FORECASTING THE FINNISH CONSUMER PRICE INFLATION USING ARTIFICIAL NEURAL NETWORK MODELS AND THREE AUTOMATED MODEL SELECTION TECHNIQUES*

ANDERS BREDAHL KOCK

CREATES, Aarhus University,
Fuglesangs Allé 4, DK-8210 Aarhus V, Denmark;
e-mail: akock@creates.au.dk

and

TIMO TERÄSVIRTA

CREATES, Aarhus University,
Fuglesangs Allé 4, DK-8210 Aarhus V, Denmark;
e-mail: tterasvirta@creates.au.dk

This paper is concerned with forecasting the Finnish inflation rate. It is being forecast using linear autoregressive and nonlinear neural network models. Perhaps surprisingly, building the models on the nonstationary level series and forecasting with them produces forecasts with a smaller root mean square forecast error than doing the same with differenced series. The paper also contains pairwise comparisons between the benchmark forecasts from linear autoregressive models and ones from neural network models using Wilcoxon's signed-rank test. (JEL: C22, C45, C51, C52, C53)

* *This paper is based on a presentation at Kansantaloustieteen päivät (Finnish Economic Association Annual Meeting), Vaasa, February 2012. Research reported here has been supported by the Danish National Research Foundation. We wish to thank Heikki Kauppi for useful remarks but retain the responsibility for any errors and shortcomings in this work.*

1. Introduction

Artificial neural network (ANN) models are popular in forecasting, and claims have been made concerning their superiority in economic forecasting as well, for references see for example Zhang, Patuwo and Hu (1998). Since these models are nonlinear, specifying their structure and estimating the relevant parameters requires iterative algorithms. The interest in the modelling and forecasting problem considered in this work has arisen from a contribution of White (2006) who devised an ingenious strategy for modelling ANNs. The computational advantage of his strategy is that it converts the specification and ensuing nonlinear estimation problem of an ANN model into a linear model selection problem. This greatly alleviates the computational effort involved in estimating these models. They are usually richly parameterised, so estimation requires both computational skills and resources. Linearising the ANN modelling problem makes it possible to consider extensive studies with the aim of finding out how well neural network models forecast. Since forecasting is the main use of ANN models when applied to time series, it is of considerable interest to investigate how well White's strategy performs in macroeconomic forecasting. In fact, there exist two other strategies based on the same idea but with a different model selection technique. They are included in the present comparison.

In this work we focus on forecasting a single series: the monthly Finnish CPI. The results can be compared with the aggregates based on a study on forecasting inflation in 11 countries, reported in Kock and Teräsvirta (2011b). We complete the results in that paper by considering the accuracy of nonlinear forecasts using Wilcoxon's signed rank test (Wilcoxon, 1945) for pairwise comparisons. Forecasts from a linear autoregressive model then form the benchmark against which the ANN-forecasts are compared.

The plan of the paper is as follows. The ANN model built and applied in this study is presented in Section 2. Forecasting techniques and the time series are discussed in Section 3. Section 4 contains the empirical results. Conclusions can be found in Section 5.

2. The model

As already indicated, our main forecasting tool is the Artificial Neural Network model. In this work it is the so-called single-hidden-layer feedforward autoregressive neural network model or single-hidden-layer perceptron

(1)
$$y_t = \boldsymbol{\beta}_0'\mathbf{z}_t + \sum_{j=1}^{q} \beta_j (1 + \exp\{\boldsymbol{\gamma}_j'\mathbf{z}_t\})^{-1} + \varepsilon_t$$

where $\mathbf{z}_t = (1, y_{t-1}, \ldots, y_{t-p})'$, $\boldsymbol{\beta}_0 = (\beta_{00}, \beta_{01}, \ldots, \beta_{0p})'$, $\boldsymbol{\gamma}_j = (\gamma_{j0}, \gamma_{j1}, \ldots, \gamma_{jp})$ and $\varepsilon_t \sim \mathrm{iid}N(0, \sigma^2)$. The weak stationarity condition of (1) is the same as that of the corresponding linear AR(p) model. What makes (1) an attractive model, at least in theory, is that neural networks are so-called universal approximators. 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 $|H(\mathbf{z}) - \beta_0 - \sum_{j=1}^{q} \beta_j (1 + \exp\{\boldsymbol{\gamma}_j'\mathbf{z}\})^{-1}| < \delta$, where $|\cdot|$ is an appropriate norm. In economic applications it may be useful to add the linear term $\boldsymbol{\beta}_0'\mathbf{z}$ to the model, as is done in (1).

