SCHOOL OF ECONOMICS AND MANAGEMENT
FACULTY OF SOCIAL SCIENCES
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric Analysis of Time Series

CREATES Research Paper 2011-28

Forecasting performance of three automated modelling techniques during the economic crisis 2007-2009

Anders Bredahl Kock and Timo Teräsvirta

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

Forecasting performance of three automated modelling techniques during the economic crisis 2007–2009

Anders Bredahl Kock and Timo Teräsvirta
CREATES, Aarhus University
Building 1322, DK-8000 Aarhus C, Denmark

August 23, 2011

Abstract

In this work we consider forecasting macroeconomic variables during an economic crisis. The focus is on a specific class of models, the so-called single hidden-layer feedforward autoregressive neural network models. What makes these models interesting in the present context is that they form a class of universal approximators and may be expected to work well during exceptional periods such as major economic crises. These models are often difficult to estimate, and we follow the idea of White (2006) to transform the specification and nonlinear estimation problem into a linear model selection and estimation problem. To this end we employ three automatic modelling devices. One of them is White's QuickNet, but we also consider Autometrics, well known to time series econometricians, and the Marginal Bridge Estimator, better known to statisticians and microeconometricians. The performance of these three model selectors is compared by looking at the accuracy of the forecasts of the estimated neural network models. We apply the neural network model and the three modelling techniques to monthly industrial production and unemployment series of the G7 countries and the four Scandinavian ones, and focus on forecasting during the economic crisis 2007–2009. Forecast accuracy is measured by the root mean square forecast error. Hypothesis testing is also used to compare the performance of the different techniques with each other.

Keywords: Autometrics, economic forecasting, Marginal Bridge estimator, neural network, nonlinear time series model, Wilcoxon's signed-rank test

JEL Classification Codes: C22; C45; C52; C53

Acknowledgements: This research has been supported by the Danish National Research Foundation. A part of the work was carried out when the first author was visiting Department of Economics, University of California, Berkeley, and the second author likewise Department of Economics, European University Institute, Florence. Kind hospitality of these institutions is gratefully acknowledged. Material from this paper has been presented at the 6th Colloquium on Modern Tools for Business Cycle Analysis: The Lessons from Global Economic Crisis, EUROSTAT, Luxembourg, September 2010, the 19th Symposium of the Society for Nonlinear Dynamics and Econometrics, Washington, DC, March 2011, the Interdisciplinary Workshop 'Econometric and Statistical Modelling of Multivariate Time Series', Louvain-la-Neuve, May 2011, and in seminars at Shaanxi Normal University, Xi'an, Southwestern University of Finance and Economics, Chengdu, and University of Alabama at Tuscaloosa. We wish to thank participants of these occasions for their comments. Errors and shortcomings in this paper are our own responsibility.

1 Introduction

Economic crises provide a useful testing ground for time series models for forecasting. It is generally not possible to forecast a crisis well in advance, unless there is information about past crises of the same type in the data, which is usually not the case. Nevertheless, it is useful to investigate how well models based on quantitative time series forecast during a crisis and in its aftermath. This puts models to a severe test, because in quantitative terms an economic crisis involves a strong decrease (in production) or increase (in unemployment), while a reverse occurs in the aftermath. Models that quickly adapt to these changing conditions would then have an advantage over less flexible parameterisations.

In this paper the attention is restricted to a well-defined class of flexible models, the so-called single hidden-layer feedforward neural network models. Neural networks or multilayer perceptrons are universal approximators that can arbitrarily accurately approximate any function satisfying rather mild regularity conditions. In a recent study, Ahmed, Atiya, El Gayar and El-Shishiny (2010) compared the forecasting ability of several 'machine learning' tools, including various neural network models. They applied them to forecasting 1045 time series included in the M3 forecasting competition, see Makridakis and Hibon (2000). The series were monthly and contained at least 80 observations. It turned out that the neural network model of the type we shall consider in this paper was the overall winner of the comparison. Our aim is to study how well this model forecasts during the recent economic crisis and compare its performance with that of a linear autoregressive model, a nonparametric model, and a simple 'no change' forecast.

A problem with these multilayer perceptrons is how to specify their structure and estimate the parameters. Recently, White (2006) presented a solution that amounted to converting the specification and nonlinear estimation problem into a linear model selection problem. This leads to a somewhat atypical situation, at least in time series econometrics, in which the number of variables may vastly exceed the number of observations. The second aim of this paper is to compare three methods for model selection capable of handling this situation. One is White's QuickNet that Ahmed et al. (2010) mentioned as a possible extension to their study. The other two are the Marginal Bridge Estimator, see Huang, Horowitz and Ma (2008), and Autometrics by Doornik (2009). White (2006) proposed comparing QuickNet with other approaches, and we take up his suggestion.

In this study we shall consider multiperiod forecasts. There are two main ways of generating them. One is to specify and estimate a single model and generate the forecasts recursively from this model. It is also possible to build a separate model for each forecast horizon and use it for obtaining the forecasts. For discussion, see for example Teräsvirta, Tjøstheim and Granger (2010, Chapter 14). Marcellino, Stock and Watson (2006) compared these two methods in a linear framework. The third aim of this paper is to do the same when the set of models mainly consists of neural network and nonparametric models but also contains linear autoregressive ones.

Nonlinear models, such as the neural network model, sometimes generate unrealistic or 'insane' forecasts, see Swanson and White (1995, 1997a,b) for discussion. This problem can at least partly be remedied by adjusting such forecasts towards more realistic values. Our fourth aim is to consider this possibility that will be called filtering and see whether it can be useful in our forecasting situation.

It is possible to test linearity of the time series before any model selection and thus preclude nonlinear models when they seem superfluous. In theory this is not necessary if linear lags of the model to be forecast are included in the set of variables to select from in building the neural network model. We shall see whether or not such pre-screening improves the accuracy of the forecasts.

These problems have already been considered in Kock and Teräsvirta (2011b). The novelty of this companion paper is its focus on the recent economic crisis and its aftermath. We shall consider forecasting two monthly macroeconomic variables that have been strongly affected by the crisis: industrial production and unemployment rate.

The plan of the paper is as follows. The neural network model is presented in Section 2 and the modelling techniques in Section 3. The two forecasting methods, recursive and direct, are briefly discussed in Section 4. The time series from 11 different countries are presented in Section 5. Section 6 is devoted to empirical results. Final remarks can be found in Section 7.

2 The model

The focus of this paper will be on forecasting with a flexible model during the recent economic crisis when the macroeconomic series to be forecast show exceptionally large fluctuations. The idea is to find out how well our flexible functional form performs in this situation. The techniques for specifying the structure of the model and estimating the parameters are the same as in Kock and Teräsvirta (2011b). The difference is that this paper concentrates on the crisis and recovery years 2007–2009.

Following Kock and Teräsvirta (2011b), our model is the so-called single-hidden-layer feedforward autoregressive neural network (ANN) model or single-hidden-layer perceptron

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

where $\mathbf{z}_t = (1, y_{t-1}, ..., y_{t-p})'$, $\boldsymbol{\gamma} = (\gamma_{j1}, \gamma_{j2}, ..., \gamma_{jp})'$, $j = 1, ..., p$, $\boldsymbol{\beta}_0 = (\beta_{00}, \beta_{01}, ..., \beta_{0p})'$,and $\varepsilon_t \sim \text{iid}\mathcal{N}(0, \sigma^2)$. As is well known, the ANN model is a so-called universal approximator. Suppose there is a functional relationship between y and $\mathbf{z}$: $y = H(\mathbf{z})$. Then for all $\delta > 0$ there exists a positive integer $q < \infty$ such that $|H(\mathbf{z}) - \sum_{j=1}^{q} \beta_j (1 + \exp\{\boldsymbol{\gamma}_j'\mathbf{z}\})^{-1}| < \delta$ where $| \cdot |$ is an appropriate norm. As explained in Kock and Teräsvirta (2011b), (1) is a flexible functional form which can be used for approximating various unknown nonlinear processes. Note that (1) is not the only available universal approximator. Nevertheless, it is a popular one, and one for which White (2006) constructed a useful specification strategy. Such a strategy is needed, because the number of the logistic functions or hidden units q is unknown a priori and has to be specified. Furthermore, the parameters of (1) have to be estimated, which is a nonlinear estimation problem. In this work we follow Kock and Teräsvirta (2011b) and linearise the nonlinear specification and estimation problem, which is what White (2006) originally suggested. Assuming the parameter vectors $\boldsymbol{\gamma}_j$, $j = 1, ..., q$, known in (1) makes the model linear. The linear model selection problem is the one of choosing a subset of variables from the set

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

where M is large. It is clear that the quality of the estimates depends on the size of S. For this reason, in a typical macroeconomic application the number of elements in S is likely to exceed the number of observations. This requires model selection techniques with which one can handle such a situation.

3 Modelling with three automatic model selection algorithms

In this section, analogously to Kock and Teräsvirta (2011b), 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, which is a development of PcGets, see Krolzig and Hendry (2001), Hendry and Krolzig (2005) and Doornik (2009), Marginal Bridge Estimator (MBE, Huang, Horowitz and Ma, 2008), and QuickNet (White, 2006). Autometrics has been built on the principle of proceeding 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 the 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

The algorithm is described in detail in Doornik (2009). Modelling begins with a linear model called the General Unrestricted Model (GUM). When the number of variables is less than the number of observations, 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. This is done by removing variables with insignificant coefficients, which can be done in various ways. 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. In the opposite case, Autometrics backtracks by adding variables until it finds a model that passes these tests. Since there are many search paths, there will in general be several terminal models as well.

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 test form a new GUM, and the general-to-specific testing procedure is repeated and a new set of terminal models is obtained. If all these are rejected against the new union model, the union will be the final model. Otherwise, modelling restarts with yet another GUM and continues until a final model has been reached.

In our case, the number of variables exceeds the number of observations. Like Hendry and Krolzig (2005) we divide the variables into subsets, each of which contains fewer variables than observations. 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, which happens in our application, 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. These will be briefly discussed in Section 4.

3.2 Marginal Bridge estimator

The MBE is designed for situations often occurring in microeconomic applications in which there is a large number of candidate variables, only a handful of which may be relevant. These variables are sorted out using the MBE. Following Huang et al. (2008), consider first the Bridge estimator (BE). This is a shrinkage estimator for a linear regression model

$$y_i = \beta_0 + \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_t \tag{3}$$

where $\mathbf{x}_i = (x_{i1}, ..., x_{ip_n})'$ is an $p_n \times 1$ observation vector, $i = 1, ..., n$, and $p_n < n$, and $\beta_0 = 0$ without loss of generality. Furthermore, $\varepsilon_i \sim \text{iid}(0, \sigma^2)$. The BE estimator is the solution to the following minimisation problem:

$$Q_n^{\text{B}}(\boldsymbol{\beta}) = \sum_{i=1}^{n} (y_i - \boldsymbol{\beta}'\mathbf{x}_i)^2 + \lambda_n \sum_{j=1}^{p_n} |\beta_j|^\gamma \tag{4}$$

where $p_n < n$. Let the true parameter vector $\boldsymbol{\beta}_0 = (\boldsymbol{\beta}_{10}', \boldsymbol{\beta}_{20}')'$ with $\boldsymbol{\beta}_{20} = \mathbf{0}$, and let $\widehat{\boldsymbol{\beta}}_n = (\widehat{\boldsymbol{\beta}}_{1n}', \widehat{\boldsymbol{\beta}}_{2n}')'$ be the corresponding estimator from (4). Under regularity conditions and $0 < \gamma < 1$, (a) the estimator $\widehat{\boldsymbol{\beta}}_{2n} = \mathbf{0}$ with probability converging to one, and (b), $\boldsymbol{\beta}_{10}$ is estimated consistently. Furthermore, the asymptotic distribution of $\widehat{\boldsymbol{\beta}}_{1n}$ is the same as if only these had been included in the model.

