Neural Network Versus Econometric Models in Forecasting Inflation

SAEED MOSHIRI¹* AND NORMAN CAMERON²
¹Allameh Tabatabie University, Iran
²University of Manitoba, Canada

ABSTRACT
Artificial neural network modelling has recently attracted much attention as a new technique for estimation and forecasting in economics and finance. The chief advantages of this new approach are that such models can usually find a solution for very complex problems, and that they are free from the assumption of linearity that is often adopted to make the traditional methods tractable. In this paper we compare the performance of Back-Propagation Artificial Neural Network (BPN) models with the traditional econometric approaches to forecasting the inflation rate. Of the traditional econometric models we use a structural reduced-form model, an ARIMA model, a vector autoregressive model, and a Bayesian vector autoregression model. We compare each econometric model with a hybrid BPN model which uses the same set of variables. Dynamic forecasts are compared for three different horizons: one, three and twelve months ahead. Root mean squared errors and mean absolute errors are used to compare quality of forecasts. The results show the hybrid BPN models are able to forecast as well as all the traditional econometric methods, and to outperform them in some cases. Copyright © 2000 John Wiley & Sons, Ltd.

KEY WORDS Forecasting; inflation; structural models; time series models; artificial neural network models

INTRODUCTION
Artificial Neural Networks (ANN) have recently been used as research tools in several fields. Investigators have been attracted by ANN’s freedom from restrictive assumptions such as linearity that are often needed to make the traditional mathematical models tractable. Most uses of ANN in economics have so far been in financial markets, in part because traditional approaches have had low explanatory power and in part because the ANN approach requires abundant data. ANN models have outperformed the traditional time series models in most cases in forecasting stock prices and exchange rates, or in classifying applications such as bond ratings (Ahmadi, 1993; Bosarge, 1993; Kamijo and Tanigavwa, 1993; Sharda and Patil, 1993; Refenes, 1993; Donaldson and Kamstra, 1994; Lachtermacher and Fuller, 1995).

* Correspondence to: Saeed Moshiri, Faculty of Economics, Allameh Tabatabie University, Shahid Beheshti, Ahmad Ghasir Avenue, Tehran, Iran. E-mail: saeedmoshiri@hotmail.com

Copyright © 2000 John Wiley & Sons, Ltd.
There has been little work on forecasting macroeconomic series using the ANN models. Maasoumi, Khotanzad, and Abaye (1996) applied a back-propagation ANN model to forecast some US macroeconomic series such as the Consumer Price Index, unemployment, GDP, money and wages. Swanson and White (1997) applied an ANN model to forecast nine US macroeconomic series and compared their results with those from traditional econometric approaches. The results were mixed, but Swanson and White nevertheless concluded that ANN models were promising even where there is no explicit non-linearity.

More work is needed, obviously. In this paper we extend the approach of Swanson and White by comparing the most commonly used type of ANN (the Back-Propagation Network (BPN) model) with six traditional econometric models (three structural models and three time series models). The macroeconomic series which all models are asked to forecast is the inflation rate for Canada.

The paper is organized as follows. In the next section we explain the specification and present the estimated version of each of the six econometric models. The third section presents the single ANN contestant. The fourth section describes the data set and the forecasting strategy being used and the rules of the competition. The fifth section presents and compares the forecasting results for each competitor, and in conclusion awards the appropriate medals.

**ECONOMETRIC MODELS**

This section presents three structural models and three time series models. The three structural models include (1) the reduced-form inflation equation that follows from a fairly standard aggregate demand–aggregate supply model with adaptive expectations, (2) the inflation equation from Ray Fair’s econometric forecasting model, and (3) a monetary model for forecasting inflation. The three time series models are an ARIMA model, a Vector Autoregression or VAR model, and a Bayesian Vector Autoregression or BVAR model.

**Specifying the structural models**

The first structural model is derived from three equations commonly used to discuss cyclical behaviour in macroeconomics. The aggregate supply equation reflects supply and demand conditions in the labour market. In general form it is (Scarh, 1988):

\[ y_t = \alpha y_{t-1} + \beta (\pi_t - \pi^*) + e_t \]  

where all variables are in log form, \( y_t \) is the deviation of output from its natural value in period \( t \), \( \pi_t \) is actual inflation, \( \pi^* \) is expected inflation, and \( e_t \) is a supply shock. Equation (1) can be interpreted by New Classicals as the original Lucas price-surprise supply curve for a market-clearing economy, with some inertia added if the coefficient on the lagged output deviation is non-zero. It can be interpreted by Keynesians as an inverted, expectations-augmented Phillips curve reflecting some price-setting inertia by producers, offset by relatively vigorous response of aggregate supply to demand shocks (Romer, 1996). The special case where the coefficient on inflation is zero represents the extreme Real Business Cycle version, but such an interpretation will be inconsistent with what follows.
The aggregate demand (AD) equation is:

\[ y_t = \delta m_t + \gamma \pi_t^* + \zeta_t \]  