Before using the model (1) for forecasting, the number of the logistic functions or hidden units q has to be specified and the parameters of the model estimated. In this work we apply the insight of White (2006). He points out that if the parameter vectors $\boldsymbol{\gamma}_j$, $j=1,\ldots,q$, are known, the model is linear in parameters. This opens up the possibility to combine specification and estimation into a single linear model selection problem. One just has to choose a subset of variables from the set

$$S = \{y_{t-i}, i = 1, \ldots, p; (1 + \exp\{\boldsymbol{\gamma}_j'\mathbf{z}_t\})^{-1}, j = 1, \ldots, M\}$$

where M is large. Since the quality of the final estimates depends on the size of S (the larger M, the more accurate the estimates), the number of variables in a typical macroeconomic application is likely to exceed the number of observations. This requires model selection techniques with which one can handle such a situation. White (2006) proposed one such technique, called QuickNet. It may be characterised as a specific-to-general-to specific procedure. This means that one starts with a small model and keeps adding variables from S before reconsidering the selection and possibly removing variables from it again. Variable selection is carried out using a form of cross-validation. But then, we shall also report results on a simplified specific-to-general version of QuickNet, in which variables are added to the model one-by-one using sequential testing.

Our considerations include two more model selection techniques. They are Autometrics, constructed by Doornik (2009), and the Marginal Bridge Estimator (MBE), see Huang, Horowitz and Ma (2008). The starting-point of both techniques involves all variables in S, but the process of selecting the final model by MBE is very different from doing that with Autometrics. This latter technique is frequently used in macroeconometric modelling, whereas statisticians and microeconometricians are more familiar with MBE. A presentation of these techniques and a discussion of how they are applied in this work can be found in Kock and Teräsvirta (2011b) and is not repeated here.

Our application is univariate, but any variable helpful in forecasting y_t can be added to $\mathbf{z}_t$ It would also be possible to consider other models than (1). However, as the ANN model is often advertised as the 'best' available forecasting tool, we use it to forecast the Finnish consumer price index or inflation and see how well it performs. It will be compared to a benchmark which in our case is a linear autoregressive model.

3. Forecasting

3.1. Two ways of generating multiperiod forecasts

There are two main ways of creating multiperiod forecasts. One can either generate the forecasts recursively, or forecast directly. 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. In forecasting with nonlinear models the recursive forecasts are computed numerically. Direct forecasting commands a computational advantage in that the forecasts are computed without recursion. It appears quite common in macroeconomic forecasting. For example, the forecasts from the nowadays popular factor models are direct without exception. In the empirical section of the paper we compare results from these two approaches.

The techniques of recursive forecasting with nonlinear models are discussed in several books, book chapters and articles; see, for example, Teräsvirta (2006), Kock and Teräsvirta (2011a) or Teräsvirta, Tjøstheim and Granger (2010, Chapter 14). For more information on recursive forecasting in the context of ANN models, the reader is referred to Kock and Teräsvirta (2011b).

3.2. Forecasts based on differences and forecast errors

The forecasts based on differences are obtained as in Kock and Teräsvirta (2011b). When forecasting recursively, first differences $\Delta y_t = y_t - y_{t-1}$ are being modelled and forecast. To obtain an h-periods-ahead forecast of y_{T+h}, the first-difference forecasts have to be cumulated[1]:

$$(2) \quad \mathrm{E}(y_{T+h}|F_T) = \sum_{j=1}^{h} \mathrm{E}(\Delta y_{T+j}|F_T) + y_T$$

where F_T is the information set available at time T. The corresponding forecast error is

[1] *The unknown $E(\Delta y_{T+j}|F_T)$ are of course replaced by their bootstrapped counterparts.*

$e_{T+h|T} = y_{T+h} - E(y_{T+h}|F_T)$. Another possibility often applied in practice is to forecast the h-period change, that is, $\Delta_h y_{T+h}$, by obtaining the forecast of Δy_{T+h} and inflating it to the h-period level.

