Local Stability Analysis of Neural Network Models with Application to Exchange-Rate Data

By

Peter Kim, Lin Pan and Tony S. Wirjanto[1]

This Version: May, 2005

Abstract: In this paper we discuss the stability property of a predictive neural network model from a deterministic point of view. In particular, the stability property of linear and nonlinear causal transmission link models of daily spot Canadian/US forein exchange rate is analyzed using a local stability analysis based on a nonlinear dynamical systems framework. This analytical result enables a numerical analysis of the stability to be fully testable on the data set. Also the stability of the interval prediction of a general neural network model is studied in this paper.

Keywords: nonlinearity, dynamical systems, eigenvalues, hyperbolic equilibrium, and stability

[1]Peter Kim is with College of Physical and Engineering Science, Department of Mathematics and Statistics, University of Guelph, Guelph, Ontario N1G 2W1, Canada. Lin Pan is with Portfolio and Financial Modeling Decision Support Services, Electronic Banking Services, Bank of Montreal, Center Tower, Toronto M8X 2X3, Canada Tony Wirjanto (corresponding author) is with Department of Economics, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada. E-mail address: twirjant@uwaterloo.ca

1 Introduction

Neural networks have been used in modeling noisy financial time series with some success. Although neural networks qualitatively provide good forecasts of such time series, they can be somewhat unstructured. In the light of this argument, a dynamical systems approach is adopted in this paper to study the local stability properties associated with the neural network modeling.

In particular, the stability property of the causal transmission link model of spot Canadian/US exchange rate, which is recently studied in Kim, Pan and Wirjanto (1999) and reproduced in Section 2, is analyzed using a local stability analysis based on linear and nonlinear dynamical systems. This is done in Sections 3 and 4. This analytical exercise enables a numerical analysis to be fully testable on the data set (Section 5). Finally the stability results for interval predictions of a general predictive model is provided in the paper (Section 6). Concluding remarks are collected in Section 7.

2 Causal Transmission Link Neural Network Model

Following Kim, Pan and Wirjanto (1999), a general causal transmission link is defined by a nonlinear function $f : R^{3+1} \longrightarrow R$, i.e.,

$$Y_t = f(X_t, X_{t-1}, X_{t-2}, N_t) \ , \tag{1}$$

where $Y_t = \log(C\$_t) - \log(C\$_{t-1})$ is the log difference of spot Canadian dollar $C\$_t$, $X_t = \nabla IS_{t-1}$ is daily changes in interest rate spread, and N_t is a stationary time series assumed to be uncorrelated with X_t and represents noise.

A general nonlinear autoregressive process of order 5, or **AR**(5) model, for X_t is expressed as

$$X_t = g(X_{t-1}, X_{t-2}, \ldots, X_{t-5}, \varepsilon_t) \ , \qquad t \in Z \ , \tag{2}$$

where $g : R^{5+1} \longrightarrow R$ is a nonlinear function and ε_t is random noise independent of X_s, for $s < t$.

From equations (1) and (2), nonlinear additive noise models can be obtained as

$$Y_t = F(X_t, X_{t-1}, X_{t-2}) + N_t \ , \qquad t \in Z \ , \tag{3}$$

and

$$X_t = G(X_{t-1}, X_{t-2}, \ldots, X_{t-5}) + \varepsilon_t \ , \qquad t \in Z \ , \tag{4}$$

respectively, where $F : R^3 \longrightarrow R$ and $G : R^5 \longrightarrow R$ are nonlinear functions.

Based on equation (3), a short term forecasting neural network model[2] can be expressed as,

$$C\$_t - C\$_{t-1} = \mathcal{N}_{Causal_1}(\nabla IS_t, \nabla IS_{t-1}, \nabla IS_{t-2}) + \varepsilon_t \ , \quad t \in Z \ , \tag{5}$$

where $\mathcal{N}_{Causal_1}$ represents the multi-layer feed-forward Network. Basically, the $\mathcal{N}_{Causal_1}$ estimates the nonlinear transmission link in equation (3), under certain stability assumptions to be elaborated in the next section.

