UC San Diego
Recent Work

Title
Introduction to M-M Processes

Permalink
https://escholarship.org/uc/item/9pk546xs

Authors
Granger, Clive W.J.
Hyung, Namwon

Publication Date
1998-07-01

98-18

UNIVERSITY OF CALIFORNIA, SAN DIEGO

DEPARTMENT OF ECONOMICS

INTRODUCTION TO M-M PROCESSES

BY

CLIVE W.J. GRANGER

AND

NAMWON HYUNG

Introduction to M-M Processes

Clive W.J. Granger
Namwon Hyung

Department of Economics
University of California, San Diego
9500 Gilman Drive
La Jolla, CA 92093-0508

July 1998

Abstract

This paper introduces a new type of nonlinear model, the min-max model, and analyzes the properties for a pair of series. Stability conditions of this system are given for the nonlinearly integrated bivariate series. Under these stability conditions, the difference of the two series has a threshold-type nonlinearity. One can construct a threshold error correction model from min-max processes. Neglected nonlinearity tests are applied, to the univariate series and to the system, to detect nonlinearity, and it turns out that the tests using the system have better power. We apply the min-max model to U.S. Treasury bill and commercial paper interest rates. The spread of these interest rates shows a threshold-type nonlinearity, and this model outperforms a linear model in terms of its predictability out-of-sample.

Keywords: Min-max process, Nonlinear error correction model, Neglected nonlinearity, Threshold

1. Introduction

A new class of non-linear models is introduced, with concentration on the bivariate process. The process has some theoretical interest, as the univariate series contains weak evidence of nonlinearity but this evidence becomes strong when the bivariate system is considered. The bivariate system can be linearly cointegrated but with a nonlinear error-correction model. The bivariate process will be generated by

$$x_{t+1} = \max(\alpha x_t + a, \beta y_t + b) + \varepsilon_{x,t+1} \qquad (1.1)$$

$$y_{t+1} = \min(\gamma x_t + c, \delta y_t + d) + \varepsilon_{y,t+1} \qquad (1.2)$$

where $e_{x,t}$, $e_{y,t}$ are independent and *i.i.d.* with variances σ_x^2 and σ_y^2 respectively. Although usually the *max-min* pair will be used, pairs such as *max-max* or *min-min* could be considered equally well, which gives the title *m-m*. All such pairs are related using the rules *A3* provided in the Appendix. Thus the *min* in (1.2) could be replaced by $max(\overset{*}{g}x_t + c^*, \overset{*}{d}y_t + d^*)$ where $\overset{*}{g} = -g\, c^* = -c$, $\overset{*}{d} = -d\, d^* = -d$. It might be noted that linear equations can be obtained by taking $b = -\infty$, $c = \infty$.

A form of particular interest has $a = b = g = d = 1$ and is called the "integrated *m-m* process", having ,

$$x_{t+1} = \max(x_t + a, y_t + b) + \varepsilon_{x,t+1} \qquad (1.3)$$

$$y_{t+1} = \min(x_t + c, y_t + d) + \varepsilon_{y,t+1} \qquad (1.4)$$

This is the bivariate version of a system discussed by Olsder and Delft (1991) but with added stochastic terms. To see how this system works in at least one case, suppose that $a < 0$, $d > 0$. Without the y_t term in the max component of (1.3), x_t would be a random walk with downward drift, but the y_t term may hold up x_{t+1} to a higher set of values, whereas in (1.4) the reverse holds. Thus, the two series are closely intertwined and the

marginal processes $E[x_{t+1}|x_{t-j}, j>0]$ and $E[y_{t+1}|y_{t-j}, j>0]$ are inclined to have quite different properties from the joint process.

Some examples are given later which show that m-m models provide better fits out-of-sample than linear models. In addition, m-m models are shown to exhibit strictly non-linear behavior which linear models cannot duplicate. Figures 1(a) and 2(a) show realizations of 200 observations from series (1.3) and (1.4), with (a, b, c, d) taking values (-0.5, 0.3, 0.3, 0.5), and (-0.01, -0.3, 0.3, -0.1) respectively. In all realizations, the distributions of $\{\varepsilon_{xt}\}$ and $\{\varepsilon_{yt}\}$ are taken to be $N(0,1)$ although this distribution has no particular relevance. It should be noted that even if the series $\{x_t\}$ and $\{y_t\}$ appear not to be stationary (see Figure 1(a) or 2(a)) the series $\{x_t-y_t\}$ may be stationary with thresholds (see Figure 1(b) or 2(b)) with some conditions depending on (a, b, c, d).

These examples are nonlinear processes because the *max* component of series $\{x_{t+1}\}$ sometimes chooses y_t+b and the *min* component of series $\{y_{t+1}\}$ sometimes chooses x_t+c. The timing of nonlinear data generation can be analyzed; if one can forecast the timing of nonlinear operation or relate it to the level of z_t ($\equiv x_t-y_t$). Figure 1(b) shows the timing at which the max operation of (1.3), upper circle, (or min operation of (1.4), lower circle) chooses y_t+b (or x_t+c). It is seen that the nonlinearity largely occurs when z_t is in the lower regime, which will be explained in detail in section 3. Figure 2(b) shows this phenomenon more clearly.

2. Equilibrium Values

A particular form of equilibrium will be considered. If there are no further stochastic shocks and if the process converges to constant values, so that $x_t \to x$, $y_t \to y$,

when does convergence occur and what values of x, y will be found? Starting with (1.1) and (1.2) with no shocks, if convergence has occurred, one gets

$$x = \max(\alpha x + a, \ \beta y + b) \tag{2.1}$$

$$y = \min(\gamma x + c, \ \delta y + d) \tag{2.2}$$

which gives

$$0 = \max((\alpha - 1)x + a, \ \beta y - x + b) \tag{2.3}$$

$$0 = \min(\gamma x - y + c, \ (\delta - 1)y + d) \tag{2.4}$$

and this can be written as

$$0 = \max(X, \ \beta \phi Y - \theta X + f_1) \tag{2.5}$$

$$0 = \min(Y, \gamma \theta X - \phi Y + f_2) \tag{2.6}$$

where

$$\text{and} \quad \left. \begin{array}{l} X = (\alpha - 1)x + a, \quad Y = (\delta - 1)y + d \\ \theta = (\alpha - 1)^{-1}, \quad\quad \phi = (\delta - 1)^{-1} \\ f_1 = \theta a - \beta \phi d + b \\ f_2 = \phi d - \gamma \theta a + c \end{array} \right\} \tag{2.7}$$

First consider the case $|\boldsymbol{a}| < 1$, $|\boldsymbol{d}| < 1$. The equilibrium possibilities are

(i) if $f_1 \leq 0, f_2 \geq 0, X = 0, Y = 0$

(ii) if $\boldsymbol{b}f_2 + f_1 \leq 0, f_2 < 0, X = 0, Y = f_2 / \boldsymbol{f} = f_2 (\boldsymbol{d}1) > 0$

(iii) if $\boldsymbol{g}f_1 + f_2 \geq 0, f_1 > 0, X = f_1 / \boldsymbol{q} = f_1(\boldsymbol{a}\text{-}1) < 0, Y = 0$

(iv) (X, Y) given by $\boldsymbol{bf}Y - \boldsymbol{q}X + f_1 = 0, \boldsymbol{g}X - \boldsymbol{f}Y + f_2 = 0$, provided $X \leq 0, Y \geq 0$. In the special case $a=b=c=d=0$ which gives $f_1 = f_2 = 0$, this equilibrium only exists if $\boldsymbol{b} = \boldsymbol{g}^{-1}$.

If $\boldsymbol{a} > 1$, then x_t will be explosive, as if $x_t > 0$

4

$$x_{t+1} > x_t$$

and so there will be no equilibrium.

For the integrated model, given by (1.3), (1.4) if there is an equilibrium then

$$
\begin{aligned}
x &= \max(x+a,\ y+b) \\
y &= \min(x+c,\ y+d)
\end{aligned}
\tag{2.8}
$$

which can be written as

$$
\begin{aligned}
0 &= \max(a,\ z+b) \\
0 &= \max(-d,\ z-c)
\end{aligned}
\tag{2.9}
$$

where $z = y - x$, using rules A1 and A3. It can be seen that there is no equilibrium if $a > 0$, $-d > 0$ as the two sides of the equations cannot equate. If $a \leq 0$, $-d \leq 0$, then the equilibrium occurs if $z + b = 0$ and $z - c = 0$, which requires that $b = -c$. Thus the integrated system only has an equilibrium if $a \leq 0$, $d \geq 0$, $b = -c$ in which case the equilibrium is $y = x + c$

If there is a shock and then no further shocks, when does convergence occur and what values of x, y will be found? If the dynamic path of z_t hits the zone *(b-c-max(a-c,b-d), max(a-c,b-d))*, then x_t and y_t start to oscillate, and so does z_t. Otherwise they converge to constant values. Note that even if the process oscillates, it may be stable under a broad conception of stability. Regardless of the behavior of the process around the equilibrium, these series display a convergent behavior back toward the equilibrium if the deviation from the equilibrium is large.

Figures 4(a) and (b) show the oscillation case and stable case, which depend on the size of the shocks, i.e., $e_{x0} = 1.0$ for (a) and $e_{x0} = 1.5$ for (b). Because of the discreteness of the data, the size of the shock will determine whether the process oscillates or not. For reference, 1(c) and (d) show the case for a stationary m-m process.

From the discussion of the equilibrium, this process does not have any equilibrium if $a >$ 1 or $d > 1$. Figure (c) shows the case of equilibrium and convergence back to it and figure (d) shows no equilibrium case, so there are no convergent properties in the series.

3. The Integrated System

  Fact: If the processes $\{x_t\}$ and $\{y_t\}$ are generated by the equations (1.3) and (1.4), then these are "nonlinearly integrated" processes.

  Proof: We can show this fact by induction. Assume $x_t = y_t = 0$ for all $t < 0$ and $e_{x,t}$, $e_{y,t}$ are independent and *i.i.d.* with variances σ_x^2 and σ_y^2 respectively. It suffices to show that $\{x_t\}$ has a nonlinearly integrated property.

$$x_T = \max(x_{T-1} + a, y_{T-1} + b) + e_{x,T}$$
$$= \begin{cases} e_{x,T} + a + x_{T-1}, & \text{if } x_{T-1} + a \geq y_{T-1} + b \\ e_{x,T} + b + y_{T-1}, & \text{if } x_{T-1} + a < y_{T-1} + b \end{cases}$$

If we do backward-induction one more step,

$$x_T = \max\left(\max(x_{T-2} + a, y_{T-2} + b) + e_{x,T-1} + a, \min(x_{T-2} + c, y_{T-2} + d) + e_{y,T-1} + b\right) + e_{x,T}$$
$$= \begin{cases} e_{x,T} + e_{x,T-1} + a + a + x_{T-2}, & \text{if } x_{T-1} + a \geq y_{T-1} + b \text{ and } x_{T-2} + a \geq y_{T-2} + b \\ e_{x,T} + e_{x,T-1} + a + b + y_{T-2}, & \text{if } x_{T-1} + a \geq y_{T-1} + b \text{ and } x_{T-2} + a < y_{T-2} + b \\ e_{x,T} + e_{y,T-1} + b + c + x_{T-2}, & \text{if } x_{T-1} + a < y_{T-1} + b \text{ and } x_{T-2} + c \geq y_{T-2} + d \\ e_{x,T} + e_{y,T-1} + b + d + y_{T-2}, & \text{if } x_{T-1} + a < y_{T-1} + b \text{ and } x_{T-2} + c < y_{T-2} + b \end{cases}$$
$$= \cdots$$

and so forth, until the starting values are reached. Then one can calculate the mean and variance of x_T.

$$E[x_T] = pa + qb + rc + sd ,$$

$$Var[x_T] = (p+q)\sigma_x^2 + (r+s)\sigma_y^2,$$

where $p + q + r + s = T$ and p, q, r, s are non-negative integers.

