



PERFORMANCE OF ARTIFICIAL NEURAL NETWORKS IN FORECASTING COSTA RICAN INFLATION

Manfred Esquivel Monge

División Económica
Departamento de Investigación Económica
DEC-DIE-029-2009-DI, junio, 2009



Abstract

The underlying processes that generate economic series such as inflation, unemployment or output gaps are potentially quite complex. Undoubtedly that makes it very difficult to forecast them and has traditionally bent attention to relatively simple linear approaches when trying to model them. Trying to capture nonlinear relationships among inflation and its determinants, this paper applies Artificial Neural Networks (ANN) to forecast Costa Rican inflation. A innovative technique that systematically discriminates among different networks in order to overcome the problem of “over-fitting” a ANN was applied. Forecasts are compared with those obtained from “thick” models and traditional linear techniques. The potentially complex nonlinear relationships between inflation and its short run determinants in an expectations-augmented Phillips Curve scheme are captured with a systematically chosen ANN. Forecasts at different horizons are computed in a rolling exercise in order to test the hypothesis of a better performance of the nonlinear parameterization. Evidence shows that linear techniques do not outperform ANN and, in the case of a Phillips Curve, networks forecasts statistically improve upon linear approaches especially for short run forecast horizons. In most cases, “thick” modeled ANN’s forecasts showed a weak performance compared with systematically chosen ANNs.

Resumen

Los procesos generadores de series temporales económicas como inflación, desempleo o brecha del producto son potencialmente muy complejos. Esto sin duda dificulta la tarea de su pronóstico y ha hecho que tradicionalmente se opte por especificaciones lineales a la hora de modelar tales procesos. En el documento aplica la técnica de redes neuronales artificiales (RNA) para pronosticar la inflación en Costa Rica capturando potenciales relaciones no lineales entre esta variable y sus determinantes. Se utiliza una innovadora técnica que discrimina sistemáticamente entre diferentes arquitecturas de red para evitar el problema de “sobrentrenamiento”. Los pronósticos obtenidos a partir de este procedimiento de selección sistemática son comparados con aquellos obtenidos al aplicar el método de modelos “densos” y con los extraídos de técnicas lineales tradicionales de estimación econométrica. Las potenciales relaciones complejas y no lineales entre la inflación y sus determinantes de corto plazo en una curva de Phillips aumentada son capturadas con redes neuronales elegidas sistemáticamente. Con el propósito de contrastar el desempeño del método no lineal de RNA con el de métodos lineales, los pronósticos se efectúan para diferentes horizontes y siguiendo un ejercicio de “rolling”. La evidencia señala que las técnicas lineales no superan a las RNA y que, en el caso de la Curva de Phillips y especialmente para horizontes cortos de pronóstico, resultados obtenidos con las redes son estadísticamente superiores. En la mayoría de los casos los modelos “densos” de RNA mostraron un débil desempeño cuando se compararon con los de redes sistemáticamente elegidas.

JEL Classification: C45, C53, E37 Key words: Forecasting, inflation, neural networks.

esquivelmm@bccr.fi.cr

1 Introduction

The underlying processes that generate economic series such as inflation, unemployment or output gaps are potentially quite complex. Undoubtedly that makes it very difficult to forecast them and has traditionally directed attention to relatively simpler linear approaches when trying to model them and when facing the task of forecasting them.

Stock, J. and Watson, M. (1999) signaled the potential improvement in forecast that incorporating nonlinear relations between inflation and the chosen independent variables will yield when compared to a traditional Phillips Curve estimated by linear techniques. Some authors¹ have reported a performance of nonlinear forecasting methodologies that is at least as good as conventional linear parameterizations. For the case of the Costa Rican economy, Solera (2005), in a first approach to forecast inflation by using Artificial Neural Networks (ANN), arrived to the same conclusion with a rather monetarist set of information that included monetary aggregates and lags of the dependant.

When estimated by means of linear methods, structural specifications of Costa Rican inflation have shown a relatively poor forecast performance compared to non structural ones. This may be due to the absence of nonlinear relationships between variables that traditional econometric theory imposes.

ANN modelling has been gaining attention as an attractive technique for estimation and forecasting in economics. The chief advantage of the ANN methods is that they are free from the assumption of linearity that is commonly imposed in order to make the traditional methods tractable. Most of the applications have been developed in financial statistics and exchange rates². Applications on macroeconomic time series have been limited in number³. However, those works usually share a common limitation: they do not implement standard practices of ANN estimation such as *early* stopping, which has shown to be quite successful in some works⁴.

This paper evaluates whether for the case of Costa Rica and based on different sets of information, the nonlinear methodology of ANN significantly improves upon inflation forecasts obtained by traditional linear methods. Three models are analyzed: an expectation-augmented Phillips Curve, a Treasury Bills Model and a Monetarist Model based on the work of Solera *op. cit.* A specific technique for each model is selected based on *rolling* root mean square errors of out-of-sample forecasts and some tests to discriminate among them. The accuracy of non-nested models is then compared based on the same criteria.

When the ANN method is implemented, several parameters need to be set. As a byproduct of the test mentioned above, the paper exposes several procedures for setting an optimal network and implements two of them for contrasting their results.

¹ Bukhari and Hanif (2007), Chen and Swanson (2001), McNeils and McAdam (2004), Moshiri and Cameron (2000) and Nakamura (2005), for example.

² Fernández-Rodríguez, González-Martel and Sosvilla-Rivero (2000) and Refenes and White (1998) are two examples of papers on financial applications.

³ See Stock and Watson (1998) Chen, Racine and Swanson (2001) and Solera (2005) for instance.

⁴ See Nakamura *op. cit.*

The document is organized as follows: in the next section some basic elements of ANN theory are exposed giving special emphasis to the problem of optimally setting a specific network and exposing what the problem of over-fitting is and why is important to control for it. After that, in section three all methodological aspects are covered; the general method for comparing forecasts is explained, the three lineal models selected for comparison are shortly exposed and, with some detail, the two techniques for optimizing forecasts from networks are covered. Additionally, section three shortly explains Diebold and Mariano (1995) and Fair and Shiller (1990) tests for discriminating forecasts which are then applied for such purposes. Section four is devoted to briefly expose the main results; firstly a forecast accuracy comparison of techniques for each model is exposed, then the best techniques from each model are weighted against each other. Section five concludes and is followed by the bibliographical references and finally all relevant appendices.

2 Theoretical background

Since ANN models take a set of inputs and produce a matching set of outputs according to some mapping relationship, they can be loosely classified as vector mappers. In spite of its suggestive name, an ANN is simply a parameterized nonlinear function that can be fitted to data in order to get a desired forecast. The nonlinear function combines several building blocks on a transfer function which can be, for example, a hyperbolic tangent or a logistic function.

When an ANN is shaped, a specific architecture must be provided. This architecture includes the number of hidden nodes (“hidden neurons” in the networks jargon) and “hidden layers”. Additionally it is possible to allow a linear component (direct connections between input and output vectors) to be part of the net.