When $p_n > n$, BE is not applicable and has to be replaced by the Marginal Bridge Estimator (MBE). The idea is to run a series of 'mini' or 'marginal' regressions, with a joint penalty. The function to be minimised equals

$$Q_n^{\text{MB}}(\boldsymbol{\beta}) = \sum_{j=1}^{p_n} \sum_{i=1}^{n} (y_i - \beta_j x_{ij})^2 + \lambda_n \sum_{j=1}^{p_n} |\beta_j|^\gamma \tag{5}$$

where $0 < \gamma < 1$, and λ_n determines the size of the penalty. Let $\widetilde{\boldsymbol{\beta}}_n = (\widetilde{\boldsymbol{\beta}}_{1n}', \widetilde{\boldsymbol{\beta}}_{2n}')'$ now be the estimator of $\boldsymbol{\beta}_0$ from (5). Under regularity conditions and $0 < \gamma < 1$, (a) holds for $\widetilde{\boldsymbol{\beta}}_{2n}$, and (b) $\Pr\{\widetilde{\beta}_{1nk} \neq 0, \widetilde{\beta}_{1nk} \in \widetilde{\boldsymbol{\beta}}_{1n}\} \to 1$, as $n \to \infty$. Thus, (a) and (b) jointly can be expected to efficiently separate the variables with nonzero coefficients from the rest.

There is yet another assumption for (a) and (b) to hold. The variables with nonzero coefficients and the ones with zero coefficients have to be either uncorrelated or only weakly correlated. This is called partial orthogonality. That assumption is clearly not satisfied in our case. Nevertheless, our aim is to see how well the MBE works in our time series example when the partial orthogonality assumption is violated.

3.3 QuickNet

QuickNet resembles an earlier modelling device called RETINA, see Perez-Amaral, Gallo and White (2003). The idea of RETINA is to find the explanatory variables in a set 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. QuickNet differs from RETINA in that the set of candidate variables is different, and the model selection criterion used for final selection is also different. QuickNet 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) and Kock and Teräsvirta (2011b) 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 absolute 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. Following White (2006) and Kock and Teräsvirta (2011b), we applied the 10-fold cross-validation criterion of Hastie, Tibshirani and Friedman (2009). We also experimented with the hv-cross validation criterion of Racine (2000), but it did not improve the forecasting performance of the resulting models.

We also considered a purely specific-to-general version of QuickNet. The variables are selected one at a time as before, but every choice is preceded by a linearity test. Parsimony is appreciated, so the significance level of the tests in the sequence decreases as the number of variables the model increases. The test is the 'economy version' of the test in Luukkonen, Saikkonen and Teräsvirta (1988). Adding variables is terminated at the first non-rejection of the linearity 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 6, this method is called QN-SG.

4 Forecasting

4.1 How to forecast?

4.1.1 Two ways of generating multiperiod forecasts

Our model-based forecasts are estimates of conditional means from a model or models. Computing a one-period-ahead forecast is no problem. There are two main ways of forecasting more than one period ahead. One can 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 of them. Next we briefly describe these two methods. Empirical results obtained from them are discussed in Section 6.

4.1.2 Recursive forecasts

In order to illuminate recursive forecasting, consider the model (1) with $p = q = 1$. These restrictions are for notational simplicity only. The one-period-ahead forecast made at time T and assuming the information set $\mathcal{F}_{T-1} = \{y_{T-j}, j \geq 1\}$ equals

$$y_{T+1|T} = \beta_{00} + \beta_{01} y_T + \beta_1 (1 + \exp\{\gamma_0 + \gamma_1 y_T\})^{-1}.$$

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

$$\begin{aligned} y_{T+2|T} &= \mathsf{E}\{\beta_{00} + \beta_{01}(y_{T+1|T} + \varepsilon_{T+1}) \\ &\quad + \beta_1 (1 + \exp\{\gamma_0 + \gamma_1 (y_{T+1|T} + \varepsilon_{T+1})\})^{-1} + \varepsilon_{T+2} | \mathcal{F}_T\} \\ &= \beta_{00} + \beta_{01} y_{T+1|T} + \beta_1 \mathsf{E}\{(1 + \exp\{\gamma_0 + \gamma_1 (y_{T+1|T} + \varepsilon_{T+1})\})^{-1} | \mathcal{F}_T\} \\ &= \beta_{00} + \beta_{01} y_{T+1|T} \\ &\quad + \beta_1 \int_{-\infty}^{\infty} (1 + \exp\{\gamma_0 + \gamma_1 (y_{T+1|T} + z)\})^{-1} \phi(z) \mathrm{d}z \end{aligned} \tag{6}$$

where $\phi(z)$ is the density of the $\mathcal{N}(0, \sigma^2)$ random variable. The integral in (6) has to be computed by numerical integration. Note that the integral becomes a multiple integral when the forecasting horizon h exceeds two. It is therefore advisable to calculate its value by simulation or by bootstrapping the residuals of the model, because that is computationally feasible even when $h > 2$. The reason we mention this is that some authors bypass this complication altogether by setting $\varepsilon_{T+1} = 0$ in the logistic function. 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 nonlinear forecasting can be found in Teräsvirta (2006), Kock and Teräsvirta (2011a) or Teräsvirta, Tjøstheim and Granger (2010, Chapter 14), among others.

4.1.3 Direct forecasts

In direct forecasting, the conditional mean estimate arises from a different model for each time horizon. 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, model selection is made using the three aforementioned techniques, but there is a 'gap' in the model in that $y_{T+h-1}, ..., y_{T+1}$ 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, which does require some computational effort. This is the case in particular when Autometrics is used to find an appropriate model.

4.1.4 Forecasts and forecast errors

The models for industrial production are built on first differences of the logarithm of the index of monthly industrial production. Unemployment rates are differenced before modelling. This implies that the forecasts are those of one-month differences $\Delta y_{T+j} = y_{T+j} - y_{T+j-1}$. In forecasting industrial production recursively $h \geq 2$ periods ahead, the forecast of interest for decision-makers is most often not $\Delta y_{T+h|T}$ but either $\Delta_h y_{T+h|T} = \sum_{j=1}^{h} \Delta y_{T+j|T}$, or $y_{T+h|T} = \sum_{j=1}^{h} \Delta y_{T+j|T} + y_T$. The corresponding h-periods-ahead forecast error is in both cases

$$e_{T+h|T} = y_{T+h} - y_{T+h|T}.$$

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 y_t are thus h-period differences, which has to be kept in mind when direct and recursive forecasts are compared with each other. The estimated model yields direct estimates of $\Delta_h y_{T+h|T} = \mathsf{E}\{y_{T+h} - y_T | \mathcal{F}_T\}$. The direct forecast of y_{T+h} then equals

$$y_{T+h|T} = \Delta_h y_{T+h|T} + y_T.$$

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 $h_{\max}$ the longest forecasting horizon. Thus,

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

The time series are monthly and the forecast horizons are $h = 1, 3, 6, 12$. We shall also compare our forecasting techniques by ranking them using squared forecast errors as the criterion.

4.2 Filters, windows and settings

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 (1995a,b, 1997) who call the procedure the insanity filter, Stock and Watson (1999) and Teräsvirta, van Dijk and Medeiros (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, use the last observed change. Swanson and White (1995) described this as 'replacing craziness by ignorance'. In our second filter, the extreme predicted change is replaced by a forecast from our benchmark linear autoregressive model: craziness is replaced by linearity.

Many forecasters assume that the parameters of their model do not remain constant over time and use a data window to take this implicitly into account. Since our forecasting exercise comprises a very extreme time period, we introduce flexibility or 'learning' into our framework by applying a rolling 12-year window when estimating our monthly neural network models. Rolling windows are quite popular among practitioners, but they have the potential drawback that the value of the observations in the series abruptly drops from a positive value equal for all observations to zero after 12 years. A more logical, albeit rarely applied, alternative would be to let the value of the past information in the series gradually decay towards zero. This can be achieved by weighted regression, see for example Törnqvist (1957) or Gilchrist (1967), but is not attempted here.

As already mentioned, Autometrics allows the user to choose settings that define different modelling strategies. We selected the Autometrics p-value for specification tests to equal 0.001. We also allowed pre-search variable reduction, which can exclude variables from consideration before the actual searches are begun. For more information, see Doornik (2009).

5 Data

The monthly industrial production and unemployment series are obtained from the OECD Main Economic Indicators. Most of them cover the period from the 1960s to the end of 2009 or early 2010. In this work, however, the observation period begins in August 1994. This is because our rolling windows consist of 12 years of monthly observations, 144 observations in all. This number may be regarded as a compromise between stability and flexibility, as it turned out that shorter observation periods had a tendency to make the ANN models more unstable and lead to a larger number of unrealistic forecasts.

The countries we consider are the G7 countries and the four Scandinavian countries, Denmark, Finland, Norway, and Sweden. The logarithmic industrial production index series of these 11 countries for the observation and prediction period are shown in Figure 1. The industrial production generally began to decline in the early 2008, in a few countries rather steeply. In many but not all cases the decline was reversed in 2009. This reversal was sharpest in Japan. Norway constitutes the most conspicuous exception to the general trend. The Norwegian industrial production has been slowly decreasing since the early 2000's, and the decline does not accelerate much during 2008.

Figure 2 contains the graphs of the 11 unemployment series for the observation and prediction period. The numbers of unemployed started to increase in 2008, and the growth has since levelled out in many but not all countries. Denmark is perhaps the most dramatic example of a country in which the unemployment rate was still rising at the end of 2009. The Norwegian unemployment rate, while also increasing, still remained on a relatively low level compared to the other ten countries in our study.

6 Results

6.1 Generating the variables

The technique for generating the potential hidden units for the ANN model (1) is described in the Appendix. We modified the original White (2006)

Filter	Hor.	AR	QuickNet	MBE	Autom.	QN-SG
	1	1	1×10^6	2.713	28.78	2.423
NF	3	1	6×10^5	4.085	2×10^7	3.561
	6	1	3×10^6	3.934	4×10^7	3.461
	12	1	3×10^6	3.526	2×10^8	3.182
	1	0.9938	1.043	1.010	1.007	1.019
SW	3	0.9946	1.060	1.053	1.026	1.068
	6	1	1.096	1.044	1.086	1.088
	12	1.020	1.149	1.056	1.128	1.137
	1	1	1.025	0.9926	1.051	1.002
AR	3	1	1.011	1.007	1.041	1.018
	6	1	1.061	1.012	1.081	1.051
	12	1	1.111	1.021	1.103	1.097

Table 1: Root mean square forecast error ratios of the recursive forecasts for the 11 industrial production series. Models respecified after each period. Forecasting begins July 2007

technique somewhat to make it more suitable to our problem; see Kock and Teräsvirta (2011b) for more discussion. Our results are based on a set of 1200 hidden units and six lags of the dependent variable. A new set was generated every time a new observation became available. We also experimented with a smaller set containing only 600 hidden units but found that the accuracy of the forecasts from the ANN model was higher on average, albeit not uniformly, for the larger set. We mainly report results based on the larger set.