There exists δ such that $\min\left(s_x^2, s_y^2\right) \geq d > 0$, so

$$Var[x_T] = (p+q)s_x^2 + (r+s)s_y^2 \geq Td = O(T) \blacksquare$$

These processes are cointegrated in the usual sense even though they come from nonlinearly integrated processes. The bivariate integrated system (1.3), (1.4) can be rewritten as

$$\Delta x_{t+1} = a + \max(0, -z_t + b - a) + \varepsilon_{x,t+1} \qquad (3.1)$$

$$\Delta y_{t+1} = d + \min(0, z_t + c - d) + \varepsilon_{y,t+1} \qquad (3.2)$$

using rule $A1$, where $z_t = x_t - y_t$. Using rule $A3$ and subtracting (3.2) from (3.1) gives

$$\Delta z_{t+1} = (a-d) + \max(0, b-a-z_t) + \max(0, d-c-z_t) + \eta_{t+1} \qquad (3.3)$$

where $h_{t+1} = e_{x,t+1} - e_{y,t+1}$. It is seen that, if z_t is stationary, then (3.1), (3.2) make up a nonlinear error-correction system, which implies that x_t, y_t will be I(1) and linearly cointegrated. For ease of presentation, to consider the properties of z_t, only the case $b = -c$ will be analyzed in detail. Writing $w_t = z_t - b$ and initially assuming $d < -a$ then (3.3) gives three regions:

Region (i): $w_t < d$, then $w_{t+1} = -w_t + \eta_{t+1}$,

so in this region w is $I_\pi(1)$, that is unit root at frequency p.

Region (ii): $d < w_t < -a$, $w_{t+1} = -d + \eta_{t+1}$,

so that w_t is $I(0)$ in this region, and

Region (iii): $-a < w_t$, $\Delta w_{t+1} = a - d + \eta_{t+1}$

so that w_t is $I(1)$ in this region. Using the equilibrium condition for this system, which includes $a \leq 0$ and $d \geq 0$, it is seen that a-d is negative, so w_{t+1} is a random walk with downward drift in region (iii). It is seen that if w_t becomes too small, i.e. negative, and so

is found in region (i), it is likely to change sign and thus go into another region. If w_t becomes large, it will be in region (i) and the downward drift will take it into another region. If it is in region (ii), when it is stationary, whether or not it stays there will depend on the width of this region compared to the standard deviation of $\mathbf{h}_{t+1}$. Overall as simulations show, w_t will appear to be $I(0)$, if the equilibrium constraints hold. The three regions change places if the assumption $d < -a$ is replaced by $d > -a$.

Figures 1(f) and 2(f) illustrate the plots of the autocorrelations of x_t, y_t and z_t. When d-a has a large positive value, the autocorrelation of z_t is declining very quickly as in Figure 1(f). However when d-a has a small positive value or is negative, so that the stability condition is violated, then the autocorrelation of z_t is declining slowly or is very similar with the autocorrelations of x_t and y_t, which appear to have long memory. Figure 1(d) shows the shape of the functional form of the cointegrated series against its lagged value when using simulated data. Under the equilibrium constraint it clearly shows different slopes with different level of lagged value of z_t.

If the equilibrium constraints do not hold, so that a-$d > 0$, for example, z_t is clearly not $I(0)$. Adding z_t to both sides of (3.3) gives

$$z_{t+1} = (a - d) - z_t + \max(z_t, b - a) + \max(z_t, d - c) + \eta_{t+1}$$

so that

$$z_{t+1} \geq (a - d) + z_t + \eta_{t+1}$$

If a-$d > 0$, z_t will be growing faster than the random walk with positive drift $\tilde{z}_{t+1} = (a - d) + \tilde{z}_t + \eta_{t+1}$, so that z_t is $I(1)$. Thus x_t, y_t will be $I(1)$ but not cointegrated which are shown in Figures 2.

Formal theorems and proofs of the conditions on the parameter $\{a, b, c, d\}$ for the process $\{z_t\}$ to be ergodic are as follow.

Theorem 1: The process $\{z_t\}$, defined by (3.3), is ergodic if $a-d < 0$.

Proof: We can rewrite (3.3) as

[Upper Regime] : $z_{t+1} = (a-d) + z_t + \eta_{t+1}$ if $z_t > \max(b-a, d-c)$

[Lower Regime] : $z_{t+1} = (b-c) - z_t + \eta_{t+1}$ if $z_t < \min(b-a, d-c)$

If $-\infty < b\text{-}a, d\text{-}c < +\infty$, then the third condition of Theorem 2.1[1] of Chan, Petruccelli, Tong and Woolford (1985) implies that $a\text{-}d < 0$ is sufficient and necessary condition for the process $\{z_t\}$ to be ergodic process ∎

Theorem 2: If $E\left[|\eta_t|^k\right] < \infty$ for some integer $k \geq 1$ and the parameters $\{a, b, c, d\}$ of (3.3) satisfy Theorem 1, the invariant probability distribution for the process $\{z_t\}$ has a finite k-th moment and the model is geometrically ergodic.

Proof: See the proof of Theorem 2.3 of Chan, Petruccelli, Tong and Woolford (1985) ∎

So far the cointegration vector is assumed to be (1,-1). This could be relaxed to a general case by using a linear transformation on the variable x_t to $\mathbf{1}x_t + \alpha$. Let $x = \lambda x' + \alpha$, still one can get a cointegrating relation between x and y but now the cointegration vector is $(\mathbf{1}, -1)$ rather than (1,-1).

---

[1]Theorem 2.1 of Chan, Petruccelli, Tong and Woolford (1985) follows as below. For any integer l, let $\infty = r_0 < r_1 < \ldots < r_l = +\infty$ and define $Z_t = f(0,k) + f(1,k)Z_{t-1} + a_t$, if $Z_{t-1} \in R_k$, where $R_k \equiv (r_{k-1}, r_k]$. $1 \leq k \leq l$. The process $\{Z_t\}$ is ergodic if only if one of the following conditions holds:
(a) $\phi(1,1) < 1, \phi(1,l) = 1, \phi(1,1)\phi(1,l) < 1$, (b) $\phi(1,1) = 1, \phi(1,l) < 1, \phi(0,1) > 0$
(c) $\phi(1,1) < 1, \phi(1,l) = 1, \phi(0,1) < 0$, (d) $\phi(1,1) = 1, \phi(1,l) = 1, \phi(0,l) < 0 < \phi(0,1)$
(e) $\phi(1,1)\phi(0,l) = 1, \phi(1,1) < 1, \phi(0,l) + \phi(1,l)\phi(0,1) > 0$

$$x'_{t+1} = max(\ x'_t + \frac{a+\alpha}{\lambda} - \alpha, \frac{y_t}{\lambda} + \frac{b}{\lambda} - \alpha\ ) + \varepsilon_{x,t+1}$$

$$= max(\ x'_t + a', \frac{y_t}{\lambda} + b'\ ) + \varepsilon_{x,t+1}$$

$$y_{t+1} = min(\ \lambda x'_t + c + \alpha, y_t + d\ ) + \varepsilon_{y,t+1}$$

$$= min(\ \lambda x'_t + c', y_t + d\ ) + \varepsilon_{y,t+1}$$

so the cointegrated series of this system has the following form,

$$z'_{t+1} = (\ a' - d\ ) - z'_t + max(\ z'_t, b' - a'\ ) + max(\ z'_t, d - c'\ ) + \eta'_{t+1}$$

where $z'_t = \lambda x'_t - y_t$, $\eta'_t = \lambda \varepsilon_{x,t} - \varepsilon_{y,t}$.

The process with an equilibrium constraint produces error-correction models that are similar to, but different from, the cointegration model which was considered by Balke and Fomby (1997). They have a threshold model, with a pull toward the center from each outer region but with x_t, y_t being random walks in the center region. Because the cointegrating relationship of m-m processes is linear, standard time series analyses used for linear cointegration will be valid asymptotically for the analysis of cointegration between m-m proceses. The nonlinearity of cointegration regression does not affect the order of integration of x_t, y_t and z_t. So we can apply conventional cointegration testing method to m-m processes.

4. The Stationary System

The system now to be considered is (1.1), (1.2) with $0 < \alpha < 1$, $0 < \delta < 1$ and $a = b = c = d = 0$. Some graphical examples are given as Figure 3 with (α, β, γ, δ) taking values (0.7,1.3,1.3,0.7) for the stationary series (panel a), and for comparison (1.01,0.3, 0.3,1.01) which produces an explosive series (panel c). The univariate series of panel (a)

does not show any clear nonlinearity. This will be tested in section 7 using single series

and the system by comparing the power of various statistics. In terms of the equilibrium

results of section 2, it follows that if $f_1 = f_2 = 0$ then the equilibrium is $x = y = 0$. Let μ_x,

μ_y be $E[x]$, $E[y]$, the unconditional means respectively from (1.1), then

$$E[x] > \alpha E[x], \text{ if } \beta \neq 0 \tag{4.1}$$

where $>$ means that the strict inequality holds for some time periods, i.e., it is assumed

that $\alpha x_t < \beta y_t$ for some t. It follows that $\mu_x > 0$. Similar assumptions, including $\gamma \neq 0$ will

give $\mu_y < 0$. A formal theorem and proof is as follows.