The basic structure of an ANN that has an input vector (X) of size m , q hidden neurons, one layer and an output vector can be written as follows:

$$F(X, W) = F \left[\omega_0 + \sum_{j=1}^q G(x \delta_j) \omega_j \right] \quad (2.1)$$

Where:

$F(X, W)$ = Output vector of the network.

F and G = Activation functions for the final step and the q hidden nodes respectively. Even though these functions can take any functional form, nonlinear sigmoidal ones⁵ are quite popular, especially for G .

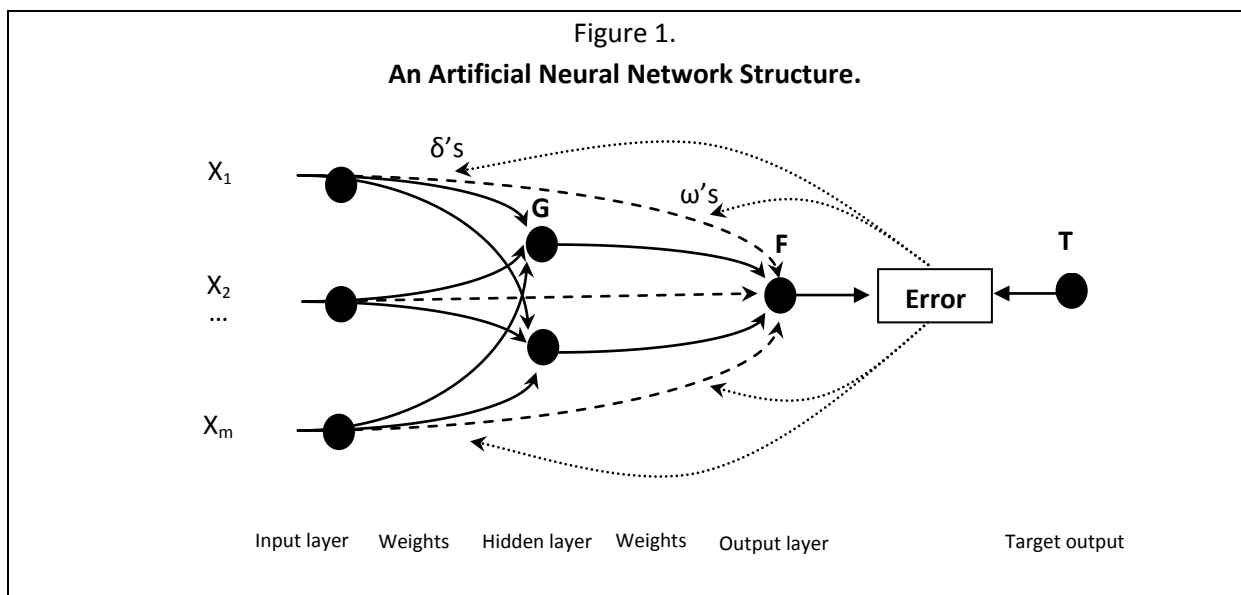
$X = 1, X_1, X_2, \dots, X_m$ = Input vector. Notice that it includes an intercept.

⁵ RATS 6.0 offers 3 options of sigmoidal functions for “squashing” the node outputs, a Logistic one: $1 / (1 + e^{-u})$; and two hyperbolic tangent: $\tanh u$ and $\tanh u / 2$. Where u is the basic output of a node. The logistic function was used in all estimations in this study.

$W = \delta_1, \delta_2, \dots, \delta_q, \omega_j$ = Matrix of parameters or weights. Each component δ_i stands for an $m \times 1$ vector of weights that maps the m input variables to one of the q hidden nodes. ω_j is a $q \times 1$ vector of weights relating each one of the q hidden nodes to the final output vector.

How exactly does an ANN fit the data for getting a forecasted value of the output series? Since a back-propagation technique is commonly used, we will be working only with back-propagation networks (BPN) models. The procedure goes as follows: Firstly, input and output vectors are read into the BPN model, then, after randomly selecting a set of parameters, the network processes the inputs and generates a predicted output vector. After that, a mean square error (MSE) is generated by comparing such output vector with the series of observed outcomes. Then the network adjusts the initial set of parameters in the direction of the negative gradient of the MSE, produces a new output vector, calculates a new MSE, adjust the parameters and so on.

Figure 1 might help to conceptualize how an ANN is built and how it fit input and output vectors.



The network in the figure is modeled with m input units (X), two hidden neurons (G) in a single hidden layer and one output vector (F). δ 's and ω 's are the hidden and output vector weights respectively. Dotted lines going directly from the X 's to the output vector represent the optional setting of direct (linear) connections.

The above described iterative process is known as *learning* or *training* process in the ANN jargon. One can *train* a net to *learn* the relationship between input and output vectors as many times as desired or until it reaches a given magnitude of the MSE.

Most of the ANN appeal rests on its flexibility to approximate a wide range of functional relationships between variables. Hornik, Stinchcombe and White (1989) explain how, given enough level of complexity (that is, a sufficiently high number of nodes and learning periods), neural networks are able to approximate any function arbitrarily well. Notice though that there might be two kinds of catches. Firstly, when an ANN is highly complex, the learning process becomes quite slow, especially if the data set is large. Secondly, one might reach a very low level of MSE in the training process, but ANNs trained a large number of times are prone to produce a good fit into the

learning sample but a disappointing performance in out-of-sample forecast. In section 3.3 there will be more to say about such problem and the *early stopping* procedures proposed to deal with it.

Having briefly explained the basic features of setting an ANN and how it works, it should be clear that given a set of information (that is a set of input and output vectors), the forecasted series obtained may vary depending on 3 factors: The number of hidden nodes (q), the number of training iterations that the network is permitted to take (Z) and whether direct connections are allowed. Additionally, even when a network is set by fixing q , Z and establishing whether there are any direct connectors, the output series might well vary due to the different set of randomly selected initial values of the weighting matrix. Yet, this last variation will be very small and not systematically biased⁶. When working with ANN, getting the most from a given set of information will consist in finding the combination of those factors that produces the best forecast performance. As it will be explained in section 3.3, there is a variety of proposals that try to accomplish this. Although none of them is theoretically superior, we can at least contrast which one is more backed by our data.

3 Methodological Approach and data.

3.1 General Approach.

The next general strategy was followed in the study. We start with three general non nested models whose dependent variable is inflation rate. An expectation-augmented Phillips Curve (PC), a Treasury Bills Model (TB) and a Monetarist Model (MM)⁷. For each model, different forecast series are computed by using at least three techniques⁸. Then, out-of-sample forecast performance statistics are computed in order to discriminate among techniques. Finally, the best forecast technique for each non nested model is compared with the other two best forecasts in order to determine whether there is a combination of model and technique that statistically outperformed the others.

For each model, one of the techniques is the traditional linear ordinary least square. The remaining two (or three) are nonlinear in the sense that they use ANN for getting the forecasts. It must be clear that the data set which is fed into each model is exactly the same among competing techniques. This is achieved by dividing the sample of available observations into two sections, a *training* section and a contrast section. The first one is used, in the case of linear models, for estimating OLS coefficients, and in the case of ANN techniques, for training the corresponding network and getting the W matrix of weights. The second one serves for comparing forecasted values with inflation realizations by means of constructing root mean square error statistic, which is the measure used for contrasting techniques and models.