[2]The short term here means that the forecast is made by using the previous information up to today or yesterday.

3 A Dynamical Systems Approach to the Nonlinear Additive Noise Model

The stability behavior of the nonlinear additive noise model in (3) and (4) is studied in this section by using the local stability analysis from a dynamical systems point of view. First we make the following assumptions: the function G in (4) is a smooth function, i.e., $G \in C^r, r > 1$, and F is a uniformly continuous function; and the random noise terms in the nonlinear models (3) and (4) are bounded almost surely in the time domain,

$$P(\mid N_t \mid < 1) = 1, \quad P(\mid \varepsilon_t \mid < 1) = 1, \quad t \in Z.$$

The above assumptions imply the following aspects of stability for the model: stability of the system in (3) exactly depends on the system in (4); and stability of the dynamical system in (4) only depends on the system without noise, that is,

$$X_t = G(X_{t-1}, X_{t-2}, \ldots, X_{t-5}). \tag{6}$$

These assumptions and aspects are supported by the analysis in the linear case as reported in Kim and Martin (1996) and the stability theory in dynamical systems - see Hartman (1964), Hirsch and Smale (1974) and Wiggins (1990).

Now, consider the nonlinear dynamical system (6), and rewrite the difference equation by setting

$$Z(t) = \begin{pmatrix} Z_1(t) \\ Z_2(t) \\ Z_3(t) \\ Z_4(t) \\ Z_5(t) \end{pmatrix} = \begin{pmatrix} X_{t-1} \\ X_{t-2} \\ X_{t-3} \\ X_{t-4} \\ X_{t-5} \end{pmatrix}.$$

System in (6) is then equivalent to a fifth order map equation of the form

$$Z(t+1) = H(Z(t)) = \begin{pmatrix} G(Z(t)) \\ Z_1(t) \\ Z_2(t) \\ Z_3(t) \\ Z_4(t) \end{pmatrix}, \tag{7}$$

with the initial state

$$Z(0) = \begin{pmatrix} X_{-1} \\ X_{-2} \\ X_{-3} \\ X_{-4} \\ X_{-5} \end{pmatrix}. \tag{8}$$

System (7) is totally determined after a random setup of the initial state. This is called a realization or a sample path of the stochastic process. An equilibrium solution of (7) is a point

$\bar{Z}(t)$, such that

$$\tilde{Z}(t) = H(\tilde{Z}(t)). \tag{9}$$

Obviously, the point $\tilde{Z} = 0$ is one of the equilibrium points. By Taylor's Theorem, the system (7) near the equilibrium ($\tilde{Z} = 0$) becomes

$$\begin{aligned} Z(t+1) &= D_Z H(0) Z(t) + \frac{1}{2} D_Z^2 H(0)(Z(t), Z(t)) + \ldots \\ &= \begin{pmatrix} G_{Z_1}(0) & G_{Z_2}(0) & G_{Z_3}(0) & G_{Z_4}(0) & G_{Z_5}(0) \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} Z_1(t) \\ Z_2(t) \\ Z_3(t) \\ Z_4(t) \\ Z_5(t) \end{pmatrix} + o(\mid Z(t) \mid), \end{aligned} \tag{10}$$

where $o(\mid Z(t) \mid)$ are higher order nonlinear terms.

4 Linearization

A good place to start in analyzing the stability of a nonlinear system such as (10) is to determine the local stability behavior of the linear part of (10) at the equilibrium ($\tilde{Z}$).

