Working Paper 99-3/ Document de travail 99-3

Forecasting GDP Growth Using Artificial Neural Networks

by
Greg Tkacz and Sarah Hu

Bank of Canada Banque du Canada

ISSN 1192-5434
ISBN 0-662-27537-3

Printed in Canada on recycled paper

Bank of Canada Working Paper 99-3

January 1999

Forecasting GDP Growth Using Artificial Neural Networks

by

Greg Tkacz and Sarah Hu

Department of Monetary and Financial Analysis
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
gtkacz@bank-banque-canada.ca

Contents

Acknowledgements

We would like to thank Kevin Clinton, Agathe Côté, Pierre Duguay, Ben Fung, Paul Gilbert, Jack Selody, and the seminar participants at the Bank of Canada for their helpful comments and suggestions. Any errors are, of course, our own.

Abstract

Financial and monetary variables have long been known to contain useful leading information regarding economic activity. In this paper, the authors wish to determine whether the forecasting performance of such variables can be improved using neural network models. The main findings are that, at the 1-quarter forecasting horizon, neural networks yield no significant forecast improvements. At the 4-quarter horizon, however, the improved forecast accuracy is statistically significant. The root mean squared forecast errors of the best neural network models are about 15 to 19 per cent lower than their linear model counterparts. The improved forecast accuracy may be capturing more fundamental non-linearities between financial variables and real output growth at the longer horizon.

Résumé

Les variables financières et monétaires sont reconnues depuis longtemps comme des indicateurs fiables de l'activité économique future. Dans cette étude, les auteurs tentent de déterminer si le recours à des réseaux neuronaux permet d'améliorer les prévisions réalisées à l'aide de ces variables. Ils constatent qu'à l'horizon d'un trimestre, les réseaux neuronaux ne produisent pas de meilleures prévisions que les modèles linéaires traditionnels. À l'horizon de quatre trimestres toutefois, on observe une amélioration significative des prévisions sur le plan statistique. Les erreurs quadratiques moyennes de prévision des meilleurs modèles neuronaux sont inférieures de 15 à 19 % à celles des modèles linéaires. Cette précision accrue des prévisions pourrait indiquer la présence de relations non linéaires fondamentales entre les variables financières et la croissance de la production réelle à l'horizon d'un an.

1. Introduction

The objective of this paper is to forecast output growth using neural networks, and to compare the forecasting performance of such non-linear models with traditional linear specifications. Neural networks can be thought of as “black box” models, as sometimes it is difficult to give economic meaning to the estimated relationships that emerge from them. Nevertheless, they have proven particularly useful as forecasting tools in the physical and natural sciences (e.g., see Ding, Canu, and Denoeux (1996)) and finance (e.g., see Angstenberger (1996)). Neural network models, due to their lack of structure, can be best viewed as indicator models. Thus, the present paper should be viewed as a contribution to, and extension of, the indicator model literature at the Bank, in the spirit of Muller (1992).

Neural networks allow for very general non-linear relationships between variables. In fact, when properly specified, they can approximate *any* non-linear function. It is known that explanatory variables such as monetary aggregates or yield spreads lead GDP growth and can therefore be used to anticipate future economic activity. If there are any non-linearities between such variables, then neural networks can exploit them to provide more accurate forecasts of economic activity.

Friedman (1968) argued that monetary policy may have an asymmetric effect on real economic activity. Tightening monetary conditions slows output growth to a greater degree than an equivalent expansion of monetary policy stimulates it. Researchers such as Cover (1992), Morgan (1993), and Rhee and Rich (1995) have found empirically that expansionary monetary policy, as measured by either money or interest rates, has a marginally lower impact on output growth than contractionary policy in the United States. These findings lend credence to Friedman’s proposition. Thus, non-linearities may very well exist in indicator models of output growth that are constructed using monetary and financial variables, and they may be representative of asymmetric effects of monetary policy on the real economy.

This paper shows that linear models are in effect constrained neural network models. Therefore, neural networks should provide forecasts that are at least as accurate as those obtained from linear indicator models. Results indicate that the neural network models perform quite well compared with linear models in a forecasting exercise conducted between 1985 and 1998, a period of about one and a half business cycles. On average, the best neural networks yield forecasts that are 15 to 19 per cent more accurate than the corresponding linear models for the 4-quarter growth rate of real output. Sequentially updating the neural network models over the forecast horizon is computationally intensive. However, this updating does yield slightly better forecasts than those from a neural net model whose weights are initially estimated using data from 1968 to 1985. At the 1-quarter horizon, no model—linear or non-linear—performs exceptionally well.

The paper is organized as follows. The next section explains the intuition behind neural networks, demonstrating how they can in fact be viewed as generalizations of linear models. Section 3 reviews some of the literature on neural networks in economics, an area that has been blossoming during the last five years. However, this literature has still been relatively scarce in the area of macroeconomics. In Section 4, linear and neural network models of output growth are constructed, using financial and monetary variables. Both types of models are compared in an out-of-sample forecasting exercise. Section 5 concludes and suggests avenues for future research.

2. Neural networks: A brief exposition

This section explains some of the intuition underlying neural networks and introduces its distinctive terminology. Neural networks are first defined and then a simple linear model is presented. This model can be viewed as a special case of the more general neural network model.

Animals (including humans) can make sense out of the large amount of visual information in their surroundings with very little effort. If this form of perception, or pattern recognition, can be transported to computing systems, then the analysis of complicated data structures could be improved. It is against this backdrop that Mehrotra, Mohan, and Ranka (1997, 1) present neural networks: “The neural network of an animal is part of its nervous system, containing a large number of interconnected neurons (nerve cells). ‘Neural’ is an adjective for neuron, and ‘network’ denotes a graph-like structure. Artificial neural networks refer to computing systems whose central theme is borrowed from the analogy of biological neural networks.” To simplify the terminology, many authors drop the word “artificial” when describing their neural network models, since any such model outside of biology must almost certainly be artificial.

Neural network models therefore attempt to emulate the impressive pattern recognition skills of animals. This is achieved by constructing a map of neurons, the simplest being the traditional linear regression model with J explanatory variables:[1]

$$Y = \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_J X_J , \tag{1}$$

or,

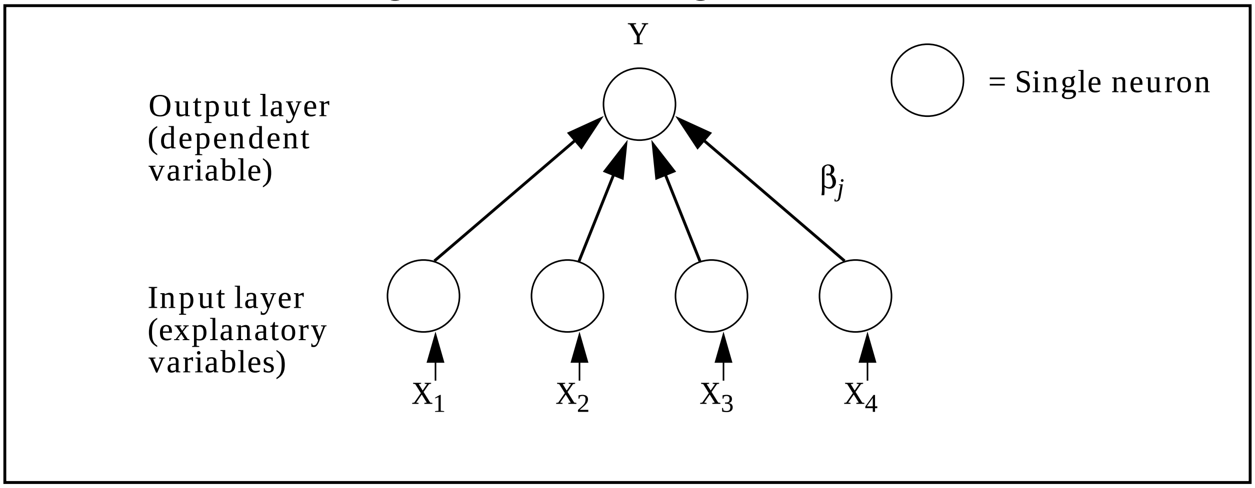
$$Y = \sum_{j=1}^{J} \beta_j X_j . \tag{2}$$

In words, (2) states that Y is a weighted sum of the X_j. This relationship is shown in Figure 2.1 with $J = 4$, where each X_j (the input neurons) is linked to Y (the output neuron) by the

1. The constant is omitted for expositional purposes, but it can easily be incorporated into neural network models; it is then referred to as a “bias term.” The time subscript is also initially omitted to simplify the notation.

parameters β_j (the input weights). In this linear model, it is easy to see the impact of changes in any of the X_j. As X_j changes by one unit, then Y will change by β_j units. For this reason, linear models are ideally suited for policy analysis as it is straightforward to perform comparative statics.