Theorem 3: If $|\alpha| < 1$, $|\delta| < 1$ then $E(x_t) > 0$, $E(y_t) < 0$

Proof: Set $f(x,y)$ be the joint density of (X,Y) then

$$E[\max(X,Y)|y] = y \int_{-\infty}^{y} f(x|y)dx + \int_{y}^{\infty} x f(x|y)dx$$

$$= y \int_{-\infty}^{y} f(x|y)dx + \int_{-\infty}^{\infty} x f(x|y)dx - \int_{-\infty}^{y} x f(x|y)dx$$

$$= \int_{-\infty}^{\infty} x f(x|y)dx + \int_{-\infty}^{y} \int_{-\infty}^{x} f(s|y)ds\,dx$$

$$E[\max(X,Y)] = E[E[\max(X,Y)|y]]$$

$$= \int_{-\infty}^{\infty} (\int_{-\infty}^{\infty} x f(x|y)dx) f_Y(y)dy + \int_{-\infty}^{\infty} (\int_{-\infty}^{y} \int_{-\infty}^{x} f(s|y)ds\,dx) f_Y(y)dy$$

$$= \int_{-\infty}^{\infty} x f_X(x)dx + \int_{-\infty}^{\infty} \int_{-\infty}^{y} \int_{-\infty}^{x} f(s,y)ds\,dx\,dy$$

By the same method,

$$E[\min(X,Y)] = \int_{-\infty}^{\infty} y f_Y(y)dy - \int_{-\infty}^{\infty} \int_{-\infty}^{x} \int_{-\infty}^{y} f(x,t)dt\,dy\,dx$$

If x_t, y_t are strictly stationary and $|\alpha| < 1$, $|\delta| < 1$, then

$$E[x_{t+1}] = E[\max(\alpha x_t, \beta y_t)] + E[\varepsilon_{x,t+1}]$$

$$= E[\alpha x_t] + \int_{-\infty}^{\infty} \int_{-\infty}^{\beta y} \int_{-\infty}^{\alpha x} f(s,y)ds\,dx\,dy$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\beta y} \int_{-\infty}^{\alpha x} f(s,y)ds\,dx\,dy \Big/ (1-\alpha) \quad > 0$$

$$E[y_{t+1}] = E[\min(\gamma x_t, \delta y_t)] + E[\varepsilon_{y,t+1}]$$

$$= E[\delta y_t] - \int_{-\infty}^{\infty}\int_{-\infty}^{\gamma x}\int_{-\infty}^{\delta y} f(x,t)\,dt\,dy\,dx$$

$$= -\int_{-\infty}^{\infty}\int_{-\infty}^{\gamma x}\int_{-\infty}^{\delta y} f(x,t)\,dt\,dy\,dx \Big/ (1-\delta) \quad < \quad 0$$

as required ∎

To consider measures of the temporal properties of this process, as

$$x_{t+1} = \max(\alpha x_t, \beta y_t) + \varepsilon_{x,t+1} \tag{4.2}$$

it follows that $E[x_{t+1}x_t] > \alpha E[x_t^2]$ since $E[\varepsilon_{x,t+1}\ x_t] = 0$ and assuming that $\beta y_t > \alpha x_t$ for some t. Substituting in (4.1) gives

$$x_{t+2} = \max(\alpha(\max(\alpha x_t, \beta y_t) + \varepsilon_{x,t+1}), \beta y_{t+1}) + \varepsilon_{x,t+2}$$

and this suggests that $E[x_{t+2}x_t] > \alpha^2 E[x_t^2]$ and generally

$$E[x_{t+k}x_t] > \alpha^k E[x_t^2] \tag{4.3}$$

Note that these do not directly involve autocovariances, as these quantities are not centered at the mean. If r_k is the kth autocorrelation then a little algebra from (4.3) gives

$$\rho_k > c_1\alpha^k - c_2$$

where $c_1 = E[x_t^2]/\text{var}(x_t^2)$, $c_2 = \mu_x^2/\text{var}(x_t^2)$.

The values of r_k for m-m processes are illustrated in Figure 3(b) and (d), the plot against k are shown for x_t and y_t. Figure 3(b) shows for stationary m-m case with $0<\alpha,\delta<1$, the autocorrelations are declining very quickly, but when $\alpha,\delta>1$ in Figure 3(d), each individual series are explosive series with slowly declining autocorrelations.

12

5. Estimation

Conventional estimation techniques of parameters cannot be directly applied to a min-max system because of the discontinuity of min-max functions. Olsder and Delft (1991) suggested three algorithms to solve the min-max problem. One of their methods makes an exponential transformation of a min-max system into one to which the conventional analysis can be applied. In order to calculate the parameters of (1.1) and (1.2), an exponential approximation for large s is used:

$$x_{t+1} = \frac{1}{s} \log[e^{s(\alpha x_t + a)} + e^{s(\beta y_t + b)}] + \varepsilon_{x,t+1} \qquad (5.1)$$

$$y_{t+1} = -\frac{1}{s} \log[e^{-s(\gamma x_t + c)} + e^{-s(\delta y_t + d)}] + \varepsilon_{y,t+1} \qquad (5.2)$$

In Olsder and Roos (1988) it has been shown that the exponential behavior of (5.1) and (5.2) as $s \to \infty$ leads exactly to Equation (1.1) and (1.2). The advantage of (5.1) and (5.2) is that conventional analysis, such as nonlinear least squares or maximum likelihood, can be used.

In the simulation study, we will control the value of s by the capacity of the computer. Each table of simulation study or empirical analysis will report the value of s that were used in the estimation procedure.

6. Linearity Testing of m-m processes

The bivariate system of m-m processes is intertwined as mentioned before, however the individual series might have quite different properties. It may not be easy to detect nonlinearity from single series, whilst the system or cointegrated series shows

nonlinearity clearly. This section discusses a simulation study of nonlinearity tests using single series and the system and compares the performance of each test.

We compare several testing methods for the null of linearity against the alternative of neglected nonlinearity. There are many tests, some of which are discussed in Granger and Teräsvirta (1993) and in Lee, White and Granger (1993). As mentioned by Granger (1995), they are all based on an assumption that the series are stationary or, in practice, at least short memory in mean. These test are clearly going to work poorly with trending, I(1) or extended memory variables. The test will be biased against rejection of the null hypothesis of no nonlinearity. It is clear that many of the standard tests for linearity cannot be directly applied to I(1) or extended memory variables.

Granger (1995) suggested testing for the null of linearity of the ECM by regressing the residuals from a cointegrating regression on their lagged values and a nonlinear function, and then performing a LM type test. A similar approach is implemented by Corradi, Swanson and White (1997) who regress the first difference of the data on the lagged value of the cointegrating vector and a nonlinear function under the maintained hypothesis of cointegration. The following two DGPs are compared,

$$\Delta \underline{X}_t = \mu + \delta \underline{Z}_{t-1} + \beta \Delta \underline{W}_t + g\left(\underline{Z}_{t-1}, \Delta \underline{W}_t\right) + e_t \tag{6.1}$$

$$\Delta \underline{X}_t = \mu_0 + \delta_0 \underline{Z}_{t-1} + \beta_0 \Delta \underline{W}_t + e_{0t} \tag{6.2}$$

where $\underline{X}_t = \left(x_t, y_t\right)'$, $\underline{W}_t = \left(x_{t-j}, y_{t-j}, j = 0,..., p\right)'$ and x_t, y_t are all I(1). There exists a constant A such that $z_t = x_t - A\,y_t$ is I(0) or at least mean reverting. Tests of linearity can be conducted by comparing the nonlinear specification (6.1) to the linear form (6.2) by performing the regression

$$e_{ot} = \mu_1 + \delta_1 \underline{Z}_{t-1} + \beta_1 \Delta \underline{W}_{t-1} + g\left(\underline{Z}_{t-1}, \Delta \underline{W}_{t-1}\right) + \varepsilon_t \tag{6.3}$$

and doing an LM test based on R^2.

The following methods[2] are used to find nonlinearity. The neural network test for neglected nonlinearity uses a single hidden layer network augmented by connections from input to output. The network output is then[3]

$$y_t = \tilde{x}_t'\theta + \sum_{j=1}^{q} \beta_j \psi(\tilde{x}_t'\gamma_j)$$

The hypothesis is that the optimal network weight b_j, has $b_j^* = 0$, $j = 1,...,q$. A Lagrange multiplier test leads to testing $E(\Psi_t e_t^*)y_t = 0$, where $\Psi_t \equiv \left(y\left(\tilde{X}_t'\Gamma_1\right),...,y\left(\tilde{X}_t'\Gamma_q\right)\right)'$, $\Gamma = (\Gamma_1,...,\Gamma_q)$ is chosen a priori, independently of X_t and $e_t^* = y_t - \tilde{X}_t'\theta$. A relevant statistic is

$$M_n = \left(n^{-1/2}\sum_{t=1}^{n}\psi_t e_t^*\right)'\hat{W}_n^{-1}\left(n^{-1/2}\sum_{t=1}^{n}\psi_t e_t^*\right) \xrightarrow{d} \chi^2(q) \text{ as } n\to\infty.$$

To avoid collinearity one can choose $q^* < q$ principal components of y_t that are not collinear with X_t, denoted y_t^*. A test statistic is

$$Neural = nR^2 \xrightarrow{d} \chi^2(q^*),$$

where R^2 is the uncentered squared correlation from a regression of $\hat{e}_t = y_t - \tilde{X}_t'\hat{\theta}$ on $\Psi_t^*, \tilde{X}_t$. The Keenan test is based on the correlation of $\hat{e}_t = y_t - \tilde{X}_t'\hat{\theta}$ and $f_t^2 = (\tilde{X}_t'\hat{\theta})^2$. A test statistic is

$$Keenan = \frac{\hat{e}'\hat{\varepsilon}(\hat{\varepsilon}'\hat{\varepsilon})^{-1}\hat{\varepsilon}'\hat{e}}{\hat{v}'\hat{v}/(n-2p-2)} \sim F(1, n-2p-2)$$

---

[2] We select the tests which have better power from the simulation study of Lee et al. (1993)

[3] In performing neural network test the logistic c.d.f $\psi(\lambda) = [1 + exp(-\lambda)]^{-1}$ is used.

where p is the number of explanatory variables, $\hat{\varepsilon}_t = f_t^2 - \tilde{X}_t'\hat{\lambda}$ and $\hat{v}_t = \hat{e}_t - \hat{\varepsilon}_t\hat{\delta}$. $\hat{\beta}$, $\hat{\lambda}$

and $\hat{\delta}$ are least square estimates from linear regressions. The Tsay test is very similar in

form to the Keenan test. Let P_t include $p(p+1)/2$ cross-product terms of the components

of X_t, of the form $y_{t-j}y_{t-k}$, $k \geq j$, $j,k=1,\ldots,p$, $\hat{\varepsilon}_t = P_t - \tilde{X}_t'\hat{\lambda}$ and $\hat{v}_t = \hat{e}_t - \hat{\varepsilon}_t\hat{\delta}$, then

$$Tsay = \frac{\hat{e}'\hat{\varepsilon}(\hat{\varepsilon}'\hat{\varepsilon})^{-1}\hat{\varepsilon}'\hat{e}/m}{\hat{v}'\hat{v}/(n-p-m-1)} \sim F(m, n-p-m-1), \text{ where m=p(p+1)/2.}$$

The Ramsey RESET test is a generalization of the Keenan test. Using one step ahead

forecast, f_t,

$$y_t = \tilde{X}_t'\theta + c_2 f_t^2 + \ldots + c_k f_t^k + v_t$$

The null hypothesis is $c_2 = \ldots = c_k = 0$. The test statistic is $nR^2 \xrightarrow{d} \chi^2(k-1)$ where R^2

is determined from a linear regression of $\hat{e}_t = y_t - \tilde{X}_t'\hat{\theta}$ on $f_t^2, \ldots, f_t^k, \tilde{X}_t$. Forming the

principal components of $(f_t^2, \ldots, f_t^k)$, choosing the p^* largest, and then regressing $\hat{e}_t$ on

these and $\tilde{X}_t$ gives R^2 statistics with

$$RESET = nR^2 \xrightarrow{d} \chi^2(p^*).$$

The White dynamic information matrix test is based on the covariance of the conditional

score function. The loglikelihood of a linear model $y_t = \tilde{X}_t'q + e_t$, $e_t \sim N(0, s^2)$, is

$$log\, f_t(X_t, \theta, \sigma) = constant - log\, \sigma - (y_t - \tilde{X}_t'\hat{\theta})^2/2\sigma^2$$

so that with $u_t = (y_t - \tilde{X}_t'\hat{q})/s$ the conditional score function is

$$s_t(X_t, q, s) = \nabla \log f_t(X_t, q, s) = s^{-1}(u_t, u_t X_t', u_t^2 - 1)'$$

The dynamic information matrix test is $n\hat{m}'_n \hat{J}_n^{-1}\hat{m}_n \xrightarrow{d} \chi^2(q)$ where $\hat{m}_n = n^{-1}\sum \hat{m}_t$,