Essentially, hyperbolic equilibria and stability properties of the nonlinear system in (10) are the same as those of the linear system based on the stable manifold theorem - see Hirsch and Smale (1974) and Wiggins (1990). Furthermore, the local behavior of the nonlinear system in (10) and its linear version at the hyperbolic equilibrium are topologically equivalent. In other words, near the hyperbolic equilibrium, the nonlinear system has the same qualitative structure as its linearized system. This follows from the Hartman-Grobman theorem - see Hartman (1964) and Wiggins (1990).

Below we state the definitions of a hyperbolic equilibrium and the stability condition associated with it, known as Lyapunov stability:

Definition 1 *Let $\tilde{X}$ be an equilibrium point of the autonomous nonlinear map, i.e., $X \longmapsto g(X), \quad X \in R^n$. Then X is called a hyperbolic equilibrium if none of the eigenvalues of $D_X g(\tilde{X})$ are on the unit circle, i.e.,*

$$\mid \lambda_i \mid \neq 1, \quad i = 1, 2, \ldots, n,$$

where $\lambda_i, \ i = 1, 2, \ldots, n$, are eigenvalues of $D_X g(\tilde{X})$.

and

Definition 2 Lyapunov Stability: *The equilibrium $\tilde{X}_t$ is said to be stable (or Lyapunov stable), if given $\varepsilon > 0$, there exist a $\delta = \delta(\varepsilon, \tilde{X}_{t_0}) > 0$ such that, for any other solution X_t, satisfying $\mid X_{t_0} - \tilde{X}_{t_0} \mid < \delta$, then $\mid X_t - \tilde{X}_{t_0} \mid < \varepsilon$ for all $t > t_0, \ t_0 \in R$. If in addition $\mid X_t - \tilde{X}_{t_0} \mid \rightarrow 0$ as $t \rightarrow \infty$, then $\tilde{X}_{t_0}$ is asymptotically stable.*

Going back to the system in (10), we rewrite the linear part as

$$Z(t+1) = AZ(t), \tag{11}$$

where

$$A = \begin{pmatrix} G_{Z_1}(0) & G_{Z_2}(0) & G_{Z_3}(0) & G_{Z_4}(0) & G_{Z_5}(0) \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}. \tag{12}$$

The corresponding characteristic equation is

$$f(\lambda) = \lambda^5 - G_{Z_1}(0)\lambda^4 - G_{Z_2}(0)\lambda^3 - G_{Z_3}(0)\lambda^2 - G_{Z_4}(0)\lambda - G_{Z_5}(0) = 0. \tag{13}$$

More generally, to each monic polynomial we have

$$f(\lambda) = \lambda^m - C_{m-1}\lambda^{m-1} - \ldots - C_1\lambda - C_0 \tag{14}$$

of degree m, we can construct an $m \times m$ matrix with minimal polynomial $f(\lambda)$. This matrix is called the *companion matrix* of $f(\lambda)$. The matrix A in (12) is the special case with $m = 5$.

Choosing an initial state $Z(0)$, the orbit of $Z(0)$ under the linear map (11) is given by the bi-infinite sequence (if the map is a C^r, $r \geq 1$, diffeomorphism)

$$\{\ldots, A^{-n}Z(0), \ldots, A^{-1}Z(0), Z(0), A^1 Z(0), \ldots, A^n Z(0), \ldots\}, \tag{15}$$

or the infinite sequence (if the map is C^r, $r \geq 1$, but noninvertible)

$$\{Z(0), A^1 Z(0), \ldots, A^n Z(0), \ldots\}. \tag{16}$$

The equilibrium point ($\tilde{Z} = 0$) of linear map (11) is asymptotically stable if all of the eigenvalues of A have modulus strictly less than one. If one of the eigenvalues has modulus strictly greater than one, the system is unstable, by the stable and unstable manifold theorem. If some of the eigenvalues lie on the unit circle, which is a degenerate case, then the centre manifold and bifurcation theorems can be used to analyze this case.