(2)

where \( m_t \) is the deviation of the log of real money supply from what would normally be held at the natural rate of output, and \( \zeta \) is a demand shock reflecting all other shift factors for total spending that would enter a standard Hicksian IS-LM model. The inflation expectations variable \( (\pi^*) \) affects aggregate demand through its effect on the real interest rate and therefore investment spending. For the expectations formation process determining expected inflation \( \pi^* \), we assume households and firms are sufficiently ill informed about the inflationary process that they use adaptive expectations, as in:

\[ \pi_t^* = \pi_{t-1}^* + \lambda(\pi_{t-1}^* - \pi_{t-1}) \quad 0 < \lambda < 1 \]

By recursive substitutions for \( \pi_{t-1}^* \) in the above equation, \( \pi_t^* \) can be approximated as a moving average of actual past inflation rates:

\[ \pi_t^* = \sum \rho_i \pi_{t-1-i} \]  

(3)

where \( \rho_i = \lambda(1 - \lambda)^i \). Solving equations (1) and (2) and substituting equation (3) for \( \pi^* \), gives us the following reduced form for inflation:

\[ \pi_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 m_t + \sum_{i=0}^{n} \rho_i \pi_{t-i} + \beta_3 \pi^m_t + \eta_t \]  

(4)

In this structural explanation, inflation is determined by the output gap of the previous period (reflecting lagged effects), by the current real money supply reflecting aggregate demand pressures, as well as by lagged values of inflation reflecting slow expectations adjustment, and by a mix of other demand and supply shocks. To adapt this standard explanation of inflation to the context of an open economy, we recognize current inflation of import prices \( (\pi^m_t) \) (over and above domestic price inflation) as an important supply shock and take it out of the residual term.

The second structural model is borrowed from Ray Fair’s medium-size structural forecasting model of the United States economy (Fair, 1984). Fair uses an equation for the price level, so we use its first difference:

\[ \pi_t = \beta_0 + \beta_1 \Delta y_{t-1} + \beta_2 \Delta w_t + \beta_3 \Delta \pi^m_t + \beta_4 \pi_{t-1} + \epsilon_t \]  

(5)

where \( w \) is the wage rate. Wage inflation is explained separately in Fair’s model, so we have approximated this jointly endogenous variable by the predicted rate of wage change as in the instrumental variables technique.\(^1\)

The third and most restrictive structural model is one in which only nominal money growth matters, though import price inflation is added to adopt the model to an open economy. When both monetary growth and import price inflation are allowed to have lagged effects through a Koyck lag, the reduced form equation is:

\[ \pi_t = \beta_0 + \beta_1 \Delta m_t + \beta_2 \Delta \pi^m_t + \beta_3 \pi_{t-1} + \epsilon_t \]  

(6)

\(^1\)As instruments we have used lagged wage inflation, lagged price inflation, and unemployment rate.
Estimating the structural models

Data are needed only for the inflation rate, GDP gap, money supply, and import price inflation. All are obtained from the CANSIM databank constructed by Statistics Canada. All data are monthly and cover the period 1970:1–1994:12. This period includes high and volatile inflation rates in the period 1973–1982 followed by low and stable inflation rates up to the 1990s. The inflation rate is calculated as the year-over-year change in the Consumer Price Index.

The GDP gap is calculated as the deviation of actual, seasonally adjusted monthly real GDP from the economy’s potential GDP for that same month. Potential GDP is estimated using the linear-quadratic trend of real GDP suggested by Armour et al. (1996). The money supply measure used is seasonally adjusted M1. The import price index is the seasonally adjusted Laspeyres index for all imported goods.

All variables are first checked for stationarity, to guard against the danger of spurious regression (Granger and Newbold, 1974). Ordinary augmented Dicky–Fuller unit root tests (Dickey and Fuller, 1979), reported in Appendix A, support the notion of unit roots for all the variables. For each variable the unit root hypothesis is rejected for the first difference, so the variables are at most I(1). Apparent non-stationarity of the series could be due to one-time shifts in the mean or trend of each series (Perron, 1989). Since Hostland (1995) has suggested that the Canadian inflation series has shifted abruptly once due to the 1973 oil price shock and then again due to the regime change in 1983, the series are all checked as well using Perron’s (1989) test for stationarity with known structural shifts.

The test equation for each variable $y$ is

$$
y_t = \mu + \delta_1 DUM1 + \delta_2 DUM2 + \beta t + \sigma y_{t-1} + \sum_{i=1}^{k} p_i \Delta y_{t-i} + \zeta_t
$$

where $DUM1$ equals 1 at and after the structural shock (in 1973:1 or 1983:1), and 0 beforehand; $DUM2$ equals $t - s$ (where $s$ is the date of the shock) at and after the shock, but 0 beforehand. Including $DUM1$ shifts the intercept of the trend, while including $DUM2$ changes the slope of the trend as well. The null hypothesis is $|\sigma| = 1$ (unit root) against the alternative hypothesis $|\sigma| < 1$. Perron (1989) shows that under the assumption of structural change, the Dickey–Fuller modified $t$–distribution is biased towards accepting non-stationarity, so he provides a more accurate distribution for the key coefficient. The results of Perron’s unit root test which are presented in Appendix B show that the unit root hypothesis still cannot be rejected for any of the three variables. We are therefore justified in first-differencing the variables before running any regression.