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 $\Delta_h y_{t-h}, \ldots, \Delta_h y_{t-h-p+1}$ and the corresponding forecast of y_{T+h} is $\mathrm{E}(\Delta_h y_{T+h}|F_T) + y_T$. The estimated model yields direct estimates of the conditional mean (2).

The main 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 forecast horizon. Thus,

$$RMSFE_h = \{(T - h_{max} - T_0 + 1)^{-1} \times \sum_{t=T_0}^{T-h_{max}} e_{t+h|t}^2\}^{1/2}$$

We also apply Wilcoxon's signed rank test (Wilcoxon, 1945) for pairwise comparisons. It tests the hypothesis that the density of $\left|e_{T+h|T}^{(1)}\right| - \left|e_{T+h|T}^{(2)}\right|$ is symmetric around zero, where $e_{T+h|T}^{(1)}$ and $e_{T+h|T}^{(1)}$ are errors of forecasting the same observation independently with two different methods or models. Some caution is required in interpreting the results when the horizon $h > 1$, because the test is based on the assumption that the forecast errors are independent over time. The test can be viewed as a way of providing an opportunity to see whether differences in the RMSFE between methods and models are significant.

3.3. Data window

Many forecasters assume that the parameters of their model do not remain constant over time and use a data window to take this into account, albeit not explicitly by modelling the change. In this work we employ an expanding window, implicitly assuming that the functional form is sufficiently flexible to accommodate potential parameter changes (in an underlying time-varying parameter linear model). A rolling window keeping the length of the time series fixed is an alternative. One can also think of an intermediate form in which the value of the past information in the series gradually decays towards zero. This can be achieved by geometrically weighted regression, see for example Törnqvist (1957).

3.4. Insanity filters

Nonlinear models sometimes generate forecasts that are deemed unrealistic in the light of the hitherto observed values of the time series. This seems not an uncommon situation in the present work due to the fact that the three model selection procedures may sometimes choose strongly correlated regressors, which adversely affects the corresponding parameter estimates. This difficulty can be remedied, at least partly, by replacing an unrealistic forecast with one that better conforms to already observed data. There are examples of this, see Swanson and White (1995, 1997a,b), who call the procedure the insanity filter, Stock and Watson (1999), Teräsvirta, van Dijk and Medeiros (2005), and Kock and Teräsvirta (2011b, in press). Following the last-mentioned authors, we apply both the Swanson and White (SW) and the linear autoregressive filter. The SW filter works as follows. If the h-step ahead predicted change exceeds the maximum h-step change observed during the estimation period, it is replaced by the most recently observed value of the variable to be predicted. Use of the linear AR filter implies that an extreme predicted change is replaced by a forecast from our benchmark linear autoregressive model.

4. Results

4.1. Data and forecasting horizons

The series to be considered is the monthly Consumer Price Index (CPI) for Finland obtained from the OECD Main Economic Indicators. The series begins March 1960 and ends in December 2009. The logarithms of the series are used for modelling and forecasting. The year-on-year inflation series (12-month differences of the logarithmic CPI) is graphed in Figure 1.