$\hat{m}_t = S\ vec\ s_t\ s'_{t-1}$, S is a nonstochastic selection matrix, and q is the dimension of m_t

$$\hat{J}_n = n^{-1}\sum_{t=1}^{n}\hat{m}_t\hat{m}'_t - \left(n^{-1}\sum_{t=1}^{n}\hat{m}_t\hat{s}'_t\right)\left[n^{-1}\sum_{t=1}^{n}\hat{s}_t\hat{s}'_t\right]^{-1}\left(n^{-1}\sum_{t=1}^{n}\hat{s}_t\hat{m}'_t\right)$$

Equivalent test statistics are

$$WHITE1 = nR^2 \xrightarrow{d} \chi^2(q)$$

with R^2 of the regression of the constant unity on the $\hat{m}_t$, $\hat{s}_t$.

$$WHITE2 = nR^2 \xrightarrow{d} \chi^2(q)$$

with R^2 of the regression of $\hat{u}_t$ on the $\tilde{X}_t$, $\hat{k}_t$, with $\hat{m}_t = \hat{k}_t\hat{u}_t$. The McLeod and Li test uses the squared residuals from a linear model and applies a standard Ljung-Box Portmanteau test for serial correlation. The test statistic is

$$McLeod = n(n+2)\sum_{i=1}^{m}\frac{\hat{r}^2(i)}{n-i} \xrightarrow{d} \chi^2(m)$$

where $\hat{r}(k) = \sum_{t=k+1}^{m}\left(\hat{e}_t^2 - \hat{\sigma}^2\right)\left(\hat{e}_{t-k}^2 - \hat{\sigma}^2\right)\bigg/ \sum_{t=k+1}^{m}\left(\hat{e}_t^2 - \hat{\sigma}^2\right)^2$, $\hat{\sigma}^2 = n^{-1}\sum_{t=1}^{n}\hat{e}_t^2$

Simulation Design: The stationary and integrated m-m processes were generated from (1.1), (1.2), (1.3) and (1.4). For the analysis of a single series of the integrated process, the series were differenced to produce stationary sequences. For the analysis of integrated m-m processes, a cointegrating regression or error correction model using a known cointegrating vector (1,-1), will be investigated in the simulation. Throughout e_{xt} and e_{yt} are drawn form $N(0,1)$. (i) Integrated m-m processes: (a, b, c, d) taking values (-1,0.7,-0.7,0.5), (-0.1,-1,1,0.1) and (0.01,0.4,0.1,0.04). (ii) Stationary m-m processes: (α,

17

β, γ, δ) taking values (0.7,1.3,1.3,0.7) and (0.7,-0.5,-0.5,0.3). For all the simulations, the information set for univariate series x is $X_t = x_{t-1}$ or $X_t = (x_{t-1}, x_{t-2})'$ and for z, $X_t = z_{t-1}$. For the bivariate case, the information set of x is $X_t = (x_{t-1}, y_{t-1})'$.

The results of the simulation can be summarized as follows. Table 6.2 shows the results of the simulation using an integrated m-m system.

(1) Using univariate series, each test has less power if more lags are adopted in the test. This suggests that a univariate series from an m-m model has less obvious evidence of nonlinearity. Tests using differences of single series have less power than tests using ECM or cointegrated series

(2) MM3 is more likely to appear linear than MM1 or MM2 since b is small and c is big. In MM3, c is not well identified because the *min* operator chooses the y series in most cases, thus series y looks like a linear process.

(3) An error correction model can improve the power of tests, but tests using a cointegrated series have the best power in the most cases. To improve the power it is necessary to add the other series or the spread to detect the nonlinearity. If one does not consider the spread, but uses the other series in a bivariate model, it can not improve the power as much as tests using cointegrated series.

(4) Overall the Neural network test has good power in many case.

Table 6.3 shows the simulation results of a stationary m-m model. In these results the benefit of the bivariate model is clear. Tests using univariate series can not improve the power even though we change the information set. However if the other series was also used in the test (XB of MM4 and MM5), then the powers were increased, especially for the Neural and Tsay tests.

8. Application

There could be several interesting practical examples of m-m processes. The transaction cost might be an example of these processes. When economic agents make decisions, they will take no action if transaction costs are larger than the benefits from that action.

Another interesting application is a nonlinear error correction model relating a pair of interest rates of different risk or different maturity periods. Suppose that these interest rate are I(1), with their spread being I(0) as found in Hall, Anderson and Granger (1992). Now consider how transaction costs might affects spread movement. A nonlinear error correction model provides an appropriate framework for this. Economic theory predicts that arbitrage and corresponding yield adjustment will occur only when the interest rate is 'sufficiently far' from equilibrium rate in the market, to imply a net gain to investors after transaction costs. This can be modeled as an "on/off" threshold error correction process. The threshold is determined by transaction costs, which deter responding to small deviation from equilibrium. One could also argue that there are two types of player in the market, the seller who want the rates to be maximized and the buyer who want the rates minimized, which thus makes the m-m model appropriate.

Data description: We examine data on the interest rates of Commercial Paper (6-months) and Treasury Bill (3-months) which reflect the risky and safe rates, using monthly observations from January 1970 to October 1997. (In-sample=1970:01~1989:12 (240 observations), out-of-sample=1990:01~1997:10 (94 observations)).

To learn about the basic structure of the data, initially a non-parametric analysis was conducted. Figure 5(c) shows the scatter plot between changes in the dependent

variable and the lagged cointegrating residual when the non-parametric kernel regression (bandwidths chosen using the leave-one-out cross validation function) is employed to estimate the ECM employing the variables in the linear ECM. Even though the scatter plot for changes in the commercial paper rate against the lagged value of spread, Figure 5(d), does not show any particular nonlinear property, because of the nonlinearity of spread, the linear ECM is a poor approximation. An interpretable parametric form to model this nonlinearity is readily apparent. The threshold ECM can be estimated since the slope coefficient of the lagged z_{t-1} is zero around the origin, and unity with negative intercept when z_{t-1} has a large value. However the plot of the univariate series, Figure 5(d), shows little evidence of nonlinearity.

Linearity tests find statistically significant evidence for each of the differences of Treasury bills and commercial paper interest rates and the spread (see Table 6.4). Virtually all of the tests except the neural network tests that use one principal component found evidence of nonlinearity of the individual series. The interest rate of Treasury bills shows less clear evidences of nonlinearity. This result will be confirmed by the results from estimation of m-m model. This rate is driving the rate of Commercial paper and is less affected by the spread between the two rates. Linearity tests using the ECM model show similar results with the tests using differenced series. All of the tests except Neural, RESET with one principal component, and Keenan, reject the null hypothesis of linearity strongly. These tests suggest that each individual series of interest rates has nonlinearity and also the spread has a strong nonlinearity.

As a basic model, (1.3) and (1.4), Case I produces the estimated results. The problem of this model is that the *min* operator of the equation (1.4) never selects series X_t

within this sample period. This means the parameter c of the equation (1.4) is not identified in this sample. To correct this problem one can use a modified m-m model by linearizing (1.3), i.e., let $c = +\infty$.[4] Case II shows the results of estimation and comparison using the partially linearized m-m model.

Another possible modification is a mixture of the integrated and stationary m-m models which allows partial adjustment for the spread in the error correction model of (3.1) and (3.2),

$$\Delta x_{t+1} = a + \max(0, -\boldsymbol{a} \, z_t + b - a) + \boldsymbol{e}_{x,t+1} \tag{3.1'}$$

$$\Delta y_{t+1} = d + \min(0, \boldsymbol{b} \, z_t + c - d) + \boldsymbol{e}_{y,t+1} \tag{3.2'}$$

where $0 < \alpha, \beta \le 1$. After a brief explanation of this model we will attempt to estimate this model using our empirical data. The x and y series are generated by the system,

$$x_{t+1} = (1 - \boldsymbol{a})x_t + \max(\boldsymbol{a}x_t + a, \boldsymbol{a}y_t + b) + \boldsymbol{e}_{x,t+1} \tag{1.3'}$$

$$y_{t+1} = (1 - \boldsymbol{b})y_t + \min(\boldsymbol{b}x_t + c, \boldsymbol{b}y_t + d) + \boldsymbol{e}_{y,t+1} \tag{1.4'}$$

If $\alpha = \beta = 1$, this system is exactly the same as the integrated m-m processes, (1.3) and (1.4). If $\alpha \le 0$ or $\beta \le 0$, then this system is not stable. Case III considers this modification with linearization of the *min* function.

---

[4] If $c = +\infty$, then
$$y_{t+1} = y_t + d + \varepsilon_{y,t+1} \tag{1.4'}$$
Using a similar method as (3.3), we could get
$$\Delta z_{t+1} = (a - d) + \max(0, b - a - z_t) + \eta_{t+1} \tag{3.3'}$$
It is seen that if z_t is stationary, then (1.3), (1.4') are a nonlinear error-correction system. (3.3') gives two regions:
(i) $z_t \ge b\text{-}a$, then $z_{t+1} = a - d + z_t + \eta_{t+1}$,
so that z_t is $I(1)$ in this region.
(ii) $z_t < b\text{-}a$, then $z_{t+1} = b - d + \eta_{t+1}$,
so in this region z_t is $I(0)$. If $a\text{-}d$ is negative, so z_t is a random walk with downward drift in region (i). It is seen that if z_t is in region (ii), when it is stationary. Overall, z_t will appear to be $I(0)$, as simulation shows, if the equilibrium constraints hold.

The specified models are estimated by nonlinear least-squares. This section presents estimated equations, and in the spirit of Granger and Anderson (1978) we also report the ratio of the residual variance of the m-m model to that of the corresponding VAR(3) model chosen by AIC. It is seen that the error variance of m-m model is a little larger than that of the VAR(3) model. s^2/s_L^2; s is the residual standard deviation of the m-m model and s_L is the corresponding statistic for the VAR(3) model. In general we need to restrict our consideration to nonlinear models for which this ratio is less than 0.9, so as to avoid models that may be spurious. In our nonlinear model, this ratio is very close to one for all of the cases, which suggests that the two models are virtually the same explanation for the in-sample period. But it turns out that our nonlinear model is better for predictability in the out-of-sample period.

Table 8.1 contains some diagnostics associate with these models. Residuals are tested against fourth-order ARCH using the LM test of Engle (1982), and checked with the Jarque-Bera normality test. The skewness and excess kurtosis of the residuals are also reported.

Table 8.1. Test statistics and *p*-values of fourth-order ARCH and Jarque-Bera normality tests, and skewness, excess kurtosis measure of residuals from the estimated nonlinear models and the ratio of the residual variance to VAR(3) model

	ARCH(4) Test	*Jarque-Bera Test*	*Skewness*	*Excess kurtosis*	s^2/s_L^2
Case II					
eq. 1	50.3 (0.00)	507.3 (0.00)	-1.04	9.95	1.0628
eq. 2	66.6 (0.00)	444.2 (0.00)	-1.11	9.40	1.0236
Case III					
eq. 1	49.5 (0.00)	482.9 (0.00)	-0.86	9.89	1.0026
eq. 2	62.3 (0.00)	439.5 (0.00)	-1.04	9.41	1.0148

Note: The values in parenthesis are *p*-values

Another way of evaluating the estimated nonlinear model is post-sample forecasting, although the insight to be gained depends on what happens in the time series

during the prediction period. If the prediction period does not contain a clear close between the CP and T-bill rate, i.e., spread in a lower regime, then the linear and nonlinear forecast will be similar, unless the specification of the m-m model is totally inadequate. The forecasts were made without re-estimating the model during the prediction period. In all cases the MSFE of the one-step-ahead forecasts were calculated. Plots of actual and predicted series by m-m and VAR models of Figure 6 suggest that the m-m model is better than the VAR model when nonlinear operation is working, i.e., *max* chooses y_t series or *min* chooses x_t series. That means an m-m model can predict more accurately when the spread is very small, i.e., when it lies in a lower regime.