Figure 2.1: The linear regression model



Now (2) is generalized by the introduction of non-linearities into the relationship. As is, the relationship between Y and the weighted sum of the inputs is the same regardless of the value of the inputs. But suppose there is reason to believe that there may be asymmetries between the inputs and the output. For example, there is Friedman's (1968) argument, "pushing versus pulling on a string," which states that contractionary monetary policy will have a relatively larger impact on GDP growth than an equivalent expansionary policy. If the X_j represent policy variables and Y output growth, then one would expect the relationship between these variables to be dependent on the magnitude or direction of the input variables. This could be accomplished through a simple threshold function, where a large value of the weighted sum of the inputs would initiate a discrete regime change. In many practical applications, the regime changes may not be so abrupt, and therefore smooth activation functions are used. Popular choices include any type of sigmoid (S-shaped) function, such as the logistic function

$$g(u) = \frac{1}{1+e^{-u}}. \qquad (3)$$

Hence, to allow for a non-linear relationship between the weighted inputs and the output, the function (3) can be applied on the linear model (2), yielding

$$Y = g\left(\sum_{j=1}^{J} \beta_j X_j\right). \qquad (4)$$

Now, suppose the link between the X_j and Y is not direct. In the case of the relationship between interest rates and output, one would expect interest-sensitive intermediate variables—such as investment—to be affected before output. For example, a drop in current investment due to higher interest rates leads to lower output in the current period. However, it also leads to lower output in future periods due to the loss of potential future income flows resulting from the capital expenditures that were not initiated.[2] More generally, the intermediate variables in neural networks need not be identified in order to forecast Y; they can simply be treated as unknown. The intermediate variables (referred to as hidden neurons in the literature) are intermediate processing stages where the inputs X_j, and their corresponding weights β_j, are subject to another re-weighting prior to affecting Y. In other words, a hidden layer of neurons is introduced. Proceeding with the hypothetical example, if an intermediate variable can be thought of as representing investment, then the neural net model can allocate larger weights for investment levels that have proportionately larger effects on output growth.

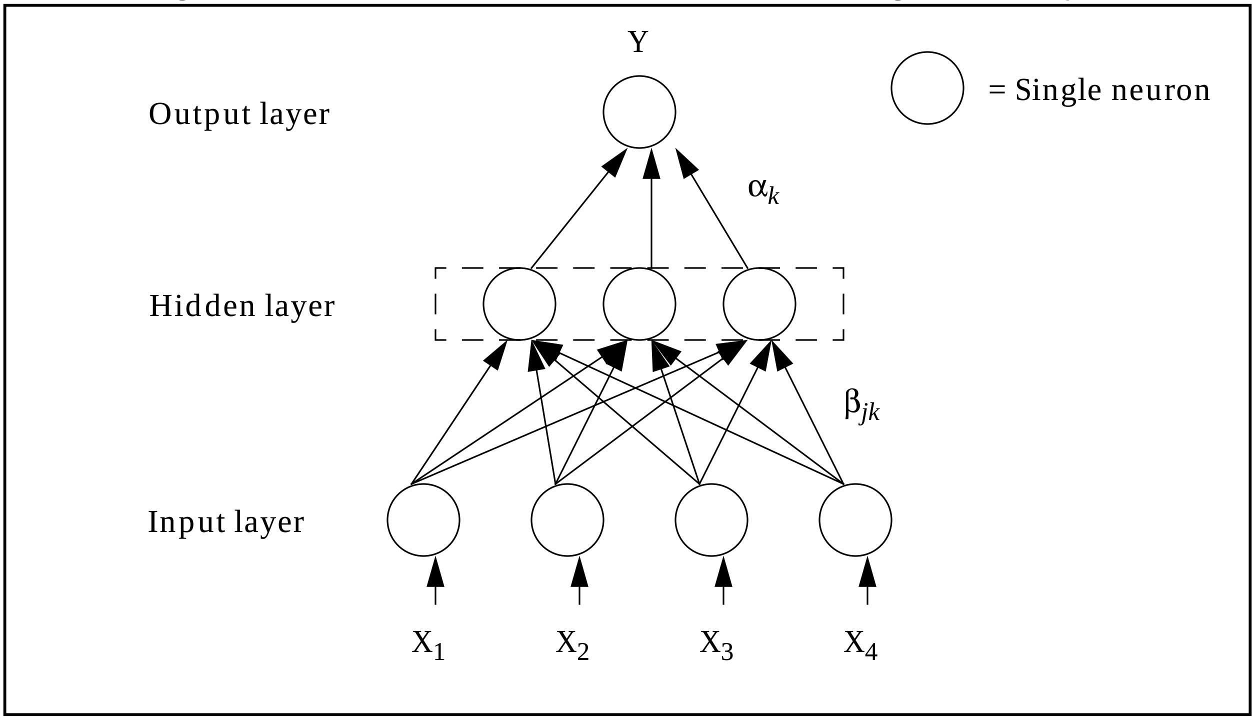
Figure 2.2 shows a feedforward neural network map with J = four input neurons, K = three hidden neurons, one hidden layer, and one output neuron. It is called a feedforward model since the X_j affect Y, while the converse is not true. (In some neural net models, one can allow for changes in Y to affect the X_j.) The connection strengths α_k, $k = 1,2,3$, link the hidden neurons to Y; the connection strengths β_{jk}, $j = 1,2,3,4$, and $k = 1,2,3$, link the input neurons to the hidden neurons. Analytically, the neural network model can be represented as

$$Y = h\left(\sum_{k=1}^{K} \alpha_k g\left(\sum_{j=1}^{J} \beta_{jk} X_j\right)\right). \tag{5}$$

Thus, it is seen that the connection strengths $\alpha_k g(\cdot)$ are summed and filtered through another activation function $h(\cdot)$. For practical purposes, $g(\cdot)$ and $h(\cdot)$ are the same. The significance of the single hidden layer of neurons between the inputs and the output may not seem apparent at this time. However, if there are sufficient hidden neurons, (5) can approximate any non-linear function to an arbitrary degree of accuracy. This is known as the universal approximation property of neural networks, and such approximation is not possible in the absence of the hidden layer (see White (1992)).

2. If the interest sensitivity of investment is not constant for all interest rate levels, then this may be a source of any non-linearities between interest rates and output growth.

Figure 2.2: The feedforward neural network with a single hidden layer



The estimation of α_k and β_{jk} in (5) (also referred to as "learning") is fairly straightforward, as one seeks to minimize the sum of squared deviations *(SSD)* between the output and the network:

$$\min_{\alpha_k,\, \beta_{jk}} SSD = \sum_{t=1}^{T} \left[Y_t - h\left(\sum_{k=1}^{K} \alpha_k g\left(\sum_{j=1}^{J} \beta_{jk} X_{jt} \right) \right) \right]^2 . \qquad (6)$$

This is accomplished by adjusting the weights α_k and β_{jk} until the desired pre-specified level of convergence is achieved. Any number of numerical algorithms can be used for this,[3] although the back-propagation method is often employed. This is a grid search method that readjusts the weights sequentially to minimize the errors.[4]

The neural network (5) should be sufficient for most economic applications, but it can be generalized further. For example, nothing prevents the existence of multiple hidden layers or multiple outputs. This could be useful if one wanted to forecast, say, output growth and inflation simultaneously. However, since this is a data-intensive technique and the number of parameters to be estimated quickly increases with every additional layer of neurons, such a practice may be

3. For example, simulated annealing and genetic algorithms are powerful, recently developed estimation methods. Chapters 7 and 8 of Masters (1993) are a readable introduction to such methods.