In some cases, the results are based on smaller sets of candidate variables. Autometrics crashed when applied to the German industrial production series with the set of 1206 candidate variables. The results for that series are based on the smaller set of 606 variables. There were similar problems in modelling unemployment series. Autometrics crashed for three countries: Finland, Germany and Norway, when the selection pool contained 1200 hidden units. For Finland and Norway, reducing the number of hidden units to 600 was enough to prevent this from happening. For Germany, it was necessary to restrict the pool to contain only 150 logistic functions and six linear lags. The results of the paper are based on these pool sizes.

6.2 Industrial production

6.2.1 Overview

Overall results of forecasting the industrial production from July 2007 onwards can be found in Tables 1–3. As already discussed, the recursive forecasts for h months ahead, $h > 1$, are not forecasts of the monthly growth rate h months ahead, because they are usually not interesting for decision-makers. They are instead forecasts of the h-month growth rate, that is, sums of the one-month growth rate forecasts.

Table 1 shows the RMSFEs of the recursive forecasts. From the no filter (NF) panel it is seen that all automatic methods sometimes choose models that yield grossly erroneous forecasts. They need not be many, as already a single such forecast has a substantial effect on the RMSFE. After filtering, differences in the accuracy between the methods are rather small. On average, the AR filter seems to be a better one of the two alternatives, but the difference between them is not large. The linear AR forecasts are superior to the nonlinear ones but as will be seen, this result does not hold uniformly on the country level. The AR-filtered MBE forecasts come closest to the linear AR ones. Interestingly, the specific-to-general version of QuickNet (QN-SG) is no worse than the original one, in which ten hidden units are selected first and the size of the model reduced thereafter. The performance of QN-SG depends, however, on the choice of the significance levels for the tests in the sequence. At least in theory, by changing them improving its performance would be possible.

Table 2 contains the RMSFEs for the same period in the case where the model is not respecified after July 2007. It appears that respecification does not generally improve forecasting accuracy. Since differences in performance are not large, however, in what follows we only report results from respecified models. Models selected by QuickNet generate some rather inaccurate forecasts as do Autometrics-based models, whereas the QN-SG version of QuickNet does not. The linear AR model remains the best alternative. The forecasting accuracy of the four ANN models is almost as good as that of the linear model. Nevertheless, none of them represents an improvement over it.

The direct forecasts are obtained from the following model:

$$\Delta_h x_{t+h} = g^{(h)}(\Delta_h x_t, \Delta_h x_{t-1}, ..., \Delta_h x_{t-p_h+1}) + \varepsilon_{t+h}^{(h)} \qquad (7)$$

where $\mathsf{E}\varepsilon_{t+h}^{(h)} = 0$. The explanatory variables are $\Delta_h x_t$ and its $p_h - 1$ lags, $h > 1$. This means that the variables in the hidden units of the ANN model are different from the ones used in recursive forecasting. The RMSFEs of the direct forecasts can be found in Table 3. Even some of the direct forecasting

Filter	Hor.	QuickNet	MBE	Autom.	QN-SG
	1	1.111	1.052	1.06	1.112
NF	3	1.774	1.079	1.104	1.101
	6	3.251	1.025	2.927	1.047
	12	7.469	0.9856	11.72	0.9936
	1	1.073	1.036	1.007	1.092
SW	3	1.076	1.053	1.070	1.109
	6	1.034	1.020	1.030	1.043
	12	1.015	0.9885	0.9915	0.9948
	1	1.084	1.037	1.056	1.102
AR	3	1.058	1.064	1.084	1.091
	6	1.033	1.019	1.033	1.039
	12	1.010	0.9829	0.9908	0.9933

Table 2: Root mean square forecast errors of the recursive forecasts for the 11 industrial production series. Models not respecified after the start of the forecasting period in July 2007

Filter	Hor.	AR	QuickNet	MBE	Autom.	QN-SG	NP	NC
	1	1	1×10^6	2.713	28.78	2.423	0.9918	1.002
NF	3	1.011	686	2456	3×10^5	684	0.9974	1.057
	6	1.029	1×10^4	1×10^6	3×10^9	7×10^4	0.9681	1.020
	12	0.9875	2×10^5	1.050	3×10^8	1.095	0.9664	0.9290
	1	0.9938	1.043	1.010	1.007	1.019	0.9905	
SW	3	0.9971	0.988	0.964	0.9671	0.9825	0.9758	
	6	1.003	0.9784	1.000	1.002	0.9941	0.9568	
	12	0.9829	1.013	1.001	1.020	1.011	0.9641	
	1	1	1.025	0.9926	1.051	1.002	1.006	
AR	3	1.011	1.007	0.9773	0.9923	0.9981	1.022	
	6	1.029	1.030	1.023	1.038	1.034	0.9763	
	12	0.9875	1.016	0.9975	1.026	1.013	0.9668	

Table 3: Root mean square forecast errors of the direct forecasts for the 11 industrial production series. Models respecified after each period. NP: nonparametric model; NC: 'no change'

models occasionally generate vastly inaccurate forecasts, and this is the case for all four automatic methods.

The nonparametric (NP) model is the overall winner. The linear AR filter makes the NP forecasts less accurate on average. After filtering, the RMSFEs of the nonlinear forecasts are quite similar to the RMSFE of the linear model. A general observation is that it is difficult to substantially improve upon the linear model using White's ANN approach and automatic modelling. In fact, the 'no change' (NC) forecasts are the best alternative in forecasting 12 months ahead, i.e., in forecasting the annual growth rate of the industrial production during the crisis period.

Another way of comparing forecasts obtained by different methods is to rank them according the size of the absolute forecast errors. We do it across all months and countries for each forecast horizon, so the results fit into a single table. We only show the results for the SW filter. The rankings for the AR-filtered forecasts and the ones obtained for SW-filtered forecasts are very similar, and the former are therefore omitted. The average ranks over months and countries can be found in Table 4. For the one-month horizon, the linear AR model is the best method (has the lowest average rank), followed by QN-SG. NP is among the worst alternatives at the shortest horizons but is the best one in forecasting 12 months ahead. For the intermediate horizons, the linear AR model does well but the direct MBE forecasts have a low average rank in forecasting three months ahead. Direct Autometrics forecasts are not good, indicating that this method is in trouble when the available variables only provide a rather poor approximation to the data-generating process. All nonlinear methods except NP have poor average ranks when it comes to direct forecasts and 12-months-ahead forecasts.

A different robust method of comparing the forecasts can be obtained by pairing individual forecasts and testing the null hypothesis that the median (or mean because the density is assumed symmetric) of the difference of the absolute forecast errors of the recursive linear AR model and an alternative equals zero. This is done using the Wilcoxon signed-rank test, see Wilcoxon (1945), in which the alternative hypothesis is that the other model generates more accurate forecasts than the linear recursive AR model. The results are based on AR-filtered forecasts and can be found in Table 5. A small p-value indicates that the null hypothesis is rejected. Subtracting the p-value from one gives a p-value of the test in which the null and the alternative have changed places. A large p-value in the table thus suggests that the linear recursive AR model generates forecasts superior to those from a direct linear AR or a nonlinear or nonparametric model. A normal approximation to the null distribution is used to obtain the p-values. In considering the results one has to keep in mind that the forecast errors for forecasting more than

Rec	Hor.	AR	QuickNet	MBE	Autom.	QN-SG
	1	3.44	3.67	3.64	3.74	3.54
SW	3	5.77	6.15	5.90	6.06	6.28
	6	5.52	6.23	6.03	6.14	6.46
	12	5.60	5.92	5.73	6.02	5.98

Rec	Hor.	AR	QuickNet	MBE	Autom.	QN-SG	NP	NC
	1	3.44	3.67	3.64	3.74	3.54	3.59	5.92
SW	3	5.86	6.23	5.76	6.15	5.93	5.96	11.3
	6	5.95	5.74	5.90	6.52	5.95	5.96	11.0
	12	5.73	6.33	6.61	6.30	6.43	5.46	11.1

Table 4: The average ranks of the methods of forecasting the industrial production (all countries). Forecasting begins July 2007. Note: The no change forecasts are unfiltered

Rec.	Hor.	QuickNet	MBE	Autom	QN-SG
	1	0.950	0.808	0.992	0.915
AR	3	0.874	0.731	0.933	0.938
	6	0.991	0.909	0.944	0.994
	12	0.865	0.409	0.880	0.918

Dir.	Hor.	AR	QuickNet	MBE	Autom	QN-SG	NP	NC
	1		0.950	0.808	0.992	0.915	0.790	0.690
AR	3	0.952	0.997	0.483	0.969	0.955	0.851	0.531
	6	0.995	0.947	0.943	1	0.985	0.997	0.500
	12	0.987	1	1	1	1	0.679	0

Table 5: p-values of the Wilcoxon signed-rank test for testing the hypothesis that the mean of the differences of absolute forecast errors of industrial production for all countries from the linear recursive AR model and another model are equal against the alternative that the other model has smaller absolute forecast errors. p-values subtracted from one are those of the test in which the null and the alternative hypothesis change places

one month ahead are not independent, which violates the assumptions of the test. Another caveat is that the average loss size varies from one country to another, which may also affect the outcomes. Nevertheless, we believe that the p-values are still indicative of the results and report them.

Table 5 shows that the only two methods for which $p < 0.95$ for all horizons are the recursive MBE and 'no change'. There is only one p-value below 0.4, that of 'no change' 12 months ahead, so on average it does not seem possible to find from our set of models one that would generate more accurate forecasts than the recursive linear AR model. It is seen that all direct methods except 'no change' underperform at the two longest horizons, which cannot be seen from Table 3. The nonparametric model in forecasting 12 months ahead constitiues an exception. This may be the case because the growth rate of many series turns from negative to positive during the forecasting period. Most models cannot forecast that turn, but the no change 'method' by construction adapts quickly.

Table 6 contains the average numbers of variables or hidden units selected by each method. For recursive forecasting models, MBE yields the most parsimonious models, followed by Autometrics. The difference between the two variants of QuickNet is small but again, the results may change if the significance levels in QN-SG are altered. Most of the variables selected are hidden units, although this does not necessarily lead to forecasts that would be more accurate than the ones from the linear AR model. Lags of Δx_t or $\Delta_h x_t$ get selected more often by MBE than by the other automatic methods. Although the general tendency becomes clear from Table 6, the results are not similar for all countries.

