n < n$. When this condition no longer holds, MBE can be applied. The idea is to run a series of 'mini' or 'marginal' regressions, with a penalty on parameters that differ from zero. The function to be minimized equals

$$Q_n(\beta) = \sum_{k=1}^{p_n} \sum_{i=1}^{n} (y_i - \beta_k x_{ik})^2 + \lambda_n \sum_{k=1}^{p_n} |\beta_k|^\gamma \tag{5}$$

Let $\widetilde{\beta}_n = (\widetilde{\beta}'_{1n}, \widetilde{\beta}'_{2n})'$ be the estimator of β_0 from (5). Under regularity conditions and $0 < \gamma < 1$, (a) the estimator $\widetilde{\beta}_{2n} = \mathbf{0}$ with probability converging to one, and (b) $P(\widetilde{\beta}_{1nk} \neq 0, \widetilde{\beta}_{1nk} \in \widetilde{\beta}_{1n}) \to 1$, as $n \to \infty$. Property (a) is similar to ii) for BE. In view of (b), the elements of $\widetilde{\beta}_{1n}$ converge to nonzero values. Thus, (a) and (b) jointly can be expected to efficiently separate the relevant variables from the rest.

Of the conditions underlying the aforementioned result the so-called partial orthogonality condition is problematic in a time series context. It states that the correlation between the relevant and irrelevant variables is not allowed to be too high. This condition can be violated if the explanatory variables are lags and functions of lags of the dependent variable as in our case. However, as we shall see in Section 5, MBE works quite well even in our context.

At this stage it is also worth mentioning that MBE is by no means the only procedure which can be used to separate relevant from irrelevant regressors. In fact, there has been a flurry of research resulting in estimators such as Lasso of Tibshirani (1996), SCAD of Fan and Li (2001), adaptive Lasso of Zou (2006) or the Dantzig selector of Candès and Tao (2007). In particular, properties of Lasso type estimators have been investigated in a time series context by, e.g., Kock and Callot (2012) and Medeiros and Mendes (2012). We choose MBE because the marginal regressions are computationally extremely efficient. On the other hand, very efficient algorithms

for the implementation of Lasso type estimators do exist and it would thus be interesting to investigate their properties in future work.

3.3. **QuickNet.** QuickNet (QN) resembles an earlier modelling device called RETINA, see Perez-Amaral et al. (2003). The idea of RETINA is to find the explanatory variables that in absolute terms are most strongly correlated with y_t. The most correlated variable is selected first, and the following ones one by one thereafter. QN differs from RETINA in that the set of candidate variables is different, as is the model selection criterion used for final selection. QN works as follows. First, the set of candidate variables S, see (2), is constructed. The variables have to be such that they show sufficient variation in the sample and are not perfectly linearly correlated; see White (2006) for details. This set of candidate variables is also used when Autometrics and MBE are applied. Once this has been done, a predetermined number of variables, $\overline{q}$, are added to the model from the set S, according to the rule that selects the variable with the strongest (positive or negative) correlation with the residuals of the previously estimated model. Then a model selection criterion is applied to choose a subset of the $\overline{q}$ variables. For this, we use 10-fold cross validation as suggested by Hastie et al. (2009).

We also experiment with a simplified unidirectional version of this method. The variables are selected one at a time as before, but the significance of the the added variable is tested at each step. Parsimony is appreciated, so the significance level of the tests is decreased as the number of variables the model increases. Adding variables is terminated at the first non-rejection of the null hypothesis, so this is a pure specific-to-general strategy. We apply this method such that the significance level of the first test in the sequence equals 0.2. Beginning with this value, the significance level is then halved at each step. In reporting results in Section 5, this method is called QN-SG.

4. Forecasting

4.1. **Two ways of generating multiperiod forecasts.** There are two main ways of creating multiperiod forecasts. One can either generate the forecasts recursively, or one may apply direct forecasting. In the former case, one and the same model is used for all forecast horizons. Direct

forecasting implies that a separate model is built for each forecast horizon. A brief discussion of these two techniques follows next.

4.1.1. *Recursive forecasts.* In order to illuminate recursive forecasting, consider the model (1) with $p = q = 1$. These restrictions are for notational simplicity only. Assuming the information set $\mathcal{F}_{T-1} = \{y_{T-j}, j \geq 1\}$ is independent of future error terms, the one-period-ahead forecast made at time T equals

$$y_{T+1|T} = E\left(y_{T+1}|\mathcal{F}_T\right) = \beta_{00} + \beta_{01}y_T + \beta_1(1 + \exp\{\gamma_{10} + \gamma_{11}y_T\})^{-1}.$$

The corresponding conditional mean $y_{T+2|T}$, that is, the two-period forecast, becomes

$$\begin{aligned} y_{T+2|T} &= E\left(\beta_{00} + \beta_{01}y_{T+1} + \beta_1(1 + \exp(\gamma_{10} + \gamma_{11}y_{T+1}))^{-1} + \varepsilon_{T+2}|\mathcal{F}_T\right) \\ &= \beta_{00} + \beta_{01}y_{T+1|T} + \beta_1 E\left(1 + \exp(\gamma_{10} + \gamma_{11}(y_{T+1|T} + \varepsilon_{T+1}))^{-1}|\mathcal{F}_T\right) \\ &= \beta_{00} + \beta_{01}y_{T+1|T} + \beta_1 \int_{-\infty}^{\infty} (1 + \exp(\gamma_{10} + \gamma_{11}(y_{T+1|T} + z)))^{-1}\phi(z)dz \end{aligned} \tag{6}$$

where $\phi(z)$ is the density of the $\mathcal{N}(0, \sigma^2)$ random variable. The integral in (6) can be computed by numerical integration. Note that it becomes a multiple integral when the forecast horizon $h > 2$. It is therefore better to calculate its value by simulation or by bootstrapping the residuals of the model, because this is computationally feasible even when $h > 2$. Some authors bypass this complication altogether by setting $\varepsilon_{T+1} = 0$ in the logistic function, and as a result their forecasts are biased estimates of the conditional mean.

In this work we apply the bootstrap. It has the advantage over simulation that unconditional heteroskedasticity of unknown form is allowed in the error process. More discussion about recursive forecasting can be found in Teräsvirta (2006), Kock and Teräsvirta (2011a) or Teräsvirta et al. (2010, Chapter 14) among others.

4.1.2. *Direct forecasts.* In direct forecasting, given the information set $\mathcal{F}_T$, the forecast for $T + h$ made at T equals

$$y^{\mathrm{D}}_{T+h|T} = g_h(y_T, y_{T-1}, ..., y_{T-p+1})$$

where g_h is a function of y_T and its lags. In our case, the models are selected using the three aforementioned techniques, but there is a gap in the model in that the variables from y_{T+1} to y_{T+h-1} do not enter the equation. The advantage of the direct method lies in its computational simplicity: no recursions are needed. But then, a separate model has to be specified for each forecast horizon.