4. Chapter 3 of Mehrotra, Mohan, and Ranka (1997) describes this algorithm in more detail.

constrained by lack of data. For forecasting purposes, neural networks typically require three different data sets. First, a training sample is needed for the initial estimation of parameters. Second, a testing sample is required to verify the accuracy of the forecasts of the trained model. If the testing sample forecasts poorly, one has to return to the training stage and re-estimate the model by modifying the number of hidden neurons and/or the duration of the training by tightening or loosening the convergence criterion. With the newly trained model, one then returns to the testing sample. Third, a forecasting sample is needed for which the model can be used in forecasting exercises. In this paper, there are only 30 years of quarterly data, which means that data scarcity becomes an issue. For this reason, the testing sample has to be omitted and the freshly trained neural net models used in the forecasting exercises. In spite of this, they still outperform the linear models at the longer forecasting horizon.

Finally, it is worth mentioning that neural networks are most convenient for forecasting purposes, not for comparative statics. Computing an elasticity is not as straightforward as it would be in a linear regression model. To determine what impact a 1 per cent increase in interest rates would have on output, one must pick two different interest rates and filter them through the model, while keeping other variables constant. In a univariate neural net model, this is fairly straightforward. One can simply read the changes in the output variable on the vertical axis, but in multivariate models, this is not feasible.[5]

3. Literature review

The use of neural networks in economics is still in its relative infancy. The article by Kuan and White (1994) is likely the definitive introduction of neural networks to the econometrics literature. Kuan and White draw many of the parallels between econometrics and neural networks, as we have attempted in Section 2. Kuan and White's theoretical contribution has been followed with some applied work by Maasoumi, Khotanzad, and Abaye (1994). These authors demonstrate that the 14 macroeconomic series analyzed in the seminal article by Nelson and Plosser (1982) can be nicely modelled using neural networks, casting strong doubt that these series follow unit root processes. Indeed, this is consistent with the conclusions of Perron (1989), who finds that allowing for one or two structural breaks renders such series stationary. The major strength of neural net models is that they are better capable of modelling breaks such as stock market crashes and oil shocks, since these may denote significant departures from linearity.

Swanson and White (1997) represent another major attempt at using neural nets to forecast macroeconomic variables. This paper compares the relative usefulness of different linear and non-linear models using a wide array of out-of-sample forecasting performance indicators. Their major

5. For additional introductions to neural networks, see McMenamin (1997), Campbell, Lo, and MacKinlay (1997), and Kuan and White (1994). The first of these references is the simplest, the last the most rigorous.

conclusion is that multivariate linear models are marginally better overall. However, they have trouble finding consistent results, as some perform better for selected variables and for selected performance statistics. Apart from a corporate bond yield, they use no other monetary or financial variables in their models.

Kohzadi et al. (1995) is a good introduction to neural networks and their uses in economics, with an application to the forecast of corn futures. Comparing a neural network to an ARIMA model, they find that the forecast error of the neural net model is between 18 and 40 per cent lower than that of the ARIMA model, using different forecasting performance criteria.

In the finance literature, Angstenberger (1996) develops a neural net model that has some success at forecasting the S & P 500 index. Hutchinson, Lo, and Poggio (1994) find that neural network models can sometimes outperform the parametric Black-Scholes option pricing formula. However, Campbell, Lo, and MacKinlay (1997) caution that this finding is model specific, and that one should not conclude that neural network models are generally superior until they have been tested against additional parametric models.

4. Models of output growth

As stated in the introduction, this paper can be viewed as an extension of traditional indicator model research at the Bank. Models such as those estimated by Muller (1992), for example, have been almost universally linear. Innovations to these models have included experimentation with the right-hand side variables and the lag structure, but no attempt has been made to modify the functional form. Meanwhile, in the neural network literature, there has been little effort thus far to experiment with the right-hand side variables, as authors have been more concerned with functional forms. The innovation of the present paper is therefore to combine the known indicator properties of selected monetary and financial variables with the promising forecast performance of neural network models. First, the forecasting performance of some linear models is assessed as this can act as a gauge of the forecasting success (or failure) of neural network models. After this assessment, neural network models are estimated.

4.1 Linear models

The variables to be forecast in this paper are the 1-quarter and 4-quarter cumulative growth rate of real GDP. The list of explanatory variables includes a long-short interest rate spread, real 90-day commercial paper and real long-term bond rates, the growth rates of narrow (real M1) and broad (real M2) monetary aggregates, and the growth rate of the real TSE 300 index. The linear models are of the form

$$\Delta^m Y_t = \alpha + \sum_{j=1}^{J} \beta_j X_{jt-m} + \varepsilon_t, \tag{7}$$

where $\Delta^m Y_t$ is the m-quarter growth rate of real output (m = 1 or 4), and X is one of the explanatory variables outlined above. In the specification search for the best linear models, these variables are allowed to enter the model both individually and in combination with other variables. Table 1 presents only the results for the best univariate and multivariate linear models taken from an extensive specification search—five models each for the 1- and 4-quarter growth rates of real output. The parameter estimates are presented for data over the entire sample, 1968:1 to 1998:1. In Table 2, the out-of-sample forecasting statistics were generated for forecasts from 1985:1+m to 1998:1. A number of forecast performance indicators are considered. The RMSE, MAD, and Theil U statistics are measures of the size of forecast deviations from the actual values; the lower these statistics, the more accurate the forecasts. The confusion rate is a measure of the directional accuracy of forecasts; the closer this statistic is to zero, the better the forecasts of the direction of output growth. The formulas for the computation of each statistic are given in the Appendix.

For 1-quarter output growth, the yield spread, real M1 growth, and short-term rates are all found to be significant explanatory variables. The model with the best fit in-sample, 1e, has the 4-quarter growth of real M1 and real R90 as regressors. In the forecasting exercise, however, this model is clearly beaten by model 1d, which includes the yield spread and real R90 as right-hand side variables. This is a case where good in-sample fit is no guarantee of good forecasting performance. Notice as well that, for the univariate models 1a, 1b and 1c, the yield spread is a better out-of-sample performer than M1, based on the performance statistics. Real money, however, seems to be a good forecaster of the direction of output growth, as the confusion rate is usually lower with the 4-quarter growth rate of M1 as a regressor. Thus, money seems to have trouble forecasting the magnitude of output growth, but does relatively better on the direction.

In all models of the 4-quarter growth rate of output, the yield spread is kept as a regressor, since it is always highly significant. Real M2 also outperforms real M1 at this horizon, so broad money is used in these models. Overall, the smoother 4-quarter growth series is found easier to forecast, as can be seen from the improved forecast performance statistics. The best model, in- and out-of-sample, is model 4e, which consists of the yield spread, real R90, and the 4-quarter growth rate of real M2. It has both the best in-sample fit and the lowest forecast errors.

4.2 Neural network models

In this section, neural network models of 1- and 4-quarter output growth are constructed. To make the comparison to linear models tractable, the same explanatory variables are used as for the 10 linear models in Table 1.[6] The models are therefore of the form of (5), with the right-hand side variables taken from the linear models. Following the recommendation of Kuan and White (1994), a single hidden layer is used as this seems to be appropriate for most economic applications.[7] The number of hidden units within the hidden layer was set to two in all cases, as this provided adequate forecasting performance. Three hidden units sometimes lowered the accuracy of forecasts. A rough rule of thumb proposed by Bailey and Thompson (1990) suggests that the number of hidden units be set to 75 per cent of the number of inputs; therefore two hidden units should suffice.

