Revista de Análisis Económico, Vol. 13, N° 2, pp. 3-24 (Noviembre 1998)

MONEY DEMAND AND SEIGNIORAGE-MAXIMIZING INFLATION IN CHILE: APPROXIMATION, LEARNING, AND ESTIMATION WITH NEURAL NETWORKS*

PAUL D. McNELIS
Georgetown University

Abstract

This paper examines money demand and the seigniorage-maximizing inflation rates in Chile, with linear error-correction models (ECM) and artificial neural network (ANN) methods. The purpose is to approximate more accurately the "true" underlying non-linear functional forms for the long-run equilibrium demand for money, to estimate the learning process in short-run monthly adjustment of money stocks, and to obtain better estimates of the seigniorage-maximizing rates of inflation.
The ANN model shows that there is a high degree of non-linearity in the long-run demand for money in Chile. and that the seigniorage-maximizing inflation rates are much lower than the predictions of previous studies.

I. Introduction

This paper examines money demand and the seigniorage-maximizing rates of inflation in Chile, with error-correction models (ECM) and artificial neural network (ANN) methods. The purpose is to approximate more accurately the "true" underlying non-linear functional forms for the long-run equilibrium demand for money, to estimate the learning process in short-run monthly adjustment of money stocks, and to obtain better estimates of the maximal seigniorage that can be

* This paper was written with the support of the Office of Chief Economist, Interamerican Development Bank. Liliana Rojas-Suárez, Principal Advisor in this office at the time of this research, provided much of the data and encouragement for this research.

extracted by higher long-run rates of inflation. Unlike most previous studies, this paper explicitly incorporates parallel-market exchange-rate uncertainty in the short-run demand for money. In contrast with the ECM approach, which shows its importance, the ANN reveals the importance –indeed dominance– of hidden non-linearities in the demand for money. Uncertainty per se does not dominate in the non-linear model.

The use of ANN methods is relatively new in macroeconomics.[1] As forecasting tools, ANN methods have been used in the field of finance, for predicting asset prices, and for evaluating credit risk, as non-linear extensions of discriminant analysis. However, there has been little application, so far, to the analysis of money demand and monetary policy.[2] The reason is that ANN models (or architectures) are usually specified with a large number of parameters. The danger has been one of "overparameterization" and very poor out-of-sample performance, even with daily observations spanning many years. However, this paper will show the comparative advantage of ANN models–parsimoniously specified– over dynamic linear cointegration and error-correction models, with relatively noisy monthly data.

The need to approximate the underlying non-linearities in money demand has been pointed out by Soto (1995). Following Soto (1995), and Arrau, De Gregorio, Reinhart, and Wickham (1995), assume that the representative household will maximize an intertemporal utility function:

$$W_t = \sum_{i=0}^{\infty} \beta^i U(c_{t+i}) \tag{1}$$

subject to the following two constraints, a transactions technology constraint and a budget constraint:

$$\begin{aligned} G(c_t , m_t , \phi_t) &= c_t \, g(\frac{m_t}{c_t} , \phi_t), \;\; g_1 < 0, g_2 > 0 \\ b_t + m_t + c_t [1 + g(\frac{m_t}{c_t} , \phi_t] &= b_{t-1} (1 + r_{t-1}) + \frac{m_{t-1}}{1 + \pi_{t-1}} + y_t \end{aligned} \tag{2}$$

where c_t represents consumption at time t, β the social discount factor, m_t real money at time t, ϕ_t a transactions or technology shock at time t, b_{t-1} the stock of interest bearing bonds at time t-1, r_{t-1} the net interest yield on bonds held at time t-1, π_{t-1} the inflation rate at time t-1, and y_t the income or payment received at time t. The first equation in system (2) is the transactions constraint, and the second is the budget constraint.

The transactions constraint is a non-linear representation of the familiar cash-in-advance constraint used by Svensson (1985) and Lucas (1984). The exact functional form of the transactions technology in this constraint is a major issue relating to the functional form of the money demand function, and the optimal inflation rate.

Guidotti and Vegh (1993) and Lucas (1993), in particular, have pointed out that the optimality of a zero rate of inflation depends on a constant-returns-to-scale transactions technology.

In this model, transactions costs increase with higher consumption relative to money holdings, and fall with higher money holdings relative to consumption. The budget constraint explicitly incorporates these costs. These higher costs mean reduced consumption or lower bond or money holdings in time t, relative to time t-1.

From the first-order conditions of this problem, Soto as well as Arrau, De Gregorio, Reinhart and Wickham obtained the following relation between the marginal costs of holding money (in terms of foregone interest) and the marginal benefits (in terms of lower transactions costs):

$$\frac{\partial g(\frac{m_t}{c_t} , \phi_t)}{\partial (\frac{m_t}{c_t})} = \frac{- i_t}{1 + i_t} \tag{3}$$

If the function g is well behaved, its inverse exists, and one can thus obtain a demand for money function in the following familiar from:

$$m_t = h(c_t , i_t , \phi_t), \textit{where} \;\; h(.) = g(.)^{-1} \tag{4}$$

As both Soto and Arrau, De Gregorio, Reinhart, and Wickham point out, in order to obtain a testable specification of the demand for money, explicit utility and transactions costs functions need to be assumed. The point these authors make is that in order to obtain linear closed-form solutions for the demand for money, very restrictive utility and transactions costs functions are required. Soto shows, for example, that the following constant elasticity of substitution (CES) specification for transactions technology:

$$g(\frac{m_t}{c_t} , \phi_t) = e^{\phi_t} [K - \frac{\sigma - 1}{\sigma} (\frac{m_t}{c_t})^{\frac{\sigma - 1}{\sigma}}] \tag{5}$$

implies the following logarithmic demand for money function:

$$\log(m_t) = \phi_t + \log(c_t) - \frac{1}{\sigma} \log(\frac{i_t}{1 + i_t}) \tag{6}$$

Soto discusses the problems stressed in recent literature when micro-foundation utility functions were put to work explaining empirical regularities in macroeconomic data. Either the utility functions that "work" do not fit the underlying theory, or implausible elasticities of substitution are required to justify the evolution of aggregate data. Soto concludes that all the departures from simple linear (or log-linear) closed-form solutions generate "non-linear structures characterized by thresholds" or "time-dependent behavior on the part of agents" [Soto (1995), p. 8].