Direct forecasts are obtained from the linear autoregressive model as well as the neural network model. In addition, to compare the forecasts from the neural network models to genuinely non-parametric ones, direct Nadaraya-Watson kernel regression forecasts (NP) are generated. Finally, 'no change' forecasts (NC), in which the variable of interest takes the same value at any future point in time as it does at the time of forecasting, are computed and compared with the rest.

4.1.3. *Forecasts based on differences and forecast errors.* The forecasts based on differences are obtained in the following way. When forecasting recursively first differences $\Delta y_t = y_t - y_{t-1}$ are being modelled and forecast. The p lags of the left hand side variable are thus $\Delta y_{t-1}, ..., \Delta y_{t-p}$ To obtain an h-periods-ahead forecast, the first-difference forecasts have to be cumulated[4]:

$$E\left(y_{T+h}|\mathcal{F}_T\right) = \sum_{j=1}^{h} E\left(\Delta y_{T+j}|\mathcal{F}_T\right) + y_T. \tag{7}$$

The corresponding forecast error is $e_{T+h|T} = y_{T+h} - E\left(y_{T+h}|\mathcal{F}_T\right)$.

In direct h-periods-ahead forecasting, the variable to be modelled is $\Delta_h y_t = y_t - y_{t-h}$. The p lags of the left-hand side variable are thus $\Delta_h y_{t-h}, ..., \Delta_h y_{t-h-p+1}$ and the corresponding forecast of y_{T+h} is $E\left(\Delta_h y_{T+h}|\mathcal{F}_T\right) + y_T$. The estimated model yields direct estimates of the conditional mean.

The measure of performance in this work is the root mean square forecast error (RMSFE). It is calculated for each time series from out-of-sample forecasts for the forecasting period beginning at T_0 and ending at $T - h_{\max}$, where T is the last available observation and h_{max} is the maximum

[4]The unknown $E\left(\Delta y_{T+j}|\mathcal{F}_T\right)$ are of course replaced by their bootstrapped counterparts.

forecast horizon. Thus,

$$\text{RMSFE}_h = \{(T - h_{max} - T_0 + 1)^{-1} \sum_{t=T_0}^{T-h_{\max}} e_{t+h|t}^2\}^{1/2}.$$

4.2. **Insanity Filters.** Nonlinear models may sometimes generate forecasts that are deemed unrealistic in the light of the hitherto observed values of the time series. This has prompted forecasters to introduce precautions in order to avoid excessive forecast errors. The idea is to replace an unrealistic forecast with a more conventional and believable one. It has been applied, among others, by Swanson and White (1995, 1997a,b) who called the procedure the insanity filter, Stock and Watson (1999) and Teräsvirta et al. (2005). We shall make use of two insanity filters. The first one works as follows: If the h-step ahead predicted change exceeds the maximum h-step change observed during the estimation period, the most recently observed value of the variable to be predicted is the forecast. Hence, in the words of Swanson and White (1995), we "replace craziness by ignorance". We shall call this filter the Swanson and White (SW) filter. In the second filter, the extreme predicted change is replaced by a forecast from our benchmark linear autoregressive model: craziness is replaced by linearity.

5. Results

The aforementioned techniques are applied to the monthly Consumer Price Index (CPI) series for the G7 countries as well as the four Scandinavian countries. Before considering these macroeconomic series we conduct a small Monte Carlo experiment. As mentioned in the introduction, the purpose of this exercise is to see how the three modelling procedures perform under controlled circumstances when the data generating process is known in two out of three designs even possible to select using these techniques.

5.1. **General methodology and data.** The technique for generating the potential hidden units for the ANN model (1) is described in the appendix of Kock and Teräsvirta (2011b). We have modified the original White (2006) technique somewhat to make it more suitable to our modelling problem. For QuickNet and MBE we use 10-fold cross validation as in Hastie et al. (2009) to determine the number of hidden units to be included. We also used the hv-Cross Validation

procedure of Racine (2000), but this did not improve the results and was omitted. Following the suggestion of White (2006), the maximum number of variables in the ANN models is equal to ten.

The macroeconomic series are obtained from the OECD Main Economic Indicators. Most series begin in the 1960s and end in December 2009 or January 2010. The CPI series were transformed to logarithms before modelling them, and the forecast errors discussed in the paper are errors in forecasting the transformed series. We also considered excluding the crisis period from July 2007 onwards. This did not change the relative accuracy of the procedures considered. For an in-depth analysis of the forecast precision during the crisis of the procedures studied in this paper the reader is referred to Kock and Teräsvirta (2014). It should also be mentioned that detailed results on forecasting the Finnish CPI can be found in Kock and Teräsvirta (2013).

5.2. **Monte Carlo.** Our simulations try to answer three questions. First, how well do the automated procedures perform in terms of variable selection and forecasting when the data is generated by a highly non-linear mechanism? Second, the same issues are investigated when the underlying model is linear. Do the automated procedures realize this or do they include redundant nonlinear hidden units? Finally, we investigate the properties of the automated procedures when the data is generated by a specific Logistic Smooth Transition Autoregressive model (LSTAR2) which is neither linear nor nested in the class of ANN models.

5.3. **A neural network DGP.** We consider a strongly nonlinear model from Medeiros et al. (2006). These authors took the well-known annual Wolf's sunspot number series because it is a distinctly nonlinear one and, after transforming the observations using the Box-Cox transformation as in Ghaddar and Tong (1981), fitted an ANN model (1) with two hidden units to the transformed series. The model has the form:

$$\begin{aligned} y_t = &- 0.17 + 0.85y_{t-1} + 0.14y_{t-2} - 0.31y_{t-3} + 0.08y_{t-7} \\ &+ 12.8G_1(\mathbf{y_{t-1}}) + 2.44G_2(\mathbf{y_{t-1}}) + \epsilon_t \end{aligned} \tag{8}$$

where the two hidden units are

$$G_1(\mathbf{y_{t-1}}) = \left(1 + \exp\left(-0.46\left(0.29y_{t-1} - 0.87y_{t-2} + .40y_{t-7} - 6.68\right)\right)\right)^{-1}$$

and

$$G_2(\mathbf{y_{t-1}}) = \left(1 + \exp\left(-1.17 \times 10^3\left(0.83y_{t-1} - 0.53y_{t-2} - 0.18y_{t-7} + 0.38\right)\right)\right)^{-1}$$

and, furthermore $\epsilon_t \sim$i.i.d. $\mathcal{N}(0,1)$. We generate 500 time series of 600 observations from this model. The set of potential variables consists of G_1, G_2, 1000 other hidden units, and ten lags of y_t, which greatly exceeds the number of observations. The forecast horizons are one, two, and five periods, and the maximum number of variables per each selected model equals ten. We report RMSFE ratios such that the denominator is the RMSFE of forecasts from (8), computed from the 500 replications.