The situation is different when direct forecasting is concerned. Table 6 only shows the numbers of variables for the 12-months-ahead models, because that already shows the general tendencies. In fact, the differences in these numbers between the three-month and the 12-month models are usually quite

Type of forecast	Recursive			Direct 12		
Number of variables	Total	Lags	HU	Total	Lags	HU
QuickNet	7.13	0.06	7.07	6.19	0.26	5.93
QN-SG	6.92	0.06	6.85	5.79	0.25	5.54
MBE	3.61	0.17	3.44	2.16	0.28	1.88
Autometrics	4.66	0.03	4.63	10.8	0.12	10.7

Table 6: The average number of linear lags and hidden units (HU) selected for ANN models for recursive, and direct 12-months ahead, forecasting of industrial production of the 11 countries from July 2007 onwards

small, although on the country level exceptions do occur. It is seen that for MBE the average number of variables in the 12-month models is smaller than that of the one-month model used for recursive forecasting. In fact, although this is not shown, the decrease is a monotonic function of the forecasting horizon. Perhaps the most striking feature is the strong increase in the number of variables selected by Autometrics. It is there already for three-month models (the number of selected variables then equals 12.3). As there is a gap in direct models between the dependent variable and the available lags, there may not be a well-fitting model to choose. Because Autometrics subjects the final model to diagnostic tests and backtracks along the search path as long as these tests reject the null hypothesis, this strategy may mean that the model that finally passes the tests contains a large number of variables.

6.2.2 Japan

The summary tables hide the fact that there is plenty of variation across individual countries. The accuracy of the forecasts varies from one country to the next, and no automatic method for generating forecasts dominates the others. For this reason, we present results for two individual countries, Japan and the UK. We also briefly touch upon some Norwegian outcomes. The remaining country-specific RMSFE results are available at `http://econ.au.dk/research/research-centres/creates/research/research-papers/supplementary-downloads/`.

It is seen from Figure 1 that the Japanese industrial production fluctuates heavily during the forecasting period. Consequently, it makes a good test case, also because the linear AR forecasts of the Japanese series are on average the least accurate among the 11 countries considered. The first four figures in the 'no filter' (NF) panel of the 'AR' column of Table 7 are the RMSFEs of the recursive forecasts for the Japanese industrial production 1, 3, 6 and 12 months ahead. Note that the RMSFE for the 12-months-ahead forecasts is a remarkable 26%. An interesting fact is that no filter is applied to any of the forecasts, except the linear AR ones. The ANN models yield more accurate one-month forecasts than the linear AR model. QuickNet has the best performance one and three months ahead, whereas MBE yields the most accurate nonlinear 12-month forecasts. If the models are not respecified during the forecasting period, the accuracy of the six- and 12-month forecasts increases for all three methods. These results are not reproduced here but, as an example, the RMSFE of the 12-month forecasts for MBE without filtering, which does not improve the forecasts, equals 0.688.

The RMSFEs of the direct forecasts can be found in Table 8. Direct

Filter	Hor.	AR	QuickNet	MBE	Autom.	QN-SG
	1	0.02232	0.7503	0.9259	0.8519	0.7322
NF	3	0.06517	0.9193	1.023	0.9965	0.9138
	6	0.1392	1.263	1.024	1.295	1.258
	12	0.2644	1.355	0.9791	1.355	1.353
	1	0.7753	0.7503	0.9259	0.8519	0.7322
SW	3	0.9194	0.9193	1.008	0.9965	0.9138
	6	1.010	1.263	1.023	1.295	1.258
	12	1.013	1.355	0.9818	1.355	1.353
	1	1	0.7503	0.9259	0.8519	0.7322
AR	3	1	0.9193	1.008	0.9965	0.9138
	6	1	1.263	1.023	1.295	1.258
	12	1	1.355	0.9818	1.355	1.353

Table 7: Root mean square forecast errors of the recursive forecasts for the Japanese industrial production series. Models respecified after each period. Forecasting begins July 2007

Filter	Hor.	AR	QuickNet	MBE	Autom.	QN-SG	NP	NC
	1	1	0.7503	0.9259	0.8519	0.7322	1.065	1.236
NF	3	1.009	0.8869	1.060	0.8044	0.8893	1.059	1.255
	6	0.8721	0.8488	0.8211	711.4	0.8375	0.9077	0.9435
	12	0.7044	0.7512	0.7356	0.7505	0.7533	0.7508	0.6863
	1	0.7753	0.7503	0.9259	0.8519	0.7322	0.8389	
SW	3	1.009	0.8869	1.060	0.8044	0.8893	0.9664	
	6	0.8721	0.8488	0.8211	0.9051	0.8375	0.9077	
	12	0.7044	0.7782	0.7632	0.7505	0.7803	0.7657	
	1	1	0.7503	0.9259	0.8519	0.7322	1.050	
AR	3	1.009	0.8869	1.060	0.8044	0.8893	1.112	
	6	0.8721	0.8488	0.8211	0.8788	0.8375	0.9077	
	12	0.7044	0.7518	0.7362	0.7505	0.7539	0.7419	

Table 8: Root mean square forecast errors of the direct forecasts for the Japanese industrial production series. Models respecified after each period. Forecasting begins July 2007

Rec.	Hor.	QuickNet	MBE	Autom	QN-SG		
	1	0.748	0.893	0.873	0.663		
AR	3	0.375	0.969	0.841	0.300		
	6	0.693	0.955	0.924	0.648		
	12	0.761	0.617	0.831	0.601		

Dir.	Hor.	AR	QuickNet	MBE	Autom	QN-SG	NP	NC
	1		0.748	0.893	0.873	0.663	0.848	0.945
AR	3	0.259	0.439	0.641	0.399	0.553	0.924	0.962
	6	0.656	0.767	0.0194	0.986	0.678	0.601	0.937
	12	0.780	0.998	0.998	0.991	0.999	0.889	0.553

Table 9: p-values of the Wilcoxon signed-rank test for testing the hypothesis that the mean of the differences of absolute forecast errors of Japanese industrial production from the linear recursive AR model and another model equals zero against the alternative that the other model has smaller absolute forecast errors. p-values subtracted from one are those of the test in which the null and the alternative hypothesis change places

three-, six- and 12-month forecasts appear more accurate than their recursive counterparts whose RMSFEs were reported in Table 7. This is also true for forecasts from the linear AR model. The three-month MBE forecasts are an exception, however. Three-month out-of-bounds Autometrics forecasts are effectively corrected by the SW filter.

Interestingly, the nonparametric model is now inferior to the other direct methods. The only exception is the 'no change' for horizons less than 12 months. (As already mentioned, Autometrics generates 'insane' forecasts at the three-month horizon, but filtering corrects them.) The SW filter, however, considerably improves the RMSFE of the nonparametric forecasts at the two shortest horizons. The 'no change' forecasts are the least accurate ones for these horizons, however, for 12 months ahead they have the smallest RMSFE of all forecasts, both recursive and direct.

Results of the Wilcoxon signed-rank test in Table 9 are similar to the general case in that it is hard to improve upon the linear recursive AR model. Only the direct six-month MBE forecasts ($p = 0.019$) seem to compare favourably with them. But then, the null hypothesis is not rejected either way for the two recursive QuickNet forecasts, the recursive Autometrics ones, the direct linear AR forecasts or the nonparametric ones, if p-values 0.05 and 0.95 are used as indicators. All direct forecasts from ANN models fail when the forecast horizon is 12 months, which strongly contrasts the picture given by Table 8. In that table, the ratios for all ANN-based forecasts

Type of forecast	Recursive			Direct 12		
Number of variables	Total	Lags	HU	Total	Lags	HU
QuickNet	8.47	0	8.47	5.40	0	5.40
QN-SG	8.20	0	8.20	3.70	0	3.70
MBE	3.43	0.03	3.40	1.90	0	1.90
Autometrics	2.53	0	2.53	21.9	0.13	21.8

Table 10: The average number of linear lags and hidden units (HU) selected for ANN models for recursive, and direct 12-months ahead, forecasting of the Japanese industrial production from July 2007 onwards

lie clearly below unity. The reason for this contradiction is that the errors of a couple of very last recursive forecasts from the linear AR model are remarkably large and have a considerable impact on the corresponding RMSFE and ratios in the table. The robust test de-emphasises this effect and in fact indicates that the recursive nonlinear forecasts are a better alternative than the direct ones when the recursive linear AR forecasts form the benchmark. These large errors also partly explain the large (26%) 12-month RMSFE in Table 7.

The average numbers of variables or hidden units selected by the four methods for models generating the recursive forecasts of the Japanese industrial production can be found in Table 10. The QuickNet is by far the least parsimonious technique with over eight hidden units selected on the average. There is little difference between the pure specific-to-general QN-SG and the original QuickNet. Autometrics is slightly more parsimonious than MBE. It may be noted that in this case, only MBE selects a linear lag or lags of Δx_t, and this happens very rarely.

The direct 12-month models tell a different story which is, however, in line with what can be concluded from Table 6. Autometrics now selects a very large number of hidden units, whereas for the other methods, the numbers are considerably smaller than in the case of the one-month models. This may be because the Japanese industrial production has been difficult to forecast and deprived of the most important lags, Autometrics struggles to find a model that would pass the diagnostic tests. MBE on the other hand, finds most of the heavily lagged hidden units irrelevant and excludes them.

These results may be contrasted with ones for Norway. The Norwegian series was easier to forecast than the Japanese one, since the industrial production of that country only decreased modestly during the crisis. This is reflected in the RMSFE of the linear AR model 12-month forecasts, which equals 0.0298, close to one tenth of the Japanese one. Table 11 contains the average variable selection numbers for the ANN models of the Norwegian industrial production. The results correlate with the RMSFEs in that the most parsimonious models, generated by MBE, yield slightly more accurate forecasts than the other. Furthermore, Autometrics that chooses the least parsimonious models also has the highest RMSFE ratio even after filtering (there are explosive recursive six- and 12-month forecasts among the unfiltered ones). However, the forecast accuracy of none of the four methods is superior to that of the linear AR model.

The result for the direct 12-month models are different in that Autometrics on the average now selects rather few variables (only 0.33 on average for the six-month models). This may have something to do with the fact that the Norwegian industrial production displays less variation during the forecasting period than the Japanese one. (The direct Autometrics forecasts for the Norwegian industrial production do not need any filtering but are, however, less accurate than forecasts from the linear AR model.) As before, MBE selects the smallest number of variables, and the share of linear lags is greater than observed elsewhere. Surprisingly, in this case QuickNet selects a relatively large number of variables.

6.2.3 United Kingdom

We shall consider the results for the UK here to in order to have an example in which many models and alternatives, the direct ones in particular, perform somewhat better than the linear AR model. The RMSFEs for the recursive forecasts can be found in Table 12. The RMSFE of the forecasts from the linear AR model equals 0.078, which is moderate in comparison to the other series.