Rather than imposing specific functional forms on the utility and transactions-technology functions, this paper will use ANN methods to approximate the non-linear functional form given in equation (4).[3]

The alternative to the ANN specification is the linear error-correction model (ECM). Arrau, De Gregorio, Reinhart, and Wickham assumed a log-linear form of money demand, and tested for long-run cointegration between real money, a scale variable, and an opportunity cost variable. To capture financial innovation, these authors either specified a trend term or a time-varying intercept in the long-run cointegrating equation. While the use of a stochastic intercept as a proxy for financial innovation did yield plausible parameter values for the long-run money demand, this proxy may simply be picking up specification error, given the restrictions these authors place on the underlying transactions technology. This paper shows that the ANN approximation outperforms the ECM model.

Unlike Arrau, De Gregorio, Reinhart, and Wickham, this paper does not attempt to isolate the effects of financial innovation in the demand for money. Given the history of inflation in Latin America, this paper makes use of GARCH transformations to capture the effects of uncertainty, and employs the neural network model to model the learning processes in short-run adjustment. While learning behavior is clearly related to financial innovation, there is no attempt to distinguish between them.

The relationship of the functional form of money demand and seigniorage was recently examined by Easterly, Mauro, and Schmidt-Hebbel (1995). Making use of an intertemporal optimizing model with money and bonds, as well as a cash-in-advance model, these authors derive a money demand function with a variable semi-elasticity of substitution:

$$\ln\left(\frac{m}{y}\right) = k + \lambda\,\pi^{\gamma} \tag{7}$$

Following Calvo and Leiderman (1992), they make use of the following measure of the opportunity cost of holding money:

$$\frac{\pi_t}{1+\pi_t} \therefore \frac{\frac{[p_t - p_{t-1}]}{p_{t-1}}}{1+\frac{[p_t - p_{t-1}]}{p_{t-1}}} \tag{8}$$

These authors estimate the money demand model in equation (7) with annual data, for eleven countries, and calculate corresponding seigniorage-maximizing inflation rates. For Chile, the rate is infinity, both in the "level" and "first-difference" version of equation (7), while the corresponding rate for Mexico ranges between 127 and 227 percent in the level version, and infinity in the first-difference version, while for Peru, the maximal rate ranges between 127 and 294 in the level model, and between 333 and 376 percent in the first-difference version.

This study addresses this issue with monthly data, since the appropriate time horizon for the demand for money is likely to decrease as inflation rises. Secondly, inflation uncertainty (proxied by the conditional variance of the parallel-market exchange rate) enters as an additional argument of the demand for money. This paper estimates both the ECM model and an ANN adaption of an ECM model, evaluate the performance of the models, and find considerably lower seigniorage-maximizing rates of inflation under the ANN monthly models than the corresponding rates found by Easterly, Mauro, and Schmidt-Hebbel (1995).

The next section is a discussion of the ECM model and the way it is adapted for approximation and short-run money demand. Section III is a discussion of the ECM and ANN results, as well as an analysis of the seigniorage-maximizing inflation rates implied by the ANN model. The last section concludes.

II. Money Demand with Error-Correction and Neural Network Approximation

Estimation of the demand for money in developing countries with a history of high and variable inflation must take into account both inflation and depreciation as well as the uncertainty associated with inflation and depreciation. Following the work by Sweeney (1988), Asilis, Honohan and McNelis (1993), and McNelis and Rojas-Suárez (1996), this paper makes use of the GARCH conditional variance of depreciation in the parallel markets as the appropriate proxy variable for uncertainty of inflation and depreciation (see Bollerslev, 1986). It is assumed that the parallel-market exchange rate s_t, in logarithmic terms, follows a random walk, with a disturbance term ε_t with mean zero and conditional variance $\hat{\sigma}^2$:

$$\begin{aligned} \Delta s_t &= \varepsilon_t,\ E(\varepsilon_t) = 0,\ Var(\varepsilon_t) = \hat{\sigma}_t^2 \\ \hat{\sigma}_t^2 &= \alpha_0 + \alpha_1 \hat{\varepsilon}_{t-1}^2 + \alpha_2 \hat{\sigma}_{t-1}^2 \end{aligned} \tag{9}$$

The econometrically-generated series $\{\hat{\sigma}^2\}$ in turn serves as an argument in the short-run money demand function. While the uncertainty proxy, itself a non-linear transformation of the prediction error of the exchange rate, may not be very important in linear regression analysis of money demand, the ANN methods capture the major importance of this variable.

Making use of unit root and cointegration properties, this paper then estimates a long-run money demand equation, in which the log of real money (m-p)

is a linear function of a scale variable (y) as well as a subset of relevant opportunity cost variables, the domestic interest rate [i/(1+i)], the foreign interest rate [i*/(1+I*)], the rate of inflation [π/(1+π)], and the rate of depreciation in the parallel market for currency, [ε/(1+ε)] ≡ [Δs/(1+Δs)]. By "relevant opportunity cost variables", this paper means those variables which have the magnitudes and signs in agreement with *a priori* theoretical expectations: the scale variable should have positive effects, with an elasticity close to unity, and the semi-elasticities associated with the opportunity cost variables should have negative values:

$$m-p=\alpha_0+\alpha_1 y+\alpha_2\frac{i}{1+i}+\alpha_3\frac{i^*}{1+i^*}+\alpha_4\frac{\pi}{1+\pi}+\alpha_5\frac{\varepsilon}{1+\varepsilon} \tag{10}$$