5 The Stability of the Nonlinear Transfer Model

Our objective in this section is to determine the stability behavior of the nonlinear transfer models in (3) and (4) at the stochastic equilibrium, about the zero-mean level. It is reasonable to begin with the deterministic system first, since the stochastic process is defined on some underlying probability space $(\Omega, \mathcal{F}, P)$ common with the time set T to the measurable space $(R^T, \mathcal{B}(R^T))$, i.e.,

$$X : T \times \Omega \longrightarrow R\,, \tag{17}$$

where $X_t = X_t(\cdot)$ is a random variable with respect to the $\mathcal{F}$ $\sigma-$field for each $t \in T$. In most cases $T = [0, \infty)$. On the other hand, for each fixed $\omega \in \Omega$, $X_t = X_t(\omega)$ is a realization or a sample path of the stochastic process.

The main result of the stability property of the model is stated in the following proposition.

Proposition 1 *The nonlinear transfer model expressed in (3) and (4) is stable, at the zero-mean level, if the changes in the spread $\{X_t,\ t \in Z\}$ are a zero-mean stationary* **AR** *process with finite degree. Moreover, the results can also apply to any constant mean level and any finite degree of nonlinearity.*

The proof of this propostion is given below.

Proof 1 *All we need to show is that none of the eigenvalues of the characteristic equation in (13) is on or outside the unit circle. Thus the nonlinear system in (6) is topologically equivalent to its linearized system in (11) or (13) at the hyperbolic equilibrium ($\tilde{Z} = 0$) in a dynamical sense, and is asymptotically stable. Thus the systems in (3) and (4) are stable according to our assumptions, i.e.*

$$\begin{aligned} \mid X_t \mid &= \mid G(X_{t-1}, X_{t-2}, X_{t-3}, X_{t-4}, X_{t-5}) + \varepsilon_t \mid \\ &\leq \mid G(X_{t-1}, X_{t-2}, X_{t-3}, X_{t-4}, X_{t-5}) \mid + \mid \varepsilon_t \mid < constant, \quad as\ t \to \infty, \end{aligned} \tag{18}$$

$$\begin{aligned} &\mid Y_t \mid = \mid F(X_t, X_{t-1}, X_{t-2}) + N_t \mid \\ &\leq \mid F(X_t, X_{t-1}, X_{t-2}) \mid + \mid N_t \mid < constant, \quad as\ t \to \infty, \end{aligned} \tag{19}$$

$$since\ uniformly\ continuous\ functions\ F\ are\ bounded\ in\ its\ bounded\ domain.$$

In order to show that the eigenvalues of (13) are within the unit circle, let $\lambda = \frac{1}{\gamma}$ and substitute this into the characteristic equation in (13). As a result, we obtain

$$\phi(\gamma) = 1 - G_{Z_1}(0)\gamma - G_{Z_2}(0)\gamma^2 - G_{Z_3}(0)\gamma^3 - G_{Z_4}(0)\gamma^4 - G_{Z_5}(0)\gamma^5 = 0, \quad (\gamma \neq 0). \tag{20}$$

The polynomial equation will be exactly the characteristic equation in the AR(5) model at the zero-mean level. Thus, if the changes in the spread $\{X_t,\ t \in Z\}$ is a zero-mean stationary process $AR(5)$, all the roots of equation (19) lie outside the unit circle, which shows that none of the eigenvalues of equation (13) is on or outside the unit circle. □

The above result is independent of the order of the AR model and how complex the causal transmission link is. However, if the model involves a feed back from the spot rate to the interest rate spread, then the stability behavior becomes more complicated. Furthermore, this proof shows that all finite order **AR** processes with a given initial state are asymptotically stable at the equilibrium point of zero mean. This is consistent with the result from time series analysis. It suggests that the data set can be modeled according to the **AR** model.