It is appropriate to test for cointegration before running the regression. The Johansen cointegration test (Johansen, 1988) produces a Likelihood Ratio of 54.016 for the notion of just one cointegrated vector of these variables, which is significant at the 5% level.$^2$ Therefore, an error correction term is included in the model to capture the influence on short-run dynamics of deviations from a long-run relationship among the variables. The lag length for the lagged first differences of inflation were chosen by the Akaike Information and Schwartz Criteria. The estimation results of the first structural model are as follows:

$$
\Delta \pi_t = \begin{array}{c}
0.004(\text{EC}) + 0.195 \Delta y_{t-1} - 0.009 \Delta m_t + 0.072 \Delta \pi_t^m + 0.20 \Delta \pi_{t-1} + 0.15 \Delta \pi_{t-2} + 0.11 \Delta \pi_{t-3} \\
\end{array}
$$

$^2$Eigenvalue = 0.096, LR = 54.016, and 5% critical value = 47.21.
$R^2 = 0.96$, AIC $= -10.22$, SC $= -10.12$, Ljung–Box $Q$-statistic for autocorrelation $= 113.07$, Breusch–Godfrey LM-statistic $= 18.68$, where EC is the error correction term.

The estimated form of the structural model derived from Fair is

$$
\pi_t = 0.002 + 0.94 \pi_{t-1} + 0.0001 \Delta y_{t-1} + 0.002 \Delta \nu_t + 0.023 \Delta \pi^m_t
$$

$R^2 = 0.98$, AIC $= -11.38$, SC $= -11.28$, Breusch–Godfrey LM-statistic $= 1.89$.

The estimated form of the purely monetary model is

$$
\pi_t = 0.002 + 0.93 \pi_{t-1} + 0.006 \Delta m_t + 0.025 \Delta \pi^m_t
$$


**Estimating the time series models**

*Autoregressive integrated moving average*

The first time series model is the single-variable model derived using the Box–Jenkins (1978) methodology. The model found to be the most suitable for the inflation series is as follows:

$$
\Delta \pi_t \equiv (\pi_t - \pi_{t-1}) = 0.089 \Delta \pi_{t-1} + 0.227 \Delta \pi_{t-2} + 0.257 \Delta \pi_{t-3} + 0.212 \Delta \pi_{t-4} - 0.023 \Delta \pi_{t-12}
$$

$R^2 = 0.51$, SSE $= 0.0028$, AIC $= -11.68$, SC $= -11.61$.

The regression includes a twelfth-order moving average term with correlation coefficient of $-0.92$ and $t$-statistic of $-52.69$. The Jarque-Bera (1987) asymptotic Lagrangian Multiplier normality test statistic of 4.29 suggests that the remaining residuals are normal. The Ljung–Box $Q$-statistic value of 26.53 for 36 lags supports the assumption of no serial autocorrelation problem for the model as specified above.

**VAR and BVAR**

The other time series approaches consider the joint behaviour of several variables. In each case the variables grouped together are the same as in the first structural model estimated above: inflation, outgap gap, real money supply, and imported inflation. The first vector autoregression model (VAR1) is in error-correction form, using the first differences of each variable (current and lagged), and the lagged error correction term. The third VAR model (VAR3) uses the levels of each variable, current and lagged, rather than first differences, on the assumption that while non-stationarity has not been rejected, neither has stationarity around a trend. For each model, the choice of number of lags is based on the AIC and SC criteria. The estimation results of the VAR models are presented in Appendix C.

In a Bayesian VAR, the VAR model is combined with prior information on the coefficients of the model and estimated using a mixed-estimation method (Theil, 1971). The prior distribution chosen for the coefficients is the Minnesota or Litterman prior (1986), shown as follows:

\[
\beta_i \sim N(1, \lambda), \quad i = 1
\]

\[
\beta_i \sim N(0, \lambda/i), \quad i = 2, \ldots, 4
\]

\[
\beta_j \sim N\left(0, \frac{\lambda}{j \sigma_j^2} \frac{\delta_j}{\sigma_j^2}\right), \quad j = 1, \ldots, 4
\]
for equation $i$, where $\beta_1$ is the coefficient of the first lag of the dependent variable with mean one and standard deviation $\lambda$. $\beta_i$ ($i = 2, \ldots, 4$) are the coefficients of the three remaining lags of the dependent variable with mean zero and standard deviation $(\lambda/i)$ and $\beta_j$ ($j = 1, \ldots, 4$) are coefficients of the lags of other variables with mean zero and standard deviation $\lambda/j(\sigma_i/\sigma_j)$.

$\lambda$ measures the tightness of the lags of the dependent variable and $\theta$ measures the tightness of the lags of the other variables. Following Litterman (1986), the values of $\lambda$ and $\theta$ are set to 0.2, which is a relatively loose value. As Sarantis and Stewart (1995) have found, the results are not very sensitive to the values of $\lambda$ and $\theta$, but the loose values produce better results. $\hat{\sigma}_i/\hat{\sigma}_j$ is a scale factor which balances the tightness of the standard deviations of the coefficients of the dependent variable with that of the other variables. $\hat{\sigma}_i$ is an estimated standard error of the residuals derived from a univariate autoregression with a constant and six lags. The estimation results of the BVAR model are also presented in Appendix C.