The short-run demand for money function represents the first difference of the dependent variable (*m-p*) as a function of its own lags, the lagged cointegration vector *v* estimated from equation (10), the lagged first differences of the arguments of equation (10), the lagged exchange rate uncertainty proxy, generated by the GARCH process, and a dummy variable for the December *aguinaldo* or bonus payment:

$$\begin{aligned}\Delta(m-p)_t &= \Sigma\beta_i\Delta(m-p)_{t-i}+\delta v_{t-1}+\Sigma\gamma_i\Delta y_{t-i}+\Sigma\lambda_i\Delta[\frac{i}{1+i}]_{t-i}\\ &+\Sigma\mu_i\Delta[\frac{i^*}{1+i^*}]_{t-i}+\Sigma v_i\Delta[\frac{\pi}{1+\pi}]_{t-i}+\Sigma\rho_i\Delta[\frac{\varepsilon}{1+\varepsilon}]_{t-i}+\kappa\hat{\sigma}^2_{t-1}+\phi DECEM_t\end{aligned} \tag{11}$$

where:

$$v=(m-p)-[\hat{\alpha}_0+\hat{\alpha}_1 y+\hat{\alpha}_2\frac{i}{1+i}+\hat{\alpha}_3\frac{i^*}{1+i^*}+\hat{\alpha}_4\frac{\pi}{1+\pi}+\hat{\alpha}_5\frac{\varepsilon}{1+\varepsilon}] \tag{12}$$

The parameters of equations (12) and (11) may be estimated sequentially, or as a non-linear system:

$$\begin{aligned}\Delta(m-p)_t &= \Sigma\beta_i\Delta(m-p)_{t-i}+\delta[(m-p)-(\alpha_0+\alpha_1 y+\alpha_2\frac{i}{1+i}+\alpha_3\frac{i^*}{1+i^*}+\alpha_4\frac{\pi}{1+\pi}+\alpha_5\frac{\varepsilon}{1+\varepsilon})\\ &+\Sigma\gamma_i\Delta y_{t-i}+\Sigma\lambda_i\Delta[\frac{i}{1+i}]_{t-i}+\Sigma\mu_i\Delta[\frac{i^*}{1+i^*}]_{t-i}\\ &+\Sigma v_i\Delta[\frac{\pi}{1+\pi}]_{t-i}+\Sigma\rho_i\Delta[\frac{\varepsilon}{1+\varepsilon}]_{t-i}+\kappa\hat{\sigma}^2_{t-1}+\phi DECEM_t\end{aligned} \tag{13}$$

Estimation of (12) and (11) in a two-step procedure, or non-linear estimation of equation (13) follows Hendry's general-to-specific approach. A liberal lag length is chosen, and the model is gradually reduced from the general equation to a specific equation, as insignificant variables are eliminated from the model.

The long-run cointegration equation in equation (13), as well as the linear error-correction mechanism, imply strong restrictions not only about the underlying transactions technology, but also about the learning process, about the way economic agents reaction to new information, and correct past "mistakes" and forecast errors.

The advantages of the artificial neural network estimation, as opposed to linear or non-linear ECM models, non-linear least squares, or higher-order polynomial approximation, are two-fold: its system of parallel processing, and the use of the logsigmod function as a representation of learning behavior.

The ANN system is a system of *parallel* processing, in addition to *sequential* processing, in forecasting. Besides the usual input variables x and an output variable y, the artificial neural network makes use of a hidden layer of "neurons", which are simply logsigmoid transformations of linear combinations of the inputs. The neurons or "squashers" of the input variables work in parallel fashion, attempting to produce accurate fits in a contemporaneous parallel way of predicted outputs to actual outputs. This approach approximates a wide class of functions more accurately and more efficiently than other methods based on a polynomial, spline, or trigonometric expansion.

The most common neural network is the feedforward network, in which each neuron in the hidden layer is a logsigmoid transformation of a linear combination of the current variables in the input layer. By contrast, the Elman network is a logsigmoid transformation of a linear combination of lagged neurons as well as the current variables in the input layer. The Elman network allows a second structure for dynamics in the model, beyond the dynamic structure captured by lags in the input layer.

Estimation of the coefficients of a neural network is by backpropagation. The weights connecting the inputs to the neurons, and the neurons to the outputs are first initialized by a random number generator. Then a gradient search method, which attempts to minimize the sum of squared residuals, updates the "weights" or coefficients after each pass through the data set until predetermined convergence criteria are met.

The use of the logsigmoid "squashing" function, $1 / [1 + \exp(-\omega x)]$, where ω is a set of weights attached to the input variables x, yields a "threshold behavior", which captures economic reactions to news, and learning in the formation of expectations. At first, changes in the level of the fundamental input variables x may produce only small changes in the output, since the neuron is not activated, but as the levels of x reach critical thresholds, the neurons "fire", and reactions in the level of y are strong.

The nonlinear threshold effect thus captures economic reactions to news and learning behavior characterized by "bounded rationality". At first a small change in inflation, for example, may have little effect on the demand for money. As the inflation effect persists or increases, however, the effects on demand for money may be more pronounced, as the economic agents incorporate the new higher inflation, through learning, into their longer-term expectations.

The ANN serves in this paper both as an approximation method (for the functional form of the long-run money demand), as well as a model of the "learning" (in the error-correction process for short-run money stock adjustment). The ANN in money demand is an example of "bounded rationality" in macroeconomics, whereby agents don't know more than the econometricians about the underlying functional forms for transactions technology or utility. As Sargent (1993) points out, even with the complement –imitation being the sincerest form of flattery– macro-econometricians have not rushed to ANN, since these methods definitely do not reduce the number of parameters to explain data. Even with the additional parameters, however, this paper shows that the ANN provides sufficient gains in explanatory power and insight in money demand to justify the increased complexity.

ANN estimation would approximate a more general form of equation (13):

$$\begin{aligned}\Delta(m-p)_t = \Phi(&\Delta(m-p)_{t-i}, [(m-p)_{t-1} - \Psi(y, \frac{i}{1+i}, \frac{i^*}{1+i^*}, \frac{\pi}{1+\pi}, \frac{\varepsilon}{1+\varepsilon})_{t-1}],\\ &\Delta y_{t-i}, \Delta[\frac{i}{1+i}]_{t-i}, \Delta[\frac{i^*}{1+i^*}]_{t-i},\\ &\Delta[\frac{\pi}{1+\pi}]_{t-i}, \Delta[\frac{\varepsilon}{1+\varepsilon}]_{t-i}, \kappa\hat{\sigma}^2_{t-1}, DECEM_t)\end{aligned} \tag{14}$$

More specifically, the ANN strategy for equation (14) consists of two stages, first by approximating Ψ, the non-linear relation implied by the transactions technology, through a "sufficiently complex" feedforward network, and secondly by estimating Φ, the short-run learning, error-correction, and adjustment process, through a fairly simple Elman recurrent networks, taking the output of the first-stage network as an input into the second-stage network. This paper thus follows a strategy used by Altman, Marco and Varetto (1994), who employed a system of interconnected networks to predict financial distress in Italian firms. The advantage of this approach is that it significantly reduces computing time and the danger of oscillating behavior in the parameter estimation of a larger network.