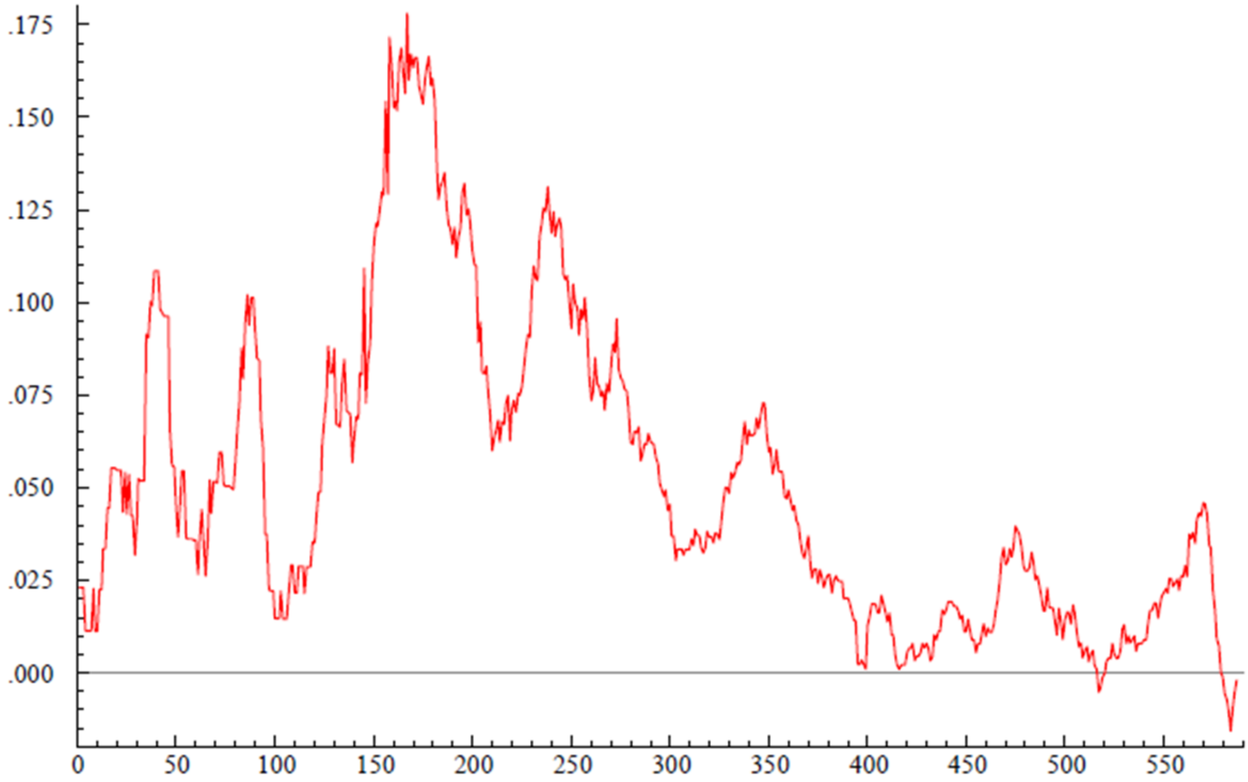

Figure 1. The Finnish year-on-year inflation, March 1961–December 2009

In building the models, the maximum lag length considered was six, and 240 forecasts based on an expanding window were generated at the 1, 3, 6, and 12-month horizon. The recursive and direct forecasts were obtained based on differences as well as the level of the series. When differences were used the forecasts were formed as described in Section 3.2. The pool of variables contained 600 hidden units and six linear lags. The models were respecified every sixth month. The benchmark for all results is the linear AR model based on differences and the basic measure of forecasting accuracy is, as already mentioned, the RMSFE.

4.2. Accuracy of forecasts from models built on differences

We first consider the recursive forecasts for differences in Table 1. The NF-AR block of the table contains the actual root mean square errors of the forecasts. The rest of the column indicates that the linear AR forecasts have not been filtered: the ratios equal unity. But then, filtering is absolutely necessary for the other methods except QN-SG. The large RMSFE-values are due to few really aberrant forecasts. In our experiment they typically amount to less than 1% of all forecasts. The reason for these forecasting failures is that the automated techniques sometimes select highly collinear variables. Their (very uncertain) parameter estimates reflect this and make the model unfit for forecasting. In 'real life' the forecaster would be able to notice this and make changes to the model, but in large-scale simulations this is hardly possible.