THE ARTIFICIAL NEURAL NETWORK (ANN) MODEL

ANN models can be classified very generally as vector mappers: they accept a set of inputs (an input vector) and produce a corresponding set of outputs (an output vector) according to some mapping relationship encoded in their structure (Wasserman, 1994). ANN models can also be viewed as non-linear input–output models with certain special features such as mass parallelism and non-linear processing of inputs which also found in biological neural networks. By trying to mimic these basic features of biological neural networks, as well as processing inputs several times in separate stages, ANN models have succeeded in doing certain jobs very well.

Back-Propagation Network (BPN) models are one of the most popular types of ANN models. They are static or feedforward-only (input vectors are fed through to output vectors, with no feedback to input vectors again); they are hetero-associative (the output vector may contain variables different from the input vector); and their learning is supervised (an input vector and a target output vector both are defined and the networks tend to learn the relationship between them through a specified learning rule). BPN models are easy to use in a variety of applications. A typical BPN model uses three vectors: input vector, one or more hidden vectors, and an output vector. After the input and output vectors are read into the BPN model, the network first selects parameters randomly and processes the inputs to generate a predicted output vector. After calculating the error between its predicted outputs and the observed outcomes, the network adjusts the parameters in directions that will reduce the error, generates a new output vector, calculates the error, adjusts its parameters again, and so on. The iteration or learning process continues until the network reaches a certain specified error. Figure 1 is a sketch of the stages of a typical BPN model with $r$ inputs, two hidden or intermediate units, and one output unit.

In a general form, the ANN output vector produced by a model or network consisting of $r$ input units, $q$ hidden units, and one output unit can be written as:

$$F(x, w) = F \left[ \beta_0 + \sum_{j=1}^{q} G(x_j)\beta_j \right]$$

(9)

3 For more detailed explanations of different ANN models see Wasserman (1994) and Haykin (1994).
where $F(x,w)$ is the network’s final output, $F$ is the activation function for the final step, $G$ is the activation function for a hidden or intermediate unit, $\hat{X} = [X_1, X_2, \ldots, X_r]$ is the input vector (including the intercept constant), and $W = (y_1, y_2, \ldots, y_q, \beta_j)$ is the parameters or weights matrix. Each term $y_i$ stands for a $r \times 1$ vector of weights relating to the $r$ input variables to one of the $q$ intermediate units. $\beta_j$ refers to a $q \times 1$ vector of weights relating each intermediate output vector to the final output vector. $F$ and $G$ can take any functional form, but the non-linear sigmoid function is a popular one, particularly for $G$.

The BPN model is able to learn all types of continuous functions provided that enough intermediate steps are allowed (Rumelhart, Hinton, and Williams, 1986; White, 1990). However, the learning in BPN is slow, particularly if the size of the network and the training data set are large.

Specifying the BPN model

To specify a BPN model we need to decide which variables go in the input vector, and how many hidden steps to include. In choosing variables for the input vector, we have matched what is used in the competing structural and time series models presented in the previous section; this will provide a level playing field for the forecasting competition. We therefore have six different BPN models using the input specifications of the structural model and the five time series models.

Choosing the number of hidden or intermediate units is more difficult. Although some formulae have been suggested in the literature for finding the optimal number of intermediate units, there is unfortunately no commonly agreed solution yet for this problem and different formulae can produce significantly different results. Therefore, we try from 1 to 10 intermediate units.

4 This is where the mass parallelism of ANN models first appears: the input vector is mapped into $q$ for different intermediate output units and then into final output unit simultaneously.

5 For instance, two of the suggested formulas for calculating the number of hidden units are as follows:

$$h = (rv)^{1/2}$$
$$h = T/[r + v]$$

where $h$ is the number of hidden units, $r$ and $v$ are the number of processing units in the input and output vectors, respectively, and $T$ is the number of patterns (observations) in the training data set (Masters, 1994; NeuralWork, 1994, p. 20). Applying the former to our problem generates $h = 4.1$ and the latter generates $h = 2.6$. 
units and choose the version that generates the minimum forecast errors. For example, in the case
where we use the same input vector as for the VAR model, the forecast error of the BPN model
decreases as the number of intermediate units increases from 1 to 5, and after that it increases, so
we chose 5 intermediate units for that version of the BPN model.

Non-linearity is an important feature of ANN models. To see how important this feature is to
the forecasting results, we also use an ANN model that is linear in its parameters: a BPN model
with no intermediate vectors at all. For the sake of comparison, we have used two different
versions of the linear BPN: the linear BPN with a linear activation (transformation) function,
and the linear BPN with a non-linear activation function. The former is exactly the same as the
linear regression model and the latter is the linear regression model with a non-linear functional
form.