The approximation of the long-run cointegration relation in the above model Ψ is done through the following feedforward system:[4]

$$\begin{aligned}(m-p)_t &= \gamma_0 + \sum_{i=1}^{i^*} \gamma_i N_{t,i}\\ N_{t,i} &= \frac{1}{1+e^{-n_{t,i}}}\\ n_{t,i} &= b_i + \sum_{j=1}^{j^*} \omega_{ij} x_{t,j}\\ x_t &= [y\ \frac{i}{1+i}\ \frac{i^*}{1+i^*}\ \frac{\pi}{1+\pi}\ \frac{\varepsilon}{1+\varepsilon}]_t\end{aligned} \tag{15}$$

For approximating the long-run cointegration function Ψ with system (15), one employs the following adaptation of sequential network construction (SNC), discussed by Moody and Utans (1995): for the given inputs x, one starts with a network architecture of j* = 2, and gradually increase the number of neurons, j*. The approach this paper follows thus goes from the simple to the more complex. Other search algorithms for network selection involve going from more complex architectures to progressively more simple ones, through gradual reduction of inputs in sensitivity-based pruning (SBP), or gradual reduction in the number of neurons in the hidden layer, through "optimal brain damage" (OBD) [Moody and Utans (1995), p. 277].

Most search algorithms for final network selection, such as SNC, SBP, or OBD, make use of out-of-sample criteria, such as the predictive stochastic complexity criterion (PSC) of Kuan and Liu (1995), based on minimizing the root mean squared error. In contrast to these approaches, this paper uses the in-sample Hanan-Quinn information criterion. With this criterion, one seeks to minimize the sum of squared residuals for a given j*, but one also penalizes oneself for greater complexity, measured by a larger parameter set. The final network is that which minimizes the following function:

$$hqif(j^*) = n \log[ssr(j^*)] + \log[\log(n)]\, k(j^*) \tag{16}$$

where n is the number of observations, and k is the number of parameters {b ω γ} in the progressively more complex networks.

This paper uses the in-sample Hanan-Quinn criterion because the work is with a model whose variables are selected by *a prior* theoretical restrictions. The goal is not to improve forecasting performance, per se, but rather to obtain coefficient estimates and elasticities implied by the data, consistent with intertemporal optimization and equilibrium conditions. Neural networks are used to approximate the underlying functional form of the transactions technology and utility functions, not to generate the best out-of-sample forecasts. Minimization of the Hannan-Quinn criterion function serves as a useful section device for the final network, since it trades-off increased explanatory power of a network architecture with increased complexity or learning effort.

The neural network approach to the analysis of cointegration is similar to the non-linear cointegration analysis of Granger (1996), who shows that it is possible to introduce wide classes of non-linearity into cointegration, the basic step of staring with I(1) processes that can be replaced with any form of persistence and a form of cointegration maintained.[Granger (1996), p. 8].

Once the feedforward network is selected and trained, the cointegration term or error becomes an argument in the short-run function:

$$\begin{aligned}\Delta(m-p)_t = \Phi(&\Delta(m-p)_{t-1}, [(m-p)_{t-1} - \hat{\Psi}(y, \frac{i}{1+i}, \frac{i^*}{1+i^*}, \frac{\pi}{1+\pi}, \frac{\varepsilon}{1+\varepsilon})_{t-1}],\\ &\Delta y_{t-i}, \Delta[\frac{i}{1+i}]_{t-i}, \Delta[\frac{i^*}{1+i^*}]_{t-i},\\ &\Delta[\frac{\pi}{1+\pi}]_{t-i}, \Delta[\frac{\varepsilon}{1+\varepsilon}]_{t-i}, \kappa\hat{\sigma}^2_{t-1}, DECEM_t)\end{aligned} \tag{17}$$

Since equation (17) is a non-linear dynamic model which embodies error-correction and learning, this paper uses an Elman (1988) recurrent network to represent this system:

$$\Delta(m-p)_t = \gamma_0 + \sum_{i=1}^{i^*} \gamma_i N_{t,i}$$

$$N_{t,i} = \frac{1}{1+e^{-n_{t,i}}} \qquad (18)$$

$$n_{t,i} = b_i + \sum_{j=1}^{j^*} \omega_{ij} z_{t,j} + \sum_{i=1}^{i^*} \delta_{t-1,i} n_{t-1,i}$$

$$z_t = [\Delta(m-p)_{t-1}, (m-p)_{t-1} - \hat{\Psi}_{t-1}, \Delta y_{t-1}, \Delta[\frac{i}{1+i}]_{t-1}, \Delta[\frac{i^*}{1+i^*}]_{t-1}, \Delta[\frac{\pi}{1+\pi}]_{t-1}, \Delta[\frac{\varepsilon}{1+\varepsilon}]_{t-1}]$$

Equation system (18) shows a dynamic process in the hidden layer as well as in the observable layer: the lagged neurons have effects on current neurons. The Elman network thus allows memory in the hidden-layer as agents learn from past errors, and thus makes the learning process at once richer and more complex. Since this network estimates the short-run adjustment, for the sake of parsimony this paper restricts the number of neurons, I*, to two. Since there are lagged effects in both the input and hidden layers, this paper limits the number of lagged input variables to no more than the number of significant lags in the linear error-correction model, since the lagged neurons may pick up additional dynamic patterns.

For obtaining the seigniorage-maximizing inflation rates for each country, equation (18) is simulated, until convergence to a long-run steady state, for a given inflation rate, and the resulting inflation tax, $[\pi/(1+\pi)]$ (M/P). The process is repeated for progressively higher inflation rates, and the long-run inflation and seigniorage combinations are plotted.

III. ECM and ANN Estimation of Money Demand and Seigniorage-Maximizing Inflation

Tables I through III contain the GARCH, ECM, and ANN estimates for Chile.

Table I shows that the GARCH coefficients for the parallel market exchange rate are significant. Moreover the GARCH process is stable.