The diagnostic statistics that are presented in Table 8.1 show clear evidence of an ARCH effect in the residuals and it is true for the VAR model. The next two tables will show diagnostic tests and forecastability of m-m models which is including GARCH (1,1). For the comparison, we estimate VAR(3) with a GARCH(1,1) specification.

Table 8.2. Mean squared forecast errors

CASE II	CP 6-months	T-bills 3-months	Spread
M-M	0.0408	0.0305	0.0098
VAR	0.0761	0.0493	0.0111
SD statist	-4.7973	-4.0953	-1.0525
CASE III			
M-M	0.0611	0.0360	0.0133
VAR	0.0761	0.0493	0.0111
SD statist	-4.9704	-4.0044	-3.9850

The *SD statistics* is the test in Granger and Newbold (1986, pp. 278-280) which is based on the correlation coefficient, *r*, of the sum and differences of the forecast errors. The null hypothesis is *r=0*, and the null distribution is based on the well-known approximation that $\frac{\sqrt{n-1}}{2}\log\left(\frac{1+r}{1-r}\right) \sim N(0,1)$, where *n* is ex-post sample size.
# of lags of VAR model by BIC : $p = 3$

Table 8.3. Test statistics and *p*-values of fourth-order ARCH and Jarque-Bera normality tests, and skewness, excess kurtosis measure of residuals from the estimated nonlinear models and the ratio of the residual variance to VAR(3) model

	ARCH(4) Test	Jarque-Bera Test	Skewness	Excess kurtosis	s^2/s_L^2
Case II,					
eq. 1	0.6773 (0.954)	0.0350 (0.983)	0.0134	2.9456	0.9989
eq. 2	2.6833 (0.612)	3.3259 (0.190)	0.1656	3.4887	1.0002
Case III,					
eq. 1	0.8154 (0.937)	0.3399 (0.844)	0.0800	3.1017	0.9986
eq. 2	2.7262 (0.605)	2.7845 (0.249)	0.1528	3.4455	1.0011

Note: The values in parenthesis are *p*-values

The estimated values of *a, d* of all models satisfy the stability condition of Theorem 1, but the difference is very small, 0.0200 in case I, 0.0449 in case II, and 0.5650 in case III. This implies that the shock to spread could be persistent within the short period. Considering the forecastability out-of-sample, the m-m model outperforms the VAR model for the series of commercial paper rate. Since the Treasury bills rate shows less nonlinearity compared to other series the m-m and VAR models have virtually the same predictive power for this series. However, the commercial paper rate and spread show very clear nonlinearity, the m-m model can outperform VAR.

Table 8.4. Mean squared forecast error

CASE II,	CP 6-months	T-bills 3-months	Spread
M-M	0.0379	0.0281	0.0096
VAR	0.0444	0.0314	0.0088
SD statist	-2.1203	-1.7912	1.0293
CASE III,			
M-M	0.0388	0.0269	0.0087
VAR	0.0444	0.0314	0.0088
SD statist	-2.4832	-2.1948	-0.5654

# of lags of VAR model by BIC : $p = 3$

In case III, the estimated value of α, 0.2737, is less than unity, which means a partial adjustment for the deviation of spread. The number of *max* operators which are

choosing the *y* variable increases from 56 of Case II Ⅽto 235, and the difference of *a, d* is very small, -0.5650, but still negative. Considering forecastibility out-of-sample the m-m model outperforms for the commercial paper series, and does marginally better for the Treasury bill and spread series.

Table 8.5. Mean squared forecast error

(1)	CP 6-months	T-bills 3-months	Spread
M-M	0.0155	0.0130	0.0014
VAR	0.0229	0.0163	0.0027
SD statist	-5.2184	-3.3913	-5.3356
(2)			
M-M	0.0152	0.0140	0.0027
VAR	0.0198	0.0173	0.0021
SD statist	-3.2923	-3.1357	2.1936
(3)			
M-M	0.0146	0.0128	0.0032
VAR	0.0193	0.0156	0.0030
SD statist	-2.9963	-2.5597	0.5236
(4)			
M-M	0.0169	0.0137	0.0032
VAR	0.0217	0.0164	0.0032
SD statist	-2.7326	-2.4847	0.1378
(5)			
M-M	0.0224	0.0152	0.0082
VAR	0.0215	0.0150	0.0061
SD statist	0.5630	0.2282	4.3068

# of lags of VAR model by BIC : $p = 3$

```
--------------------------
 t-1    t*   t+1    t+2      * nonlinear operation at t
 (1)   (2)   (3)    (4)
--------------------------
```

(1) MSFE when nonlinear operation of m-m function is working (spread <b-a)
(2) MSFE of 1-step forecast right after knowing that nonlinear operation of m-m function was working
(3) MSFE of 1-step forecast of 1 month after that nonlinear operation of m-m function was working
(4) MSFE of 1-step forecast of 2 month after that nonlinear operation of m-m function was working
(5) MSFE when spread is above lower bound (b-a)
MSE ratio (s^2/s_L^2) when (spread <b-a): eq1=1.0280, eq2=1.0249
MSE ratio (s^2/s_L^2) when (spread <b-a): eq1=0.9929, eq2=0.9946

This m-m model has linear specification before the nonlinear operation is working. If we focus on these events when nonlinear operations are working, then we would get stronger support of the m-m model compared to overall sample period

comparison. In Figures 6, prediction errors of m-m and VAR models using Case II′ are useful in assessing the benefits of the m-m model. It is seen that the movement of spread in the lower regime area is much better explained by the m-m model than by the VAR(3) model. The following table 8.5 shows forecast comparisons around these events. And for the reference, we reported the s^2/s_L^2 values, which do not show much difference between different regimes.

9. Conclusions

The purpose of this paper has been to introduce a nonlinear model motivated by economic agents' minimizing or maximizing behavior. It turns out this system can have an error correction model with thresholds, which can account for the presence of a fixed cost of adjustment. This model would appear to have potential applicability when this type of nonlinear behavior is thought to be important and the system has equilibrium even though each individual series is not stationary. To simplify our argument we have considered only either an integrated or stationary bivariate m-m process, not mixtures of two processes. However, as mentioned briefly in section 8, a mixture of these two processes would be a natural extension and have interesting properties, which were not fully investigated in this paper. Another underlying assumption is that the cointegration vector is known. In the actual data with an unknown cointegration vector, except the spread of interest rates, we need to estimate the cointegrating relationship. Even though Balke and Fomby (1997) partly investigated this problem, there are still open questions, such as how to estimate a nonlinear cointegration and how to test for this.

Appendix

Some rules for max, min operators

A1 $\quad \max(X+a, Y+a) = a + \max(X,Y)$

A2 $\quad \max(\alpha X, \alpha Y) = \alpha \max(X, Y) \qquad$ if $\alpha > 0$

A3 $\quad \min(-X, -Y) = - \max(X, Y)$

which is an example, with $\alpha = -1$ of A4

A4 $\quad \max(\alpha X, \alpha Y) = \alpha \min(X, Y) \qquad$ if $\alpha < 0$

A5 $\quad \max(\max(a_1, a_2), \max(a_3, a_4)) = \max(a_1, a_2, a_3, a_4)$

A6 $\quad \max(a_1, a_2) + \max(a_3, a_4) = \max(a_1 + a_3, a_1 + a_4, a_2 + a_3, a_2 + a_4)$

Reference

Anderson, H.M., 1996, Transaction Costs and Nonlinear Adjustment Towards Equilibrium in the US Treasury Bill Market, University of Texas at Austin, mimeo.

Balke, N.S., and T.B. Fomby, 1997, Threshold Cointegration, forthcoming in *International Economic Review.*

Chan, K.S., J. Petruccelli, H. Tong and S. Woolford, 1985, A Multiple Threshold AR(1) Model, *Journal of Applied Probability*, 22, 267-279.

Corradi, V., N.R. Swanson and H. White, 1995, Testing for Stationarity-Ergodicity and for Comovement Between Nonlinear Discrete Time Markov Processes, University of Pennsylvania, mimeo.

Granger, C.W.J., 1995, Modelling Nonlinear Relationships Between Extended-Memory Variables, *Econometrica*, 63, 265-279.

Granger, C.W.J., and T. Teräsvirta, 1993, *Modelling Non-linear Economic Relationships*, Oxford University Press, New York.

Lee, T.H., H. White and C.W.J. Granger, 1993, Testing for Neglected Nonlinearity in Time Series Models - A Comparison of Neural Network Methods and Alternative Tests, *Journal of Econometrics*, 56, 269-290.

Olsder, G.J., and G. Delft, 1991, Eigenvalues of Dynamic Max-Min Systems, *Discrete Event Dynamic Systems: Theory and Applications* 1, 177-207.

Olsder, G.J., and C. Ross, 1988, Cramer and Cayley-Hamilton in the Max-Algebra, *Linear Algebra and Its Applications*, 101:87-108.

Table 6.1. Critical Values (5%) for Univariate Models

Test	Univariate 1	Univariate 2	Bivariate 1
NEURAL	5.60 (5.99)	5.53 (5.99)	5.84 (5.99)
KEENAN	3.69 (3.84)	3.66 (3.84)	3.53 (3.84)
TSAY	3.69 (3.84)	2.51 (2.37)	2.56 (2.37)
WHITE1	10.27 (9.49)	16.56 (15.51)	18.23 (16.92)
WHITE2	10.17 (9.49)	15.14 (15.51)	16.52 (16.92)
MCLEOD	31.35 (31.41)	31.14 (31.41)	31.24 (31.41)
RESET1	3.42 (3.84)	3.55 (3.84)	3.61 (3.84)

(i) The first numbers in columns of univariate 1 and univariate 2 are simulated critical values from AR(1), $y_t = 0.6 y_{t-1} + e_t$, and the first number in columns of bivariate 1 from VAR(1), $x_t = 0.6 x_{t-1} + e_t$, $y_t = 0.6 y_{t-1} + e_t$ with 200 sample size and 6000 replications.

(ii) The second number in parenthesis is the asymptotic critical value.

(iii) Univariate 1 denotes a univariate model with lag 1, Univariate 2 denotes a univariate model with lag 2 and Bivariate 1 denotes a bivariate model with lag 1.

(iv) We choose $q=10$ and $q^*=2$ largest principal components (excluding the first principal components). The input to hidden unit weights Γ_{ij} were randomly drawn from uniform distribution on [-2,2]. The variables X_t, Y_t are rescaled onto [0,1]. For the White dynamic information matrix tests, $m_t' = \sigma^{-2} \tilde{X}_t' \tilde{X}_{t-1} u_t u_{t-1}$ were constructed without any identical columns to secure full rank of m_t matrix. For RESET1 $k=5$, $p^*=1$, RESET2 $k=10$, $p^*=2$.