For forecasting purposes with neural networks, we proceed in two different ways. In the first instance, static forecasts are computed, based on the estimation of the neural network model from 1968:1 to 1985:1 (i.e., the training sample). Estimates of the parameters α_k and β_{jk} in (5) are thus obtained. This model is then used to forecast 1- and 4-quarter output growth from 1985:1+m to 1998:1. For example, if the only explanatory variable is the 4-quarter lag of the yield spread, an in-sample fitted curve as in Figure 1 (see page 19) is obtained. Note that the neural network curve is concave, implying that *positive values of the yield spread have marginally lower impacts on output growth than negative values*. If monetary policy were to have some control over this spread, then a tight policy (represented by an inverted yield curve) would have a relatively larger impact on economic activity than an expansionary policy (as denoted by an upward-sloping yield curve). The finding of non-linearities between financial variables and output growth can be interpreted as being generally consistent with the asymmetry findings of Cover (1992) and others. In Figure 2 (see page 19), where the models are estimated over the entire sample period from 1968:1 to 1998:1, the asymmetry becomes even more pronounced as the neural net curve has a steeper slope for negative spreads, and a flatter slope for positive spreads.

For a univariate model such as 4a, a forecast of the 4-quarter growth rate of real output in 1986:1 consists of taking the yield spread on the horizontal axis at 1985:1, and reading the corresponding output growth rate on the vertical axis. The spread for 1985:2 is then taken and the corresponding growth rate for 1986:2 obtained. This practice is continued until an output growth forecast for 1998:1 is obtained. For the multivariate models, the forecasting procedure is analogous, as the input data is filtered through the model at time t to obtain forecasts at t+4.

A problem with this methodology is that the parameters α_k and β_{jk} are not updated to account for newly available data. The parameters of the linear model (7) are updated each quarter to produce a forecast, but this is not the case for the static neural network model. Ideally, it would be useful to perform the same kind of sequential updating exercise for the neural network model,

6. Also used were variables, such as stock prices, that were found to contain little forecasting power in the linear models. In the current forecasting context, these variables did not improve neural network forecasts either, and therefore those results are not presented.

7. There were also experiments with two hidden layers, but the results were not significantly different.

but unfortunately some logistical problems are encountered. The greatest problem is that the neural network model is computationally intensive. It may sometimes require several hundreds of thousands of epochs (replications, or passes through the data) during the back-propagation search technique to find parameters that will meet the pre-specified convergence criterion. And sometimes the search technique may become “stuck” in a plateau or valley, causing the procedure to continue indefinitely.[8]

However, such an intensive updating of the neural network model is impractical, since several factors need to be specified by the user each time the parameters are estimated. One can choose to update the model less frequently than each quarter. For example, for the 4-quarter growth rate forecasts, 49 forecasts are required between 1986:1 to 1998:1. Thus the model can be sequentially updated every seven quarters, with the implication that it requires updating only seven times. This is far more manageable as the pre-specified convergence criterion need not be changed with each update. This is a fairly common practice and was used by, for example, Kohzadi et al. (1995). Our model is therefore estimated with data from 1968:1 to 1985:1; 4-quarter forecasts for the 4-quarter output growth rate for 1986:1 to 1987:3 are performed. The neural net model is then re-estimated with data from 1968:1 to 1986:4, and forecasts for 1987:4 to 1989:2 are obtained. This is repeated until a forecast for 1998:1 is found. The forecasts from models that are sequentially updated in such a manner are referred to as dynamic forecasts. For the 1-quarter models, there are 52 observations to forecast from 1985:2 to 1998:1, implying the model can be updated every four quarters for a total of 13 updates.

Tables 3 and 4 present the forecasting performance statistics for the static and dynamic neural network models.[9] For forecasts of the 1-quarter growth rate, the static model performs relatively poorly compared with the linear model, as it only marginally outperforms the latter in one instance. The dynamic forecasts in Table 4 are noticeably better, as the neural network model outperforms the linear model by just over 5 per cent in terms of forecast accuracy for most specifications. The confusion rates for the dynamic forecasts are also marginally lower, implying improved forecast directions. The poor forecasting performance of the static model is likely because output growth is highly volatile and the training sample is too small to map the pattern of the data adequately. This can be seen in Figure 3 (see page 20) where no model is a stellar

---

8. Estimating neural network models requires some care, as the modeller must use a pre-specified convergence criterion for the back-propagation search. If the criterion is too tight, the model can overfit in-sample, causing it to miss the trend of the output variable, and thereby producing poor forecasts. If the criterion is too loose, the model may not capture the underlying pattern of the data. The results in Tables 3 and 4 were obtained after searching for the optimal convergence criteria. Furthermore, since the parameter search begins with random values, the results can differ slightly each time the estimation procedure is performed.

9. The estimated parameters are not presented, since they have little economic meaning. A large value for β_{11} should not imply that the first input variable has a strong impact on output growth, since its impact could be mitigated by the activation functions and α_1 and α_2. Similarly, small values of β_{11} could be amplified by the activation functions and α_1 and α_2, making direct interpretation difficult. If one wants to perform comparative statics using neural network models, there is no alternative but to vary the chosen input by, say, 1 per cent, and to observe the percentage change in output growth.

performer. All models appear to adequately capture the general trend of this series but are unable to properly capture the volatility. It can be concluded therefore that financial variables are relatively poor predictors of economic activity at the 1-quarter horizon, as average forecast errors of over 2.5 per cent should not be acceptable to policy-makers.[10]

At the 4-quarter horizon, the neural net models perform much better. The output growth series is far less volatile, and financial variables are known to be better indicators at this horizon. The two best linear models are 4d and 4e, both of which include the yield spread and real R90, and the 4-quarter growth of real M2 in the case of model 4e. These models are also the best for the static and dynamic neural network models—the root mean squared errors are between 15 and 19 per cent lower than those for the linear models, with the dynamic models outperforming the static models. Figure 4 (see page 20) shows that the linear model has been overpredicting output growth by about 0.5 to 1 per cent in many periods since 1991. In fact, linear model forecasts are usually higher than neural net forecasts at all peaks between 1985 and 1998. Meanwhile, in the troughs, the models produce nearly identical forecasts. This is again mild evidence of asymmetry between the financial variables and output growth, consistent with the pictures of the kinked relationship between the yield spread and GDP growth in Figures 1 and 2 (see page 19).

It is also interesting to note that the directions of the forecasts in Figures 3 and 4 (see page 20) are very similar for all three models. In other words, there does not appear to be any gain in directional accuracy of forecasts. Any forecast gains from neural nets emerge only in the magnitudes of the forecasts. It is not clear whether this is a general feature of neural networks, or whether this is specific to the present models.

For growth up to 1999:1, all three models agree on the direction of GDP growth, as it is clearly falling. The linear model has the 4-quarter growth rate of real GDP falling to 2.0 per cent in 1999:1, the dynamic neural net model to 2.8 per cent, and the static neural net model as the most optimistic with forecasted growth of 3.2 per cent for the same period.

5. Conclusion and discussion

This paper seeks to determine whether more accurate indicator models of output growth based on monetary and financial variables can be developed using neural networks. No indicator model appears to perform exceptionally well for 1-quarter forecasts, as such variables are unable to explain adequately the volatility of the 1-quarter growth rate of real output. At this horizon, dynamic neural network models outperform the linear models in terms of forecast accuracy by about 5 per cent, while the static models fail to outperform the linear models. However, at the

10. If such variables are indicators of policy, then they should be more useful for forecasting output growth at longer horizons, consistent with the lag with which monetary policy is thought to affect the economy.