TABLE I

GARCH PARAMETERS

Dependent variable: conditional variance		
Argument	Estimate	T-Stat
constant	0.00013	2.92
e-sq(-1)	0.2261	3.82
sig(-1)	0.7329	14.53

Table II, Panel A contains the estimates for the long-run cointegrating vector for the log level of M1, for the estimation period 1983.03 through 1994.08. The coefficient of output, y, is close to unity, and the coefficients for the domestic interest, foreign interest, and inflation opportunity-cost variables are all negative.[5]

Table II, Panel B gives the ECM estimates for short-run money demand. This paper presents the coefficient estimates given by the two-step method, and for the non-linear system method, in which the long-run cointegration coefficients and the short-run ECM coefficients are jointly obtained. This panel shows that the lagged first differences of the money stock have a negative effect on the current first difference. The error-correction mechanism is significant and negative, while the lagged GARCH variable, and the lagged first differences of income and the interest rate are significant under the two-step estimation but not in the system estimation. However, both systems show that the December *aguinaldo* dummy is significant. Finally, the diagnostics show that the overall explanatory power of the system method is appreciably greater than the two-step method.

TABLE II

LINEAR ECM MODEL

Panel A: Long-Run Cointegration Vector

Dependent variable: (m-p)
Sample: 83.03 - 94.08

Argument	Estimate	T-Stat
y	0.9657	6.64
i/(1+i)	-1.0498	3.51
i*/(1+i*)	-0.6679	0.48
inf/(1+inf)	-1.5576	1.05
constant	3.6548	4.91
R-SQ	0.69	
DW	0.51	

Panel B: Error-Correction Estimates

Dependent variable: D(m-p)

	Two-Step Method		System Estimation	
Argument	Estimate	T-Stat	Estimate	T-Stat
D[m-p](t-1)	-0.2099	2.1	-0.2064	2.33
cv(-1)	-0.1015	2.12	-0.1095	2.13
y			0.8008	1.58
il(1+i)			1.3944	0.8175
i*/(1+i*)			-4.1306	0.6897
inf/(1+infl			-40.28	-2.03
constant			4.73	1.76
sig(-1)	-1.21	1.98	-0.3913	0.2245
D[y(-1)]	-0.2071	1.88	-0.1223	1.22
D[i/(1+i)](t-1)	-0.3753	2.74	-0.2324	1.34
decem	0.1591	5.3	0.1341	4.39
R-SQ	0.2914		0.3863	
DW	2.24		2.21	
SSE	1.1746		1.1048	

Table III, Panel A contains the diagnostics and the neural network estimates for the approximation of the cointegrating equation, or long-run money demand, by the feedforward network.[6] The diagnostics show, by the Hannan-Quinn criterion, the optimal number of neurons for approximating the long-run function is seven. The R-squared under the neural network with seven neurons is .81, as opposed to .69 under the linear model. The second block in Panel A presents the weight estimates from the inputs to the seven neurons in the hidden layer of the feedforward network. In relative size, measured by the absolute value of the weights, the foreign interest rate variable is the most important "stimulus" to the neurons in the hidden layer, followed by the inflation variable and the domestic interest variable.[7] The weight estimates from the hidden layer to the output variable, the log first-difference of money, are all positive.[8]

Panel B gives the diagnostics and weight estimates for the recurrent Elman network estimation of the short-run ECM money demand. The diagnostics show that the ANN model outperforms the linear ECM, by the Hannan-Quinn criterion.[9] The weight estimates from the inputs to the two hidden neurons show that the error-correction coefficient is negative. The two GARCH coefficients are opposite in sign, but the dominate one is negative. Similarly, the dominate or large *aguinaldo* coefficient is positive, the dominant income coefficient is positive, and the two interest opportunity cost variables are negative. The last set of coefficients show that the two hidden neurons have about equal and positive effects on the output.

Finally, the last row in Panel B shows the partial derivatives computed from the neural network model, calculated on the basis of the last observation. These derivatives are calculated numerically, on the basis of a gradient step.

In terms of relative size, the most important variables are the error-correction vector, the December *aguinaldo*, and the interest rate. The non-linear uncertainty proxy, while highly significant in the linear error-correction model, falls in relative importance in the neural network model. This phenomenon suggests that the GARCH uncertainty variable may be a proxy for a hidden non-linearity in the ECM model.

Figure 1 pictures the long-run inflation and inflation tax combinations found in steady-state equilibrium generated by the estimated error-correction model for Chile. The inflation tax measures on the y-axis are based on the initial values of the money supply, which this paper sets at the mean value of the sample. This graph shows that the inflation tax revenue reaches a maximum at an inflation rate of 100 per cent, but at rates higher than this, the inflation tax quickly falls.

While the non-linear model shows that there is an inflation tax "Laffer curve" for Chile, the relatively high degree of non-linearity shows that the shape of the Laffer curve is not symmetric. After specific threshold rates of inflation, the seigniorage quickly drops. The results from Chile suggest that high inflation may not only be a very inefficient form of taxation, but after a certain point, a self-defeating form of taxation.

TABLE III

NEURAL NETWORK ESTIMATION

Panel A: Neural Network Estimation for Long-Run Money Demand
Dependent Variable: (m-p)

Diagnostics and Hannan-Quinn Information Criteria

Neurons:	OLS	4	5	6	7
R-SQ	0.693428	0.68597	0.702836	0.776304	**0.811376**
S$R	3.62774	3.719586	3.57646	2.699269	**2.293612**
HQIF	185.8021	221.148	225.3018	196.0383	**183.1333**

Weight Estimates from Inputs to Hidden Layer

Neuron:	Input: y	i/(1+i)	i*/(1+i*)	inf/(1+inf)
1	2.375904	-14.5364	20.32542	-68.6367
2	-3.04134	-11.157	-76.6879	-56.4916
3	5.827766	14.95165	67.21106	23.39697
4	5.23793	-9.62214	-84.3207	40.8856
5	-0.46111	10.11591	100.3114	51.43672
6	-5.04738	-13.8306	-62.3577	42.8267
7	-6.01068	-4.54219	-73.8226	59.42897

Weight Estimates from Hidden Layer to Output
Neuron:

1	2	3	4	5	6	7
0.626776	2.855637	0.548891	0.675053	0.078305	0.595207	2.467755

Panel B: Neural Network Estimates for Short-Run Adjustment
Dependent variable: D(m-p)

Diagnostics of Network Model and Hannan-Quinn Criteria

	ECM	ANN
R-SQ	0.3345	0.396725
SSE	1.1048	0.936805
HQIF	31.29628	24.4818

Weight Estimates from Inputs to Hidden Layer

Neuron:	[m-p](-1)	err(-1)	sig(-1)	decem	D[y(-1)]	D[i/1+i](-1)	nur1(-1)	nur2(-1)
1	-1.62864	-1.6508	5.333658	1.146784	2.242341	-0.75546	1.125894	0.59233
2	1.077017	-1.06926	-18.2404	-0.29233	-1.83142	-1.25303	-0.19198	-0.12923

Weight Estimates from Hidden Layer to Output

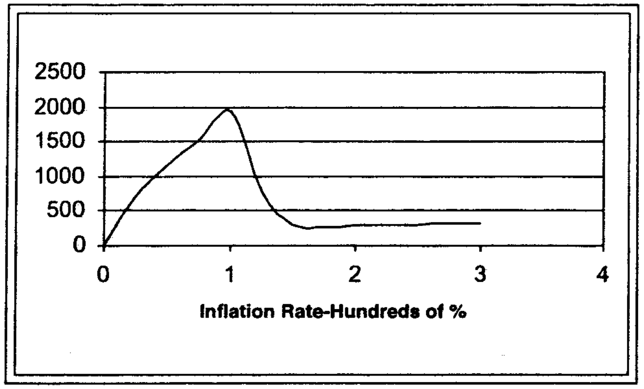
Neuron:	1	2
	0.268957	0.280999

Partial Derivatives

[m-p](-1)	err(-1)	sig(-1)	decem	D[y(-1)]	D[i/1+i](-1)
-10.1112	-287.032	-29.2967	202.4729	27.71179	-325.076

FIGURE 1

INFLATION LAFFER CURVE



IV. Policy Implications and Conclusion

The neural network approach to money demand indicates that non-linearity plays a major role in the demand for money. While this result is not particularly startling, the evidence from the linear ECM approach puts more weight on an uncertaintly proxy generated by a GARCH model.

To be sure, the coefficients of an ANN model are not readily interpreted. To make sense of the estimation results, one must calculate the partial derivatives based on a numerical gradient method. The ANN model is also more time consuming to estimate. However, it does «detect» higher degrees of non-linearity, and as the analysis of the inflation Laffer curve, paying attention to such non-linearities may be policy-makers understand the "optimal" rate of inflation may be much lower than previously expected.

While the results of the ANN are not spectacular, it should be noted that all of the estimates were obtained from relatively noisy monthly data. While longer-term quarterly models would surely give better results, more accurate monthly forecasts in times of high volatility or crisis may be more useful for the practice of monetary policy. This is where the ANN seems to have the comparative advantage.

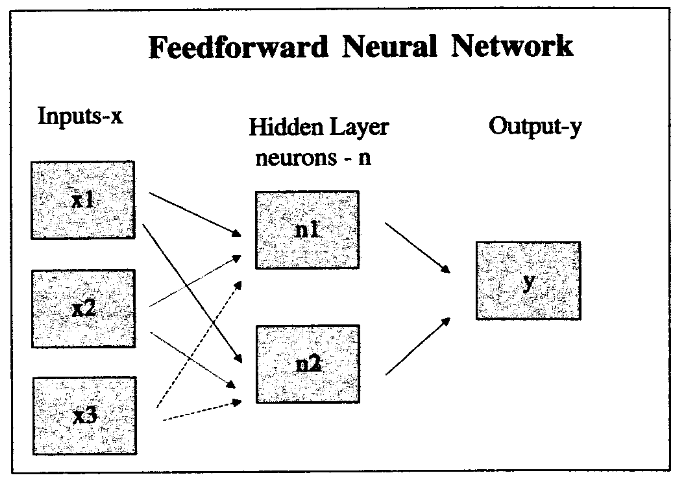
APPENDIX

ARCHITECTURE OF THE NEURAL NETWORK

The neural network is a specific model of how perceptions are formed, as people observe input variables such as interest rates, and make decisions about setting prices, making investments, or trading currency. What this approach adds to a conventional model of inputs x and outputs y is a hidden layer of unobserved neurons, n, which are functions of the observed inputs x. The typical model or "architecture" of a neural network appears in Figure A-1. As one can see, the neural network links the inputs x to the outputs y through the neurons. The neurons "activate" the impulses of linear combinations of the observed inputs within the system (as inputs pass to the hidden layer), and pass on as output a linear combination of one, two, or more nonlinear transformations of the weighted inputs. The arrows passing from the inputs x to the hidden layer, and from the hidden layer to the output represent coefficient values which link inputs to neurons, and neurons to output at each stage of the network:

FIGURE A-1

FF NETWORK

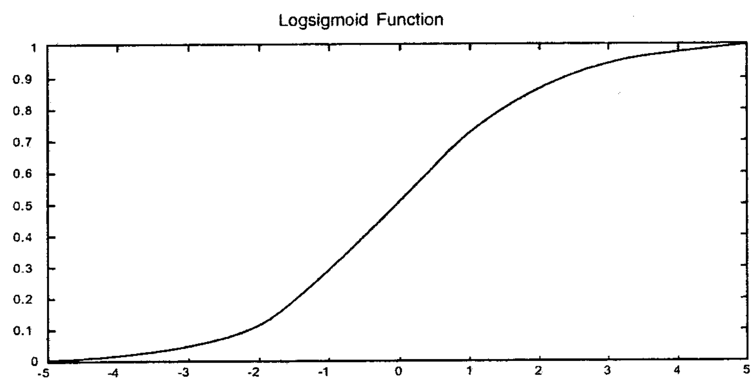


What gives the neural network its forecasting power is *parallel processing*: in addition to the *sequential processing* of typical linear systems, in which only observed inputs are used to predict an observed output. With a neural network, the neurons process the inputs in a *parallel* fashion, in order to improve the predictions.