Recursive	Hor.	DGP	AR	QN	MBE	Autom.	QN-SG
	1	1.82	1.456	1.343	1.730	1.105	1.805
NF	2	2.739	1.536	3.282	1.659	1.073	1.568
	5	4.172	1.337	$9 \cdot 10^4$	1.394	4023	1.283
	1	1	1.456	1.513	1.730	1.105	1.855
SW	2	1.001	1.536	1.532	1.658	1.074	1.552
	5	1.001	1.392	1.218	1.395	1.028	1.269
	1	1	1.456	1.322	1.730	1.105	1.776
AR	2	1.001	1.536	1.366	1.658	1.074	1.552
	5	1.001	1.337	1.214	1.395	1.028	1.269

TABLE 1. Average root mean square forecast error ratios for the recursive forecasts of the simulated sunspot series. DGP: Data generating process, AR: Autoregression, QN: QuickNet, MBE: Marginal Bridge Estimator, Autom.: Autometrics, QN-SG: QuickNet specific to general. NF: No Filter (for the DGP the NF subcolumn contains the actual root mean square forecast error from forecasting with the DGP), SW: Swanson-White filter, AR: Insane forecasts replaced by linear autoregressive ones.

Table 1 contains these ratios for the recursive forecasts. The first three entries in the column named DGP contain the RMSFE for the forecasts from the true model (8). As expected, all RMSFE ratios exceed unity. Autometrics-selected models generate by far the most accurate forecasts of the alternatives to the DGP, indicating that the method works well when there is a true model that can be selected from the set of variables available for the purpose. The other methods lead to models whose forecasts are of more or less the same quality. The forecasts from

Direct	Hor.	AR	QN	MBE	Autom.	QN-SG	NP	NC
	1	1.456	1.343	1.730	1.105	1.805	1.546	3.560
NF	2	1.518	9.575	1.549	1.652	1.436	1.332	4.226
	5	1.306	1.353	1.241	1.359	1.293	1.124	3.984
	1	1.456	1.513	1.730	1.105	1.855	1.658	
SW	2	1.518	1.52	1.549	1.532	1.733	1.424	
	5	1.363	1.326	1.241	1.322	1.293	1.124	
	1	1.456	1.322	1.730	1.105	1.776	1.555	
AR	2	1.518	1.35	1.549	1.355	1.444	1.335	
	5	1.306	1.219	1.241	1.246	1.293	1.124	

TABLE 2. Average root mean square forecast error ratios for the direct forecasts of the simulated sunspot series. NP: Non-parametric, NC: 'no change' forecasts. NF: No Filter, SW: Swanson-White filter, AR: Insane forecasts replaced by linear autoregressive ones.

MBE-selected models do not need filtering but are nevertheless slightly more inaccurate than the other (filtered) ones.

The performance of direct models is reported in Table 2. Models selected by Autometrics no longer generate more accurate forecasts than the other nonlinear models. Every possible direct model is misspecified by definition because the shortest lag (two-year model) or lags (five-year model) of y_t cannot be used, and Autometrics clearly suffers from this. Note the good performance of the nonparametric model in forecasting five years ahead. The kernel autoregression seems to make most of the available information, and the forecasts hardly need filtering. In fact, the SW filter has a negative effect on the accuracy of the forecasts from this model. As may be expected, the 'no change' forecast does not perform well in predicting these strongly cyclical realisations.

The results from different models and methods are also compared by two other devices. We calculate the average ranks of the absolute forecast errors over the forecasts for the three forecast horizon. We also use Wilcoxon's signed-rank test for comparing the forecasts from the DGP with the others. This yields a set of pairwise comparisons. For space reasons, the results from these two approaches are not reported here but are available at `https://sites.google.com/site/andersbkock/`.

Table 3 offers some background to the results in Tables 1 and 2. It contains information about the size and variable types in the nonlinear models for recursive forecasting. The average number of variables in every type of model is greater than the size of (8) which is six variables when the intercept is not counted. It is worth noting that Autometrics, while selecting the largest models,

Recursive	Total	Linear	Nonlinear	DGP units
QN	9.55	0.348	9.21	1.64
MBE	9.22	0.756	8.47	0.77
Autom	11	1.5	9.51	3.47
QN-GS	5.3	0.324	4.98	1.12

TABLE 3. Average number of variables selected for the recursive forecasts of the sunspot series. "Total" indicates total number of variables included, "Linear" indicates the number of linear units included, "Nonlinear" gives the number of hidden units included, and DGP units gives the number of units included from the data generating process.

picks up elements of the true model more frequently than the other model selection techniques. This is probably the most important factor in explaining its success. Moreover, Autometrics on average chooses more linear lags than the other models, although fewer than their number in the true model. The average number of linear lags in the other models is rather small. The specific-to-general QN-SG is clearly more parsimonious than QuickNet, but this result is not invariant to the choice of significance levels in the test sequence. QuickNet-based recursive forecasts are somewhat more accurate than QN-SG ones at one-and five-year horizons.

5.4. **A linear DGP.** The linear model is a simple first-order AR model

$$y_t = 0.9y_{t-1} + \epsilon_t$$

where the ϵ_t are i.i.d. and follow the standard Gaussian distribution. As in the nonlinear example we generate 500 data sets with 600 observations and the pool of hidden units consists of 1000 elements. The number of linear lags in the pool of variables is 10. The forecast horizons are the same as before as is the maximum number of variables (linear as well as non-linear) included in the models (ten). The forecasts from the linear AR model form the benchmark.

Forecasting performance of the different techniques is reported in Tables 4 and 5 of which the former contains the results for recursive forecasting. It is seen that measured by forecasting accuracy, Autometrics is now the worst performing automated procedure. At the top end, MBE-based forecasts are as precise as the ones from the linear AR model, and they are not filtered. The results on direct forecasting in Table 5 are similar to the recursive ones in the sense that MBE is still the most reliable method, whereas Autometrics generates the most inaccurate forecasts.