Some recursive forecast errors from both QuickNet and Autometrics are very large, but filtering changes the situation. Note that the one-way QN-SG

Type of forecast	Recursive			Direct 12		
Number of variables	Total	Lags	HU	Total	Lags	HU
QuickNet	3.63	0	3.63	9.83	0.83	9.00
QN-SG	6.20	0	6.20	8.00	0.83	7.17
MBE	4.26	0.03	4.23	1.27	0.83	0.43
Autometrics	6.83	0.03	6.80	3.97	0.03	3.93

Table 11: The average number of linear lags and hidden units (HU) selected for ANN models for recursive, and direct 12-months ahead, forecasting of the Norwegian industrial production from July 2007 onwards

Filter	Hor.	AR	QuickNet	MBE	Autom.	QN-SG
	1	0.01317	1.625	0.936	301.7	1.014
NF	3	0.0267	171.9	1.011	6228	1.053
	6	0.04575	8871	1.003	3×10^7	0.9518
	12	0.07799	2×10^4	1.001	6×10^8	1.041
	1	1	0.9949	0.936	0.7121	1.014
SW	3	1	1.084	1.011	0.7863	1.029
	6	1	0.9551	1.003	1.053	0.9451
	12	1	1.007	1.001	1.235	1.035
	1	1	0.9934	0.936	0.9444	1.014
AR	3	1	1.080	1.011	0.9759	1.029
	6	1	0.9556	1.003	1.001	0.9451
	12	1	1.006	1.001	0.9863	1.035

Table 12: Root mean square forecast errors of the recursive forecasts for the UK industrial production series. Models respecified after each period. Forecasting begins July 2007

does not yield such forecasts. The SW filter makes Autometrics a superior performer at short horizons. MBE forecasts do not need filtering and are at longer horizons comparable in quality to the linear AR forecasts.

The RMSFEs of direct forecasts can be found in Table 13. As for Autometrics, direct forecasts, after SW filtering, are slightly more accurate than the recursive ones. Some improvement is also apparent for MBE and the linear AR forecasts. Filtered MBE forecasts are also more accurate than their recursive counterparts. Some of the direct six-month forecasts are out-of-bounds, and this is true for all four methods. Filtering corrects the problem. The nonparametric model generates more accurate forecasts than the linear recursive AR one, and filtering is almost never applied. Note, however, that the direct linear AR forecasts are also more accurate than the corresponding recursive ones. The 'no change' alternative is superior to the recursive linear AR model, but in this case it is not the best method for forecasting 12 months ahead.

The results of the Wilcoxon test appear in Table 14. Direct forecasts are more accurate than they seem in the summary table (Table 5). Direct linear AR forecasts are more accurate than the recursive linear ones, as are direct Autometrics and nonparametric forecasts at longest horizons. It is also seen that, due to the exceptional period to be forecast, the results vary from one horizon to the other. According to the test, all direct forecasts are relatively accurate at the six-month horizon, whereas QuickNet fails when

Filter	Hor.	AR	QuickNet	MBE	Autom.	QN-SG	NP	NC
	1	1	1.625	0.936	301.7	1.014	0.8867	0.8909
NF	3	0.8926	0.9696	0.6225	0.7366	0.7609	0.8727	0.9447
	6	0.9305	8203	2433	2.198	2305	0.8603	0.9588
	12	0.9854	1.043	0.9854	0.9666	1.024	0.9312	0.9743
	1	1	0.9949	0.936	0.7121	1.014	0.8867	
SW	3	0.8926	0.9696	0.6225	0.7366	0.7609	0.8727	
	6	0.8635	0.7829	0.8764	0.8835	0.8537	0.8603	
	12	0.9763	1.009	0.9854	0.9666	1.000	0.9318	
	1	1	0.9934	0.936	0.9444	1.014	0.8867	
AR	3	0.8926	0.9696	0.6225	0.7366	0.7609	0.8727	
	6	0.9305	0.9309	0.8961	0.9031	0.8437	0.8603	
	12	0.9854	1.031	0.9854	0.9666	1.022	0.9438	

Table 13: Root mean square forecast errors of the direct forecasts for the UK industrial production series. Models respecified after each period. Forecasting begins July 2007

the horizon is 12 months. All p-values for recursive forecasts are below 0.95, so one cannot reject the null hypothesis in favour of the recursive linear AR model. But then, one p-value for recursive Autometrics forecasts is less than 0.05, suggesting a rejection in the opposite rejection.

The average numbers of selected variables can be found in Table 15. MBE is again the most parsimonious choice for models for recursive forecasting with 3.3 variables per equation. QuickNet and Autometrics select largest models on average, more than eight variables per model. There is now no clear correlation, positive or negative, between the size of the model and the accuracy of the forecasts.

Among the models for generating direct forecasts, the MBE-selected ones show the same tendency as before: the number of variables in 12-month models is less than in one-month ones. The largest deviation from the general results is that for 12-month models. Remarkably, there is not a single occasion in which Autometrics would choose even one variable. The 12-month forecasts from Autometrics thus consist of means of the annual difference of the (log) industrial production. It is seen from Tables 13 and 14 that these forecasts do well in comparison. Only the forecasts from the nonparametric model have a smaller RMSFE, and the null hypothesis of the Wilcoxon test is rejected in favour of Autometrics at the 5% level (but remember caution in interpreting these p-values). In comparison, it may be mentioned that the average number of variables in the three-month Autometrics-based

Rec.	Hor.	QuickNet	MBE	Autom	QN-SG
	1	0.573	0.322	0.608	0.617
AR	3	0.803	0.877	0.0859	0.455
	6	0.786	0.921	0.375	0.877
	12	0.803	0.864	0.0469	0.700

Dir.	Hor.	AR	QuickNet	MBE	Autom	QN-SG	NP	NC
	1		0.573	0.322	0.608	0.617	0.728	0.831
AR	3	0.0214	0.728	0.131	0.149	0.149	0.111	0.159
	6	0.00219	0.0393	0.0449	0.068	0.0627	0.068	0.0958
	12	0.0184	0.994	0.577	0.0376	0.990	0.0393	0.103

Table 14: p-values of the Wilcoxon signed-rank test for testing the hypothesis that the mean of the differences of absolute forecast errors of the UK industrial production from the linear recursive AR model and another model equals zero against the alternative that the other model has smaller absolute forecast errors. p-values subtracted from one are those of the test in which the null and the alternative hypothesis change places

Type of forecast	Recursive			Direct 12		
Number of variables	Total	Lags	HU	Total	Lags	HU
QuickNet	8.83	0	8.83	6.60	0	6.60
QN-SG	6.47	0	6.47	6.77	0	6.77
MBE	3.30	0.17	3.13	2.50	0	2.50
Autometrics	8.13	0.07	8.07	0	0	0

Table 15: The average number of linear lags and hidden units (HU) selected for ANN models for recursive, and direct 12-months ahead, forecasting of the UK industrial production from July 2007 onwards

models equals 17.2. According to Table 13, forecasts from them are quite competitive as well, less accurate than the MBE forecasts (the average size of the corresponding model is 3.0 variables) but better than the other direct alternatives.

6.2.4 Summary

Since it is not possible to discuss all empirical results in detail, we shall briefly list results common to most series and mention aspects in which the countries differ from each other. The most conspicuous common result is that the accuracy of the 'no change' 12-month forecasts, measured in RMSFEs, is superior to that of all other forecasts in six countries and better than the linear AR in all but one. This country is Norway for which, as already mentioned, the RMSFE of the linear AR model 12 months ahead is lowest of all. But then, the 'no change' forecast is not usually among the most accurate ones at shortest (one- and three-month) horizons. These results are reflected in the NC column of Table 1.

Another common outcome is that the nonparametric model most often has a lower RMSFE than the linear AR model. This is not always case for all forecasting horizons, but in four cases out of 11, it has the lowest RMSFE for 12-month forecasts.

Third, the sets of forecasts for many countries contain insane recursive forecasts. This occurs in seven countries out of 11. Only two countries, Denmark and Finland, are completely free of them. Even the direct method generates occasional insane forecasts, due to highly correlated variables in the selected models.

Fourth, filtering is useful for a majority of the series. It does not only adjust the insane forecasts but in many cases also improves the accuracy overall and leads to an RMSFE less than that of the forecasts from the linear AR model. It seems that QuickNet and Autometrics profit more from filtering than MBE.

Fifth, we also considered forecasts in a situation in which the model was not respecified after July 2007. In a majority of cases, forecasts from this model are inferior to the ones obtained from respecified and re-estimated models. The only distinctly different case is the Japanese industrial production when the forecast horizon is at least six months. Then all automatic techniques profited from the original model not being respecified.

Sixth, we tested linearity of the series before modelling and only employed the ANN model when the test rejected. The effect of this pre-testing was mixed, producing minor improvements in some cases but also making the accuracy of the forecasts worse in others. The gains, when they occurred, were not systematic in the sense that sometimes only a subset of techniques were positively helped by this pre-screening.

Seventh, as is already obvious from Table 6, MBE selects the smallest models. It appears that there is no clear correlation between the size of the model and the forecasting accuracy. Autometrics-built direct forecasting models tend to contain the largest number of variables, although exceptions to this rule do exist.

The country that differs most from the rest is the United States. For the US, all ANN models generate insane recursive forecasts, and the recursive linear AR ones are usually more accurate than their ANN counterparts. The recursive QuickNet 12 months ahead forecasts constitute the only exception and, if the model is not respecified after June 2007, MBE forecasts for the

Filter	Hor.	AR	QuickNet	MBE	Autom.	QN-SG
	1	1	3×10^5	1.022	1.254	1.107
NF	3	1	1×10^5	1.010	1.251	1.048
	6	1	7×10^6	1.008	1.343	1.027
	12	1	1×10^7	1.018	1.865	1.005
	1	1.028	1.116	1.044	1.115	1.133
SW	3	0.9923	1.041	0.9945	1.076	1.037
	6	0.9931	1.015	0.9867	1.059	1.018
	12	0.9962	1.007	0.9926	1.031	1.002
	1	1	1.077	1.023	1.100	1.107
AR	3	1	1.044	1.002	1.077	1.045
	6	1	1.024	0.9928	1.061	1.023
	12	1	1.014	0.9932	1.032	1.002

Table 16: Root mean square forecast error ratios of the recursive forecasts for the 11 unemployment rate series. Models respecified after each period. Forecasting begins July 2007

same horizon form another one. The 'no change' forecasts of the US industrial production are inaccurate, and the nonparametric forecasts are better than their linear AR counterparts only at the 12-month horizon. The SW filter considerably improves direct Autometrics three- and six-month forecasts.

6.3 Unemployment rate

6.3.1 Overview

It is seen from Figure 2 that unemployment rates have not fluctuated as strongly during the crisis period as the industrial production. In most countries, the economic recovery did not yet show in the unemployment rates at the end of 2009. The results of recursive forecasts appear in Table 16. They show that the linear AR model yields the most accurate forecasts overall. After filtering, however, MBE is fully competitive with the linear model at the three-month and longer forecasting horizons. Kock and Teräsvirta (2011b) also found that MBE-selected ANN models generated the most accurate forecasts which, however, on average did not improve upon what was obtained from the linear AR model.