The neurons process the input data in two ways: first by forming linear combinations of the input data (the third equation in system 1), and then by "squashing " these linear combinations through the logsigmoid function (the second equation in system 1). Figure A-2 illustrates the operation of the logsigmoid activation function or "squasher" on a series ranging from -5 to +5.[10]

FIGURE A-2

LOGSIGMOID OR LOGISTIC SQUASHER



The appeal of the logsigmoid squasher function comes from its "threshold behavior", which characterizes many types of economic responses to changes in fundamental variables. For example, if interest rates are already very low or very high, small changes in this rate will have very little effect on investment or money demand. However, within critical ranges, small changes may signal possibilities of significant upward or downward movements, so that the response of investment or money demand may be very pronounced.

An alternative rationale for the logsigmoid function is "learning behavior". At very low or very high levels, small upward or downward movements in the interest rate will trigger little response in investment or money-demand behavior. As interest rates continue to increase, from low levels, or to fall from high levels, economic decision-makers will gradually "learn" and form expectations about the future state of policy and the economy. As the interest rate rises to or falls below a critical threshold level, investment or money demand will start to respond in significant ways to continuing changes in this rate. Thus, the nonlinear logsigmoid function captures a threshold response characterizing "bounded rationality" or a "learning process" in the formation of expectations.

The following system describes in a more general way the most commonly used feed-forward network:

FF NETWORK

$$y_t = \gamma_0 + \sum_{i=1}^{i^*} \gamma_i N_{t,i}$$

$$N_{t,i} = \frac{1}{1 + e^{-n_{t,i}}}$$

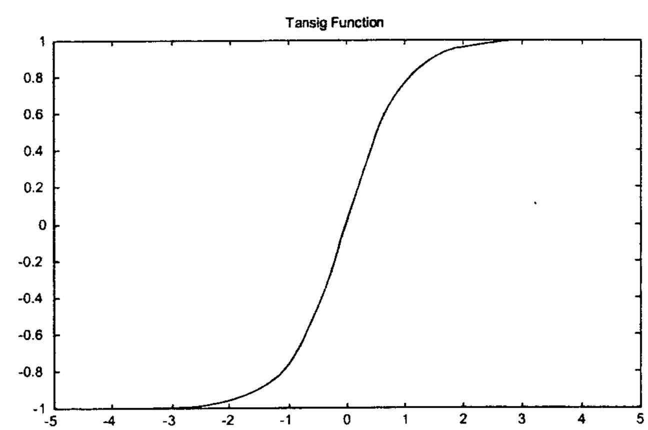
$$n_{t,i} = \omega_0 + \sum_{j=1}^{j^*} \omega_{ij} x_{t,j}$$

in which there are j* input variables and i* neurons. The neurons themselves are logsigmoid functions of linear combinations of the input data.

An alternative "activation function" for the neurons in a neural network is the hyperbolic tangent function. It is also known as the "tansig" or "tanh" function. It squashes the linear combinations of the inputs within the interval [-1,1], rather than [0,1] in the logsigmoid function. The following figure shows the behavior of this alternative function:

FIGURE A-3

TANSIG SQUASHER



The mathematical representation of the feedforward network with the tansig activation functions is given by the following system:

FF NET WITH TANSIG SQUASHERS

$$y_t = \gamma_0 + \sum_{i=1}^{i^*} \gamma_i N_{t,i}$$

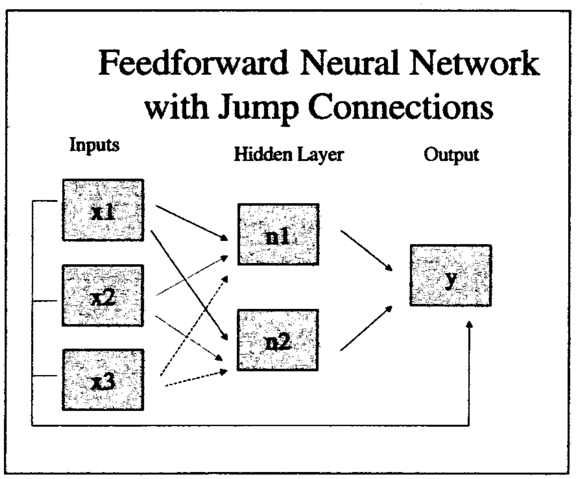
$$N_{t,i} = \frac{e^{n_{t,i}} + e^{-n_{t,i}}}{e^{n_{t,i}} - e^{-n_{t,i}}}$$

$$n_{t,i} = \omega_0 + \sum_{j=1}^{j^*} \omega_{ij} x_{t,j}$$

One alternative to the pure feedforward network is a feedforward network with jump connections, in which the inputs x has direct linear links to output y, and as links to the output through the hidden-layer of squashed functions. Figure A-4 pictures a feedforward jump connection network with three inputs, one hidden layer, and two neurons (j* =3, i* =2):

FIGURE A-4

FF NET WITH JUMP CONNECTIONS



The mathematical representation of the network in Figure A-4 , for logsigmoid activation functions, is given by the following system:

FF NET WITH JUMP CONNECTIONS

$$y_t = \gamma_0 + \sum_{i=1}^{i^*} \gamma_i N_{t,i} + \sum_{j=1}^{j^*} \beta_j x_{t,j}$$

$$N_{t,i} = \frac{1}{1 + e^{-n_{t,i}}}$$

$$n_{t,i} = \omega_0 + \sum_{j=1}^{j^*} \omega_{ij} x_{t,j}$$

Note that the feedforward network with the jump connections increases the number of parameters in the network by j*, the number of inputs.