Table 1. Root mean square forecast error ratios of the recursive 1, 3, 6 and 12-month forecasts of the CPI series

Filter	Hor.	AR	QN	MBE	AM	QN-SG
	1	0.003687	1.006	1.034	1.475	1.036
NF	3	0.00758	9490	1.059	$2.1x10^{8}$	1.068
	6	0.01337	$1.4x10^{6}$	7.992	$1.2x10^{10}$	1.076
	12	0.02817	$5.2x10^{6}$	35.3	$1.8x10^{10}$	1.338
	1	1	1.006	1.034	1.081	1.036
SW	3	1	1.043	1.054	1.397	1.068
	6	1	1.050	1.079	1.240	1.077
	12	1	1.041	1.099	1.202	1.050
	1	1	1.006	1.034	1.080	1.036
AR	3	1	1.043	1.054	1.399	1.068
	6	1	1.050	1.079	1.242	1.077
	12	1	1.041	1.099	1.194	1.050

Models based on differences and respecified every six months. Forecasting begins March 1989. AR: linear AR, QN: QuickNet, MBE: Marginal Bridge Estimator, AM: Autometrics, QN-SG: QuickNet with specific to general. NF: No filter, SW: Swanson and White filter, AR: Linear AR filter. The NF-AR block contains the actual Root Mean Square Forecast Errors of forecasts from the linear AR model. '

A disappointing fact is that none of the modelling techniques leads to ANN models that would outperform the linear AR model, not even after filtering. QuickNet comes perhaps closest, and Autometrics is the worst performer. These results are in line with the general ones in Kock and Teräsvirta (2011b) with one exception: when the forecasts from all 11 countries are considered jointly, MBE seems to yield the best performing ANN models.

Next we consider the direct forecasts from models based on CPI differences. We add a nonparametric model containing six lags of Δy_t and a 'no change' forecast. The results are in Table 2. It is seen that at six- and twelve-month horizons direct forecasts are more accurate than the recursive ones. This is not, however, the case for nonparametric or 'no change' forecasts. On the other hand, even the direct linear AR forecasts are more accurate than their recursive counterparts. If the nonlinear forecasts are compared with them, it seems that the ANN models do not outperform the linear AR ones at any horizon. One can also observe that filtering is not needed for any model forecasts. These results do not deviate very much from the general 11-country outcomes in Kock and Teräsvirta (2011b): the main findings are quite similar to what is reported here.

Table 2. Root mean square forecast error ratios of the direct 1, 3, 6 and 12-month forecasts of the CPI series

Filter	Hor.	AR	QN	MBE	AM	QN-SG	NP	NC
	1	1	1.006	1.034	1.475	1.036	1.162	1.042
NF	3	0.9618	2.540	1.010	1.276	1.068	1.155	1.079
	6	0.8070	0.8301	0.8076	0.8220	0.8222	1.060	1.021
	12	0.7316	0.7309	0.7210	0.9572	0.7320	0.9607	0.8552
	1	1	1.006	1.034	1.081	1.036	1.162	
SW	3	0.9618	1.074	1.010	1.192	1.068	1.155	
	6	0.8070	0.8301	0.8076	0.8220	0.8222	1.060	
	12	0.7316	0.7309	0.7210	0.8628	0.732	0.9607	
	1	1	1.006	1.034	1.080	1.036	1.162	
AR	3	0.9618	1.061	1.010	1.193	1.068	1.155	
	6	0.8070	0.8301	0.8076	0.8220	0.8222	1.060	
	12	0.7316	0.7309	0.7210	0.8659	0.7320	0.9607	

Models based on differences and respecified every six months. Forecasting begins January 1989. AR: linear AR, QN: QuickNet, MBE: Marginal Bridge Estimator, AM: Autometrics, QN-SG: QuickNet with specific to general, NP: nonparametric, NC: no change, NF: No filter, SW: Swanson and White filter, AR: Linear AR filter.

In order to see whether the differences in performance observed in Tables 1 and 2 are significant we apply the Wilcoxon signed-rank test, see Wilcoxon (1945). As already mentioned, the results for the longer horizons have to be treated with caution because the forecast errors for horizons longer than one month are not independent. Under the null hypothesis, when all differences $|e^{AR}_{T+h|T}| - |e^{ANN}_{T+h|T}|$ are ordered in ascending order, the expected value of the rank sums of the positive and negative differences is zero. The test is based on AR-filtered forecasts. The results are in Table 3. Small p-values indicate rejection in favour of the ANN-forecasts. Large p-values (or small values of $1-p$) suggest rejecting the null hypothesis in the other direction.