Table 6.2. Power for integrated min-max processes

TEST	MM1					MM2					MM3				
	DX1	*DX2*	*DXB*	*ECM*	*Z*	*DX1*	*DX2*	*DXB*	*ECM*	*Z*	*DX1*	*DX2*	*DXB*	*ECM*	*Z*
NEURAL	37.2	13.8	92.7	100	100	11.3	9.1	18.3	85.5	95.0	7.2	5.8	8.0	43.9	50.4
	(38.9)	(16.1)	(93.4)	(100)	(100)	(12.7)	(11.1)	(19.1)	(87.1)	(95.2)	(8.2)	(7.5)	(8.8)	(46.0)	(51.7)
KEENAN	45.7	44.3	97.2	100	100	11.1	9.2	21.1	62.1	76.5	6.3	5.5	8.1	20.8	21.5
	(47.0)	(46.9)	(97.5)	(100)	(100)	(11.4)	(10.1)	(23.6)	(63.1)	(77.1)	(6.6)	(6.4)	(9.6)	(21.3)	(21.5)
TSAY	45.7	44.6	95.2	100	100	11.1	10.8	23.4	62.1	76.5	6.3	7.3	11.0	20.8	21.5
	(47.0)	(41.0)	(94.2)	(100)	(100)	(11.4)	(9.3)	(19.3)	(63.1)	(77.1)	(6.6)	(6.6)	(8.8)	(21.3)	(21.7)
WHITE1	65.5	31.4	98.8	98.4	100	10.8	9.6	16.2	41.4	81.4	7.2	8.2	10.3	15.7	25.8
	(60.5)	(25.9)	(97.8)	(97.8)	(100)	(8.3)	(6.7)	(11.3)	(37.5)	(78.5)	(5.2)	(6.0)	(7.3)	(12.6)	(22.8)
WHITE2	64.6	30.0	98.9	97.9	100	11.5	8.1	16.8	50.2	87.1	6.9	6.7	6.9	19.3	34.0
	(66.6)	(31.6)	(97.8)	(98.4)	(100)	(12.8)	(9.1)	(18.6)	(51.9)	(87.5)	(7.5)	(7.6)	(8.4)	(20.0)	(34.7)
MCLEOD	6.4	5.8	14.0	16.5	64.1	6.2	6.2	6.0	6.1	15.1	6.4	5.6	5.7	5.1	9.1
	(6.5)	(6.4)	(14.1)	(16.5)	(64.3)	(6.2)	(6.6)	(6.2)	(6.2)	(15.2)	(6.4)	(6.2)	(5.8)	(5.1)	(9.1)
RESET1	19.9	25.6	64.6	99.5	99.3	7.9	8.5	17.6	75.3	44.9	6.1	4.9	6.4	34.1	12.2
	(23.7)	(28.9)	(66.8)	(99.5)	(99.4)	(10.1)	(9.1)	(18.4)	(77.1)	(48.2)	(6.9)	(5.8)	(7.6)	(37.2)	(15.5)

Power using 5% asymptotic critical values is shown, and size-corrected power using simulated critical value is shown in parenthesis.
Sample size = 200. Replications = 1000.
MM1: a=-1,b=0.7,c=-0.7,d=0.5, *MM2*: a=-0.1,b=-1,c=1,d=0.1, *MM3*: a=0.01,b=0.4,c=0.1,d=0.04,
DX1, DX2 denote an equation of differenced X_t using AR(1) or AR(2)
DXB denotes a bivariate AR(1) model for differenced X_t
ECM denotes an error correction equation of X_t without any lags of **DX_t** , **DY_t**
Z uses a cointegrated series with (1,-1)

Table 6.3. Power for stationary min-max processes

	MM4			*MM5*		
TEST	*X1*	*X2*	*XB*	*X1*	*X2*	*XB*
NEURAL	8.5	5.4	90.3	44.6	16.9	82.2
	(9.5)	(6.6)	(90.8)	(47.8)	(19.7)	(83.2)
KEENAN	7.9	5.5	4.7	52.5	44.7	14.1
	(8.1)	(6.2)	(6.0)	(54.2)	(46.4)	(16.7)
TSAY	7.9	15.3	97.8	52.5	19.6	82.9
	(8.1)	(13.5)	(97.2)	(54.2)	(20.8)	(80.0)
WHITE1	21.6	11.0	44.2	40.4	73.3	52.8
	(17.6)	(7.7)	(36.9)	(33.6)	(67.9)	(44.5)
WHITE2	19.8	8.2	44.1	38.0	66.9	49.3
	(21.0)	(9.4)	(46.7)	(39.6)	(68.7)	(51.6)
MCLEOD	4.4	4.1	5.2	4.6	4.1	5.3
	(4.6)	(4.4)	(5.4)	(4.6)	(4.5)	(5.3)
RESET	5.1	4.3	6.9	4.3	3.5	5.1
	(6.2)	(4.9)	(7.3)	(5.6)	(4.6)	(5.6)

Power using 5% critical values simulated with AR(1) is shown.
Sample size = 200. Replications = 1000.
MM4: $\alpha=0.7$, $\beta=1.3$, $\gamma=1.3$, $\delta=0.7$, *MM5*: $\alpha=0.7$, $\beta=-0.5$, $\gamma=-0.5$, $\delta=0.3$
X1 and *X2* denote an equation of X_t using AR(1) or AR(2)
XB denotes a bivariate AR(1) model for X_t.

Table 6.4. Tests on interest rates of T-bill, CP and spread

Test	Treasury bill	Commercial paper	ECM : T-bill	ECM : CP	Spread
NEURAL1	SB= 0.6919	0.0054	0.0787	0.0000	0.0217
	HB= 0.5298	0.0054	0.0583	0.0000	0.0217
NEURAL2	SB= 0.0053	0.0001	0.0002	0.0000	0.0363
	HB= 0.0053	0.0001	0.0002	0.0000	0.0363
NEURAL3	SB= 0.0000	0.0000	0.0002	0.0000	0.0000
	HB= 0.0000	0.0000	0.0002	0.0000	0.0000
NEURAL4	SB= 0.0000	0.0000	0.0000	0.0000	0.0000
	HB= 0.0000	0.0000	0.0000	0.0000	0.0000
NEURAL5	SB= 0.0000	0.0000	0.0000	0.0000	0.0000
	HB= 0.0000	0.0000	0.0000	0.0000	0.0000
KEENAN	0.0005	0.0669	0.0001	0.0292	0.0681
TSAY	0.0000	0.0000	0.0000	0.0000	0.0000
WHITE1	0.0000	0.0000	0.0000	0.0000	0.0000
WHITE2	0.0000	0.0000	0.0000	0.0000	0.0000
MCLEOD	0.0000	0.0000	0.0000	0.0000	0.0000
RESET1	0.0450	0.6035	0.0000	0.0000	0.0000
RESET2	0.0005	0.0026	0.0000	0.0000	0.0000

(1) Treasury Bill, AR(9) determined by BIC; Commercial Paper, AR(8); Spread, AR(7)
(2) *ECM* of T-bill and CP with 2 lags
(3) Number of principal components of *NEURALi = i, i*=1,2,...,5.
(4) Number of principal components in *RESET*1 = 1, *RESET*2 = 2
(5) SB denotes Simple Bonferroni and HB Hochberg Bonferroni

CASE I: Forecast comparison between M-M and VAR

(1) In-sample (1970:01~1989:12, 240 observations),
   Out-of-sample (1990:01~1997:10, 94 observations).
(2) x_t = 6-months Commercial Paper; y_t = 3-months Treasury Bill.

I. M-M Model I: Unrestricted form

$$x_t = \max(x_{t-1} - 0.0166, \ y_{t-1} + 0.3432) + \hat{u}_{xt}$$
$$\qquad\quad (0.0479) \qquad\quad (0.1061)$$

$$R^2 = 0.940, \ D/W = 1.99, \ \text{Sum of Squared residuals} = 116.10$$

$$y_t = \min(x_{t-1} + 5.0000, \ y_{t-1} + 0.1485) + \hat{u}_{yt}$$
$$\qquad\quad (—) \qquad\qquad (0.1388)$$

$$R^2 = 0.947, \ D/W = 2.00, \ \text{Sum of Squared residuals} = 92.73$$

$s = 40$
of (the max operation selects y_t+b) = 40
of (the min operation selects x_t+c) = 0

II. Test statistics and *p*-values of ARCH(4) and Jarque-Bera normality tests

	ARCH(4) Test	Jarque-Bera Test	Skewness	Excess kurtosis	s^2/s_L^2
eq. 1	50.28 (0.000)	507.27 (0.000)	-1.0415	9.9545	1.0628
eq. 2	66.57 (0.000)	442.28 (0.000)	-1.1104	9.4047	1.0236

Note: The values in parenthesis are *p*-values

III. Mean squared forecast error

	CP 6-months	T-bills 3-months	Spread
M-M	0.0408	0.0305	0.0098
VAR	0.0761	0.0493	0.0111
SD statist	-4.7973	-4.0953	-1.0526

# of lags of VAR model by BIC : p = 3
The *SD statistics* is the test in Granger and Newbold (1986, pp. 278-280) which is based on the correlation coefficient, *r*, of the sum and differences of the forecast errors. The null hypothesis is *r=0*, and the null distribution is based on the well-known approximation that $\dfrac{\sqrt{n-1}}{2} \log\left(\dfrac{1+r}{1-r}\right) \sim N(0,1)$, where *n* is ex-post sample size.

CASE II ⮕Forecast comparison between M-M and VAR

(1) In-sample (1970:01~1989:12, 240 observations),
 Out-of-sample (1990:01~1997:10, 94 observations).
(2) x_t = 6-months Commercial Paper; y_t = 3-months Treasury Bill.

I. M-M method: Restricted form

$$x_t = \max(x_{t-1} - 0.0038, \ y_{t-1} + 0.3984) + \hat{u}_{xt}$$
$$\qquad\quad (0.0689) \qquad\quad (0.0716)$$

$$R^2 = 0.879, \text{ D/W} = 1.86, \text{ Sum of Squared residuals} = 234.00$$

$$y_t = y_{t-1} + 0.0411 + \hat{u}_{yt}$$
$$\qquad\quad (0.0302)$$

$$R^2 = 0.866, \text{ D/W} = 2.03, \text{ Sum of Squared residuals} = 234.28$$

$$\begin{bmatrix} \hat{u}_{xt} \\ \hat{u}_{yt} \end{bmatrix} = \begin{bmatrix} 0.31(0.13) & 0.26(013) \\ 0.06(0.10) & 0.41(0.11) \end{bmatrix} \begin{bmatrix} \hat{u}_{xt-1} \\ \hat{u}_{yt-1} \end{bmatrix} + \begin{bmatrix} 0.06(0.11) & -0.17(0.12) \\ 0.16(0.09) & -0.29(0.11) \end{bmatrix} \begin{bmatrix} \hat{u}_{xt-2} \\ \hat{u}_{yt-2} \end{bmatrix} + \begin{bmatrix} \hat{\varepsilon}_{xt} \\ \hat{\varepsilon}_{yt} \end{bmatrix}$$

$$h_{xt} = 0.0173 + 0.5462 e_{xt}^2 + 0.4532 h_{xt-1}$$
$$\quad (0.0078) \ (0.0857) \ \ (0.1187)$$

$$h_{yt} = 0.0185 + 0.5774 e_{yt}^2 + 0.3686 h_{yt-1}$$
$$\quad (0.0103) \ (0.1379) \ \ (0.1286)$$

$s = 40$
of (the max operation selects $Y_t + b$) = 56

CASE III乙Forecast comparison between M-M and VAR

(1) In-sample (1970:01~1989:12, 240 observations),
    Out-of-sample (1990:01~1997:10, 94 observations).
(2) x_t = 6-months Commercial Paper; y_t = 3-months Treasury Bill.