As in the case of industrial production, QuickNet-based ANN models generate a number of insane forecasts. But then, the specific-to-general version QN-SG does not do that. It appears that allowing QuickNet to first select ten variables and then ask it to reduce the size of the model sometimes leads

Filter	Hor.	AR	QuickNet	MBE	Autom.	QN-SG	NP	NC
	1	1	3×10^5	1.022	1.254	1.107	1.052	1.225
NF	3	1.010	1.248	1.083	3×10^4	1.149	1.109	1.259
	6	1.017	1.110	1.072	1×10^5	1.109	1.087	1.158
	12	1.023	2×10^4	1.046	7530	0.9855	1.069	1.012
	1	1.028	1.116	1.044	1.115	1.133	1.075	
SW	3	1.033	1.091	1.090	1.216	1.117	1.109	
	6	1.015	1.078	1.070	1.100	1.077	1.063	
	12	1.023	0.9842	1.046	1.031	0.9775	1.067	
	1	1	1.077	1.023	1.100	1.107	1.075	
AR	3	1.010	1.103	1.082	1.230	1.117	1.109	
	6	1.017	1.090	1.070	1.130	1.088	1.070	
	12	1.023	0.9927	1.046	1.047	0.9787	1.067	

Table 17: Root mean square forecast errors of the direct forecasts for the 11 unemployment rate series. Models respecified after each period. Forecasting begins July 2007

to overfitting. After filtering the differences in the RMSFE between these two versions of QuickNet practically disappear.

Autometrics also generates some forecasts that have to be filtered. A comparison of filtered forecasts suggests that on average, this method yields the least accurate recursive unemployment rate forecasts. As may be expected, this result does not hold for all countries in the sample.

The main effect of not respecifying the model during the forecasting period is that MBE is less accurate than it is when updating the model takes place. (These results are not reported in detail here.) Forecasts from the Autometrics-based models without filtering improve, and filtering has a minimal effect on them. They still remain slightly less accurate than forecasts obtained by the other selection techniques.

The RMSFE ratios of direct forecasts can be found in Table 17. The MBE forecasts are now clearly inferior to the recursive ones. The same is true for Autometrics and even the linear AR model. The only gain can be found in QuickNet and QN-SG 12-months-ahead forecasts, for which the RMSFE ratio is slightly below unity. The nonparametric forecasts perform less well than in the case of industrial production and have, after filtering, the worst 12-months-ahead performance. The 'no change' forecast is not a good choice either. The reason can be seen from Figure 2. Since the unemployment rates mostly increase during the forecasting period and no recoil takes place, this alternative works less well than it does in forecasting industrial production

Rec	Hor.	AR	QuickNet	MBE	Autom.	QN-SG		
	1	3.61	3.91	3.53	3.88	3.92		
SW	3	6.06	6.43	5.89	6.43	6.46		
	6	5.78	6.18	5.65	6.45	6.08		
	12	6.03	6.41	5.85	6.67	6.19		
Dir	Hor.	AR	QuickNet	MBE	Autom.	QN-SG	NP	NC
	1	3.61	3.91	3.53	3.88	3.92	3.93	4.80
SW	3	6.02	6.23	6.23	6.89	6.48	6.51	7.84
	6	5.78	6.68	6.09	6.92	6.55	6.47	8.95
	12	6.49	6.35	7.11	6.66	6.28	7.19	6.32

Table 18: The average ranks of the methods of forecasting the unemployment rate (all countries). Forecasting begins July 2007

during the crisis.

The average ranks of different methods computed as the ones for the industrial production forecasts can be found in Table 18. They are quite different from those in Table 4. Recursive MBE forecasts now have the lowest average rank for all four horizons, followed by the recursive linear AR forecasts. The direct AR forecasts perform equally well as the recursive ones except for the longest forecasting horizon. The 'no change' forecasts now have the highest average rank at the three shortes horizons and the nonparametric forecasts at the 12-month horizon. The direct Autometrics and MBE 12-month forecasts have the highest ranks of all ANN-forecasts. As far as Autometrics-based forecasts are concerned, the reason for this is probably the same as in the industrial production case: difficulties in finding a satisfactory approximation to the data-generating process when the most relevant lags are not allowed to enter the model.

From Table 19 it can be concluded that on average recursive forecasts are more accurate than the direct ones, when the yardstick is the Wilcoxon signed-rank test. Direct linear AR forecasts are the only exception, whereas the other direct models, including 'no change', perform less well than the recursive linear AR model. The other method for which all p-values are less than 0.95 is the recursive MBE. At the 12-month horizon the p-value of the test is as low as 0.054. Autometrics is the worst performer, which accords with the results in Tables 16 and 17.

The number of variables selected by different methods appear in Table 20. MBE and Autometrics select the fewest variables for models of recursive forecasting. QuickNet that generated many insane forecasts has the highest average number selected. For direct models, the situation changes radically.

Rec.	Hor.	QuickNet	MBE	Autom	QN-SG
	1	0.994	0.830	0.997	1
AR	3	0.887	0.307	0.999	0.983
	6	0.802	0.248	1	0.854
	12	0.684	0.0543	1	0.459

Dir.	Hor.	AR	QuickNet	MBE	Autom	QN-SG	NP	NC
	1		0.994	0.830	0.997	1	0.976	1
AR	3	0.311	0.985	0.990	1	0.987	0.997	1
	6	0.838	1	0.988	1	0.999	1	0.999
	12	0.604	0.860	0.985	0.977	0.700	1	0.100

Table 19: p-values of the Wilcoxon signed-rank test for testing the hypothesis that the mean of the differences of absolute forecast errors of unemployment rate for all countries from the linear recursive AR model and another model are equal against the alternative that the other model has smaller absolute forecast errors. p-values subtracted from one are those of the test in which the null and the alternative hypothesis change places

Type of forecast	Recursive			Direct 12		
Number of variables	Total	Lags	HU	Total	Lags	HU
QuickNet	7.08	0.18	6.90	6.27	0.01	6.26
QN-SG	5.02	0.18	4.84	6.06	0.01	6.05
MBE	2.88	0.24	2.64	1.47	0.28	1.46
Autometrics	2.91	0.09	2.82	15.8	0.26	15.5

Table 20: The average number of linear lags and hidden units (HU) selected for ANN models for recursive, and direct 12-months ahead, forecasting of unemployment rate of the 11 countries

Filter	Hor.	AR	QuickNet	MBE	Autom.	QN-SG
	1	0.1648	0.9341	1.007	1.129	0.9461
NF	3	0.3614	11.5	0.9447	1.061	0.7723
	6	0.8164	9.462	0.9378	1.058	0.8042
	12	1.941	9.082	0.9769	1.014	0.9282
	1	1	0.9341	1.007	1.129	0.9461
SW	3	1	0.7660	0.9447	1.061	0.7723
	6	1	0.8113	0.9369	1.049	0.8036
	12	1	0.9198	0.9751	1.011	0.9270
	1	1	0.9341	1.007	1.129	0.9461
AR	3	1	0.7660	0.9447	1.061	0.7723
	6	1	0.8113	0.9369	1.049	0.8036
	12	1	0.9198	0.9751	1.011	0.9270

Table 21: Root mean square forecast error ratios of the recursive forecasts for the US unemployment rate series. Models respecified after each period. Forecasting begins July 2007

The number of variables selected by MBE decreases whereas the same number for Autometrics strongly increases. A similar pattern was already found for industrial production models. When a number of important lags are removed from consideration, Autometrics has problems in finding a model that satisfies the diagnostic criteria. On average it selects 19.0 variables for three-month and 19.6 variables for six-month ANN models. For QuickNet, the differences in the average model size between recursive (one-month) and direct forecasting models are relatively small.

6.3.2 United States

Similarly to the industrial production case, we consider in detail the country (the United States) with the largest 12-month RMSFE for recursive linear AR forecasts. The RMSFE ratios of recursive forecasts for the US appear in Table 21. As in the summary table, models chosen by QuickNet generate some infeasible forecasts. This is avoided by QN-SG, but after filtering the results for the two variants of QuickNet are quite close to each other and give the best results. MBE-based ANN models also yield more accurate forecasts than the linear AR model, and filtering is hardly used. The Autometrics forecasts are not filtered either, but they are less accurate than even the linear AR forecasts.

If the models are not respecified during the forecasting period (detailed results not given here), the situation changes. MBE becomes the worst alter-

Filter	Hor.	AR	QuickNet	MBE	Autom.	QN-SG	NP	NC
	1	1	0.9341	1.007	1.129	0.9461	1.065	1.301
NF	3	1.016	1.453	1.155	1.504	1.322	1.266	1.575
	6	1.056	1.319	1.176	1.390	1.339	1.355	1.473
	12	1.045	0.9944	1.066	1.018	0.9953	1.173	1.224
	1	1	0.9341	1.007	1.129	0.9461	1.065	
SW	3	1.016	0.9598	1.155	0.9683	0.8892	1.266	
	6	1.056	1.138	1.176	1.028	1.184	1.095	
	12	1.045	0.9594	1.066	0.9931	0.9608	1.151	
	1	1	0.9341	1.007	1.129	0.9461	1.065	
AR	3	1.016	1.039	1.155	1.087	0.974	1.266	
	6	1.056	1.195	1.176	1.142	1.226	1.171	
	12	1.045	0.9711	1.066	0.9992	0.9725	1.156	

Table 22: Root mean square forecast errors of the direct forecasts for the US unemployment rate series. Models respecified after each period. Forecasting begins July 2007

native, whereas Autometrics generates more accurate results than the linear AR model for six-and 12-month horizons. QuickNet also gains from respecification, although, like Autometrics, it generates more accurate forecasts than the linear AR model for the two longest horizons when the same model is used for the whole forecasting period.

The results from direct forecasting models are reported in Table 22. They follow the general tendency in that direct linear AR forecasts are less accurate than the recursive ones. Furthermore, neither the nonparametric model nor the 'no change' forecast are useful alternatives. QuickNet and Autometrics (after filtering) perform best, but the recursive forecasts from the QuickNet-based ANN models are more accurate than the direct ones. The same is true for MBE. The overall conclusion is that for this series, recursive forecasts are clearly superior to the direct ones, only Autometrics constitutes an exception.

Table 23 contains the results of the Wilcoxon signed-rank test. All recursive methods except Autometrics perform reasonably well. Compared to their recursive counterparts, direct linear AR forecasts are also acceptable. It seems that forecasting six months ahead using the direct method is particularly difficult. In accord with the results in Tables 21 and 22, forecasts from the nonparametric model and the 'no change' forecast are clearly inferior to recursive linear AR ones.