Recursive	Hor.	AR	QN	MBE	Autom.	QN-SG
	1	0.9687	2.255	0.9983	1.114	1.024
NF	2	1.328	1.627	0.9989	1.133	1.035
	5	1.885	1.37	0.9975	1.084	1.036
	1	1	1.021	0.9983	1.114	1.024
SW	2	1	1.046	0.9989	1.133	1.035
	5	1	1.06	0.9976	1.074	1.035
	1	1	1.022	0.9983	1.114	1.024
AR	2	1	1.046	0.9989	1.133	1.035
	5	1	1.059	0.9976	1.074	1.035

TABLE 4. Average root mean square forecast error ratios for the recursive forecasts of the simulated AR(1) series. DGP: Data generating process, AR: Autoregression, QN: QuickNet, MBE: Marginal Bridge Estimator, Autom.: Autometrics, QN-SG: QuickNet specific to general. NF: No Filter (for the AR(1) the NF subcolumn contains the actual root mean square forecast error from forecasting with the AR(1)), SW: Swanson-White filter, AR: Insane forecasts replaced by linear autoregressive ones. The entries of the column AR, rows NF section are the average root mean square forecast errors of the AR(1) model without any insanity filter.

Nonparametric forecasts are quite accurate, which is not surprising as the data are generated from a linear AR(1) model. Since the linear model is quite persistent, it is no wonder that the one- and two steps-ahead forecasts are rather accurate.

Table 6 sheds light to the switch in the behaviour of Autometrics when one moves from the nonlinear to the linear AR model. Autometrics only manages to find the correct variable (the first lag of y_t) in 13% of the replications, whereas the other procedures find it every time. Besides, the average number of variables in Autometrics-based models is very large. In fact, Autometrics selects ten times as many variables as MBE that generates the smallest models.

It is also seen that QuickNet generates less accurate forecasts than the pure specific-to-general version QN-SG. It seems that in this simulation one has better control on what is selected when the selection is carried out by sequential testing than when it is done by a pure QuickNet procedure. The number of variables selected by QN-SG is not much smaller than what is achieved by QuickNet, but forecasts are clearly more accurate at shortest horizons. Besides, filtering them is not necessary. As already pointed out, the number of variables selected by QN-SG can be controlled by the significant levels of the tests. In this study, the procedure is rather 'generous', in that the significance level of the first test as high as 0.2. In this particular example, a lower significance levels would no doubt have given better forecasts.

Direct	Hor.	AR	QN	MBE	Autom.	QN-SG	NP	NC
	1	1	2.255	0.9983	1.114	1.024	1.009	1.013
NF	2	1.002	1.108	0.9976	4.912	1.076	1.008	1.041
	5	1.004	1.078	0.9957	1.105	1.074	1.008	1.11
	1	1	1.021	0.9983	1.114	1.024	1.009	
SW	2	1.002	1.07	0.9976	1.155	1.076	1.008	
	5	1.004	1.078	0.9957	1.09	1.074	1.008	
	1	1	1.022	0.9983	1.114	1.024	1.009	
AR	2	1.002	1.07	0.9976	1.155	1.076	1.008	
	5	1.004	1.078	0.9957	1.092	1.074	1.008	

TABLE 5. Average root mean square forecast error ratios for the direct forecasts of the simulated AR(1) series. DGP: Data generating process, AR: Autoregression, QN: QuickNet, MBE: Marginal Bridge Estimator, Autom.: Autometrics, QN-SG: QuickNet specific to general, NP: Non-parametric, NC: 'no change'. NF: No Filter (for the AR(1) the NF subcolumn contains the actual root mean square forecast error from forecasting with the AR(1)), SW: Swanson-White filter, AR: Insane forecasts replaced by linear autoregressive ones.

Recursive	Total	Linear	Nonlinear	DGP units
QN	8.07	1	7.07	1
MBE	1.84	1.13	0.708	1
Autom	19.8	0.236	19.5	0.132
QN-GS	7.65	1	6.65	1

TABLE 6. Average number of variables selected for the recursive forecasts from the simulated AR(1) model. "Total" indicates total number of variables included, "Linear" indicates the number of linear units included, "Nonlinear" gives the number of hidden units included, and DGP units gives the number of units included from the data generating process.

The most important reason for the inferior performance of Autometrics and QuickNet in this example appears to be that they often select highly correlated variables. This leads to imprecise coefficient estimates and, consequently, inferior forecasts. It would seem a good idea to furnish these procedures with checks that would prevent this from happening. As things are now, MBE, being inherently parsimonious, dominates those two in this simulation.

It is obvious that in this artificial experiment it would have been very useful to test linearity before the actual variable selection. With 600 observations, linearity would have been rejected with a very high probability, and the model building problem would have been reduced to selecting the relevant linear lags for the AR model. As mentioned in the introduction, in the forecasting exercise with real economic variables performing a linearity test before selecting the variables did not, however, seem to increase the accuracy of the forecasts.

Recursive	Hor.	DGP	AR	QN	MBE	Autom.	QN-SG
NF	1	0.975	1.041	$2 \cdot 10^3$	1.022	1.112	1.063
	2	1.221	1.044	$7 \cdot 10^4$	1.020	1.179	1.055
	5	1.439	1.006	$2 \cdot 10^5$	0.9897	$1 \cdot 10^7$	1.057
SW	1	1	1.041	1.055	1.022	1.034	1.063
	2	1	1.044	1.045	1.020	1.048	1.054
	5	0.9991	1.006	1.037	0.9897	1.044	1.044
AR	1	1	1.041	1.051	1.022	1.032	1.063
	2	1	1.044	1.041	1.020	1.047	1.054
	5	0.9991	1.006	1.033	0.9897	1.044	1.044

TABLE 7. Average root mean square forecast error ratios for the recursive forecasts of the simulated LSTAR2 series. DGP: Data generating process, AR: Autoregression, QN: QuickNet, MBE: Marginal Bridge Estimator, Autom.: Autometrics, QN-SG: QuickNet specific to general. NF: No Filter (for the AR(1) the NF subcolumn contains the actual root mean square forecast error from forecasting with the AR(1)), SW: Swanson-White filter, AR: Insane forecasts replaced by linear autoregressive ones. The entries of the column AR, rows NF section are the average root mean square forecast errors of the AR(1) model without any insanity filter.

5.5. **An intermediate case.** As a third example we generate data from a special case of an LSTAR2 model which is not nested in the neural network structure. The data were generated from

$$y_t = 0.9y_{t-1} + 2\left(1 + \exp\left\{-[y_{t-1} - 0.5][y_{t-1} + 0.5]\right\}\right)^{-1} + \epsilon_t \tag{9}$$

where $\epsilon_t \sim N(0, 1)$. In the model the intercept fluctuates symmetrically around zero as a function y_{t-1} and varies between zero and two. The logistic function takes its minimum at zero and tends to one as $y_{t-1} \to \pm\infty$.

We again generate 500 data sets with 600 observations. The number of linear lags in the pool of variables is 10 while 1000 hidden units are included. The benchmark forecasts are the ones from (9) assuming that the parameters are known.