Other parameters are set as follows: initial learning rate equals 0.01, where it changes with the
learning pace, and the number of training iterations allowed is 100,000. The common inter-
mediate vector activation function is the non-linear sigmoid function which returns the values
between zero and one. Since there are periods with a negative inflation rate in the sample data set,
we use the non linear tan-sigmoid function which returns the values between \(-1\) and \(+1\). The
output vector activation function is linear. The weights are set randomly at the start of the very
first forecast exercise (from 1991:1 on), but for later forecasts the initial weights are set equal to
those of the immediately previous forecast.

A specific learning rule commonly used in the BPN model is the ‘generalized delta rule’, which
updates the weight for each unit as follows:

\[
w(t + 1) = w(t) + \eta \nabla \]

where \(w(t)\) is the connection weight at time \(t\), \(\eta\) is the learning rate (typically less than 1), and \(\nabla\) is
the gradient vector associated with the weights. The gradient vector is the set of derivatives for all
weights with respect to the output error. BPN calculates the gradient vector on a vector-by-vector
basis using the chain rule for partial derivatives.

Training the BPN with the gradient descent method is very slow. In our problem, training each
BPN model and recursive forecasting take about 24 hours on a Pentium 166 machine. To get
better performance, we use the Levenberg–Marquardt (LM) learning method instead. The LM
learning method is an approximation of Gauss–Newton’s optimization rule and is more powerful
than gradient descent method. It is

\[
\Delta W = (H' H + \nu I)^{-1} H' E
\]

where \(H\) is the Jacobian matrix of derivatives of each error to each weight, \(\nu\) is the scalar, and \(E\) is
an error vector. The LM update rule approximates gradient descent if \(\nu\) is very large, and is
equivalent to the Gauss–Newton’s method if \(\nu\) is small. In the LM method, \(\nu\) changes as the
network trains. Since the Gauss–Newton method is faster and more accurate around the mini-
mum error, the network shifts the learning rule from the gradient descent to Gauss–Newton by
decreasing \(\nu\) when the error declines. The LM method is very fast, but memory intensive. In our
problem, training a BPN model and recursive forecasting takes 10 minutes on average. We use
Matlab’s Neural Network Toolbox to train the BPN models for inflation and to generate out-of-
sample forecast.
DATA AND FORECASTING STRATEGY

To test the robustness of the models, the dataset for 1970:1 to 1994:12 is divided into two parts: Initially 1970:1 to 1990:12 is used for estimation (training), and the remainder for out-of-sample forecasting (generalization). That is, there is a sample of twenty-one years with 252 observations for the estimation period and a sample of four years with 48 observations for the forecasting period, a ratio of 5.25 to 1. The estimation period starts with a relatively low inflation rate (around 5%), then turns to a high rate (including two subperiods of double-digit inflation rates in the mid-1970s and the early 1980s). The forecasting period includes the years of very low inflation rates for the early 1990s.

The forecasting results are all for out-of-sample forecasting, for periods whose behaviour the model has not been able to take into consideration. Our choice of forecasting period should provide a discriminating test of the competing models in that the forecasting period (1991 to 1994) contains some inflation behaviour quite unlike previous experience. For instance, introduction of the Goods and Services Tax in January 1991 imposed a one-time, 2.5% jump in the inflation rate. Similarly, two relatively large drops in the inflation rate during the forecast periods (1991 and 1993) are unprecedented in the previous ten years. Models that merely extrapolate from past behaviour without being sensitive to possible changes in regime will do poorly in these circumstances, while models that adapt quickly to new data will do better.

We apply a recursive forecasting strategy: after making the first set of forecasts with a model estimated using data for 1970:1 to 1990:12, we then extend the estimation data set by one period, re-estimate the model, and make a new set of forecasts. This process is repeated until the sample is exhausted. Each set of forecasts includes forecasts for one period, three periods, and twelve periods ahead (where the remaining sample allows). Therefore, there will be 48 estimations and sets of forecasts for each model, producing for each model 48 one-period-ahead, 46 three-periods-ahead, and 36 twelve-periods-ahead forecasts. At the final stage the forecast errors for all horizons are calculated and used to produce two summary criteria: Root Mean Square Errors (RMSE) and Mean Absolute Errors (MAE).

The forecasts can be either static or dynamic; static forecasts use actual values of lagged dependent variables, where these are required, whereas dynamic forecasts use the previously forecast values of these variables. Since the errors made in dynamic forecasting for periods early in the forecasting horizon contaminate forecasts for later periods as well, forecast errors in dynamic forecasts are expected to be greater than those in static forecasts. All estimation and forecasting is done using either E-views or Matlab (Demuth and Beale, 1995).

COMPARISON OF FORECASTING PERFORMANCE

Figure 2 shows the performance of the three econometric models vs the BPN model with the ARIMA specification. Table I summarizes the dynamic forecasting results obtained from the structural, time series, and BPN models and Table II their static forecasting results.