The average numbers of variables selected for ANN models of the US unemployment series follow the general pattern evident in Table 20. It is

Rec.	Hor.	QuickNet	MBE	Autom	QN-SG
	1	0.648	0.908	0.986	0.792
AR	3	0.0627	0.322	0.798	0.149
	6	0.169	0.259	0.932	0.314
	12	0.383	0.0602	0.886	0.537

Dir.	Hor.	AR	QuickNet	MBE	Autom	QN-SG	NP	NC
	1		0.648	0.908	0.986	0.792	0.882	0.996
AR	3	0.233	0.488	0.897	0.601	0.314	0.973	0.999
	6	0.641	0.999	0.921	0.969	0.998	1	1
	12	0.693	0.214	0.352	0.488	0.191	1	0.998

Table 23: p-values of the Wilcoxon signed-rank test for testing the hypothesis that the mean of the differences of absolute forecast errors of the US unemployment rate from the linear recursive AR model and another model equal zero against the alternative that the other model has smaller absolute forecast errors. p-values subtracted from one are those of the test in which the null and the alternative hypothesis change places

Type of forecast	Recursive			Direct 12		
Number of variables	Total	Lags	HU	Total	Lags	HU
QuickNet	6.73	0	6.73	8.67	0	8.67
QN-SG	6.30	0	6.30	7.53	0	7.53
MBE	3.93	0	3.93	1.40	0	1.40
Autometrics	3.20	0	3.20	21.5	0.27	21.2

Table 24: The average number of linear lags and hidden units (HU) selected for ANN models for recursive, and direct 12-months ahead, forecasting of the US unemployment rate

Filter	Hor.	AR	QuickNet	MBE	Autom.	QN-SG
	1	0.09843	1.055	1.022	1.123	1.048
NF	3	0.1664	10.88	1.013	1.205	1.064
	6	0.2798	94.17	0.9782	1.136	1.024
	12	0.5117	226	0.9324	0.9876	0.9511
	1	1	1.055	1.022	1.123	1.048
SW	3	1	1.218	1.013	1.205	1.064
	6	1	1.074	0.9782	1.136	1.024
	12	1	0.9938	0.9324	0.9876	0.9511
	1	1	1.055	1.022	1.123	1.048
AR	3	1	1.218	1.013	1.205	1.064
	6	1	1.074	0.9782	1.136	1.024
	12	1	0.9938	0.9324	0.9876	0.9511

Table 25: Root mean square forecast error ratios of the recursive forecasts for the Norwegian unemployment rate series. Models respecified after each period. Forecasting begins July 2007

seen from Table 24 that Autometrics and MBE are the most parsimonious methods for recursive models. Likewise, MBE selects very few variables for direct 12-month models, whereas Autometrics does exactly the opposite. It is the only method that occasionally selects a linear lag, and this happens in constructing 12-month models. The numbers for QuickNet lie between these two extremes.

6.3.3 Norway

The Norwegian unemployment rate has not been much affected by the crisis, and the RMSFE for a linear 12-month forecast equals only 0.520. Table 25 shows that for recursive forecasts, the linear AR model is the most accurate for the shortest horizons but is inferior to all ANN forecasts at the 12-month horizon. QuickNet generates some out-of-bounds forecasts, whereas the other three methods do not need filtering. MBE is the most accurate one of them. If the model is not respecified during the forecasting period, Autometrics is the best performer in forecasting one and 12 months ahead (results not reported here in detail). Linearity testing has a positive effect on the forecasts from MBE and QN-SG but a negative one on Autometrics-based forecasts.

Results on direct forecasts can be found in Table 26. They indicate that for short horizons, the recursive forecasts from the linear AR model are superior to the other forecasts. In 12-month forecasts the situation changes quite drastically, and all the methods, except 'no change' and Autometrics

Filter	Hor.	AR	QuickNet	MBE	Autom.	QN-SG	NP	NC
	1	1	1.055	1.022	1.123	1.048	1.000	1.066
NF	3	0.9837	1.162	1.022	1.374	1.129	1.046	1.237
	6	1.011	1.015	1.019	1.149	1.060	0.9247	1.192
	12	0.9538	0.7795	0.8881	1.048	0.7058	0.9519	1.011
	1	1	1.055	1.022	1.123	1.048	1.000	
SW	3	0.9837	1.162	1.022	1.374	1.129	1.046	
	6	1.011	1.015	1.019	1.149	1.060	0.9247	
	12	0.9538	0.7795	0.8881	1.048	0.7058	0.9519	
	1	1	1.055	1.022	1.123	1.048	1	
AR	3	0.9837	1.162	1.022	1.374	1.129	1.046	
	6	1.011	1.015	1.019	1.149	1.060	0.9247	
	12	0.9538	0.7795	0.8881	1.048	0.7058	0.9519	

Table 26: Root mean square forecast errors of the direct forecasts for the Norwegian unemployment rate series. Models respecified after each period. Forecasting begins July 2007

yield models that forecast better than the linear model. The two QuickNet methods are the most accurate ones, with QN-SG having the smallest RMSFE ratio. The direct 12-month forecasts are even more accurate than the corresponding recursive ones. No forecasts are being filtered.

Results of the Wilcoxon signed-rank test in Table 27 deviate from the general trend in that many direct models generate relatively acceptable forecasts. They include both QuickNet methods, MBE and the nonparametric model. This is true in particular when the forecasting horizon is 12 months, QN-SG being ahead (the p-value equals 0.0225). Recursive MBE is also a good performer: the lowest p-value (at the 12-month horizon) equals 0.053. Autometrics does not do well in this test if the 12-month horizon is omitted from comparison.

The average number of variables selected for Norwegian ANN models can be found in Table 28. Linear lags are very rarely selected. QuickNet, the only method that generates insane forecasts, is the one choosing the largest amount of hidden units for models recursive forecasting. There is a big difference between it and QN-SG. The latter also produces more accurate forecasts of the two. Note, however, that this difference disappears when direct 12-month models are considered. In this case, neither QuickNet nor QN-SG forecasts need filtering.

MBE is still the most parsimonious method. Autometrics has the same property as before: the models for direct multiperiod forecasting contain a

Rec.	Hor.	QuickNet	MBE	Autom	QN-SG
	1	0.431	0.714	0.978	0.911
AR	3	0.982	0.700	0.984	0.984
	6	0.961	0.609	0.996	0.908
	12	0.809	0.0532	0.735	0.359

Dir.	Hor.	AR	QuickNet	MBE	Autom	QN-SG	NP	NC
	1		0.431	0.714	0.978	0.911	0.617	0.911
AR	3	0.0602	0.741	0.820	0.993	0.455	0.869	0.983
	6	0.767	0.367	0.663	0.656	0.439	0.245	0.886
	12	0.265	0.0735	0.0958	0.447	0.0225	0.322	0.447

Table 27: p-values of the Wilcoxon signed-rank test for testing the hypothesis that the mean of the differences of absolute forecast errors of Norwegian unemployment rate from the linear recursive AR model and another model are equal against the alternative that the other model has smaller absolute forecast errors. p-values subtracted from one are those of the test in which the null and the alternative hypothesis change places

Type of forecast	Recursive			Direct 12		
Number of variables	Total	Lags	HU	Total	Lags	HU
QuickNet	9.53	0	9.53	4.80	0	4.80
QN-SG	2.77	0	2.77	5.57	0	5.57
MBE	1.87	0	1.87	1.57	0	1.57
Autometrics	3.43	0	3.43	16.9	0.07	16.8

Table 28: The average number of linear lags and hidden units (HU) selected for ANN models for recursive, and direct 12-months ahead, forecasting of the Norwegian unemployment rate from July 2007 onwards

large number of hidden units. Although the Norwegian unemployment rate has been relatively easy to forecast, Autometrics still favours large models when the shortest lags are barred from the selection pool. This becomes obvious already in three-month models that on average contain 22.6 hidden units (and no linear lags). This tendency to generate large models may lie behind the crashes of Autometrics in our experiments.

The remaining country-specific RMSFE results can be found at `http://econ.au.dk/research/research-centres/creates/research/research-papers/supplementary-downloads/`.

6.3.4 Summary

One of the most consistent features of the ANN models for recursive forecasting is that QuickNet-based models sometimes generate very inaccurate forecasts. This happens for eight countries out of eleven. One of the remaining three is Germany, whose models were based on only 150 hidden units. It may be noted that the specific-to-general version QN-SG did not generate a single completely inaccurate forecast in that no forecast was filtered. The QN-SG forecasts were also often more accurate than the QuickNet ones.

Another common outcome is that the benchmark linear AR model is quite competitive against the ANN ones when recursive forecasts are compared. It is most frequently surpassed by MBE which also selects the most parsimonious models for both recursive and direct forecasting. This happens for more than one half of the countries at least at two forecast horizons. For two countries, the linear AR model is always better than any of its ANN competitors. They are France and Germany. Note that for the latter, the selection pool only contained 150 hidden units, 1/8 of the total forming the starting-point of the experiment. It may be that this is too little for the ANN models estimated through linear model selection to have a fair chance to perform better than the linear model.

Third, filtering is most often applied to QuickNet (as is obvious from above) and Autometrics recursive forecasts. There are only two countries for which the unemployment rate forecasts are never filtered: Sweden and Germany. For the latter, this may again have to do with the size of the selection pool.

Fourth, it seems that not respecifying the model during the forecasting period beginning in July 2007 was most beneficial for Autometrics (five cases out of 11). In general, however, respecifying seems a better idea. Furthermore, Autometrics in some cases seems to have gained from prescreening via linearity testing, but in general this did not seem useful. For Japan, however, linearity was never rejected. Nevertheless, for this country all four methods (after filtering) generated more accurate forecasts than the linear AR model.

Fifth, even some direct models occasionally yield vastly inaccurate forecasts. This occurs less frequently than in the case of recursive forecasting, but both some QuickNet and Autometrics-based models do that. Results of the Wilcoxon signed-rank test indicate that direct forecasts are on average not superior to their recursive counterparts, although country-specific results may contain exceptions. In comparing recursive and direct linear forecasts using the same test the results are similar to the previous ones: the recursive AR forecasts in general appear more accurate than direct ones.

7 Conclusions

The results of our forecasting experiment show that when it comes to forecast during a severe economic crisis, there is no dominant method for doing that among the ones considered here. The behaviour of the series may affect the results. In most cases, the industrial production growth rates turn negative and then positive again. The unemployment rates typically increase without a clear turning-point during the forecasting period. A method may perform well for one type of series but not for the other. A case in point is the nonparametric model that yields reasonable forecasts overall for industrial production series but fares less well in forecasting unemployment rates. The same is true for the 'no change' forecast. A general conclusion is that on average, it is not easy to improve upon the linear AR model.

Concerning ANN selection methods, a tentative conclusion is that parsimonious methods work best. This is in agreement with the results reported in Kock and Teräsvirta (2011b). In comparing QuickNet with QN-SG it is seen that the latter generates fewer completely erroneous forecasts. These forecasts are likely to emerge from models with a number of very strongly correlated variables. The possibility of choosing such models is greater in QuickNet than in QN-SG but could be controlled by keeping the maximum number of variables to be selected small in QuickNet. In this experiment it has equalled ten.

It seems that the modelling philosophy of MBE is suitable for this type of model selection problem. MBE works to first remove all irrelevant variables from consideration and build a model on the remaining ones. The result is most often a parsimonious model. Judging from the results, this is a reasonable strategy.

Autometrics tries to find a model that is parsimonious but also passes a battery of misspecification tests. This strategy works well when the model is a reasonable approximation to reality but less well when it is not. In direct forecasting one is facing the latter situation, because the most relevant lags are excluded from consideration. The result is a heavily parameterised ANN model that can sometimes forecast well but often is not competitive against the other ANN models. It appears that Autometrics may not be an appropriate tool for building models for direct multiperiod forecasting. It can be an excellent choice when the data-generating process is well approximated by a subset of variables in the data set of the researcher. A simulated example in Kock and Teräsvirta (2011b) demonstrates this fact very clearly.

References

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

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

Gilchrist, W. G.: 1967, Methods of estimation involving discounting, *Journal of the Royal Statistical Society B* **29**, 355–359.