6 Numerical Analysis of Stability

Our numerical analysis is based on the linear model in (11) from a dynamical point of view. The procedure of the numerical study is as follows:

a. Use a linear regression to estimate the parameters in following linear model, i.e.,

$$X_t = \sum_{i=1}^{5} \alpha_i X_{t-i} + \phi_t \ , \tag{21}$$

where $X_t = IS_t - IS_{t-1}$ is daily changes in the interest rate spread and ϕ_t is white noise.

The data considered here is part of the entire data set, approximately one year (300) daily data points.[3] In other words, the local linear regression model is used over the entire data set.

b. Using equation (13), calculate the eigenvalues corresponding to equation (20).

Table 1: **Moduli of Eigenvalues in Test I.**

Time Index	$\mid \lambda_1 \mid$	$\mid \lambda_2 \mid$	$\mid \lambda_3 \mid$	$\mid \lambda_4 \mid$	$\mid \lambda_5 \mid$
1-300	0.76289	0.76289	0.77169	0.77169	0.68466
301-600	0.47784	0.47784	0.48492	0.27543	0.27543
601-900	0.61408	0.61408	0.51435	0.51435	0.54126
1801-2100	0.58866	0.58866	0.50036	0.50036	0.42070
2401-2700	0.64703	0.61950	0.61950	0.43827	0.43827
2701-2300	0.40139	0.40139	0.43974	0.43974	0.29529
3001-3300	0.58636	0.58636	0.60730	0.60730	0.42899

Table 1 shows that all moduli of eigenvalues are strictly less than 1. Therefore, the series $X_t = \nabla IS_t$ is asymptotically stable after the initial observed state, even in a short time period.

Next, we use the same procedure on the series IS_t directly, i.e.,

$$IS_t = \sum_{i=1}^{6} \hat{\alpha}_i IS_{t-i} \ . \tag{22}$$

where $\hat{\alpha}_i$ are estimates of α_i.

Part of the results are listed in Table 2. It shows that there is one eigenvalue numerically very close to 1, and others are strictly less than 1. Consequently it is very hard to make the conclusion that the time series is stable or unstable during the time period that was tested. The obvious reason for this is that the eigenvalues are calculated from the estimates of $\alpha_i, \ i = 1, \ldots, 6$. Nevertheless, it is possible that the series has one cyclic component in these six lags.

[3]The data set used in this study is generously supplied by Financial Markets Research of CIBC Wood Gundy. It is part of the data set used in Kim, Pan and Wirjato (1999), which contains 3327 daily prices for the Canadian/US foreign exchange rate, covering the period from January 2, 1984 to October 1, 1996. The spot Canadian/US foreign exchange rate ($C\$$) is measured in US cents, while the daily insterest rate spread (IS) is measured as the difference, in percent, between 90 day Canadian treasury bill rate and 90 day US treasury bill rate.

Table 2: **Moduli of Eigenvalues in Test II.**

Time Index	$\mid \lambda_1 \mid$	$\mid \lambda_2 \mid$	$\mid \lambda_3 \mid$	$\mid \lambda_4 \mid$	$\mid \lambda_5 \mid$	$\mid \lambda_6 \mid$
1-300	1.00262	0.76211	0.76211	0.77115	0.77115	0.67712
301-600	0.99987	0.47779	0.47779	0.49288	0.29211	0.29211
601-900	0.99914	0.61705	0.61705	0.51447	0.51447	0.54918
1801-2100	0.99798	0.57787	0.57787	0.47619	0.47619	0.32552
2401-2700	0.99699	0.65112	0.62327	0.62327	0.43854	0.43854
2701-3000	0.99565	0.39710	0.39710	0.43276	0.43276	0.28906
3001-3300	0.99165	0.58267	0.58267	0.60061	0.60061	0.42238

Although the above analysis provides only a partial conclusion, nevertheless, this analysis does provide us with the idea to define stability of the interval prediction of the model. On the other hand, for a nonstationary time series, this method can be used to determine the d-th differencing necessary in an integrated model , i.e.,

$$\phi(B) = \varphi(B)(1 - B)^d \ , \tag{23}$$

where $\phi(B)$ is the generalized autoregression operator; $\varphi(B)$ is the autoregression operator; and B is the backward shift operator.