In general the values and the range of the RMSE of both static and dynamic forecasts rise with the change in the forecast horizon from one period to twelve periods, entirely as expected. However, since the static forecasts use the actual values of the lagged dependent variable over the forecasting period, the forecast errors do not change as much as they do in the case of the dynamic forecasts. The overall results shown in Tables I and II indicate that the ANN models are...
Figure 2. Static and dynamic forecasts using econometric and Back-Propagation Network models
able to generate forecasts as well as the time series models and, in some cases, to outperform them. In the dynamic forecasting contest between the BPN and the econometric models, all models did almost the same job in the one- and three-periods-ahead forecasting, but the BVAR outperformed the others in twelve-periods-ahead forecasting. In the static forecasting contest between the BPN and the econometric models, the BPN with the ARIMA and VAR model specification produced results almost as well as the time series models in one and three periods ahead, but outperformed them in forecasting twelve periods ahead.

Among the BPN models, the BPN model with the ARIMA specification produced the best results in all three forecast horizons. A single-vector BPN with the ARIMA specification (BPN2) even produced the least error among all other models in forecasting twelve periods ahead.

It is useful to compare each BPN model directly with the corresponding econometric model whose input specification it uses. Table III reports the ratio of the RMSE from the BPN model to the RMSE from the econometric model. The BPN model with the structural model specification outperformed the structural model in forecasting one and three periods ahead. The BPN model with the VAR specification outperformed VAR2 (the best of the VAR models) only in forecasting three periods ahead. The BPN model which uses the VAR specification but uses no hidden vectors is as good as VAR2 only in forecasting one period ahead; it is worse for the other two forecast horizons. Finally, all three of the BPN models which use the same input vector as the ARIMA specification (BPN1, BPN2, and BPN3) outperformed the ARIMA model by a wide margin in forecasting twelve periods ahead, but were outperformed by the ARIMA model in
forecasting only one period ahead. The BPN models had a slight edge over the ARIMA models in forecasting three periods ahead. It seems a safe generalization that the BPN model has been able to outperform econometric models over longer forecast horizons.

Fair and Shiller (1990) have suggested the information test as another method of comparing forecasting models. This test can discriminate between competing forecasting models even when they have very close RMSE values. The test compares the useful forecasting information contained in one model relative to another by regressing the actual values of the inflation rate on the forecast inflation values generated by a BPN and by an econometric model. If standard t-tests show the coefficient of the econometric model’s forecast to be insignificantly different from zero,

Table II. Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) of static forecasts for three horizons using Econometric and ANN models

<table>
<thead>
<tr>
<th>Criterion</th>
<th>RMSE</th>
<th>MAE</th>
<th>1 month</th>
<th>3 months</th>
<th>12 months</th>
</tr>
</thead>
<tbody>
<tr>
<td>Horizon:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Structural</td>
<td>0.6c</td>
<td>0.6</td>
<td>0.6</td>
<td>0.5</td>
<td>0.4</td>
</tr>
<tr>
<td>ARIMA</td>
<td>0.4</td>
<td>0.2</td>
<td>0.2</td>
<td>0.4</td>
<td>0.2</td>
</tr>
<tr>
<td>VAR1a</td>
<td>0.6</td>
<td>0.5</td>
<td>0.6</td>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td>VAR2</td>
<td>0.6</td>
<td>1.2</td>
<td>6.2</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>VAR3</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>BVAR</td>
<td>0.5</td>
<td>0.4</td>
<td>0.5</td>
<td>0.3</td>
<td>0.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Modelb (input vector used)</th>
<th>Artificial Neural Network (Back-Propagation Network) models</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPN1 (ARIMA)</td>
<td>0.5 0.6 0.7 0.4 0.4 0.5</td>
</tr>
<tr>
<td>BPN2 (ARIMA)</td>
<td>0.5 0.5 0.5 0.3 0.3 0.3</td>
</tr>
<tr>
<td>BPN3 (ARIMA)</td>
<td>0.5 0.5 0.5 0.3 0.3 0.4</td>
</tr>
<tr>
<td>BPN4 (VAR)</td>
<td>0.5 0.5 0.5 0.3 0.3 0.3</td>
</tr>
<tr>
<td>BPN5 (VAR)</td>
<td>0.6 0.5 0.9 0.5 0.5 0.8</td>
</tr>
<tr>
<td>BPN6 (Structural)</td>
<td>0.5 0.5 0.6 0.3 0.3 0.4</td>
</tr>
</tbody>
</table>

aVAR1: VAR with first-differences variables and an error correction term. VAR2: VAR with the first-differences variables only. VAR3: VAR with the non-stationary, undifferenced variables.
bBPN1: Normal BPN with ARIMA specification. BPN2 = BPN with ARIMA specification and no intermediate vector. BPN3 = BPN with ARIMA specification and linear activation function. BPN4 = Normal BPN with VAR specification. BPN5 = BPN with VAR specification and no intermediate vector. BPN6 = Normal BPN with structural model specification.
cAll numbers are in percentage values.