⁶ Most modern software includes a “seed” instruction for controlling the random selection of parameters. RATS offers such instructions which ensures, at least into a single program execution, that the initial set of random weights is fixed, so it is possible to control for this variation.

⁷ Although the name used for this model is quite similar to the *Monetary Model* which is currently used at the Central Bank, it has to be clear that they are not the same.

⁸ In the case of the MM model, a fourth set of forecasts series was also estimated by using an ANN set according to Solera (2005). This was in order to quantify a possible gain from the *early stopping* procedure that is being firstly applied in this work to Costa Rican data.

Forecasts were obtained in an 8-step rolling scheme for one, two, four and eight quarters ahead horizons (h). Following this, in order to maximize the number of data points available, the date that determines the limit between *training* and *contrast* samples varies according to h and the specific rolling step. In any case, all model's first-rolling-step was set to start in 2007Q3, 2007Q2, 2006Q4 and 2005Q4 for $h=1,2,4$ and 8 respectively⁹. All series are quarterly and all forecasts are real time-static.

3.2 The Three Lineal Models.

Exposing the linear version of the three models might be a friendly way to clarify which information was provided and why those models were chosen in the first place. The set of information supplied in each case was based firstly on the possibility to compare results among non nested models (this is why quarterly data was used); secondly it was based on recent studies that identify variables, lags and samples for each model, this was subject to find a lineal version that did not have specification problems when quarterly data are used¹⁰.

3.2.1 Phillips Curve.

Most macroeconomic models include a version of a Phillips Curve describing the short run dynamics of prices. Despite its theoretical economic appeal, its forecast performance generally has not been superior to other non-structural specifications.

Shocks on expectations or output gap may well affect inflation quite differently across different economics situations, levels of the variables or its current variance. Then a likely reason for such a lack of forecast power may be that traditional econometrics directly assumes linear relationships among variables involved in the Phillips Curve when they might be nonlinearly related.

The linear version of the Phillips Curve analyzed in this study is expectation-augmented similar to the one that is part of the Quarterly Projection Macroeconomic Model (QPMM). The stochastic equation can be written down as follows:

$$\pi_t = \beta_1 E_t \pi_{t+4} + \beta_2 Y_{t-6}^b + \mu_t \quad (3.1)$$

Where:

π_t is year to year quarterly inflation rate in quarter t .

$E_t \pi_{t+4}$ is one year ahead inflation expectation conditional to information available on quarter t .

Y_{t-6}^b is the output gap in quarter $t-6$.

μ_t is a normal, zero-mean, constant-variance disturbance term.

⁹ Notice that according with such a scheme, each forecast series is build of eight point forecasts corresponding to eight steps in the rolling procedure.

¹⁰ The corresponding test of specification for each lineal version of the models can be found in appendix A.

From expression (3.1) the set of information that was given to the OLS and ANN's procedures corresponds to π_t (output vector) and $\left[E_t \pi_{t+4}, Y_{t-6}^b \right]$ (input vector). In this case the total sample goes from 1991Q1 through 2008Q4.

3.2.2 Treasury Bills Model.

Since October 2001, the Central Bank of Costa Rica carries out a forecast combination of inflation through grouping a set of models and producing a single forecast that combines all common and complimentary information. One of these models is the so-called *Treasury Bills Model*¹¹.

The latest version of the *Treasury Bills Model* uses monthly data and the stochastic version of its equation is the following:

$$\pi_t = \beta_0 + \sum_{i=1}^3 \beta_i \pi_{t-i} + \sum_{i=2}^3 \delta_i F_{t-i} + \varepsilon_t \quad (3.2)$$

Where:

π_t is year to year monthly inflation rate in quarter t .

F_{t-i} is year to year percentage change in the net balance of bond debt of the Costa Rican Government valued at market prices in quarter t .

ε_t is a normal, zero-mean, constant-variance disturbance term.

More recently, in Mora and Rodríguez (2009), the authors found that the Treasury Bills Model seems to show a better forecast performance among all models in the forecast combination. Additionally, they show that in some cases this model seems to encompass the other models when computing an inflation forecast.

Due to the evidence mentioned above, one may wonder whether is it possible to improve upon such a successful forecast performance by trying a nonlinear estimation that uses the same set of information as the Treasury Bills Model. Additionally, it would be valuable to know whether the best forecast from this model outperforms the one that is identified as the best from the Phillips Curve Model. In order to carry out this comparison, it was decided to estimate the *Treasury Bills* equation with quarterly data just as the Phillips Curve of the last section was fitted. Then, the equation finally estimated in this work was:

$$\pi_t = \beta_0 + \beta_1 \pi_{t-1} + \delta_1 F_{t-1} + \varepsilon_t \quad (3.3)$$

where all variables are the same as in (3.2) but quarterly measured.

¹¹ Details of how this model is incorporated to the forecast combination of inflation can be found in Saborío and Solera (2004).

It follows then that the set of information that was fed into the statistical procedures is formed by π_t (output vector) and π_{t-1}, F_{t-1} (input vector). In this case the total sample goes from 1997Q3 through 2008Q4 as this is the same sample used in the last documented estimation of this model in the Central Bank¹².

3.2.3 Monetarist Model.

The first work in which the ANN technique was implemented to forecast inflation in Costa Rica is Solera (2005). In this study the author compares the forecast performance of ANN with 6 models then used at the Central Bank to forecast inflation. The author tries three different sets of information for training his networks and, based on RMSE measures, concludes that the best set contains a series of inflation as output vector and the series of lags 1 through 4 of inflation and lags 1 through 24 of M1 as input vector.

The linear version of the model that was used to contrast results here is the following:

$$\pi_t = \alpha_0 + \alpha_1 \pi_{t-1} + \sum_{i=1}^8 \delta_i M1_{t-i} + \eta_t \quad (3.4)$$

Where:

π_t is year to year inflation rate in quarter t .

$M1_t$ is year to year percentage change of M1 in quarter t .

Again, in order to compare results with those obtained from the Phillips Curve and the Treasury Bills Model, quarterly information was used. This is different from what Solera (2005) reports since he uses monthly data. The sample of the estimation goes from 1992Q1 through 2008Q4.

3.3 Choosing Among Networks.

When an ANN is specified, several factors need to be set, all of which determine how accurate the network adjusts realizations of the output vector. Those factors comprise the selection of variables to be included into the input and output vectors, how many hidden layers the system will have, the number of hidden nodes (hereafter Q), whether to include direct connections and how many iterations the network is allowed to take (Z henceforth).

In choosing variables for the input vector, those included into the competing linear models were matched. Such a strategy ensures a sort of *fair play* for the forecasting competition.

As for the number of hidden layers, there were not many alternatives. The available software to manage ANN was RATS 6.0, which offers a single option of one hidden layer. Aside was the question of whether to allow for direct connection between input and output vectors. All estimations were

¹² See Durán and Rojas (2007).

carried out under both options (with and without direct connections) and consistently better results were achieved under the last assumption. Hereafter all ANN results will be from networks that do not include such direct relations.