Table 3. p-values of the Wilcoxon test of the hypothesis that absolute errors of linear AR forecasts based on the differenced series are of the same size as ones from an ANN mode built on differences

Filter	Hor.	QN	MB	AM	QN-SG
	1	0.756	0.953	0.837	0.949
AR	3	0.980	0.999	1	1
	6	0.991	1	1	1
	12	0.981	1	1	0.976

Filter	Hor.	AR	QN	MB	AM	QN-SG	NP	NC
	1		0.756	0.953	0.837	0.949	1	0.621
AR	3	0.00139	0.487	0.336	0.996	0.658	1	0.709
	6	0	0	0	2×10^{-9}	0	1	0.409
	12	0	0	0	5×10^{-7}	0	0.0117	0.000173

Upper panel: recursive forecasts; lower panel: direct forecasts.

The results show that the recursive linear AR forecasts are significantly more accurate than the nonlinear ones at horizons exceeding one month (p-values close to one). The lower panel indicates that at six- and twelve-month horizons direct forecasts are superior to recursive linear ones, whereas the results are mixed for the three-month horizon. Note, however, that direct forecasts from the nonparametric model are strongly inferior to recursive linear AR forecasts at the three shortest horizons, whereas there is some evidence of a significant difference in the other direction for 12-month forecasts. These outcomes are in line with the information in Table 2.

In order to shed light on the question why the direct forecasts appear more accurate than recursive forecasts we report results on the composition of the selected models in Table 4. It is seen that the recursive (or one month direct) forecasting ANN models almost exclusively contain hidden units (logistic functions in $\mathcal{S}$). For longer horizons we have chosen MBE to represent ANN models. The tendency is quite clear: the share of hidden units decreases when the forecasting horizon increases. This explains the fact that the linear autoregressive and MBE-based forecasts are almost equally accurate at six- and 12-month horizons.

Table 4. The average number of variables (Total), linear lags (Linear) and hidden units (HU) selected for ANN models for recursive forecasting and direct MBE-based models

Recursive	Total	Linear	HU	Direct, MBE	Total	Linear	HU
QN	6.10	0.05	6.05	1 month	5.72	0	5.72
MBE	5.72	0	5.72	3 months	4.38	1.30	3.08
AM	9.93	0.225	9.70	6 months	4.3	4.05	0.25
QN-SG	5.15	0	5.15	12 months	2.8	2.70	0.10

A comparison of the model selection techniques in the left-hand panel shows that Autometrics on average tends to select the least parsimonious models. The same tendency is visible in the 11-country study in Kock and Teräsvirta (2011b). Combining these results with the RMSFE outcomes in Tables 1 and 2 suggests that the ANN models without linear components may fit the data better than the linear AR models (the selection techniques prefer hidden units) but yield less accurate forecasts.

4.3. Accuracy of forecasts from models built on levels

Since the consumer price index and its logarithm are positively trending variables, it may at first seem strange to build models based on log-levels of the index. But then, because six lags of the (log-)index are available to be included in the models without any parameter restrictions we give level models a try. We begin by reporting results of recursive forecasting. To facilitate comparisons with forecasts discussed in the preceding section, the benchmark is still the same: the recursive linear AR forecasts from the model built on first differences.

The results for recursive forecasts can be found in Table 5. The top left-hand NF-AR panel contains the same benchmark RMSFEs as Table 1 and form the denominators of the ratios in this table. These ratios are completely different from the ones in Table 1. In particular, Autometrics is now the best performer and outperforms the recursive linear AR forecasts based on first differences at every horizon. Their superiority increases with the length of the horizon, and filtering is only required at the 12-month horizon. Even the other methods, QN-SG excepted, perform better than the linear AR. MBE is a steady performer, and forecasts from it now need no filtering at any horizon. The good performance of models built by Autometrics is somewhat atypical in the light of the aggregate 11-country results in Kock and Teräsvirta (2011b). They show that both QuickNet and MBE-based models yield more accurate forecasts than Autometrics.