Table III. Comparing BPN and econometrics models with the same input vectors

<table>
<thead>
<tr>
<th>Models/horizons</th>
<th>1 months</th>
<th>3 months</th>
<th>12 months</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPN6 vs Structural</td>
<td>0.83a</td>
<td>1.6</td>
<td>0.38</td>
</tr>
<tr>
<td>BPN4 vs VAR2</td>
<td>1.2</td>
<td>0.75</td>
<td>1.8</td>
</tr>
<tr>
<td>BPN5 vs VAR2</td>
<td>1.1</td>
<td>1.12</td>
<td>1.18</td>
</tr>
<tr>
<td>BPN1 vs ARIMA</td>
<td>1.25</td>
<td>1.17</td>
<td>0.23</td>
</tr>
<tr>
<td>BPN2 vs ARIMA</td>
<td>1.25</td>
<td>0.83</td>
<td>0.15</td>
</tr>
<tr>
<td>BPN3 vs ARIMA</td>
<td>1.5</td>
<td>1</td>
<td>0.3</td>
</tr>
</tbody>
</table>

aThe numbers are ratios of the RMSE of each BPN to the RMSE of the corresponding econometric models. A number greater than one indicates that the econometric model has better performance and vice versa.
while the coefficient of the BPN model is significant, we infer that the information contained in the econometric model is completely contained in the BPN model and that the BPN model contains further information as well; that is, the BPN forecast is superior. If both coefficients are insignificantly different from zero, we infer that neither model contains useful information, or at least that the two forecasting models are equally useful. Finally, if both models have non-zero coefficients, we can conclude that both BPN and econometric models contain independent information, so neither is clearly superior.

The results of the information test for dynamic inflation forecasts coefficients and \( t \)-values for BPN and econometric model forecasts, in \( y_t = \beta_0 + \beta_{BPN} y_{BPN,t-1} + \beta_{ECON} y_{ECON,t-1} + \mu_t \),

<table>
<thead>
<tr>
<th>Forecast/horizon</th>
<th>1 Month</th>
<th>3 Month</th>
<th>12 Month</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPN (( t )-statistic)</td>
<td>1.010 (13.08)</td>
<td>-0.033 (0.48)</td>
<td>-0.014 (17.77)</td>
</tr>
<tr>
<td>&amp; ARIMA</td>
<td>0.039 (0.48)</td>
<td>0.97 (17.77)</td>
<td>0.101 (1.14)</td>
</tr>
<tr>
<td>BPN</td>
<td>1.06 (10.86)</td>
<td>0.68 (5.72)</td>
<td>0.47 (3.57)</td>
</tr>
<tr>
<td>&amp; VAR</td>
<td>-0.15 (-0.17)</td>
<td>0.30 (2.91)</td>
<td>0.106 (1.98)</td>
</tr>
<tr>
<td>BPN</td>
<td>1.028 (9.18)</td>
<td>0.013 (0.10)</td>
<td>0.39 (2.45)</td>
</tr>
<tr>
<td>&amp; VAR</td>
<td>0.018 (0.162)</td>
<td>0.982 (7.60)</td>
<td>0.46 (8.95)</td>
</tr>
<tr>
<td>BPN</td>
<td>1.07 (12.18)</td>
<td>0.79 (5.037)</td>
<td>0.459 (3.29)</td>
</tr>
<tr>
<td>&amp; Structural</td>
<td>-0.020 (-0.25)</td>
<td>0.178 (0.127)</td>
<td>0.059 (0.96)</td>
</tr>
</tbody>
</table>

The numbers in parentheses are \( t \)-statistics.

The results of the information test for dynamic forecasting, presented in Table IV, can be summarized as follows. In one-period-ahead dynamic forecasting, the information contained in the econometric models is contained in the BPN and the BPN contains further information; the BPN models are superior for all four comparisons. Over a three-period forecast horizon the BPN models are superior in two comparisons and inferior in two. Over a twelve-period forecast horizon, the BPN models are superior in two comparisons and equally good or bad in two.

The results of the information test, in general, are consistent with those obtained using the RMSE and MAE as criteria. In some cases where RMSEs are very close to each other, the information test can shed more light on the comparison between the models. For instance, according to the RMSE and MAE criteria, the performance of the BPN and econometric models are almost the same in one- and three-periods-ahead dynamic forecasting, but information test results indicate that BPN outperforms all other models in one-period-ahead dynamic forecasting. The information test results for twelve-periods-ahead forecasting are also consistent with previous results, in that the BPN models are shown to be equal or better than their econometric counterparts in dynamic forecasting.
APPENDIX A: ADF UNIT ROOT TEST RESULTS

\[ \Delta y_t = \mu + \gamma y_{t-1} + \beta t + \sum_{i=1}^{i} \rho_i \Delta y_{t-i} + \xi_t \]