I. <u>M-M method:</u> Restricted form

$$x_t = (1-\hat\alpha)x_{t-1} + \max(\hat\alpha x_{t-1} - 0.5282,\ \hat\alpha y_{t-1} + 0.1951) + \hat u_{xt},\ \hat\alpha = 0.2737$$
$$\qquad\qquad\qquad (0.4419)\qquad\quad (0.0926)\qquad\qquad (0.1303)$$

$$R^2 = 0.879,\ \text{D/W} = 1.87,\ \text{Sum of Squared residuals} = 233.94$$

$$y_t = y_{t-1} + 0.0368 + \hat u_{yt}$$
$$\qquad\quad (0.0290)$$

$$R^2 = 0.866,\ \text{D/W} = 2.02,\ \text{Sum of Squared residuals} = 234.49$$

$$\begin{bmatrix}\hat u_{xt}\\ \hat u_{yt}\end{bmatrix} = \begin{bmatrix}0.39\,(0.15) & 0.19\,(0.15)\\ 0.05\,(0.10) & 0.41\,(0.11)\end{bmatrix}\begin{bmatrix}\hat u_{xt-1}\\ \hat u_{yt-1}\end{bmatrix} + \begin{bmatrix}0.13\,(0.10) & -0.24\,(0.11)\\ 0.15\,(0.10) & -0.28\,(0.11)\end{bmatrix}\begin{bmatrix}\hat u_{xt-2}\\ \hat u_{yt-2}\end{bmatrix} + \begin{bmatrix}\hat\varepsilon_{xt}\\ \hat\varepsilon_{yt}\end{bmatrix}$$

$$h_{xt} = 0.0170 + 0.5818\varepsilon_{xt}^2 + 0.4008 h_{xt-1}$$
$$\quad (0.0079)\ (0.0870)\qquad (0.1076)$$

$$h_{yt} = 0.0174 + 0.5786\varepsilon_{yt}^2 + 0.3751 h_{yt-1}$$
$$\quad (0.0099)\ (0.1375)\quad (0.1317)$$

$s = 50$
of (the max operation selects $\alpha y_t + b$) = 235

Figure 1. The Integrated System 1

(a) Min−Max Series with (−0.5,0.3,0.3,0.5): Integrated system

(b) Cointegrated Series : $x_t - y_t$

(c) Nonparametric kernel estimation of dynamic functional form between $x_t - x_{t-1}$ and $x_{t-1} - x_{t-2}$

(d) Nonparametric kernel estimation of dynamic functional form between $x_t - y_t$ and $x_{t-1} - y_{t-1}$

(e) Nonparametric kernel estimation of dynamic functional form between $y_t - y_{t-1}$ and $y_{t-1} - y_{t-2}$

(f) Autocorrelations of x, y and x−y

series x_t
series y_t

series x_t
series y_t
series $x_t - y_t$

Figure 2. The Integrated System 2

(a) Min-Max Series with (−0.01,−0.3,0.3,−0.1): Integrated system

(b) Cointegrated Series : $x_t - y_t$

(c) Nonparametric kernel estimation of dynamic functional form between $x_t - x_{t-1}$ and $x_{t-1} - x_{t-2}$

(d) Nonparametric kernel estimation of dynamic functional form between $x_t - y_t$ and $x_{t-1} - y_{t-1}$

(e) Nonparametric kernel estimation of dynamic functional form between $y_t - y_{t-1}$ and $y_{t-1} - y_{t-2}$

(f) Autocorrelations of x, y and x−y

Figure 3. The Stationary System

(a) Min-Max Series with (0.7,1.3,1.3,0.7)
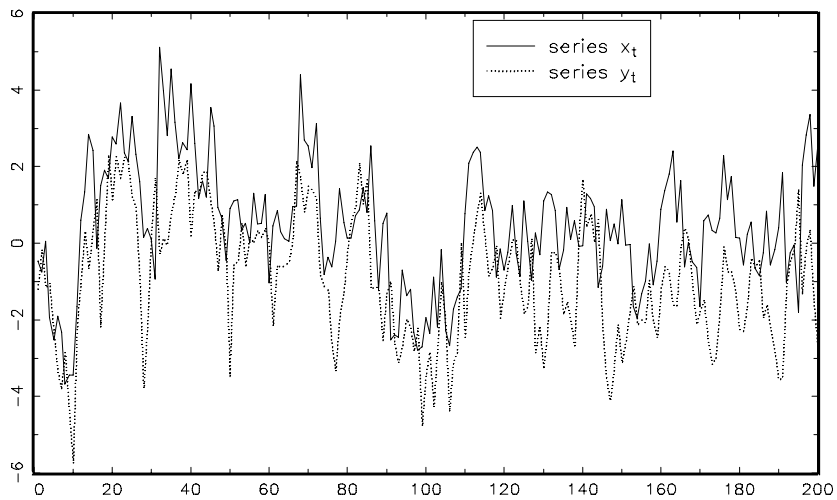
(b) Autocorrelations of x, y
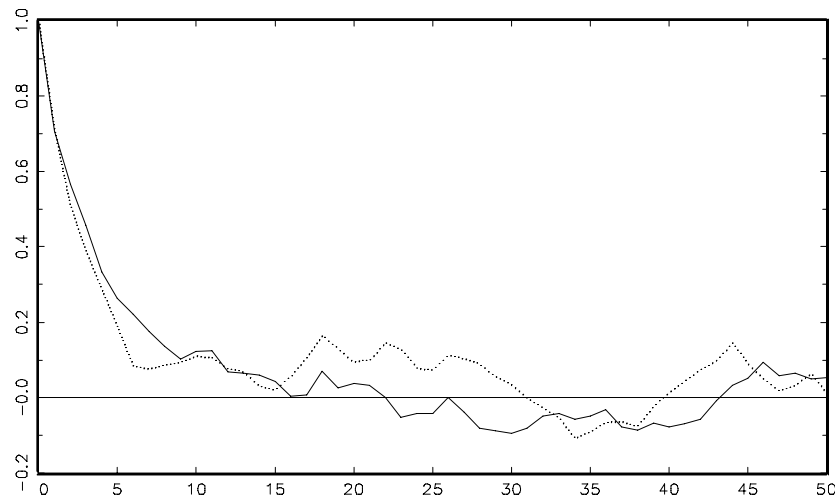
(c) Min-Max Series with (1.01,0.3,1.01,0.3,1.01)
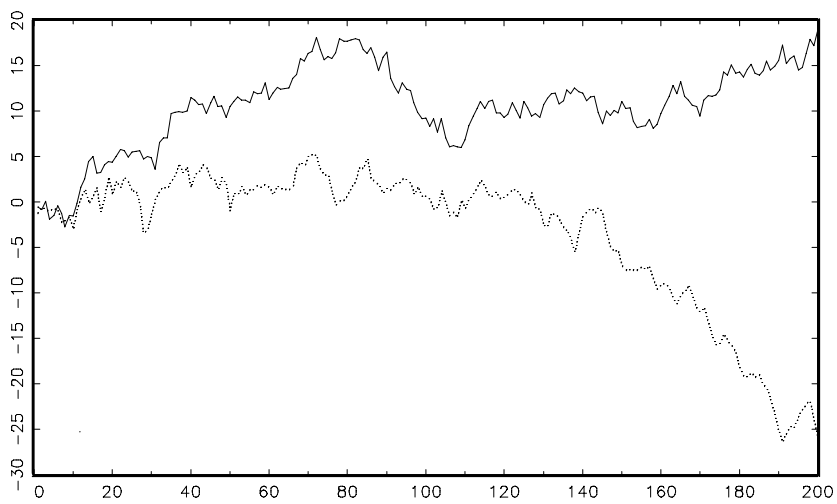
(d) Autocorrelations of x, y
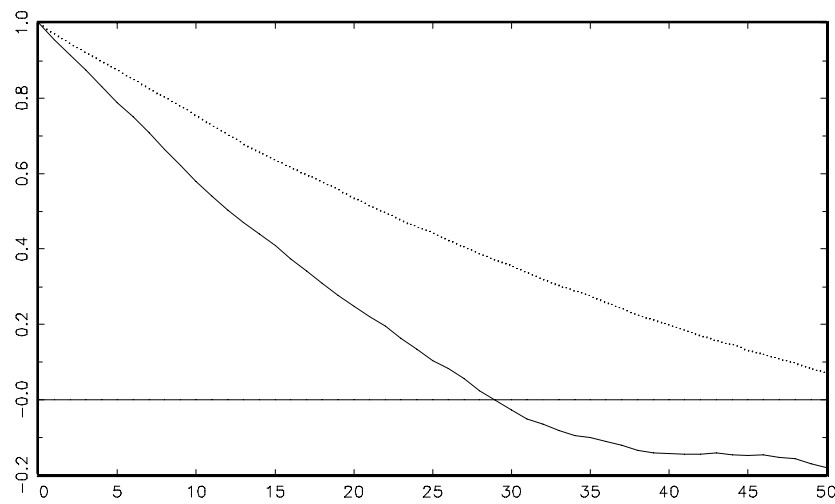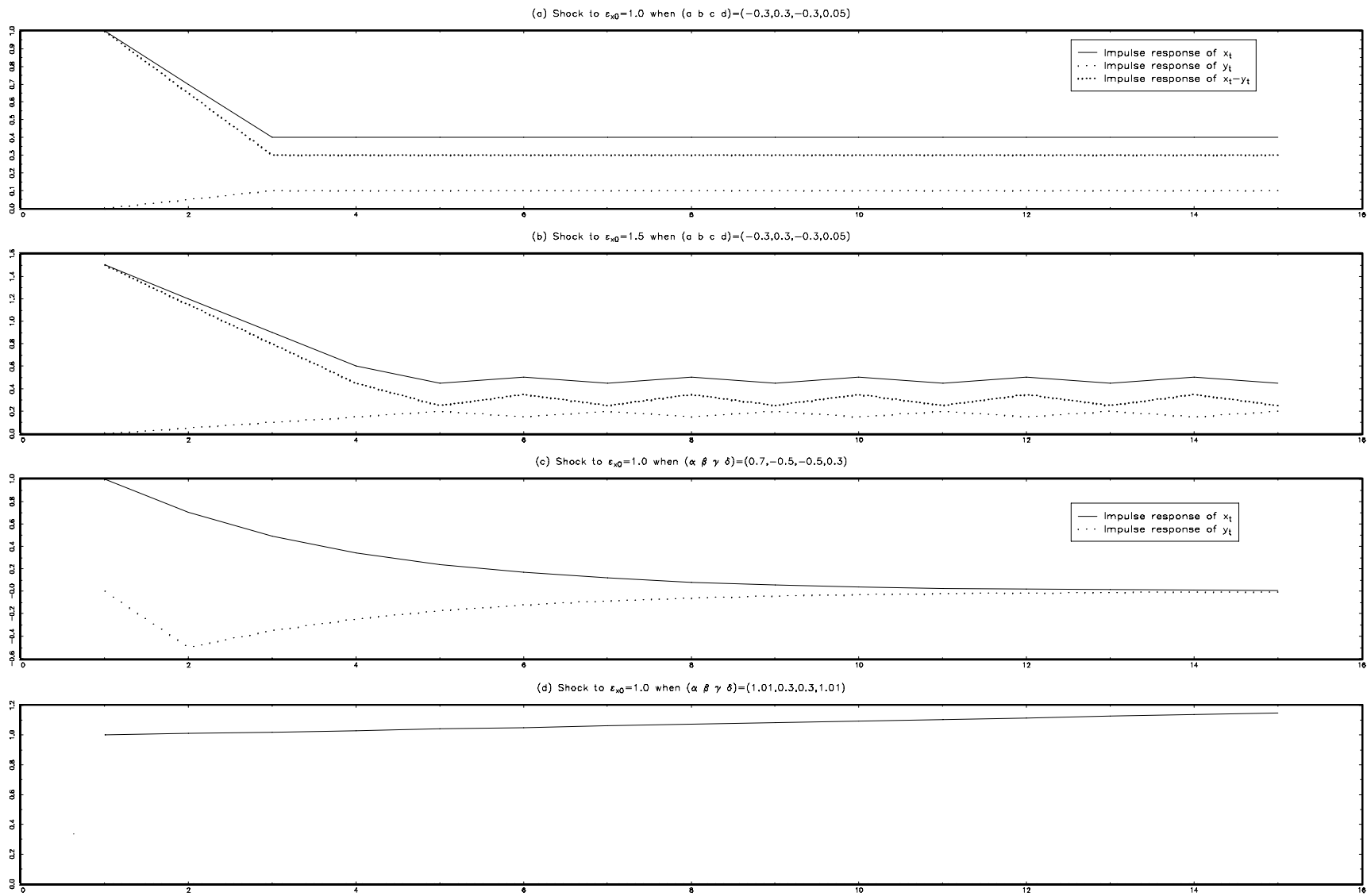
Figure 4. The Impulse Responses of M−M Processes