When selecting Q , a more difficult question is faced. For finding the optimal number of hidden nodes, several formulae have been proposed in literature¹³, yet there is no theoretic solution for this problem. It has been found that significantly different results are obtained across formulae.

For dealing with the problem of which Q to use, a more empirical approach has been the rule. Solera (2005) and Nakamura (2005) for instance, select the number of nodes that minimizes the MSE into the training sample. When implementing this procedure, it is common that MSE will decrease up to certain Q , and then grow or become quite volatile for larger numbers of hidden nodes.

An interesting way to deal with the issue was proposed by Granger and Jeon (2004), they called it *thick modelling* strategy and it combines several ANN's forecasts which are based on different Q 's. For combining all such forecasts they construct, for each forecasted period, a *trimmed mean* of the forecasts that result from each network specification. McNeils and Mc Adam (2004) applied this *thick modelling* strategy by means of ranking the predicted values and then removing the $100 \cdot \alpha\%$ smallest and largest estimations. The remaining $100 \cdot (1 - 2\alpha\%)$ are averaged. These authors set $\alpha = 5\%$ and argue that the trimmed mean avoid the problem of finding optimal weights, which has to be faced in a more standard process of forecast combination.

The selection of Z is a more obscure subject because it has been little explored at least in macroeconomic applications of ANNs. This is not a trivial election due to the potential problem of over-fitting that networks undergo when they are allowed to iterate too many times. As explained above, these systems are able to approximate functions arbitrarily well given enough complexity. By means of increasing Z , it is possible to achieve smaller mean square errors into the *training* sample. Eventually an excessively trained network might produce a near perfect fit in-sample but quite disappointing out-of-sample predictions.

When dealing with the problem of over-fitting, which is not common in macroeconomic applications of ANNs but standard in other fields, several *early stopping* procedures are applied. Among the most common is the following *validation based* approach¹⁴. Firstly, the sample is split into a training set and a validation set. Then the training algorithm is executed only with the training set just until the MSE starts to grow on the validation set. If there is a problem of over-fitting, such turn will occur long before the minimum MSE is reached on the training set.

In view of these issues, we propose a method that systematically identifies a network based on its forecast performance; henceforth the method is referred to as *systematic*. Firstly we select Q by means of increasing it until the MSE starts to grow into the training set. For this we allowed Z to take a value of 10.000¹⁵. Then with Q fixed, we select Z following the *validation based* approach of

¹³ For example, Masters (1994) suggests the following: $Q^* = rv^{1/2}$ and $Q^* = T / [5r + v]$, where Q^* is the optimal number of hidden nodes, r and v are the number of series into the input and output vectors respectively, and T is the sample size of the training period.

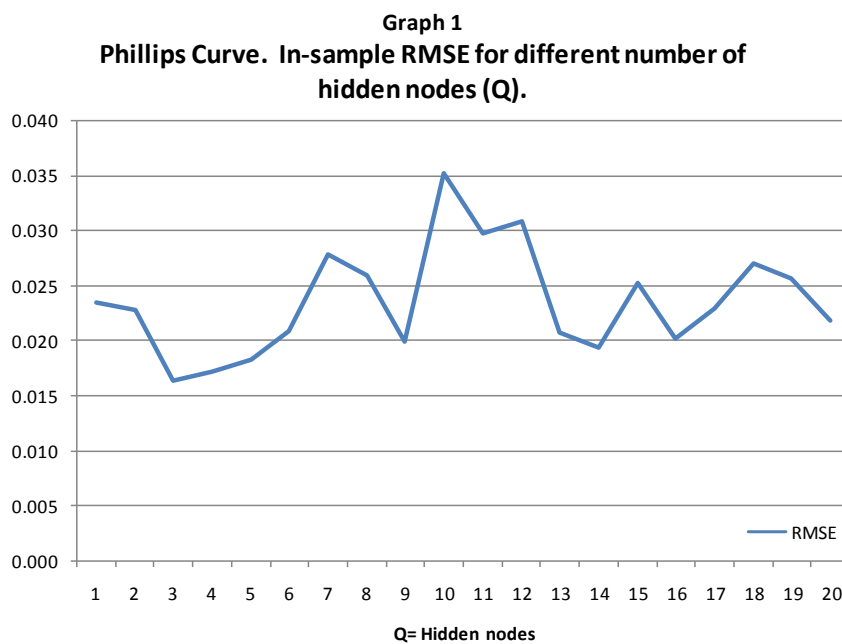
¹⁴ See Nakamura op. cit. for further details.

¹⁵ When selecting Z , this value was as large as possible for not causing over-fitting problems.

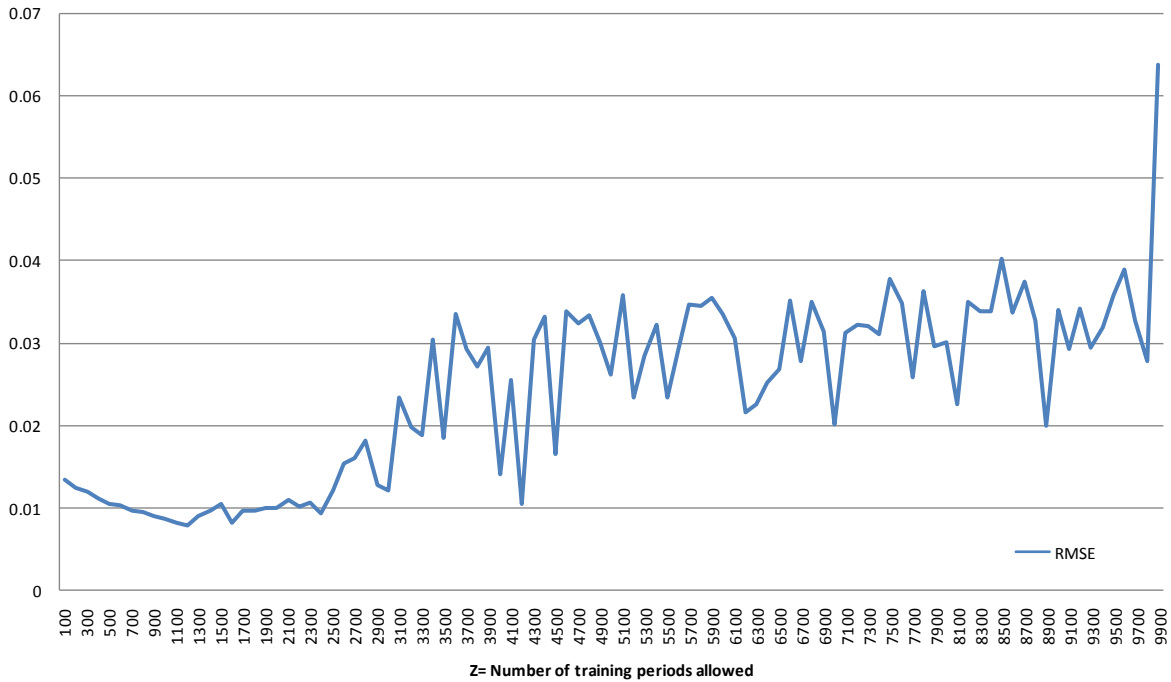
Nakamura (2005). In an iterative-nested-rolling process we set an initial value of $Z=100$, train the network and produce an h step ahead static forecast, we roll the train-forecast process for a two years period increasing the training set one period at the time, compute the MSE statistic and then repeat the process for $Z=200, 300, \dots 10.000$. We run a separate program for $h=1, 2, 4$ and 8 . This systematic procedure was applied to the three alternative models listed above. For each model and forecast horizon (h), it was then easy to identify from which level of Z the MSE starts to grow.