4-quarter horizon, monetary and financial variables are found to be much better predictors of output growth than at the 1-quarter horizon. Furthermore, the best neural network models outperform the best linear models by between 15 and 19 per cent at this horizon, implying that neural network models can be exploited for noticeable gains in forecast accuracy. The gains in forecast accuracy seem to originate from the ability of neural networks to capture asymmetric relationships. The data reveal that positive yield spreads (which may be an indicator of an expansionary monetary policy) have a relatively lower impact on output growth than negative yield spreads (which may be an indicator of contractionary monetary policy).[11]

In future work, we hope to further refine the best neural network models in this paper (by considering additional layers, different training periods, etc.) for use as forecasting tools to exploit readily available monetary and financial data in order to gauge future economic activity. Forecast comparisons with more complex models, such as a vector error-correction model or the Bank's own Quarterly Projection Model, may also be undertaken. As well, future projects may involve the construction of neural net models to forecast inflation. Projects could also examine the possibility of combining forecasts from different models, as Donaldson and Kamstra (1996) have had some success combining forecasts of linear and neural network models for stock prices.

As a broader guide for future research, it is perhaps constructive to identify some of the areas in economics where neural networks could be expected to be of use, as well as some of their potential limitations. Based on our experiments and reading of the literature, we believe that neural networks may be of use in the following situations:

- *Generally poor linear models:* If linear models have traditionally performed poorly in forecasting exercises in spite of well-developed economic theory, then non-linearities may exist in the data, so neural networks can be potentially useful. The classic example here involves the modelling of exchange rates, as it is generally understood that exchange rates should be functions of, among other factors, domestic and foreign price levels and interest rates. Empirically, however, such models perform poorly (e.g., see Amano and van Norden (1993) for a discussion).

- *Unknown parametric form:* Economic theory does not always yield a specific functional form that is to be used for empirical verification of the theory. Researchers have traditionally relied on linear models because of their ease of estimation, but with the advent of modern computing technology, one can venture beyond the bounds of linearity.

11. The improvement over linear models is even more impressive when it is recalled that the linear models used in the comparison exercise were among the best univariate and multivariate indicator models identified in the specification search. Poorer linear models, such as those based on stock prices, might in fact be poor due to marked non-linearities that the linear models were unable to capture. In such instances, one can suspect that neural network models could provide even greater forecast accuracy over linear models.

- *Structural breaks:* If linear models suffer from structural breaks, and if the training period is sufficiently long to incorporate different interesting episodes in the history of a variable, then the estimated parameters in the neural net model will have incorporated unusual movements in the data. This implies that neural net models may be more robust to structural breaks than linear models. A structural break is, in essence, just a special form of non-linearity, to which the neural net can adapt.

- *Benchmarks:* When assessing the forecasting performance of a new (linear) model, researchers have traditionally compared the model to the forecasting ability of a simple ARIMA model. In many applications, it has been found that neural networks outperform ARIMA models (e.g., see Kohzadi et al. (1995)). Hence, one can "raise the bar" by using neural network forecasts as a benchmark instead of the ARIMA model. The neural network benchmark can either be univariate (using lags of the independent variable as inputs) or multivariate (incorporating the variables of the model whose forecast performance the researcher wishes to assess).

In spite of their numerous strengths, neural nets suffer from a number of weaknesses that researchers should keep in mind.

- *Sample size:* Neural networks, more than linear models, need larger samples in order to be estimated properly. This is due to the large number of parameters introduced in such models that link the inputs to the hidden neurons, which are then linked to the output variable. As the data available for the training and testing of the model increase, one would then expect the marginal gains in forecast accuracy over linear models to increase. There is no rule of thumb for the "optimal" sample size for which one can expect neural nets to improve noticeably over linear models A "guesstimate" at this point would be around 300 observations. This explains why, for example, there have been few macroeconomic applications of neural networks thus far in the literature. Forty years of quarterly data are usually judged insufficient to learn the relationships between the input and output variables properly. As such, neural networks should noticeably outperform linear models in forecasts of higher-frequency variables.

- *Lack of economic structure:* Due to the black box nature of neural networks, users of forecasts may feel some uneasiness if they are unable to give proper economic interpretation to the estimated relationships. At the same time, it is difficult to determine which of the explanatory variables are driving the bulk of the forecasts, as comparative statics are difficult to perform.

Table 1: Linear models, in-sample, 1968:1 to 1998:1

$$\Delta^m Y_t = \alpha_0 + \sum_{j=1}^{J} \alpha_j X_{jt-m} + \varepsilon_t \text{ for } m = 1, 4$$

Model	Regressors	Constant	α_1	α_2	α_3	$\bar{R}^2$
1a	$\text{Spread}_{t\text{-}1}$	2.711 (7.825)	0.713 (4.456)	--	--	0.112
1b	$\Delta\text{RM1}_{t\text{-}1}$	2.739 (8.669)	0.185 (4.787)	--	--	0.166
1c	$\Delta^4\text{RM1}_{t\text{-}1}$	2.553 (8.019)	0.261 (5.160)	--	--	0.185
1d	$\text{Spread}_{t\text{-}1}$ $\text{RR90}_{t\text{-}1}$	4.108 (7.289)	0.543 (3.513)	-0.371 (-3.299)	--	0.170
1e	$\text{RR90}_{t\text{-}1}$ $\Delta^4\text{RM1}_{t\text{-}1}$	3.787 (7.829)	-0.332 (-3.281)	0.220 (4.559)	--	0.234
4a	$\text{Spread}_{t\text{-}4}$	2.610 (13.491)	0.807 (7.543)	--	--	0.348
4b	$\text{Spread}_{t\text{-}4}$ $\Delta^4\text{RGDP}_{t\text{-}4}$	2.025 (6.010)	0.796 (7.688)	0.188 (2.480)	--	0.379
4c	$\text{Spread}_{t\text{-}4}$ $\Delta^4\text{RM2}_{t\text{-}4}$	1.385 (3.253)	0.962 (9.176)	0.233 (3.302)	--	0.449
4d	$\text{Spread}_{t\text{-}4}$ $\text{RR90}_{t\text{-}4}$	3.924 (17.52)	0.650 (7.419)	-0.348 (-6.173)	--	0.479
4e	$\text{Spread}_{t\text{-}4}$ $\text{RR90}_{t\text{-}4}$ $\Delta^4\text{RM2}_{t\text{-}4}$	2.803 (5.956)	0.792 (8.206)	-0.294 (-5.103)	0.171 (2.478)	0.540

Newey-West corrected *t*-statistics are in parentheses.

Table 2: Linear model forecast performance statistics

Model	Regressors	RMSE	MAD	Theil U	Confusion rate
1a	$Spread_{t-1}$	2.925	2.359	0.830	0.48
1b	$\Delta RM1_{t-1}$	3.088	2.625	0.877	0.58
1c	$\Delta^4 RM1_{t-1}$	2.966	2.527	0.842	0.42
1d	$Spread_{t-1}$ $RR90_{t-1}$	*2.611*	*2.156*	*0.741*	0.50
1e	$RR90_{t-1}$ $\Delta^4 RM1_{t-1}$	2.691	2.261	0.764	*0.36*
4a	$Spread_{t-4}$	2.459	2.126	0.822	*0.32*
4b	$Spread_{t-4}$ $\Delta^4 RGDP_{t-4}$	2.420	2.065	0.809	*0.32*
4c	$Spread_{t-4}$ $\Delta^4 RM2_{t-4}$	2.214	1.868	0.740	*0.32*
4d	$Spread_{t-4}$ $RR90_{t-4}$	2.042	1.736	0.682	0.34
4e	$Spread_{t-4}$ $RR90_{t-4}$ $\Delta^4 RM2_{t-4}$	*1.899*	*1.580*	*0.635*	*0.32*

Best statistics for each performance criterion are highlighted. The model is initially estimated from 1968:1 to 1985:1; a single forecast is generated; the model is then re-estimated with one additional observation; and another forecast is generated. The model's parameters are sequentially updated until a forecast is obtained for 1998:1.