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

Hendry, D. F. and Krolzig, H.-M.: 2005, The properties of automatic Gets modelling, *Economic Journal* **115**, C32–C61.

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

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

Kock, A. B. and Teräsvirta, T.: 2011b, Nonlinear forecasting of macroeconomic variables using automated model selection techniques, *CREATES Research Paper 2011-27*, Aarhus University.

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

Luukkonen, R., Saikkonen, P. and Teräsvirta, T.: 1988, Testing linearity against smooth transition autoregressive models, *Biometrika* **75**, 491–499.

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

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

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

Racine, J.: 2000, A consistent cross-validatory method for dependent data: *hv*-block cross validation, *Journal of Econometrics* **99**, 39–61.

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

Swanson, N. R. and White, H.: 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.

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

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

Teräsvirta, T.: 1998, Modeling economic relationships with smooth transition regressions, *in* A. Ullah and D. E. Giles (eds), *Handbook of Applied Economic Statistics*, Dekker, New York, pp. 507–552.

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

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

Teräsvirta, T., van Dijk, D. and Medeiros, M. C.: 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* G. Elliott, C. W. J. Granger and A. Timmermann (eds), *Handbook of Economic Forecasting*, Vol. 1, Elsevier, Amsterdam, pp. 459–512.

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

Appendices

A Creating the pool of hidden units

We use the following modification of the strategy of White (2006); see Kock and Teräsvirta (2011b):

1. Rewrite the argument of the logistic function in (1) as follows:

$$\boldsymbol{\gamma}'\mathbf{z}_t = (\gamma_1/\hat{\sigma}_z)(\boldsymbol{\gamma}_2'\tilde{\mathbf{z}}_t - \gamma_0) \tag{8}$$

   where $\mathbf{z}_t = (1, \tilde{\mathbf{z}}_t')'$ and $\hat{\sigma}_z$ is the sample standard deviation of $\boldsymbol{\gamma}_2'\tilde{\mathbf{z}}_t$. Choose $\boldsymbol{\gamma}_2$ as in White (2006).

2. Next obtain γ_0. Consider the values $x_t = \boldsymbol{\gamma}_2'\tilde{\mathbf{z}}_t$, $t = 1, ..., T$. Let $x_{\min}$ and $x_{\max}$ denote the minimum and maximum values of this sequence. Let $d = x_{\max} - x_{\min}$. Now draw γ_0 from a uniform $(x_{\min} + \delta d, x_{\max} - \delta d)$ distribution for $\delta \in [0, 0.5]$. We choose $\delta = 0.1$. In this way we make sure that the hidden units are not centred at very small or large values of $\boldsymbol{\gamma}_2'\tilde{\mathbf{z}}_t$. As a result of the parameterization (8), demeaning $\widetilde{\mathbf{z}}_t$ is not necessary.

3. Finally, the slope parameter γ_1 is chosen uniformly at random from the set $\{1.25^j : j = 0, 1, ..., 20\}$. The range of possible values is $(1, 87)$. The set is deliberately constructed to be denser for small values since the slope of the logistic function changes more for changes in γ_1 when γ_1 is small than when it is large. For large values of γ_1 changes in this parameter will not affect the slope of the logistic function much, and so it is less important to have a dense grid there.

The decisive difference between White's strategy and ours lies in §3. In the former, γ_1 is not a scale-free parameter. That is, a change of units in $\widetilde{\mathbf{z}}_t$ affects the set of possible slopes that can be selected, which is a disadvantage. In (8), γ_1 is a scale-free parameter due to the division of the exponent by $\widehat{\sigma}_z$, for discussion, see for example Teräsvirta (1998). This makes it possible for the user to define a reasonable range for this parameter. The minimum value of the scale-free γ_1 is set to unity in order to avoid logistic functions with too little sample variation.

B Figures

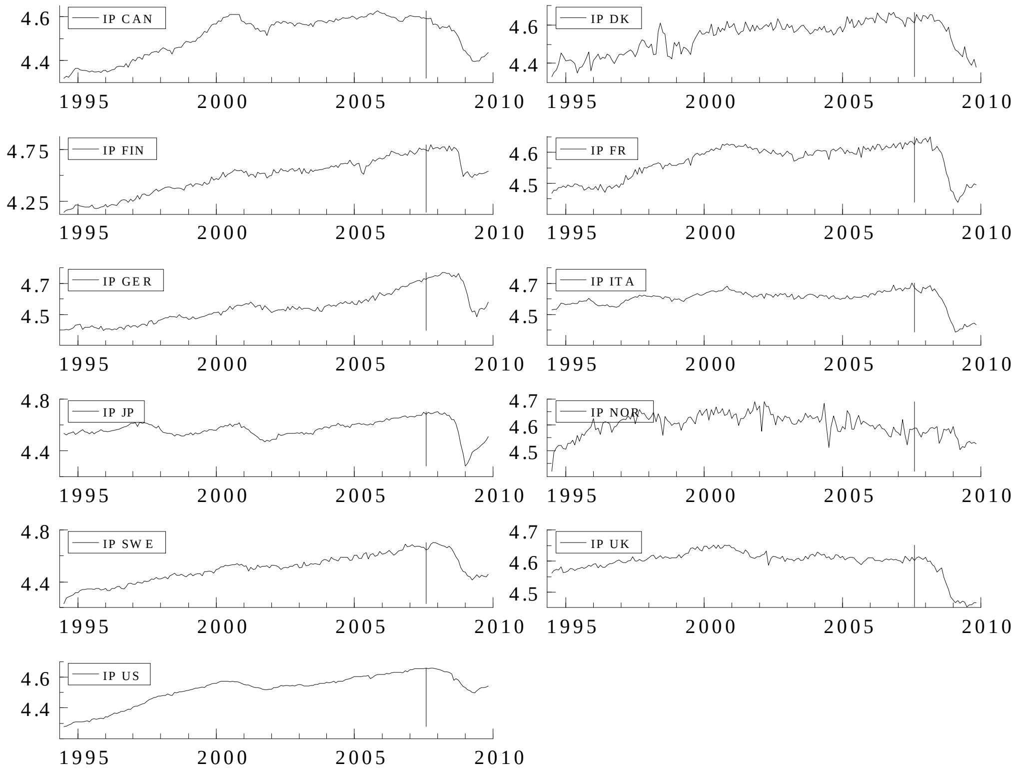

Figure 1: Logarithm of the industrial production index for the G7 and the four Scandinavian countries, 1994(8)–2009(12)

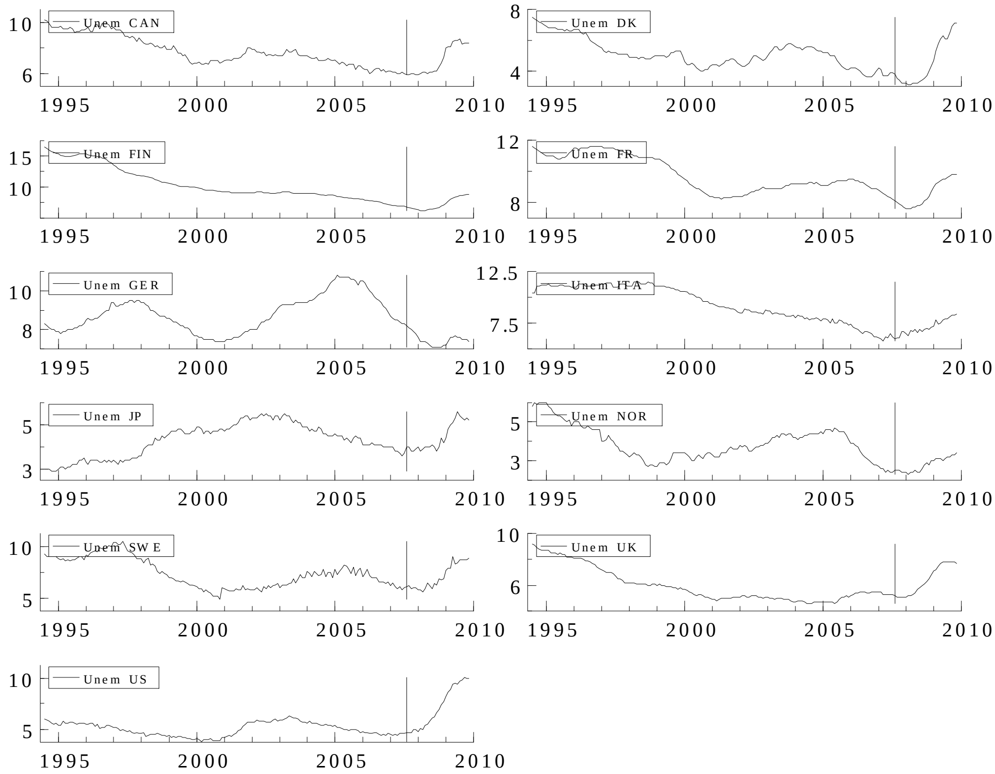

Figure 2: Unemployment rate series for the G7 and the four Scandinavian countries, 1994(8)–2009(12)

Research Papers 2011

2011-14: Stefano Grassi and Paolo Santucci de Magistris: When Long Memory Meets the Kalman Filter: A Comparative Study

2011-15: Antonio E. Noriega and Daniel Ventosa-Santaularia: A Simple Test for Spurious Regressions

2011-16: Stefano Grassi and Tommaso Proietti: Characterizing economic trends by Bayesian stochastic model specification search

2011-17: Søren Johansen and Theis Lange: Some econometric results for the Blanchard-Watson bubble model

2011-18: Tom Engsted and Thomas Q. Pedersen: Bias-correction in vector autoregressive models: A simulation study

2011-19: Kim Christensen, Roel Oomen and Mark Podolskij: Fact or friction: Jumps at ultra high frequency

2011-20: Charlotte Christiansen: Predicting Severe Simultaneous Recessions Using Yield Spreads as Leading Indicators

2011-21: Bent Jesper Christensen, Olaf Posch and Michel van der Wel: Estimating Dynamic Equilibrium Models using Macro and Financial Data

2011-22: Antonis Papapantoleon, John Schoenmakers and David Skovmand: Efficient and accurate log-Lévi approximations to Lévi driven LIBOR models

2011-23: Torben G. Andersen, Dobrislav Dobrev and Ernst Schaumburg: A Functional Filtering and Neighborhood Truncation Approach to Integrated Quarticity Estimation

2011-24: Cristina Amado and Timo Teräsvirta: Conditional Correlation Models of Autoregressive Conditional Heteroskedasticity with Nonstationary GARCH Equations

2011-25: Stephen T. Ziliak: Field Experiments in Economics: Comment on an article by Levitt and List

2011-26: Rasmus Tangsgaard Varneskov and Pierre Perron: Combining Long Memory and Level Shifts in Modeling and Forecasting of Persistent Time Series

2011-27: Anders Bredahl Kock and Timo Teräsvirta: Forecasting Macroeconomic Variables using Neural Network Models and Three Automated Model Selection Techniques

2011-28: Anders Bredahl Kock and Timo Teräsvirta: Forecasting performance of three automated modelling techniques during the economic crisis 2007-2009