As an example for illustrating the systematic procedure, Graph 1 shows how Q was selected for the case of the Phillips Curve Model. In this case the RMSE decreases when the number of hidden nodes is increased, this happens up to $Q=3$, then this is the chosen number of nodes. Now, with the same Phillips Curve Model example, and after fixing $Q=3$, Graph 2 shows the selection of Z when $h=4$ in the rolling procedure. Notice that RMSE stop decreasing after $Z=1200$ which is then the number of training periods allowed in this specific case. Appendix B contains the entire collection of network's architectures and training periods for all the cases analyzed in the paper that were selected through this systematic approach..

Additionally, in order to contrast results of the systematic method, *thick models* combined through trimmed mean forecast were also computed for the three models and $h=1,2,4,8$. In this case Q was allowed to take values ranging from 1 up to 10 and the number of training periods (Z) was selected on the basis of what was found for each model- h case in the systematic scheme.



Graph 2.
Phillips Curve. Out-of-sample RMSE in rolling forecasts when h=4 for different number of training periods.



3.4 Comparing Predictive Accuracy.

As explained above, there will be three non-nested models: Phillips Curve, Treasury Bills and Monetarist. Then for each model there will be three (four in the case of the Monetarist Model) forecasting techniques: lineal (OLS), systematic ANN and thick ANN models. And for each technique, four forecast horizons: 1,2,4 and 8 quarters ahead. In each specific case, a two-year rolling procedure is applied in order to get the RMSE statistic¹⁶, which is the measure we use to discriminate among competing forecasts.

In the case of the Monetarist Model, a fourth set of rolling forecasts were also carried out with a network chosen as in Solera (2005), that is, selecting the Q that minimizes in-sample forecast performance and allowing as many training periods as necessary to reach an R^2 of 0.95. The idea is to determine whether there is any improvement in forecasting accuracy from applying the systematic or the thick models approach, which has not been tried before with Costa Rican data.

Firstly, the three (or four) techniques are compared into each non-nested model. This yields three *winning* techniques. After that, those winning techniques among non-nested models are compared yielding the model-technique combination that gives the best out-of sample forecast. Notice also that there are 4 of such separate competitions, one for each forecast horizon.

Here comes up the question of whether a given forecast is statistically better than another one. For answering this when dealing with the above explained forecast comparison, Diebold and Mariano (1995) and Fair and Shiller (1990) tests were used.

¹⁶ Notice that in all cases this will yield a series of 8 forecasted observations.

The first of those tests aims to contrast the null hypothesis of equal expected forecast accuracy against the alternative of different forecasting ability across models. If $L \bullet$ is some loss function defined over the errors of prediction, then the null hypothesis of the test can be written as:

$$H_0 : D_t = E \left[L \varepsilon_t^A - L \varepsilon_t^B \right] = 0 \quad (3.5)$$

where ε_t^i corresponds to the error of model i when performing an h -step ahead forecast. The test also uses the autocorrelation-corrected sample mean of D_t when testing (3.5). If m observations and forecasts are available, the test statistic is therefore:

$$S = \left[\hat{V} \bar{D} \right]^{-1/2} \bar{D} \quad (3.6)$$

Where:

$\hat{V} \bar{D} = m^{-1} \left[\hat{\gamma}_0 + s \sum_{k=1}^{h-1} \hat{\gamma}_k \right]$, the $\hat{\gamma}_0$ parameter is the variance of the D series and $\hat{\gamma}_k$ is the

corresponding covariance term which is obtained with the following formula:

$$\hat{\gamma}_k = \frac{1}{m} \sum_{t=k+1}^m (D_t - \bar{D})(D_{t-k} - \bar{D})$$

The authors show that under the null hypothesis of equal forecast accuracy S is asymptotically normally distributed. Then H_0 is rejected at 5% if $|S| > 1.96$.

On the other hand, the Fair and Shiller (1990) test is based on the argument that the information contained in one model's forecast compared to that in another can be discerned by means of a regression of actual values on predicted values from the two models. Such a test has two main advantages over the standard procedure of computing RMSEs when comparing alternative forecasts. Firstly, when the RMSEs of the two models are quite close, little can be concluded about the relative advantages of one model over the other. Secondly, it is possible that even when the RMSE of one model is significantly smaller, the other may contain information not included in the first one. The simple comparison of RMSE cannot test for this.

The procedure is as simple as follows. If ${}_{t-h}\hat{Y}_t^A$ is model's A forecast of Y_t made with information up to $t-h$ and ${}_{t-h}\hat{Y}_t^B$ is the same forecast but coming from model B. Then the next regression is run:

$$Y_t - Y_{t-h} = \alpha_0 + \alpha_A ({}_{t-h}\hat{Y}_t^A - Y_{t-h}) + \alpha_B ({}_{t-h}\hat{Y}_t^B - Y_{t-h}) + \varepsilon_{t+h} \quad (3.7)$$

Now, if none of the two models contains useful information for h -periods-ahead forecast, then we will not reject separately neither $H_0 : \alpha_A = 0$ nor $H_0 : \alpha_B = 0$. If both models have independent information for the forecast, then those last null hypotheses will be separately rejected. These two cases do not allow discriminating among the models. On the other hand, if for instance α_A is

nonzero, then model A contains further information than model B and it should be considered as a better forecaster.

4 Results.

For exposing our main results, first we will go across the three models exposed in section 3.2 showing forecast statistics for each of the techniques. That is, nested model comparison is shown first. Then non-nested comparison is carried out by showing forecast statistics of the best technique of each model. Diebold-Mariano and Fair-Shiller are shown in order to test consistency across specifications.

4.1 Phillips Curve Model.

Chart 1 compares forecast performance when Phillips Curve set of information is provided in the training (estimation) sample. Notice that for all horizons the smaller RMSE is reached by means of the systematic ANN technique. Yet, according with Diebold-Marino test, the difference is statistically significant only when compared with the thick model forecast. When forecast horizon is one quarter, the three techniques are equally effective.