Table 3: Static neural network forecast performance statistics

Model	Inputs	RMSE	MAD	Theil U	Confusion rate	%(RMSE^{sta} - RMSE^{linear})
1a	Spread$_{t\text{-}1}$	3.181	2.672	0.903	0.44	+8.8%
1b	ΔRM1$_{t\text{-}1}$	2.970	2.470	0.843	0.58	-3.8%
1c	Δ^4RM1$_{t\text{-}1}$	2.976	2.455	0.845	0.40	+0.3%
1d	Spread$_{t\text{-}1}$ RR90$_{t\text{-}1}$	*2.639*	2.209	*0.749*	0.46	+1.1%
1e	RR90$_{t\text{-}1}$ Δ^4RM1$_{t\text{-}1}$	2.733	*2.166*	0.776	*0.38*	+1.6%
4a	Spread$_{t\text{-}4}$	2.345	2.023	0.784	*0.32*	-4.6%
4b	Spread$_{t\text{-}4}$ Δ^4RGDP$_{t\text{-}4}$	2.225	1.906	0.744	*0.32*	-8.1%
4c	Spread$_{t\text{-}4}$ Δ^4RM2$_{t\text{-}4}$	2.174	1.839	0.727	*0.32*	-1.8%
4d	Spread$_{t\text{-}4}$ RR90$_{t\text{-}4}$	1.710	1.410	0.572	0.34	-16.3%
4e	Spread$_{t\text{-}4}$ RR90$_{t\text{-}4}$ Δ^4RM2$_{t\text{-}4}$	*1.620*	*1.251*	*0.542*	0.34	-14.7%

Best statistics for each performance criterion are highlighted. Negative values for the percentage difference in the RMSE between the static neural network forecasts and the linear forecasts indicate that the neural network forecasts are more accurate; positive signs imply the opposite. The model is estimated from 1968:1 to 1985:1, and these parameters are used to generate forecasts for each period until 1998:1.

Table 4: Dynamic neural network forecast performance statistics

Model	Inputs	RMSE	MAD	Theil U	Confusion rate	%($RMSE^{dyn}$ - $RMSE^{linear}$)
1a	$Spread_{t-1}$	2.775	2.207	0.788	0.42	-5.1%
1b	$\Delta RM1_{t-1}$	2.825	2.264	0.802	0.54	-8.5%
1c	$\Delta^4 RM1_{t-1}$	2.768	2.221	0.786	0.38	-6.7%
1d	$Spread_{t-1}$ $RR90_{t-1}$	*2.456*	*1.909*	*0.697*	0.48	-5.9%
1e	$RR90_{t-1}$ $\Delta^4 RM1_{t-1}$	2.627	2.043	0.746	*0.34*	-2.4%
4a	$Spread_{t-4}$	2.097	1.750	0.701	*0.32*	-14.4%
4b	$Spread_{t-4}$ $\Delta^4 RGDP_{t-4}$	2.106	1.733	0.704	*0.32*	-13.0%
4c	$Spread_{t-4}$ $\Delta^4 RM2_{t-4}$	2.095	1.762	0.700	*0.32*	-5.4%
4d	$Spread_{t-4}$ $RR90_{t-4}$	1.645	1.304	0.550	0.36	-19.4%
4e	$Spread_{t-4}$ $RR90_{t-4}$ $\Delta^4 RM2_{t-4}$	*1.575*	*1.267*	*0.526*	0.36	-17.1%

Best statistics for each performance criterion are highlighted. Negative values for the percentage difference in the RMSE between the dynamic neural network forecasts and the linear forecasts indicate that the neural network forecasts are more accurate. The model is initially estimated from 1968:1 to 1985:1. These parameters are used to generate forecasts for the next four quarters for models 1a to 1e, and for seven quarters for models 4a to 4e. The network is then trained anew, and forecasts are again generated. The model's parameters are updated thirteen times between 1985:1 and 1998:1 for models 1a to 1e, and seven times for models 4a to 4e.

Figure 1: Linear and neural network fitted curves, 1968: 1 to 1985:1 (model 4a)

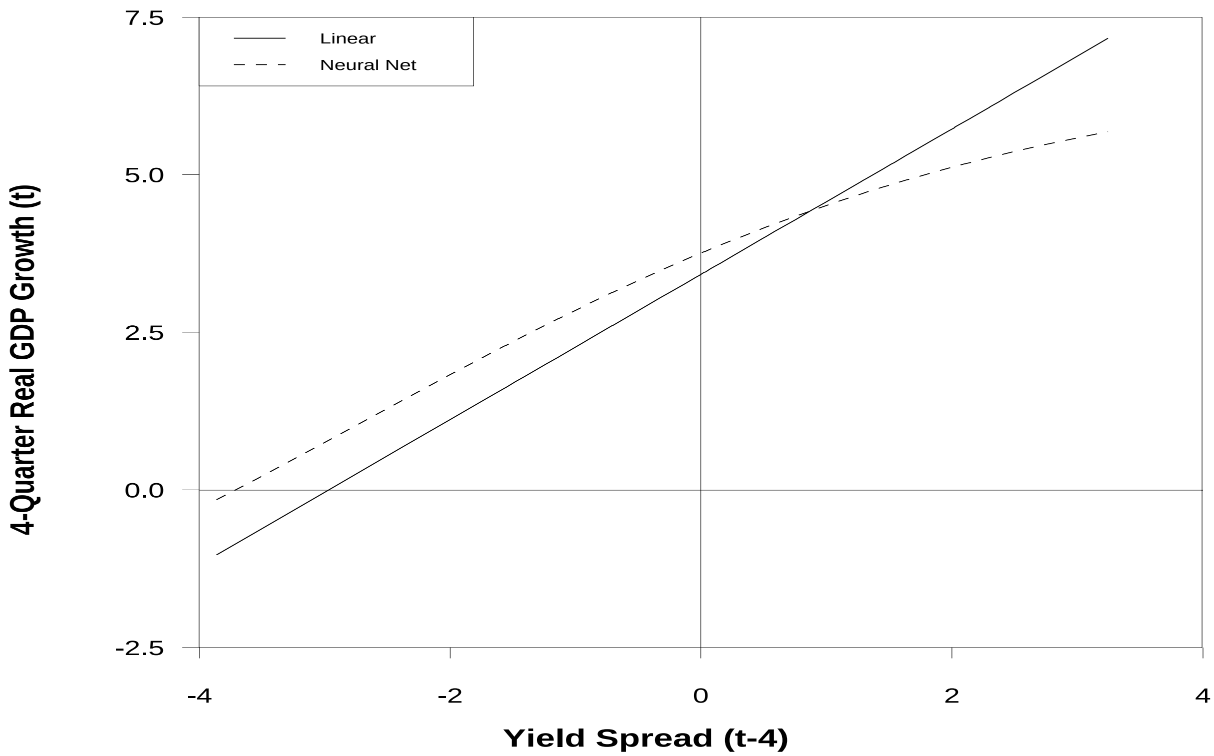


Figure 2: Linear and neural network fitted curves, 1968: 1 to 1998:1 (model 4a)