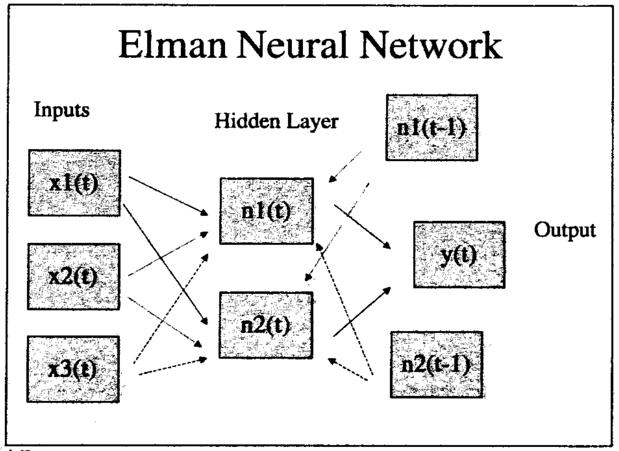
The advantage of the feedforward network with jump connections is that it nests the pure linear model as well as the feedforward neural network. If the underlying relationship between the inputs and the output is a pure linear one, than only the direct jump connectors, given by the coefficients $\{\beta_j\}$, are significant. However, if the true relationship is a complex nonlinear one, then we would expect that the coefficient sets $\{\omega_j\}$, $\{\gamma_i\}$ would be highly significant, while the coefficient set $\{\beta_j\}$ would be relatively insignificant. Finally, if the underlying relationship between {x,y} can be decomposed into linear and non-linear components, then we would expect all three sets of coefficients, $\{\omega_j\}$, $\{\gamma_i\}$ and $\{\beta_j\}$ to be significant.

Another commonly used neural architecture is the Elman "recurrent" network. This network allows the neurons to depend not only on the input variables x, but also on their own lagged values. Thus the Elman network builds in "memory" into the evolution of the neurons.

This type of network is similar to the commonly used moving average (MA) process in time series analysis. In the MA process, the dependent variable y is a function of observed inputs x as well as current and lagged values of an unobserved disturbance term or random shock, ε. Thus, a q-th order MA procress has agruments, ε_t, ε_{t-1}, ..., ε_{t-q}. In a similar fashion, the Elman network makes use of lagged as well as current values of unobserved neurons. In the estimation of both Elman networks and MA processes, it is necessary to use a multi-step estimation procedure: one has initializes the vector of lagged neurons or disturbance terms with arbitrary values in order to obtain parameter values, and the one recalculates the vector of lagged neurons or disturbance terms and re-estimates the parameter values. The process continues until convergence takes place. Figure A-5 pictures this system:

FIGURE A-5

ELMAN RECURRENT NET



Note that the inputs, neurons, and output boxes have time labels for the current period, t, or the lagged period, t-1. The Elman network is a network specific to data which have a time dimension. The feedforward network, on the other hand, may be used for cross-section data, which is not dimensioned by time, as well as time-series data.

The following equation system expresses the Elman network pictured in Figure A-5:

ELMAN RECURRENT NET

$$y_t = \gamma_0 + \sum_{i=1}^{i^*} \gamma_i N_{t,i}$$

$$N_{t,i} = \frac{1}{1 + e^{-n_{t,i}}}$$

$$n_{t,i} = \omega_0 + \sum_{j=1}^{j^*} \omega_{ij} x_{t,j} + \sum_{i=1}^{i^*} v_i N_{t-1,i}$$

Notes

1 The appendix contains a brief review of commonly used neural network models or "architectures".

2 An important exception is Dorsey (1995), who uses ANN-methods to compare the forecasting performance of alternative money aggregates for inflation.

3 While any non-linear function g(x) may be approximated by a polynomial expansion, the ANN achieves greater accuracy with few parameters. This issue is treated in greater detail in the next section.

4 Cho and Sargent (1995) have pointed out that any continuous function Ψ can be approximated arbitrarily well if one chooses j*, the number of hidden neurons, large enough. They also point out that the feedforward network in system (15) uses only linearly increasingly more parameters, while polynomial, spline, and trigonometric expansions use parameters that grow exponentially for a given approximation.

5 For the sake of brevity, this paper does not show the unit root tests for the variables in Panel B. All are I(1), and the residual is I(0).

6 The network was "trained" for 10,000 "epochs" or iterations through the data set, under each specification.

7 It is important to remember that the weight values do not present partial derivatives. Such derivatives would have to be calculated numerically, based on mean values of the inputs.

8 Computationally, one may assess the statistical significance of the network weights by calculating the inverse of the Hessian matrix of second-order derivatives. The approximate t-statistics were highly significant. Similarly, rough measures of the relative contributions of different inputs to the determination of the output are based on the relative sizes of the absolute values of the network weights.

9 The "cointegrating vector" in this estimation is the residual from the ANN estimation of the long-run equation in Table II-Panel A. Like the linear cointegrating vector generated by the parameters of II- A, it is I(0).

References

ALTMAN, EDWARD I., MARCO GIANCARLO and FRANCO VARETTO (1994). "Corporate Distress Diagnosis: Comparisons Using Linear Discriminant Analysis and Neural Networks (the Italian Experience)", *Journal of Banking and Finance* 18: 505-529.

ARRAU, PATRICIO, JOSE DE GREGORIO, CARMEN M. REINHART, and PETER WICKHAM (1995). "The Demand for Money in Developing Countries: Assessing the Role of Financial Innovation", *Journal of Development Economics* 46: 317-40.

ASILIS, CARLOS, PATRICK HONOHAN and PAUL MCNELIS (1993). "Money Demand During Hyperinflation and Stabilization: Bolivia 1980-88", *Economic Inquiry* 31: 262-73.

BOLLERSLEV, TIM (1986). "Generalized Conditional Heteroskedasticity", *Journal of Econometrics* 31: 307-27.

CALVO, GUILLERMO and LEONARDO LEIDERMAN (1992). "Optimal Inflation Tax under Precommittment: Theory and Evidence", *American Economic Review* 82: 179-95.

CHO, IN-KOO and THOMAS J. SARGENT (1995). "Neural Networks for Encoding and Adapting in Dynamic Economies". Manuscript.

DORSEY, ROBERT E. (1995). "Neural Networks with Divisa Money: Better Forecasts of Future Inflation? More Evidence on the Superior Information Content of Divisa Monetary Aggregates". *Working Paper*, University of Mississippi, Department of Economics.

EASTERLY, WILLIAM R., PAOLO MAURO and KLAUS SCHMIDT-HEBBEL (1995). "Money Demand and Seigniorage-Maximizing Inflation", *Journal of Money, Credit, and Banking* 27: 583-603.

ELMAN, JOHN (1988). "Finding Structure in Time", CRL Report 8801, Center for Research in Language, University of California at San Diego.

GRANGER, CLIVE (1996). "Introducing Non-Linearity into Cointegration". *Working Paper*, University of California, San Diego.