Chart 1
Phillips Curve Model. RMSE in Rolling Forecast for Different Horizons

	1 quarter	2 quarters	4 quarters	8 quarters
ANN (systematic)	0.00863	0.00728*	0.00835*	0.00761*
ANN (thick model)	0.01010	0.01055	0.01053	0.01042
OLS	0.00892	0.00904	0.00919	0.00966

* Differs from Thick model, ** differs from OLS, *** differs from both techniques. (5% Diebold-Mariano)

Results from Fair and Shiller test is shown in Chart 2 which contains p-values of the corresponding parameters from a regression like (3.7). In this case Y_t is actual inflation rate, \hat{Y}_{t-h}^A is the series of rolling forecast from the systematic technique, and \hat{Y}_{t-h}^B is, alternatively, the rolling forecast from the thick model or the OLS method. Notice that, since the smaller RMSE statistic was obtained with the systematically selected ANN, it is being used as benchmark in the comparison. Henceforth, all Fair and Shiller charts of results will follow the same format, in all of them the t-statistic from the systematic coefficient will be shown as benchmark.

Chart 2
Phillips Curve. Fair and Shiller P-values for Different Methodologies and Forecast Horizons

	1 quarter	2 quarters	4 quarters	8 quarters
Systematic	0.3770	0.095*	0.2360	0.1120
Thick model	0.9920	0.9990	0.9346	0.7030
Systematic	0.5240	0.0711*	0.1771	0.023*
OLS	0.8000	0.6911	0.9895	0.6430

* Indicates a superior forecast performance (10%)

From Chart 2 it is possible to confirm that when the horizon is two quarters the systematically chosen ANN outperforms the thick modeled ANN. For all other horizons it is not possible to discriminate among those methods. Additionally, by means of this test, it is also possible to distinguish between Systematic ANN forecasts and OLS's at least for horizons of two and eight quarters, in both cases the first technique performed better in the rolling exercise.

In a nutshell, there is evidence suggesting that in the case of the Phillips Curve, systematically chosen ANN improve out-of-sample forecasts over thick and OLS's when the forecast horizon are two and eight quarters. Systematic ANN is also a better forecaster than OLS when h is two quarters or longer.

4.2 Treasury Bills Model.

What is shown in Chart 3 is equivalent to Chart 1 but it uses Treasury Bills Model and its set of information for generating forecasts. Again the smaller RMSE are achieved in the case of the systematic ANN method in all four horizons. In this case though, according to Diebold-Mariano test RMSE significantly differ only with respect to OLS technique when the forecast horizon is one year. In all other cases it is not possible to reject the null of an equal RMSE statistic.

Chart 3

Treasury Bills Model. RMSE in Rolling Forecast for Different Horizons

	1 trimestre	2 trimestres	4 trimestres	8 trimestres
ANN (systematic)	0.01113	0.01423	0.01280**	0.01357
ANN (thick model)	0.01247	0.01399	0.01314	0.01400
OLS	0.01378	0.01430	0.01455	0.01396

* Differs from Thick model, ** differs from OLS, *** differs from both techniques. (5% Diebold-Mariano)

When the other contrast is applied, there is not too much coincidence among results. Chart 4 includes P-values from applying Fair-Shiller test. Those results show that the systematic technique outperforms thick and OLS approaches only when the forecast horizon is one quarter.

Chart 4

Treasury Bills Model. Fair and Shiller P-values for Different Methodologies and Forecast Horizons

	1 quarter	2 quarters	4 quarters	8 quarters
Systematic	0.0058*	0.5559	0.9222	0.4481
Thick model	0.1116	0.0062	0.0000	0.0427
Systematic	0.0539*	0.1574	0.0656	0.1774
OLS	0.3575	0.1018	0.0139	0.0235

* Indicates a superior forecast performance (10%)

The case of Treasury Bills Models can be summarized by saying that there is evidence supporting the hypothesis that the systematic method is a better forecaster than the thick model and the OLS only when the horizon is one quarter. Additionally, when one year ahead forecast is made, OLS is outperformed by the systematic method.

4.3 Monetarist Model.

When the monetarist model is used the smaller RMSE for all horizons is still reached by means of using the systematic approach. This is exposed in Chart 5 which also shows that such technique achieves a statistically significant difference in all horizons when compared with the Solera op. cit. method. When horizons are less than 8 quarters, Diebold-Mariano test indicates that systematic RMSE statistic significantly outperformed also OLS. In this case, the thick models technique gets no different measure of out-of-sample forecast performance.

Chart 5
Monetarist Model. RMSE in Rolling Forecast for Different Horizons

	1 trimestre	2 trimestres	4 trimestres	8 trimestres
ANN (systematic)	0.01154**	0.01112**	0.01107**	0.01039*
ANN (thick model)	0.01245	0.01258	0.01351	0.01270
ANN (Solera 2005)	0.02012	0.01828	0.01777	0.02135
OLS	0.01317	0.01311	0.01287	0.01287

* Differs from Solera, ** differs from OLS and Solera, *** differs from Thick and Solera, ****differs from all techniques (5% Diebold-Mariano)

If we turn to what can be said from the Fair-Shiller test, Chart 6 shows that when the forecast horizon is two quarters or longer, the systematic approach outperformed Solera's method, this is consistent with chart 5. Forecast from the Systematic technique can also be distinguished as superior to OLS's when h is two or eight quarters. Finally, Fair-Shiller test cannot discriminate between Systematic and Thick modeled ANNs for neither horizon.

Chart 6
Monetarist Model. Fair and Shiller P-values for Different Methodologies and Forecast Horizons

	1 quarter	2 quarters	4 quarters	8 quarters
Systematic	0.0264	0.8698	0.5914	0.1544
Thick model	0.0259	0.2481	0.5232	0.9408
Systematic	0.1402	0.000*	0.000*	0.000*
Solera	0.6821	0.9075	0.5299	0.1785
Systematic	0.2525	0.006*	0.1313	0.0341*
OLS	0.5955	0.1623	0.7765	0.4935

* Indicates a superior forecast performance (10%)

In order to summarize the case of the Monetarist Model it is only necessary to say that the systematic approach outperforms Solera and OLS method and does not statistically differ from the Thick model technique.

4.4 Non-Nested Model Comparison.

From the above analysis, it is possible to conclude that, although not statistically different in all cases, at least in terms of absolute RMSE value the systematic ANN forecast shows higher accuracy in out-of-sample forecast. This is valid not only across models, but also for all horizons. Then by selecting this technique and comparing its forecast accuracy among non-nested models, it will be possible to determine which set of information, Phillips Curve, Treasury Bills or Monetarist, generates better results.

For those three sets of information, Chart 7 contains RMSE statistics from the same 8-step rolling exercise. The shorter indicator is achieved by the Phillips Curve in all four horizons. According with the Diebold-Mariano contrast, the difference is statistically significant when compared with the Treasury Bills method across all horizons. When the forecast is for shorter periods ahead, such difference becomes also meaningful when compared with what the Monetarist Model produces.

Chart 7
Non-Nested Models. RMSE in Rolling Forecasts for Different Horizons

	1 quarter	2 quarters	4 quarters	8 quarters
Phillips Curve (Systematic ANN)	0.00863**	0.00728**	0.00835*	0.00761*
Treasury Bills (Systematic ANN)	0.01113	0.01423	0.01280	0.01357
Monetarist (systematic ANN)	0.01154	0.01112	0.01107	0.01039

* Differs from Treasury Bills, ** differs from Treasury Bills and monetarist. (5% Diebold-Mariano)

When Fair-Shiller test is applied both the Phillips Curve and the Treasury Bills Model ANN's forecasts contain useful and independent information for all but the two quarters horizon, this means that it is not possible to identify one of the models as a superior forecaster than the other except when we consider a half-year ahead forecasts. On the other hand, the Phillips Curve set of information outperforms Monetarist's for horizons of two and four quarters, otherwise it is not possible to discriminate between them by means of this test. All this is shown in Chart 8.