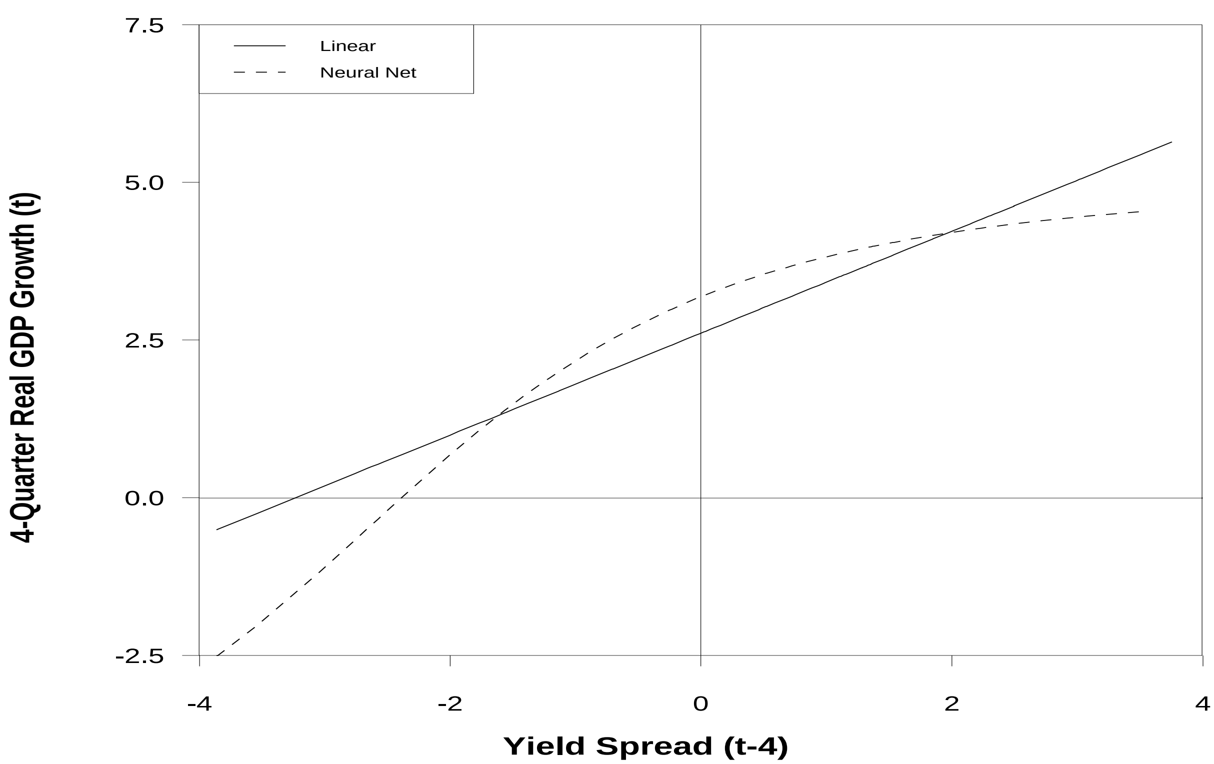


Figure 3: Forecasts of 1-quarter real GDP growth rate, 1985:2 to 1998:2 (model 1d)

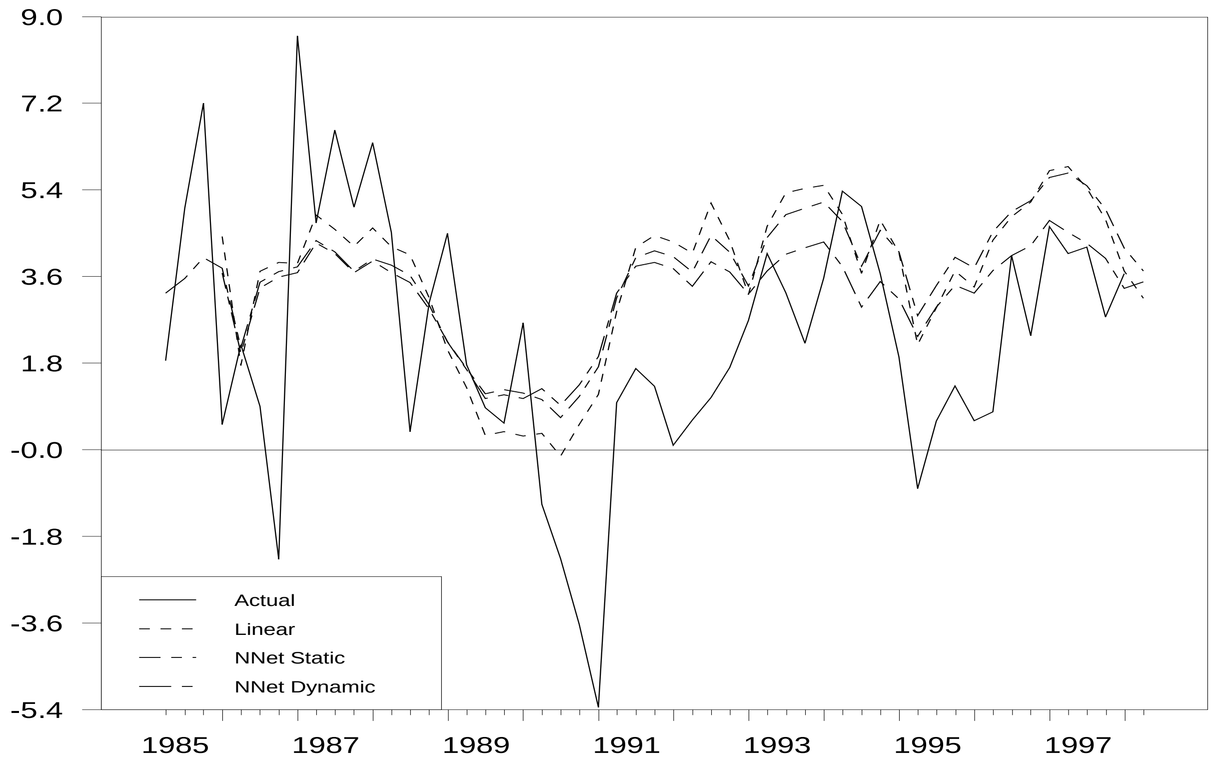
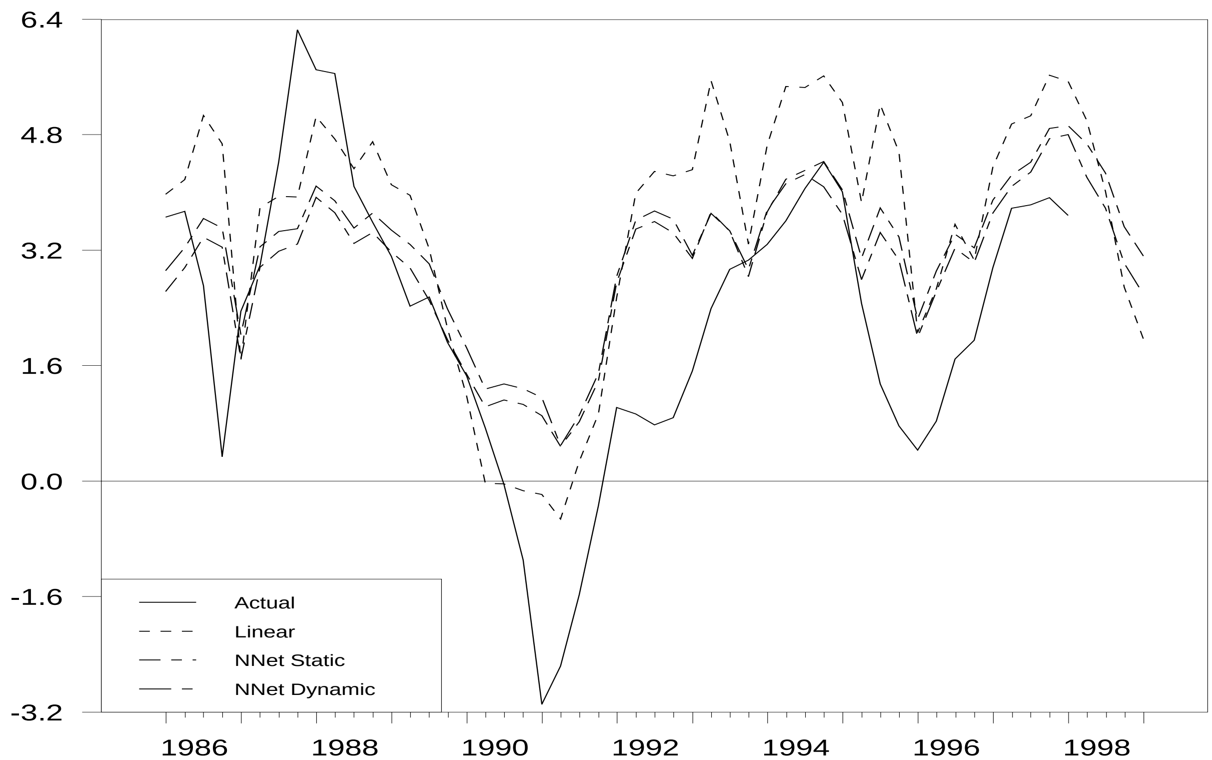


Figure 4: Forecasts of 4-quarter real GDP growth rate, 1986:1 to 1999:1 (model 4e)



Appendix

Forecast performance statistics

The criteria used to evaluate the accuracy of the out-of-sample forecasts are fairly well-known. The first three can be found in Holden, Peel, and Thompson (1990), while the fourth is discussed in Swanson and White (1997). In what follows, *A* represents the actual observation, and *F* the forecast.