Table 5. Root mean square forecast error ratios of the recursive forecasts for the CPI series

Filter	Hor.	AR	QN	MBE	AM	QN-SG
	1	0.003687	0.9534	0.9737	0.9423	1.125
NF	3	0.00758	153.9	0.9203	0.8691	1.276
	6	0.01337	$3x10^6$	0.8152	0.7613	1.313
	12	0.02817	$9x10^6$	0.6826	$5x10^6$	1.281
	1	0.9874	0.9534	0.9737	0.9423	1.125
SW	3	0.9623	0.9406	0.9203	0.8691	1.276
	6	0.9159	0.8545	0.8150	0.7612	1.311
	12	0.8576	0.7644	0.6826	0.6592	1.206
	1	0.9874	0.9534	0.9737	0.9423	1.125
AR	3	0.9623	0.9406	0.9203	0.8691	1.276
	6	0.9159	0.8545	0.8150	0.7612	1.311
	12	0.8576	0.7644	0.6826	0.6592	1.210

Models based on levels and respecified after six months. Forecasting begins March 1989. AR: linear AR, QN: QuickNet, MBE: Marginal Bridge Estimator, AM: Autometrics, QN-SG: QuickNet with specific to general. NF: No filter, SW: Swanson and White filter, AR: Linear AR filter.

Before studying possible significance of differences between the models we report results obtained by direct forecasting. They appear in Table 6. The general tendency is the same as in Table 5 in that the ANN-forecast ratios are all below unity. Obviously, the nonparametric model is not a feasible choice as it is now used to forecast a trending time series. The 'no change' forecasts are the same as in Table 2. It may be noticed that Autometrics-based forecasts are no longer among the most accurate ones. MBE now seems the best performer, followed by QuickNet. Direct AR models generate more accurate forecasts than the recursive ones even when models based on levels are compared. The results in Table 6 do not deviate much from the aggregates in Kock and Teräsvirta (2011b).

Table 6. Root mean square forecast error ratios of the direct forecasts for the CPI series

Filter	Hor.	AR	QN	MBE	AM	QN-SG	NP	NC
	1	0.9874	0.9534	0.9737	0.9423	1.125	16.4	1.042
NF	3	0.9404	0.9882	0.8849	0.9739	1.282	8.244	1.079
	6	0.8612	0.8511	0.7519	0.9967	1.348	4.893	1.021
	12	0.6829	0.5990	0.5495	0.7240	1.112	2.536	0.8552
	1	0.9874	0.9534	0.9737	0.9423	1.125	4.009	
SW	3	0.9404	0.9882	0.8849	0.9739	1.282	5.976	
	6	0.8612	0.8511	0.7519	0.9967	1.348	4.893	
	12	0.6829	0.5990	0.5495	0.7240	1.112	2.536	
	1	0.9874	0.9534	0.9737	0.9423	1.125	3.996	
AR	3	0.9404	0.9882	0.8849	0.9739	1.282	5.962	
	6	0.8612	0.8511	0.7519	0.9967	1.348	4.893	
	12	0.6829	0.5990	0.5495	0.7240	1.112	2.536	

Models based on levels and respecified after six months. Forecasting begins March1989. AR: linear AR, QN: QuickNet, MBE: Marginal Bridge Estimator, AM: Autometrics, QN-SG: QuickNet specific to general, NP: nonparametric, NC: no change, NF: No filter, SW: Swanson and White filter, AR: Linear AR filter.

Table 7. *p*-values of the Wilcoxon test of the hypothesis that absolute errors of linear AR forecasts based on the differenced series are of the same size as the ones from an ANN model built on levels

Filter	Hor.	QN	MB	AM	QN-SG
	1	0.0646	0.0597	0.0544	1
AR	3	0.0700	0.00931	0.00576	1
	6	0.000649	0.000491	0.00137	1
	12	$5x10^{-6}$	$3x10^{-9}$	$1x10^{-6}$	1

Filter	Hor.	AR	QN	MB	AM	QN-SG	NP	NC
	1		0.0646	0.0597	0.0544	1	0.999	0.685
AR	3	$2x10^{-5}$	0.7520	$1x10^{-6}$	0.0921	1	1	0.901
	6	$4x10^{-10}$	0.00843	$2x10^{-7}$	0.6190	1	1	0.873
	12	0	0	0	$8x10^{-5}$	1	1	0.130

Upper panel: recursive forecasts; lower panel: direct forecasts.