<table>
<thead>
<tr>
<th>Variable</th>
<th>( \pi )</th>
<th>( \Delta \pi )</th>
<th>( y )</th>
<th>( \Delta y )</th>
<th>( m )</th>
<th>( \Delta m )</th>
<th>( \pi^m )</th>
<th>( \Delta \pi^m )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma )</td>
<td>-0.01</td>
<td>-0.663</td>
<td>-0.04</td>
<td>-0.776</td>
<td>-0.015</td>
<td>-1.34</td>
<td>-0.037</td>
<td>-0.7</td>
</tr>
<tr>
<td>ADF test statistic</td>
<td>1.166</td>
<td>-6.638</td>
<td>-2.491</td>
<td>-5.851</td>
<td>-1.07</td>
<td>-8.81</td>
<td>-2.77</td>
<td>-6.8</td>
</tr>
</tbody>
</table>

\[ a: \text{inflation}, y: \text{output gap}, m: \text{real money supply}, \pi^m: \text{excess import inflation}, \Delta: \text{first difference}. \]

\[ b: \text{ADF: The Augmented Dickey–Fuller test statistic.} \]

\[ c: \text{MacKinnon 1\% and 5\% critical values for rejection of the hypothesis of a unit root are} -3.99 \text{ and} -3.43, \text{ respectively.} \]

APPENDIX B: ADF UNIT ROOT TEST STATISTICS ASSUMING STRUCTURAL CHANGES

\[ y_t = \mu + \delta_1(DUM1) + \delta_2(DUM2) + \beta_t + \sigma y_{t-1} + \sum_{i=1}^{i} \rho_i \Delta y_{t-i} + \xi_t. \]

\( DUM1 = 1 \text{ and} DUM2 = \text{t from date of shock onwards; both} = 0 \text{ beforehand} \)

<table>
<thead>
<tr>
<th>Variables and trend specification</th>
<th>( \pi )</th>
<th>( y )</th>
<th>( m )</th>
<th>( \pi^m )</th>
<th>Critical values ( 1% ) (2.5%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( t^b )</td>
<td>0.689</td>
<td>0.969</td>
<td>0.99</td>
<td>0.96</td>
<td>-4.30 (-3.93)</td>
</tr>
<tr>
<td>( t_1 )</td>
<td>156.79</td>
<td>65.005</td>
<td>59.44</td>
<td>72.09</td>
<td>\text{null}</td>
</tr>
<tr>
<td>2</td>
<td>0.989</td>
<td>0.968</td>
<td>0.99</td>
<td>0.96</td>
<td>\text{null}</td>
</tr>
<tr>
<td>( t_2 )</td>
<td>156.96</td>
<td>65.015</td>
<td>59.06</td>
<td>66.96</td>
<td>-4.65 (-4.32)</td>
</tr>
<tr>
<td>3</td>
<td>0.974</td>
<td>0.980</td>
<td>0.98</td>
<td>0.96</td>
<td>\text{null}</td>
</tr>
<tr>
<td>( t_3 )</td>
<td>174.65</td>
<td>66.060</td>
<td>67.4</td>
<td>66.02</td>
<td>-4.32 (-4.01)</td>
</tr>
<tr>
<td>4</td>
<td>0.968</td>
<td>0.976</td>
<td>0.97</td>
<td>0.95</td>
<td>\text{null}</td>
</tr>
<tr>
<td>( t_4 )</td>
<td>102.284</td>
<td>63.685</td>
<td>57.93</td>
<td>63.30</td>
<td>-4.90 (-4.53)</td>
</tr>
</tbody>
</table>

\[ a: \text{Variables are} \pi: \text{inflation}, y: \text{output gap}, m: \text{real money supply}, \pi^m: \text{excess imported inflation.} \]

\[ b: \text{Specifications are} \#1: \text{permanent shift in intercept in} 1973:1. \#2: \text{permanent shift in intercept and change in slope in} 1973:1. \#3: \text{permanent shift in intercept in} 1983:1. \#4: \text{permanent shift in intercept and change in slope in} 1983:1. \]

\[ c: \text{Perron’s critical values with} \hat{\lambda} = 0.2 \text{ (for specifications 1 and 2) and} 0.5 \text{ (for specifications 3 and 4), where} \hat{\lambda} \text{ is the ratio of sample size before break point to total sample size.} \]

\[ \text{The number of lags} k \text{ (of the dependent variable) is} 3. \text{ The results are not sensitive to changes in} k. \]
ACKNOWLEDGEMENTS

We would like to thank Wayne Simpson, David Scuse, and Robert Dorsey for their valuable comments. The first author also thanks participants and discussants for their comments in the 5th Symposium of the Society for Non-linear Dynamics and Econometrics which was held in Washington, DC, in April 1997, and in the Canadian Economic Association conferences which were held in St Catherine and St John’s, Newfoundland, in June 1996 and 1997. All remaining errors are ours.

REFERENCES


White, H. ‘Connectionist nonparametric regression: multivector feedforward networks can learn arbitrary mappings’, *Neural Networks*, **3** (1990), 525–549.
Author’s biographies:
Saeed Moshiri is an assistant professor at the Faculty of Economics, Allameh Tabatabie University. He has been teaching macroeconomics and quantitative methods in economics at the University of Manitoba and the University of Winnipeg since 1994.

Norman Cameron, Professor of Economics, has been teaching at the University of Manitoba since 1969 in macro- and monetary economics.

Author’s Addresses:
Saeed Moshiri, Faculty of Economics, Allameh Tabatabie University, Shahid Beheshti, Ahmad Ghasir Avenue, Tehran, Iran.

Norman Cameron, Department of Economics, University of Manitoba, Winnipeg, Manitoba, R3T 2N2 Canada.