Table 7 reports the forecasting performance of the recursive forecasts. As expected, no procedure systematically outperforms the forecasts from the data generating process. Most of the root mean square forecast error ratios are above one. An exception is provided by MBE at the longest forecasts horizon. This procedure is also very stable as none of its forecasts is deemed insane. QuickNet and Autometrics are less stable than MBE. The forecasts from the plain linear autoregressions do also seem to work quite well when the forecasting horizon is long. It seems that

Direct	Hor.	AR	QN	MBE	Autom.	QN-SG	NP	NC
	1	1.041	$2 \cdot 10^3$	1.022	1.112	1.063	1.012	1.114
NF	2	1.048	1.062	1.042	1.459	1.054	1.013	1.196
	5	1.006	$4 \cdot 10^3$	1.018	352.2	1.024	1.003	1.245
	1	1.041	1.055	1.022	1.034	1.063	1.012	
SW	2	1.048	1.062	1.042	1.059	1.054	1.013	
	5	1.006	1.030	1.012	1.050	1.024	1.003	
	1	1.041	1.051	1.022	1.032	1.063	1.012	
AR	2	1.048	1.062	1.042	1.057	1.054	1.013	
	5	1.006	1.044	1.026	1.062	1.024	1.003	

TABLE 8. Average root mean square forecast error ratios for the direct forecasts of the simulated LSTAR2 series. DGP: Data generating process, AR: Autoregression, QN: QuickNet, MBE: Marginal Bridge Estimator, Autom.: Autometrics, QN-SG: QuickNet specific to general, NP: Non-parametric, NC: 'no change'. NF: No Filter (for the AR(1) the NF subcolumn contains the actual root mean square forecast error from forecasting with the AR(1)), SW: Swanson-White filter, AR: Insane forecasts replaced by linear autoregressive ones.

for other procedures than Autometrics, the LSTAR2 model is easier to forecast than the neural network model in Section 5.3.

The results for the direct forecasts are contained in Table 8. These forecasts are less precise than their recursive counterparts and not even MBE has any root mean square forecast error ratio below one. Autometrics and QuickNet forecasts need filtering even in this case. The nonparametric forecasts perform similarly to the neural network based ones while the "no change" forecasts are clearly inferior to all the others once the latter are filtered.

Table 9 reports the structure of the models chosen for the recursive forecasts. All procedures include one linear lag. This is in accordance with the structure of the data generating process. On the other hand, models selected by QuickNet and Autometrics contain more hidden units than QN-SG and, in particular, MBE. Note, however, that as opposed to the previous experiment Autometrics does not choose models which are much larger than the ones selected by the other procedures.

Computational speed. It is appropriate to briefly comment on the computational speed of the procedures considered. The most important observation is that Autometrics is in general at least 100 times slower than the other automated procedures. To be more precise, Autometrics typically takes a few minutes to perform one estimation while QuickNet, MBE and QN-SG generally

Recursive	Total	Linear	Nonlinear
QN	6.88	1	5.88
MBE	2.05	1.04	1.01
Autom	6.75	0.966	5.78
QN-GS	4.2	1	3.2

TABLE 9. Average number of variables selected for the recursive forecasts from the simulated LSTAR2 model. "Total" indicates total number of variables included, "Linear" indicates the number of linear units included.

take less than a second. The nonparametric estimation lies between these two extremes. These computations were carried out on a desktop DualCore with 1.83GHz and 1.49GB RAM.

5.6. **Forecasting the Consumer Price Index.** We now turn to forecasting the Consumer Price Index of our 11 countries. We have also done the same for unemployment rates, but to conserve space the complete results from that exercise are not reported here. See instead Kock and Teräsvirta (2011b). The CPI series are forecast 1, 3, 6, and 12 months ahead. The series are transformed into logarithms, and 240 forecasts based on an expanding window are generated for each horizon[5]. Forecasts from models of differenced series are formed as described in Section 4.1.3. The pool of variables contains 600 hidden units with $p = 6$ in (1) and the first six linear lags of the dependent variable. The models are respecified every six months. This is because Autometrics is quite slow: otherwise respecification could easily be done every month.

The RMSFE ratios for recursive CPI forecasts from models of differenced series can be found in Table 10. The denominator in the RMSFE ratio is now the RMSFE of the recursive linear AR forecasts. It is seen from the table that filtering the forecasts is necessary. All four model selection techniques lead to ANN models that generate some very inaccurate forecasts. This is the case already for one-month forecasts and is, as in the AR(1) simulation, due to the fact that some models contain very strongly correlated variables. A pair of them typically has large (in absolute value) coefficients with opposite signs. Forecasting with such a model yields inaccurate forecasts and cumulating them in forecasting more than one month ahead makes the situation even worse. Furthermore, all RMSFE ratios exceed one, which means that on average no ANN model, not

[5]For some of the shorter data sets the number of forecasts is less than 240, because the first window was set to include at least 200 observations.

Recursive	Hor.	AR	QN	MBE	Autom.	QN-SG
NF	1	0.0037	16.82	1.02	257.9	1.043
	3	0.0080	$5 \cdot 10^4$	$2 \cdot 10^6$	$2 \cdot 10^9$	1.052
	6	0.0132	$4 \cdot 10^5$	$1 \cdot 10^6$	$6 \cdot 10^9$	2.411
	12	0.0245	$1 \cdot 10^6$	$1 \cdot 10^6$	$1 \cdot 10^{10}$	$3 \cdot 10^5$
SW	1	1	1.040	1.020	1.074	1.047
	3	1.004	1.033	1.020	1.075	1.061
	6	1.003	1.055	1.020	1.085	1.076
	12	1.011	1.107	1.034	1.172	1.091
AR	1	1	1.042	1.019	1.072	1.044
	3	1	1.025	1.014	1.058	1.052
	6	1	1.036	1.017	1.047	1.071
	12	1	1.066	1.032	1.105	1.088

TABLE 10. Average root mean square forecast error ratios for the recursive forecasts of the CPI series based on differences. NF: No Filter, SW: Swanson-White filter, AR: Insane forecasts replaced by linear autoregressive ones. The first four figures in the column AR are RMSFEs from the autoregressive model.

Direct	Hor.	AR	QN	MBE	Autom.	QN-SG	NP	NC
NF	1	1	16.82	1.02	257.9	1.043	1.148	1.133
	3	0.976	2.699	0.9893	2464	1.02	1.074	1.169
	6	0.8123	20.77	0.8239	1869	0.8362	0.9335	1.159
	12	0.7336	3.286	0.7284	20.08	0.7436	0.8203	1.134
SW	1	1	1.040	1.020	1.074	1.047	1.150	
	3	0.976	1.039	0.9893	1.059	1.030	1.081	
	6	0.8123	0.8452	0.8239	0.8987	0.836	0.9335	
	12	0.7336	0.7584	0.7284	0.8355	0.7397	0.8203	
AR	1	1	1.042	1.019	1.072	1.044	1.147	
	3	0.976	1.020	0.9893	1.042	1.019	1.075	
	6	0.8123	0.840	0.8239	0.8819	0.835	0.9335	
	12	0.7336	0.7591	0.7284	0.8371	0.7395	0.8203	

TABLE 11. Average root mean square forecast error ratios for the direct forecasts of the CPI series based on differences. NF: No Filter, SW: Swanson-White filter, AR: Insane forecasts replaced by linear autoregressive ones.

even after filtering, generates more accurate recursive forecasts than the linear AR model. Models selected by MBE perform slightly better than the other nonlinear models.