The results of the Wilcoxon test in Table 7 agree with what the RMSFE ratios are suggesting. The upper panel shows that the difference between the absolute forecast errors of the recursive forecasts from the linear AR model based on differences and those from ANN models built on levels is significantly different from zero in favour of the latter. The QN-SG forecast errors constitute an exception. The results are somewhat more mixed for direct forecasts as there are cases in which the null hypothesis is not rejected. As expected, it is strongly rejected in the opposite direction for the nonparametric forecasts.

Table 8. The average number of variables (Total), linear lags (Linear) and hidden units (HU) selected for ANN models for recursive forecasting and direct MBE-based models in levels

Recursive	Total	Linear	HU
QN	5.42	1	4.42
MBE	6.70	6	0.700
AM	22.5	1.15	21.4
QN-SG	1.35	1	0.35

Direct, MBE	Total	Linear	HU
1 month	6.70	6	0.700
3 months	6.42	6	0.425
6 months	6.70	6	0.700
12 months	6.75	6	0.750

The average number of variables selected for the ANN models can be found in Table. The results reveal that MBE almost exclusively selects linear models, but it can be seen from Tables 5 and 6 that hidden units, when selected, do make a difference. Another notable fact is that Autometrics on average selects a remarkably large number of variables, only very few of them being linear lags. This outcome is not specific for the Finnish data, see Kock and Teräsvirta (2011b). Despite this, the Autometrics-based model is very successful in recursive forecasting when the data are in levels. Finally, QN-SG selects very few variables, which may explain its inferior performance. But then, this can be changed by changing the significance levels in the test sequence used to select the variables.

5. *Final remarks*

It appears difficult to draw general conclusions from the results. One seems to be that it is not worthwhile to difference the CPI series before building a model for forecasting. Another is that direct forecasts are somewhat more accurate than their recursive counterparts. These results do not generalize to other series, however. For example, in forecasting Finnish unemployment using monthly time series (results not discussed in this paper), building the models on differenced series and forecasting with them yields more accurate forecasts than what is obtained from models built on levels. Furthermore, recursive forecasts are slightly more accurate than the direct ones, which is not the case for the CPI forecasts. But then, in unemployment forecasts there are only few occasions in which the Wilcoxon test rejects the null hypothesis in favour of an ANN model.

A clear conclusion, also obvious from Kock and Teräsvirta (2011b), is that no single model or model-building technique dominates others. Another one is that nonlinearity may be an advantage in forecasting some macroeconomic series but not all of them. Finally, the results would probably change if other quantitative information than the past values of the series itself were used for forecasting. Using ANN models for forecasting with an information set extended with strongly exogenous variables would be an interesting research topic which is left for future research.

References

Doornik, J.A. (2009). "Autometrics." In *The Methodology and Practice of Econometrics: Festschrift in Honour of David F. Hendry*, 88–121. Eds. J. L. Castle and N. Shephard. Oxford: Oxford University Press.

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

Kock, A.B., and T. Teräsvirta (in press). "Forecasting performance of three automated modelling techniques during the economic crisis 2007-2009." *International Journal of Forecasting*.

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

Kock, A.B., and T. Teräsvirta (2011b). Nonlinear forecasting of macroeconomic variables using neural network models and automated model selection techniques. CREATES Research Paper 2011-27, Aarhus University.

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

Swanson, N.R., and H. White (1997a). "Forecasting economic time series using flexible versus fixed specifcation and linear versus nonlinear econometric models." *International Journal of Forecasting* 13, 439–461.

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

Swanson, N.R., 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 the Business and Economic Statistics* 13, 265–275.

Teräsvirta, T. (2006). "Forecasting economic variables with nonlinear models." In *Handbook of Economic Forecasting*. Vol. 1, 413–457. Eds. G. Elliott, C.W.J. Granger and A. Timmermann. Amsterdam: Elsevier.

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

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

Törnqvist, L. (1957). "A method for calculating changes in regression coefficients and inverse matrices corresponding to changes in the set of available data." *Skandinavisk Aktuarietidskrift* 40, 219–226.

White, H. (2006). "Approximate nonlinear forecasting methods." In *Handbook of Economic Forecasting*. Vol. 1, 459–512. Amsterdam: Elsevier.

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

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