Once the numerical study gives two eigenvalues which are close to 1, this means that a second order of differencing is needed to transfer the nonstationary time series to a stationary time series.

7 Stability of the Interval Prediction

In this section we propose a stability result for the interval prediction of a general predictive neural network model. Consider the following predictive model

$$Y_t = f(\cdot) + N_t \ . \tag{24}$$

We make the following definition.

Definition 3 *Let $\hat{Y}_t = f(\cdot)$ be the predictor of the true value Y_t. Then Y_t is said to be stable relative to the predictor $\hat{Y}_t$, at $1 - \alpha$ confidence if for any given $\varepsilon > 0$, and an $0 < \alpha < 1$, there exists an $0 < \eta(\varepsilon, \alpha)$, such that,*

$$P[\mid \hat{Y}_{t_0} - Y_{t_0} \mid < \eta] = 1 - \alpha \ . \tag{25}$$

Then

$$P[\mid \hat{Y}_t - Y_t \mid < \varepsilon] \geq 1 - \alpha \ . \tag{26}$$

for all $t > t_0$, $t_0 \in R$.

Obviously, the above definition is associated with the concept of $1-\alpha$ confidence interval of the predictor $\hat{Y}_t$. Thus, our concern is that if the model is stable, then the length of confidence interval does not increase over time.

Finally, we make the following remark.

Remark 1 *Consider the nonlinear model*

$$Y_t = f(\cdot) + N_t \ , \qquad t \in Z \ . \tag{27}$$

For a confidence interval of length 2ε *and confidence coefficient* α*, the predictor* $\hat{Y}_t = f(\cdot)$ *is said to be stable at* $1-\alpha$*, if the residual of the model* N_t *is a zero mean process with a suitable small variance over the time period* Z*, such that,*

$$VarN_t \leq \alpha \times \varepsilon^2 \ , \quad t \in Z \ . \tag{28}$$

The proof of this remark is straightforward and given below.

Proof 2 *By using the Chebyshev inequality,*

$$P[\mid \hat{Y}_t - Y_t \mid < \varepsilon] \geq 1 - \frac{E[(\hat{Y}_t - Y_t)^2]}{\varepsilon^2} = 1 - \frac{Var[N_t]}{\varepsilon^2} \geq 1 - \alpha \tag{29}$$

for all $t \in Z$*.* □

Therefore, the stability of any predictive model is related to the initial length $1-\alpha$ of the confidence interval and the variance of the predictive model. In other words, if the initial $1-\alpha$ confidence interval is small and the model variance is small over the modeling time period, then that predictive model is stable. Therefore, the measurement of how small can be calibrated to trader limitations is determined by risk managers.

8 Conclusion

In this paper we have adopted a dynamical systems approach to analyze the stability property of a causal predictive neural network model. This analytical exercise enabled us to conduct a numerical analysis of the stability of the model that is fully testable on the data set. The stability results of the interval prediction for a general predictive model was also provided in this paper.

References

Hartman, P. (1964): *Ordinary Differential Equations*, New York: John Wiley & Son.

Hirsch. M. W. and S. Smale (1974): *Differential Equations, Dynamical Systems, and Linear Algebra.*

Kim, P. and P. Martin (1996): "On the Relationship Between the Interest Rate Spread and The Spot Canadian Dollar", *Lecture Notes in Time Series Analysis*, University of Guelph, 1-21.

Kim, P., L. Pan and T. S. Wirjanto (1999): "Neural Network Models of the Spot Canadian/U.S. Exchange Rate," Mimeo, University of Waterloo, Waterloo, Ontario, Canada.

Wiggins, S. (1990): *Introduction to Applied Nonlinear Dynamical Systems and Chaos*, New York: Springer-Verlag.