Chart 8
Non Nested Models. Fair and Shiller P-values.

	1 quarter	2 quarters	4 quarters	8 quarters
Phillips Curve (Syst. ANN)	0.0000	0.000*	0.0000	0.0000
Treasury Bills (Syst. ANN)	0.0049	0.8836	0.0000	0.0487
Phillips Curve (Syst. ANN)	0.000*	0.000*	0.0000	0.0000
Monetarist Model (Syst. ANN)	0.2444	0.1152	0.0116	0.0000

* Indicates a superior forecast performance (10%)

In order to summarize the non-nested model comparison we can say that, either from Diebold-Mariano or Fair-Shiller test, for all horizons there is evidence of a better out-of-sample forecast accuracy of the Phillips Curve set of information over the Treasury Bills one. When shorter horizons are considered, there is also evidence that support Phillips Curve superiority over Monetarist Model. When we are interested in one year ahead or longer forecasts, the Phillips Curve and Monetarist Model are statistically alike.

5 Conclusions

Since main objective of this paper is examining whether to allow nonlinearities in some economic models of inflation estimated with Costa Rican data yields any improvement in forecast performance, conclusions will firstly cover such issue. It was found that ANN selected following systematic approach reduces RMSE of out-of-sample forecasts when compared with OLS method applied over the same set of information. Such reduction is statistically meaningful for two out of four horizons in the case of Phillips Curve and Treasury Bills Models and for all horizons in the case of the Monetarist Model. The conclusion supports the existence of relations among variables that are not fully captured by standard lineal econometric methods due to the nonlinear nature of some of them.

When comparing among techniques for selecting the best ANN forecast, the systematic approach achieved smaller RMSE statistics than the Thick models method, yet such difference is statistically important basically in the case of the Phillips Curve Model. Based on this, and considering that the Thick model technique requires specifying several factors in an *ad-hoc* manner (for instance: how many different architectures will be averaged, how many forecast will be excluded from the trimmed average, etc), the author would rather recommend the systematic approach over the Thick models technique.

When choosing among different architectures for ANNs, a systematic approach that controls not only for in sample fit of different number of hidden nodes but also for the number of training periods yields a significant improvement in forecasting accuracy. This comes from comparing in the case of the Monetarist Model, the performance of what we have called Solera's ANN with the systematic technique. In all horizons there is a statistically meaningful difference between those methods. In other words, it is worth controlling for over-fitting.

It was mentioned at the beginning of the paper that it is common for Phillips Curve relationships not to offer quite as good forecasts of inflation as some other non-structural specification when Costa Rican data are used. It was shown in the paper that this changes when nonlinear relations between variables are allowed. Phillips Curve processed by means of systematic ANN outperformed the Monetarist and Treasury Bills sets of information and the difference is statistically important when short run forecast horizons are considered.

Results from this paper suggest that systematic ANN's forecast of inflation may be used confidently for instance in early stages of the macroeconomic simulation exercises as a base scenario to feed into the Quarterly Projection Macroeconomic Model. Additionally, short run quarterly forecasts of inflation can also be obtained through systematic ANN Phillips Curve estimation, such forecasts significantly improve upon what can be obtained from other models or methodologies analyzed in this paper.

6 Bibliographical References.

- Chen, Xiaohong; J. Racine and N. Swanson (2001). *Semiparametric ARX Neural Network Models with an Application to Forecast Inflation*. IEEE Transactions on Neural Networks, 12, 674-83.
- Dielbod, F and R. Mariano (1995). *Comparing Predictive Accuracy*. Journal of Business and Economic Statistics, 3, 253-63
- Durán, R. and M. Rojas (2007). *Validación del modelo de títulos fiscales para el pronóstico de la inflación*. Informe técnico DIE-42-2007-IT. Banco Central de Costa Rica.
- Fair, R. and R.J. Shiller (1990). *Comparing Information in Forecasts from Econometric Models*. American Economic Review, 80, 375-389.
- Fernández-Rodríguez, F.; C. González-Martel and S. Sosvilla-Rivero (2000). *Neural Networks for Macroeconomic Forecasting: A Complementary Approach to Linear Regression Models*. Finance Canada Working Paper 2000-07
- Granger, C. and Y. Jeon (2004). *Thick Modelling*. Economic Modelling, 21, 2, 323-43.
- Hornik, K.; M. Stinchcombe and H. White (1989). *Multilayer Feedforward Networks are Universal Approximators*. Neural Networks, 2, 359-366.
- McNeils, P. and P. Mc Adam (2004). *Forecasting Inflation with Thick Models and Neural Networks*. Working Paper Series No. 352. European Central Bank.
- Masters, t. (1994). *Signal and Image Processing with Neural Networks*. John Wiley & Sons Inc.
- Mora, C. and A. Rodríguez (2009). *Combinación de proyecciones de inflación: Nuevas metodologías*. Banco Central de Costa Rica. Documento de Investigación DEC-DIE-10-2009-DI.
- Nakamura, E. (2005). *Inflation Forecasting using a Neural Network*. Economic Letters, 86, 3, 373-78.
- Refenes, A. and H. White (1998). *Neural Networks and Financial Economics*. International Journal of Forecasting, 6 (17).
- Saborío G. and A. Solera (2004). *Incorporación del modelo de saldos de títulos de propiedad del Gobierno Central en la combinación de proyecciones de inflación*. Nota Técnica DIE-DCS-01-2004-NT. Banco Central de Costa Rica.
- Solera, A. (2005). *Pronóstico de inflación en Costa Rica: Una estimación con redes neuronales artificiales*. Documento de investigación DCS-155-2005. Banco Central de Costa Rica.
- Stock, J. and M. Watson (1998). *A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series*. NBER Working Paper 6607
- Stock, J. and M. Watson (1999). *Forecasting Inflation*. Journal of Monetary Economics, 44, 293-335.