- Root mean squared error:

$$RMSE = \sqrt{\frac{\sum (F_t - A_t)^2}{n}} \tag{A1}$$

- Mean absolute deviation:

$$MAD = \frac{\sum |F_t - A_t|}{n} \tag{A2}$$

- Theil U statistic:

$$U = \frac{RMSE^2}{(\sum A_t^2)/n} \tag{A3}$$

- Confusion rate: This measures the accuracy of the *directions* of the forecasts. This is the ratio of the number of incorrect forecast directions and the total number of forecasts. It lies between 0 and 1, being equal to 0 if all the forecast directions are correct, and 1 when all the forecast directions are incorrect.

References

Amano, R. and S. van Norden. 1993. “A Forecasting Equation for the Canada-U.S. Dollar Exchange Rate.” In *The Exchange Rate and the Economy*, 207–65. Proceedings of a conference held at the Bank of Canada, June 1992. Ottawa: Bank of Canada.

Angstenberger, J. 1996. “Prediction of the S&P 500 Index with Neural Networks.” In *Neural Networks and Their Applications*, edited by J. G. Taylor, 143–152. Chichester: Wiley and Sons.

Bailey, D. L. and D. M. Thompson. 1990. “Developing Neural Network Applications.” *AI Expert* (September): 34–41.

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay. 1997. *The Econometrics of Financial Markets*. Princeton: Princeton University Press.

Cover, J. P. 1992. “Asymmetric Effects of Positive and Negative Money Supply Shocks.” *Quarterly Journal of Economics* 107: 1261–83.

Ding, X., S. Canu, and T. Denoeux. 1996. “Neural Network Based Models for Forecasting.” In *Neural Networks and Their Applications*, edited by J. G. Taylor, 153–167. Chichester: Wiley and Sons.

Donaldson, R. G. and M. Kamstra. 1996. “Forecast Combining with Neural Networks.” *Journal of Forecasting* 15: 49–61.

Friedman, M. 1968. “The Role of Monetary Policy.” *American Economic Review* 58: 1–17.

Holden, K., D. A. Peel, and J.L. Thompson. 1990. *Economic Forecasting: An Introduction*. Cambridge: Cambridge University Press.

Hutchinson, J., A. Lo, and T. Poggio. 1994. “A Nonparametric Approach to the Pricing and Hedging of Derivative Securities Via Learning Networks.” *Journal of Finance* 49: 851–99.

Kohzadi, N., M. S. Boyd, I. Kaastra, B. S. Kermanshahi, and D. Scuse. 1995. “Neural Networks for Forecasting: An Introduction.” *Canadian Journal of Agricultural Economics* 43: 463–74.

Kuan, C.-M. and H. White. 1994. “Artificial Neural Networks: An Econometric Perspective.” *Econometric Reviews* 13: 1–91.

Maasoumi, E., A. Khotanzad, and A. Abaye. 1994. “Artificial Neural Networks for Some Macroeconomic Series: A First Report.” *Econometric Reviews* 13: 105–22.

Masters, T. 1993. *Practical Neural Network Recipes in C++*. San Diego: Academic Press.

McMenamin, J. S. 1997. “A Primer on Neural Networks for Forecasting.” *Journal of Business Forecasting* (Fall): 17–22.

Mehrotra, K., C. K. Mohan, and S. Ranka. 1997. *Elements of Artificial Neural Networks*. Cambridge: MIT Press.

Morgan, D. P. 1993. “Asymmetric Effects of Monetary Policy.” *Federal Reserve Bank of Kansas City Economic Review* 78: 21–33.

Muller, P. 1992. “The Information Content of Financial Aggregates During the 1980s.” In *Monetary Seminar 90*, 183–304. A Seminar Sponsored by the Bank of Canada, May 1990. Ottawa: Bank of Canada.

Nelson, C. and C. Plosser. 1982. “Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications.” *Journal of Monetary Economics* 10: 139–62.

Perron, P. 1989. “The Great Crash, the Oil Price Shock and the Unit Root Hypothesis.” *Econometrica* 57: 1361–1401.

Rhee, W. and R. W. Rich. 1995. “Inflation and the Asymmetric Effects of Money on Output Fluctuations.” *Journal of Macroeconomics* 17: 683–702.

Swanson, N. R. and H. White. 1997. “A Model Selection Approach to Real-Time Macroeconomic Forecasting Using Linear Models and Artificial Neural Networks.” *Review of Economics and Statistics* 79: 540–50.

White, H. 1992. *Artificial Neural Networks: Approximation and Learning*. Cambridge: Blackwell.

Bank of Canada Working Papers
Documents de travail de la Banque du Canada

Working papers are generally published in the language of the author, with an abstract in both official languages. ***Les documents de travail sont publiés généralement dans la langue utilisée par les auteurs; ils sont cependant précédés d'un résumé bilingue.***

1999

99-3	Forecasting GDP Growth Using Artificial Neural Networks	G. Tkacz and S. Hu
99-2	Capital Gains and Inflation Taxes in a Life-cycle Model	C. Leung and G.-J. Zhang
99-1	Dynamic Employment and Hours Effects of Government Spending Shocks	M. Yuan and W. Li

1998

98-23	Résultats empiriques multi-pays relatifs à l'impact des cibles d'inflation sur la crédibilité de la politique monétaire	P. St-Amant et D. Tessier
98-22	A Non-Paradoxical Interpretation of the Gibson Paradox	S. Coulombe
98-21	Une nouvelle méthode d'estimation de l'écart de production et son application aux États-unis, au Canada et à l'Allemagne	R. Lalonde, J. Page et P. St-Amant
98-20	Evaluating Alternative Measures of the Real Effective Exchange Rate	R. Lafrance, P. Osakwe, and P. St-Amant
98-19	Can a Matching Model Explain the Long-Run Increase in Canada's Unemployment Rate?	A. Hornstein and M. Yuan
98-18	The Sale of Durable Goods by a Monopolist in a Stochastic Environment	G. Srour
98-17	La politique monétaire a-t-elle des effets asymétriques sur l'emploi	L. Pichette
98-16	Consumer Attitudes, Uncertainty, and Consumer Spending	D. Côté and M. Johnson
98-15	On the Believable Benefits of Low Inflation	C. Ragan
98-14	Non-Linearities in the Output-Inflation Relationship: Some Empirical Results for Canada	C. Dupasquier and N. Ricketts
98-13	Le PIB potentiel des États-Unis et ses déterminants : la productivité de la main-d'oeuvre et le taux d'activité	R. Lalonde
98-12	Un examen de la crédibilité de la politique monétaire au Canada	P. Perrier
98-11	Liquidity Effects and Market Frictions	S. Hendry and G. Zhang
98-10	Fundamentals, Contagion and Currency Crises: An Empirical Analysis	M. Kruger, P. Osakwe, and J. Page
98-9	Buying Back Government Bonds: Mechanics and Other Considerations	T. Gravelle

Copies and a complete list of working papers are available from:
Pour obtenir des exemplaires et une liste complète des documents de travail, prière de s'adresser à:

Publications Distribution, Bank of Canada
234 Wellington Street Ottawa, Ontario K1A 0G9

Diffusion des publications, Banque du Canada
234, rue Wellington, Ottawa (Ontario) K1A 0G9

E-mail / Adresse électronique: publications@bank-banque-canada.ca
WWW: http://www.bank-banque-canada.ca/
FTP: ftp.bank-banque-canada.ca (login: anonymous, to subdirectory/pub/publications/working.papers/)