These results may be compared with the ones in Table 11. This table contains the RMSFE ratios for direct forecasts from models based on differenced series. Models built using QuickNet and Autometrics still generate a few forecasts that require filtering, whereas MBE-based forecasts no longer do. After filtering, the six- and 12-month forecasts from the ANN models are more accurate than the benchmark ones. This is also the case for forecasts from direct linear AR models. Their RMSFE ratios are comparable to those obtained from models constructed by

Recursive	Hor.	AR	QN	MBE	Autom.	QN-SG
NF	1	1.011	1.013	0.977	1.062	1.139
	3	1.001	20.39	0.9315	6311	1.195
	6	0.9728	$3 \cdot 10^5$	0.8535	$1 \cdot 10^7$	1.223
	12	0.9372	$3 \cdot 10^6$	0.787	$3 \cdot 10^8$	1.309
SW	1	1.011	1.013	0.977	1.062	1.139
	3	1.001	0.9685	0.9314	1.003	1.184
	6	0.9728	0.896	0.8532	0.9299	1.187
	12	0.9372	0.823	0.7871	0.8489	1.185
AR	1	1.011	1.013	0.977	1.062	1.139
	3	1.001	0.9661	0.9314	1.003	1.181
	6	0.9728	0.8923	0.8532	0.9299	1.187
	12	0.9372	0.8143	0.7871	0.8489	1.167

TABLE 12. Average root mean square forecast error ratios for the recursive forecasts of the CPI series based on levels. NF: No Filter, SW: Swanson-White filter, AR: Insane forecasts replaced by linear autoregressive ones.

MBE which is the best-performing model selection technique. The forecasting performance of the nonparametric model is below average, and the 'no change' forecasts are less accurate than even the corresponding recursive ones.

The RMSFE ratios in Table 12 refer to recursive forecasts from models built on CPI levels. Filtered forecasts are more accurate on average than the corresponding forecasts in Table 10. MBE-based forecasts are the most accurate ones and models built by QN-SG generate the least accurate recursive forecasts: all ratios remain above one. Recursive linear AR models built on levels are somewhat superior to the ones built on differences. The RMSFE ratios lie below one for the two longest horizons but are greater than the corresponding ratios for forecasts from models obtained by MBE, QuickNet and Autometrics.

Table 13 contains the RMSFE ratios for direct forecasts from models specified and estimated from the level series. It appears that MBE is the best model-building method when RMSFE is used as the criterion. The ratios are even smaller than the ones found in Tables 10–12. Direct models selected by QuickNet also perform better than the recursive ones, whereas the same cannot be said of models based on Autometrics or QN-SG. In the light of these results, going from specific to general and back again (QuickNet) is a better idea than going from specific to general only (QN-SG), but this finding cannot be generalized. The simulations with an AR(1) model yielded the opposite result. It may be noted that the nonparametric model built on levels generates much

Direct	Hor.	AR	QN	MBE	Autom.	QN-SG	NP	NC
	1	1.011	1.013	0.977	1.062	1.139	16.77	1.133
NF	3	0.9661	0.9418	0.9057	0.9761	1.198	8.037	1.169
	6	0.9053	3.401	0.8114	0.9982	1.204	5.072	1.159
	12	0.7771	0.7205	0.6928	0.9416	1.173	3.119	1.134
	1	1.011	1.013	0.977	1.062	1.139	3.783	
SW	3	0.9661	0.9418	0.9057	0.9761	1.198	5.172	
	6	0.9053	0.8305	0.8114	0.954	1.204	4.907	
	12	0.7771	0.7205	0.6928	0.9416	1.173	3.119	
	1	1.011	1.013	0.977	1.062	1.139	3.675	
AR	3	0.9661	0.9418	0.9057	0.9761	1.198	5.136	
	6	0.9053	0.8303	0.8114	0.9564	1.204	4.904	
	12	0.7771	0.7205	0.6928	0.9416	1.173	3.119	

TABLE 13. Average root mean square forecast error ratios for the direct forecasts of the CPI series based on levels. NF: No Filter, SW: Swanson-White filter, AR: Insane forecasts replaced by linear autoregressive ones.

less accurate forecasts than the same model estimated from differenced series. Its RMSFE ratios are remarkably larger than any other ratio. Summing up, it seems that direct forecasts are on average more accurate than the recursive ones. Exceptions do exist: compare Autometrics-based six- and 12-month RMSFE ratios in Tables 12 and 13. It should be pointed out that these results are aggregate ones and do not necessarily hold for all 11 countries.

We also compare the forecast performance of the different methods by using the ranks and Wilcoxon tests as documented in the online appendix at `https://sites.google.com/site/andersbkock/`. This is done for all countries and forecast horizons. Forecasts from models built on differences as well as the ones based on levels are included in the same comparison. The results and comments can be found in the online appendix at the above address.

As in the simulated example, it is interesting to see whether the size of the model and the accuracy of the forecasts from it are related. Table 14 contains information about the size and composition of models based on differenced series. When forecasting recursively, it is seen from the left panel that QN-SG selects the most parsimonious models which do not, however, yield the most accurate forecasts. MBE selects somewhat less parsimonious models that on average do yield the most accurate recursive forecasts. It also chooses the largest fraction of linear lags, although their average number remains below one. As in the AR(1) simulations, models selected

Recursive	Total	Linear	Nonlinear
QN	6.35	0.298	6.05
MBE	5.51	0.818	4.69
AM	15.5	0.393	15.1
QN-SG	4.03	0.195	3.83

Direct MBE	Total	Linear	Nonlinear
1 mth	5.51	0.818	4.69
3 mths	5.48	2.45	3.03
6 mths	5.29	3.55	1.74
12 mths	2.69	1.72	0.964

TABLE 14. Left panel: Average number of variables selected for the models generating recursive forecasts of the CPI based on differences. "Total" indicates total number of variables included, "Linear" indicates the number of linear units included, and "Nonlinear" gives the number of hidden units included. Right panel: Average number of variables selected for the direct forecasts of the CPI based on differences by MBE.