(a) Shock to $\varepsilon_{x0}=1.0$ when (a b c d)=(−0.3,0.3,−0.3,0.05)

Impulse response of x_t
Impulse response of y_t
Impulse response of x_t-y_t

(b) Shock to $\varepsilon_{x0}=1.5$ when (a b c d)=(−0.3,0.3,−0.3,0.05)

(c) Shock to $\varepsilon_{x0}=1.0$ when (α β γ δ)=(0.7,−0.5,−0.5,0.3)

Impulse response of x_t
Impulse response of y_t

(d) Shock to $\varepsilon_{x0}=1.0$ when (α β γ δ)=(1.01,0.3,0.3,1.01)

Figure 5. Commercial Paper and T−bills rates and Spread

(a) US CP and T−bills interest rates: 1970/01−1997/10

Commerical Paper, 6−month
Treasury Bills, 3−month
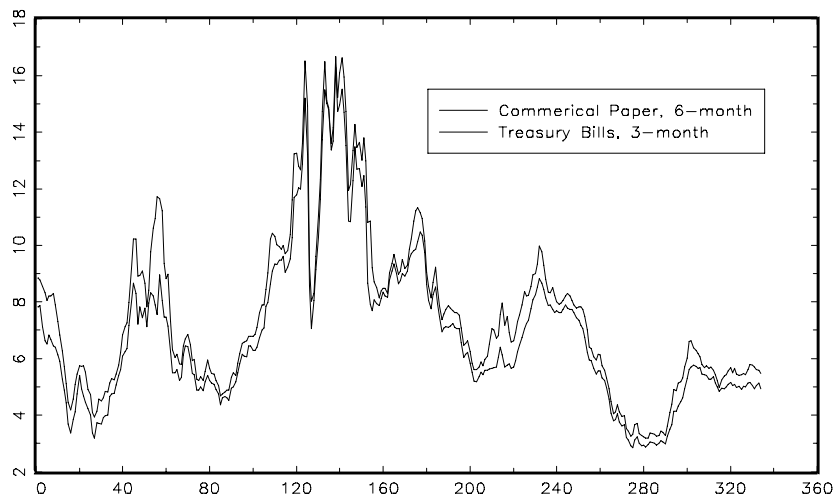
(b) Spread between CP rate and T−Bills rate
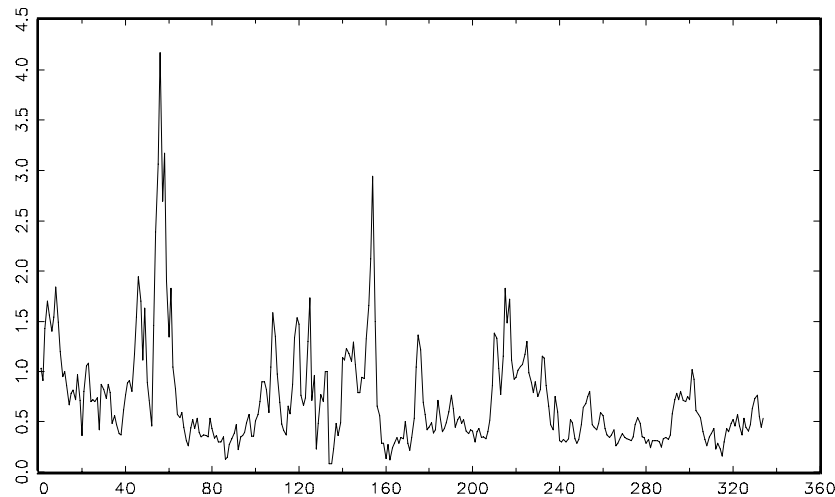
(c) Nonparametric kernel estimation of dynamic functional
form between spread and its lagged value
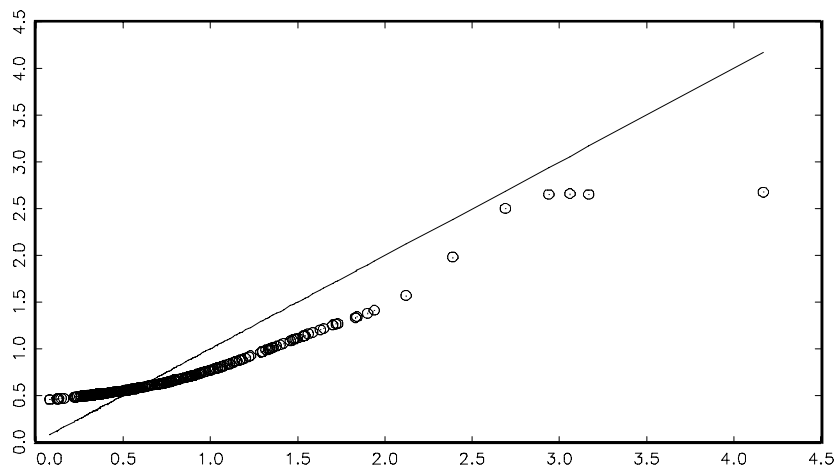
(d) Nonparametric kernel estimation of dynamic functional
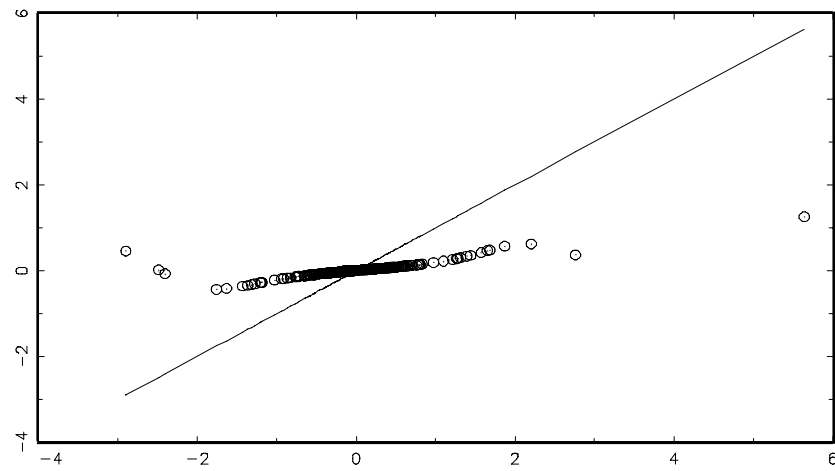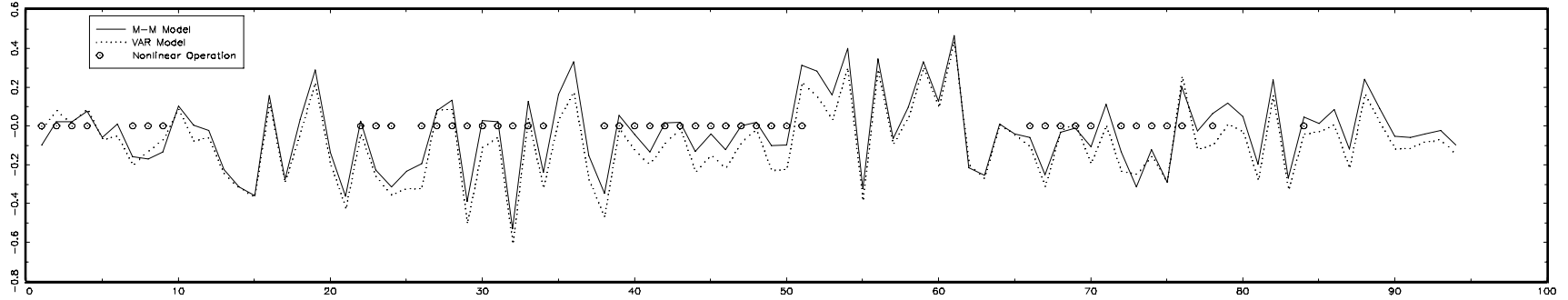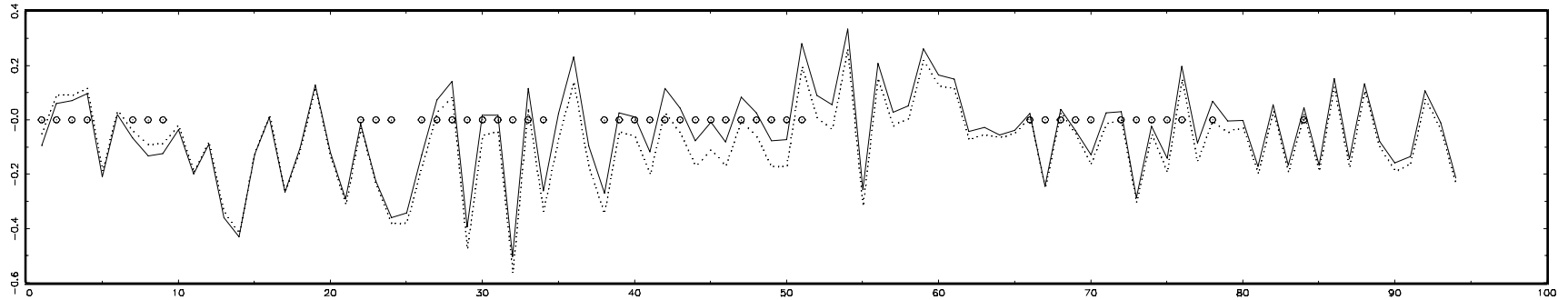form between difference of CP and its lagged value

Figure 6. Forecast Error Comparison of the M–M and VAR Models

(a) 1–Step Forecast errors of CP rate

(b) 1–Step Forecast errors of T–bills rate

(c) 1–Step Forecast errors of Spread