7 Appendixes

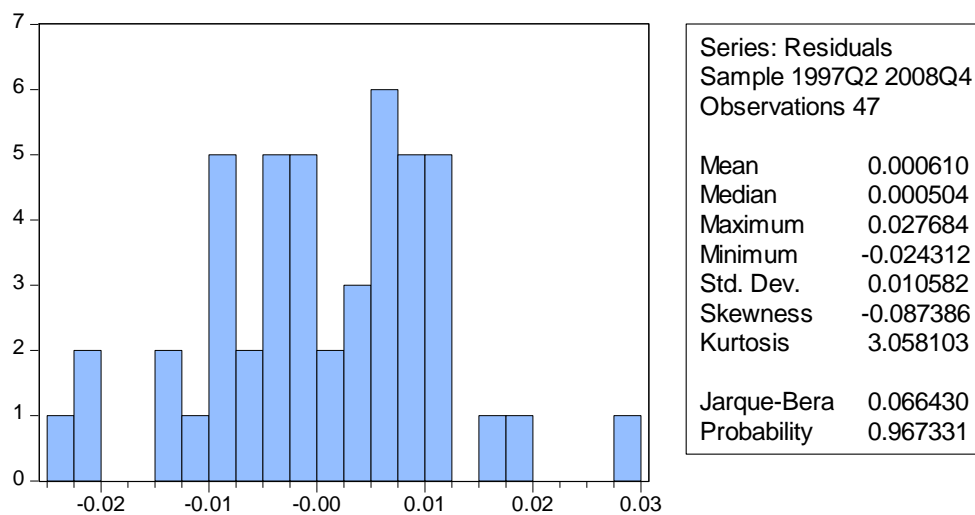
Appendix A.

Linear Models Adjustments and Test Specification Diagnosis

Phillips Curve.

Estimated equation: $\hat{\pi}_t = 1.006386 E_{t-1} \pi_{t+4} + 0.496096 Y_{t-6}^b$

Normality Test



Breusch-Godfrey Serial Correlation LM Test:

F-statistic	0.688531	Prob. F(2,42)	0.5079
Obs*R-squared	1.337765	Prob. Chi-Square(2)	0.5123

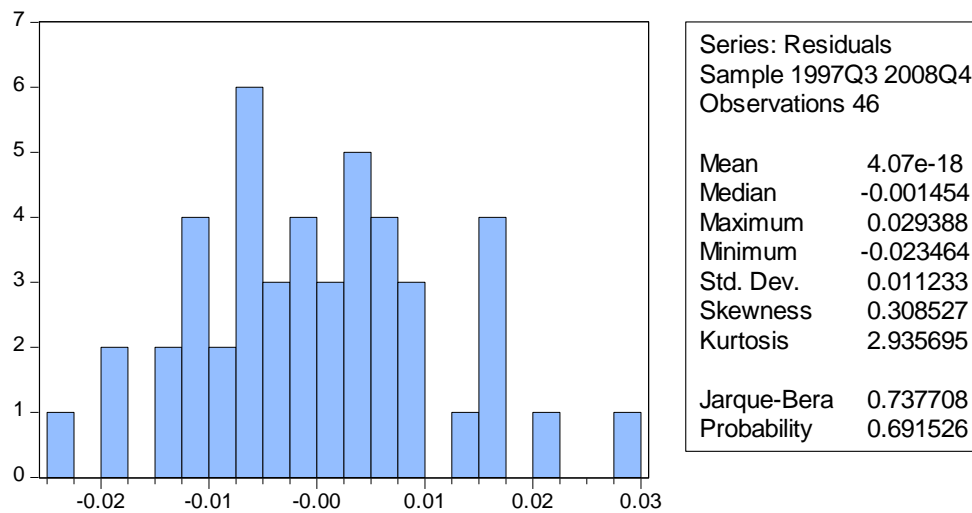
Heteroskedasticity Test: White

F-statistic	0.950438	Prob. F(2,43)	0.4248
Obs*R-squared	2.922748	Prob. Chi-Square(3)	0.4037
Scaled explained SS	2.609530	Prob. Chi-Square(3)	0.4558

Treasury Bills.

Estimated equation: $\hat{\pi}_t = 0.029622 + 0.766945 \pi_{t-1} - 0.02766 F_{t-1}$
0.011499 0.098733 0.016043

Normality Test



Breusch-Godfrey Serial Correlation LM Test:

F-statistic	1.450597	Prob. F(2,41)	0.2462
Obs*R-squared	3.039892	Prob. Chi-Square(2)	0.2187

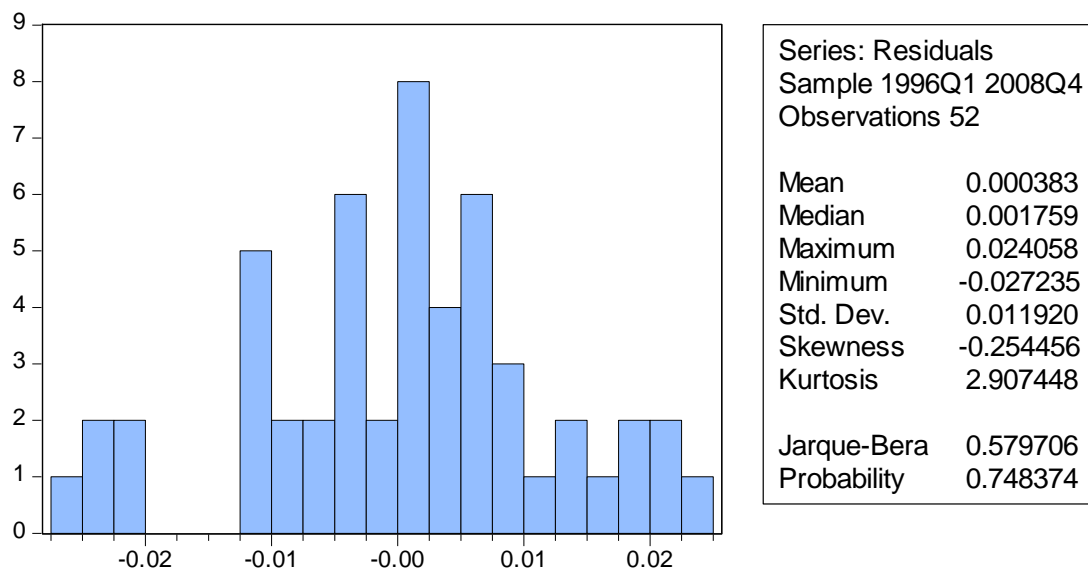
Heteroskedasticity Test: White

F-statistic	1.063498	Prob. F(5,40)	0.3949
Obs*R-squared	5.397574	Prob. Chi-Square(5)	0.3693
Scaled explained SS	4.564853	Prob. Chi-Square(5)	0.4713

Monetarist Model.

Estimated equation: $\pi_t = 0.84897 \pi_{t-1} + 0.039372 M1_{t-4} + 0.044385 M1_{t-8}$
0.030595 0.013364 0.01523

Normality Tets



Breusch-Godfrey Serial Correlation LM Test:

F-statistic	0.108350	Prob. F(2,47)	0.8975
Obs*R-squared	0.184310	Prob. Chi-Square(2)	0.9120

Heteroskedasticity Test: White

F-statistic	2.016472	Prob. F(3,48)	0.1241
Obs*R-squared	5.820036	Prob. Chi-Square(3)	0.1207
Scaled explained SS	4.844175	Prob. Chi-Square(3)	0.1836

Appendix B.

Chosen Neural Network Architectures in Systematic Method

Forecast horizon	Phillips Curve		Treasury Bills		Monetarist Model	
	Q	Z	Q	Z	Q	Z
1	3	1200	3	2200	5	600
2	3	900	3	1600	5	600
4	3	1400	3	1700	5	900
8	3	1300	3	1900	5	600