Recursive	Total	Linear	Nonlinear
QN	5.35	1.09	4.27
MBE	7.19	5.64	1.55
AM	19.1	1.34	17.7
QN-SG	1.39	1	0.386

Direct MBE	Total	Linear	Nonlinear
1 mth	7.19	5.64	1.55
3 mths	7.24	5.74	1.49
6 mths	7.42	6	1.42
12 mths	7.21	6	1.21

TABLE 15. Left panel: Average number of variables selected for the models generating recursive forecasts of the CPI based on levels. "Total" indicates total number of variables included, "Linear" indicates the number of linear units included, and "Nonlinear" gives the number of hidden units included. Right panel: Average number of variables selected for the direct forecasts of the CPI based on levels by MBE.

by Autometrics are by far the largest ones. There does not seem to be a clear connection between the model size and forecast accuracy.

The right-hand panel of Table 14 contains the average size and composition of models based on differenced series and selected by MBE for direct forecasting. The average number of variables is halved when one moves from six- to 12-month models, whereas the share of linear lags of the total increases up to six-month models and remains about the same for 12-month ones.

Table 15 contains the same information for models built on levels. All methods now select more linear variables than in the previous case. QN-SG is still the most parsimonious technique, and even QuickNet selects fewer variables than MBE. As Tables 13 and A.7 (to be found in the online appendix) indicate, forecasts from MBE are still the most accurate ones on average. The use of Autometrics leads to largest models. They perform better than QN-SG-selected models but less well than ones specified using MBE. The right panel of Table 15 shows that MBE selects a large number of linear lags for all direct models. In fact, every MBE-model built for the two longest horizons contains all six lags and only a small number of hidden units. A comparison of the RMSFE ratios in Tables 10 and 12 on the one hand and Tables 11 and 13 on the other (indirectly)

suggests that direct models based on data in levels and selected by MBE may be slightly superior to the same type of model, selected by the same technique, but based on differenced series. Whether or not this is due to the larger amount of linear lags in the former models is not clear, however.

Individual countries. This section gives a brief review of some results for the individual countries since the aggregate results may sometimes hide differences between the individual countries. A detailed description of the results for the individual countries can be found in Kock and Teräsvirta (2011b). Examining the RMSFE ratios for the individual countries reveals that direct forecasts tend to be more accurate than their recursive counterparts. This is true for forecasts based on differences as well as levels. Furthermore, MBE is rather stable. MBE is the only procedure which has RMSFE ratios below one at all horizons for Italy, Japan, and the US. However, a different selection technique is dominant for each country. The nonparametric forecasts are very inaccurate for all three countries.

Summing up, results on forecasting the CPI series suggest that forecasts based on levels are superior to their counterparts based on differences. Furthermore, direct forecasting is preferable to recursive forecasts and MBE is the most stable forecasting procedure. This last observation may be attributed to the high number of linear units supplemented by a few relevant nonlinear units in the MBE-built models.

5.7. **Forecasting unemployment rates.** To save space, we shall not give the full details of the results for the unemployment series since no dominant method was found for those series. However, a couple of findings deserve to be mentioned. First, except for MBE the nonlinear procedures produce many insane forecasts such that filtering is needed. Second, the direct forecasts no longer outperform their recursive counterparts. Finally, MBE in general chooses much smaller models than Autometrics. For detailed results, see Kock and Teräsvirta (2011b).

6. Conclusions

In this paper we consider macroeconomic forecasting with a flexible nonlinear model, the single-hidden layer feedforward neural network model that is a universal approximator. We apply the idea of White (2006) of transforming the specification and estimation problem of this model to a linear

model specification problem. This leads to a situation in which the number of candidate variables to choose among vastly exceeds the number of observations. Three modelling techniques, White's QuickNet among them, that can handle this difficulty are compared and the models selected are used for forecasting.

The benchmark in our forecast comparisons is, with two exceptions, the linear AR model with recursive forecasts. It turns out to be difficult to improve upon its forecasting precision using recursive forecasting, while the direct method seems to be a more successful approach. It appears that the Marginal Bridge Estimator of Huang et al. (2008) yields the best performing ANN models overall, but the results do vary from one country to the other. Autometrics of Doornik (2009) selects models with excellent forecasting performance when there is a well-fitting nonlinear model to be discovered but does poorly when no potential model fits the data sufficiently well or when the DGP is a simple first-order AR model. QuickNet selects models whose average forecasting performance lies between that of the two others. Parsimony plays a role since MBE often selects models with the fewest variables of the available alternatives. The purely nonparametric model generates relatively accurate forecasts for inflation series in differences but is much less successful in forecasting CPI series in levels or unemployment rates (for the latter the results are not shown). The performance of the models may also vary as a function of the forecasting horizon.

All three techniques often produce models that yield some very erroneous or 'insane' forecasts, which makes filtering them necessary. The two insanity filters considered in this paper perform almost equally well, although the AR filter may have a slight edge over the filter that Swanson and White (1995) introduced. Multicollinearity is the main reason for insane forecasts, and it might be a good idea to develop all three modelling strategies further in order to reduce the probability of the outcomes in which the final model contains very strongly linearly correlated variables.

Another finding is that that testing linearity before variable selection (results not reported here) does not help in choosing useful models. It may do so for certain countries and variables but may lead to weakened forecasting performance in some others. For this reason it cannot be recommended as a part of any of the three modelling strategies under consideration.

Forecasts are generated using both the recursive and the direct method. Overall, direct forecasting is somewhat superior to the recursive technique, but it does not dominate the latter. The results vary from one country and variable to the other. This is also true in comparing the accuracy of recursive and direct forecasts just from linear AR models: on average direct forecasts are more accurate than the recursive ones.

When it comes to choosing between models based on first differences of the series and ones specified and estimated using levels it turns out that in forecasting the CPI, models built on levels tend to generate more accurate forecasts on average than the corresponding models constructed using first differences. It is not clear why that is the case. In forecasting unemployment rates (results not shown) the outcome is less clear: the models based on levels cannot be viewed as superior to models built on first differences.

A general conclusion is that the ANN model can be useful in macroeconomic forecasting but that the linear AR model is a serious competitor. In practice, the forecaster may experiment with several models and methods before settling for one, if the final goal is to find a model with the best performance for a given country and variable. Another possibility left for further work would be to combine recursive and direct forecasts obtained with various linear AR and ANN models.

Finally, the purpose of this work has not been to compare the forecasting performance of different nonlinear models. Doing so in a satisfactory fashion would require a vast amount of computational and human resources. It would also shift the focus away from our main aim: comparing different modelling techniques for the single-hidden layer ANN model made possible by the work of White (2006), and has therefore not been attempted here.

References

N. K. Ahmed, A. F. Atiya, N. El Gayer, and H. El-Shishiny. An empirical comparison of machine learning tools for time series forecasting. *Econometric Reviews*, 29:594–621, 2010.

U. Anders and O. Korn. Model selection in neural networks. *Neural Networks*, 12:309–323, 1999.

E. Candès and T. Tao. The dantzig selector: Statistical estimation when p is much larger than n. *Annals of Statistics*, 35:2313–2351, 2007.

E. J. Candès. *Ridgelets: theory and applications*. PhD thesis, Dept. of Statistics, Stanford University, 1998.

E. J. Candès. Ridgelets: estimating with ridge functions. *Annals of Statistics*, 31:1561–1599, 2003.

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

J. A. Doornik. Autometrics. In J. L. Castle and N. Shephard, editors, *The Methodology and Practice of Econometrics*, pages 88–122. Oxford University Press, Oxford, 2009.

J. Fan and R. Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96:1348–1360, 2001.

T. L. Fine. *Feedforward neural network methodology*. Springer Verlag, New York, 1999.

A. R. Gallant. The Fourier flexible form. *American Journal of Agricultural Economics*, 66:204–208, 1984.

D. K. Ghaddar and H. Tong. Data transformation and self-exciting threshold autoregression. *Applied Statistics*, 30:238–248, 1981.

W. L. Goffe, G. D. Ferrier, and J. Rogers. Global optimization of statistical functions with simulated annealing. *Journal of Econometrics*, 60:65–99, 1994.

T. Hastie, R. Tibshirani, and J. H. Friedman. *The elements of statistical learning: data mining, inference, and prediction*. Springer Verlag, New York, 2009.

D. F. Hendry and H. M. Krolzig. The properties of automatic Gets modelling. *Economic Journal*, 115:32–61, 2005.

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

J. Huang, J. L. Horowitz, and S. Ma. Asymptotic properties of bridge estimators in sparse high-dimensional regression models. *Annals of Statistics*, 36:587–613, 2008.

J. T. G. Hwang and A. A. Ding. Prediction intervals for artificial neural networks. *Journal of the American Statistical Association*, pages 748–757, 1997.

A. B. Kock and L. F. A. Callot. Oracle inequalities for high dimensional vector autoregressions. *Journal of Econometrics (forthcoming)*, 2012.

A. B. Kock and T. Teräsvirta. Forecasting with nonlinear time series models. In M. P. Clements and D. F. Hendry, editors, *Oxford Handbook of Economic Forecasting*, pages 61–87. Oxford University Press, Oxford, 2011a.

A. B. Kock and T. Teräsvirta. Forecasting macroeconomic variables using neural network models and three automated model selection techniques. *CREATES Research Paper 2011-27, Aarhus University*, 2011b.

A. B. Kock and T. Teräsvirta. Forecasting the Finnish consumer price inflation using artificial neural networks and three automated model selection techniques. *Finnish Economic Papers*, 26:13–24, 2013.

A. B. Kock and T. Teräsvirta. Forecasting performance of three automated modelling techniques during the economic crisis 2007-2009. *International Journal of Forecasting*, 30:616–631, 2014.

H. M. Krolzig and D. F. Hendry. Computer automation of general-to-specific model selection procedures. *Journal of Economic Dynamics and Control*, 25:831–866, 2001.

C.-M. Kuan and T. Liu. Forecasting exchange rates using feedforward and recurrent neural networks. *Journal of Applied Econometrics*, 10:347–364, 1995.

T.-H. Lee, H. White, and C. W. J. Granger. Testing for neglected nonlinearity in time series models: A comparison of neural network methods and alternative tests. *Journal of Econometrics*, 56:269–290, 1993.

E. Maasoumi, A. Khontanzed, and A. Abaye. Artificial neural networks for some macroeconomic series: A first report. *Econometric Reviews*, 13:105–122, 1994.

S. Makridakis and M. Hibon. The M3-Competition: results, conclusions and implications. *International Journal of Forecasting*, 16:451–476, 2000.

M. Marcellino. Instability and non-linearity in the EMU. Discussion Paper No. 3312, Centre for Economic Policy Research, 2002.

M. Marcellino, J. H. Stock, and M. W. Watson. A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics*, 135:499–526, 2006.

M. C. Medeiros and E. Mendes. Estimating high-dimensional time series models. *CREATES Research Paper*, 37, 2012.

M. C. Medeiros, T. Teräsvirta, and G. Rech. Building neural network models for time series: A statistical approach. *Journal of Forecasting*, 25:49–75, 2006.

T. Perez-Amaral, G. M. Gallo, and H. White. A flexible tool for model building: the relevant transformation of the inputs network approach (RETINA). *Oxford Bulletin of Economics and Statistics*, 65:821–838, 2003.

J. Racine. Consistent cross-validatory model-selection for dependent data: hv-block cross-validation. *Journal of Econometrics*, 99:39–61, 2000.

G. Rech. Forecasting with artificial neural network models. SSE/EFI Working Paper Series in Economics and Finance 491, Stockholm School of Economics, 2002.

H. Simon. *Neural networks: a comprehensive foundation*. Prentice Hall, Upper Saddle River, NJ, 1999.

J. H. Stock and M. W. Watson. A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. In R. F. Engle and H. White, editors, *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W.J. Granger*, pages 1–44. Oxford University Press, Oxford, 1999.

N. R. Swanson and H. White. A model-selection approach to assessing the information in the term structure using linear models and artificial neural networks. *Journal of Business & Economic Statistics*, 13:265–275, 1995.

N. R. Swanson and H. White. A model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks. *Review of Economics and Statistics*, 79:540–550, 1997a.

N. R. Swanson and H. White. Forecasting economic time series using flexible versus fixed specification and linear versus nonlinear econometric models. *International Journal of Forecasting*, 13:439–461, 1997b.

T. Teräsvirta. Forecasting economic variables with nonlinear models. In G. Elliott, C. W. J. Granger, and A. Timmermann, editors, *Handbook of Economic Forecasting*, volume 1, pages

413–457. Elsevier, North-Holland, 2006.

T. Teräsvirta, C. F. Lin, and C. W. J. Granger. Power of the neural network linearity test. *Journal of Time Series Analysis*, 14:209–220, 1993.

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

T. Teräsvirta, C. W. J. Granger, and D. Tjøstheim. *Modelling Nonlinear Economic Time Series*. Oxford University Press, Oxford, 2010.

R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58:267–288, 1996.

H. White. Approximate nonlinear forecasting methods. In G. Elliott, C. W. J. Granger, and A. Timmermann, editors, *Handbook of Economic Forecasting*, volume 1, pages 459–512. Elsevier, Amsterdam, 2006.

F. Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1:80–83, 1945.

G. Zhang, B. E. Patuwo, and M. Y. Hu. Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, 14:35–62, 1998.

H. Zou. The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476):1418–1429, 2006.