

Nonlinear Time Series and Financial Applications

Gerald P. Dwyer, Jr.
Clemson University

April 2014

Abstract

This is a preliminary, very brief summary of nonlinear time series useful for finance. The purpose of these notes is to provide an overview of nonlinear time series and their financial applications. The notes cover the basics of linear and nonlinear difference equations, chaos, and linear and nonlinear time series, all in a bit over 20 pages! This is very brief. I make no pretense that the notes are complete, although I do think that they are an informative introduction to nonlinear time series if you have some familiarity with linear time series analysis. Comments or suggestions are welcome.

Acknowledgement 1 *This summary is based on lectures given at the University of Rome at Tor Vergata in November 2000. Linda Mundy transcribed the notes, no small task. This draft corrects some typographical errors in an earlier version. I expect to update the draft shortly. This will not change anything fundamental.*

1 Introduction

Nonlinear time series analyses are appearing more and more often in finance. Nonlinear time series analysis, however, seems very complicated and you may think that it is comprehensible only to sophisticated econometricians. In these

notes, I provide some suggestion that nonlinear time series analysis can be informative about financial markets. I also hope to dispel the notion that the subject is that all that complex. Like a lot of things, once you understand it, it's easy. A key in nonlinear time series is to learn how different things are related and remember that. In particular, it is important to see how linear time series analysis is related to nonlinear time series analysis, which suggests when nonlinear time series analysis is likely to be informative.

2 Chaos

Nonlinear time series often is confused with chaos, which is unfortunate. I think that chaos theory has a bad reputation in economics and finance. I am inclined to think that this bad reputation is due to the fact chaos plays a particular role in the physical sciences which is relatively unimportant in economics and finance. That role is to show that equations summarizing a small number of factors can appear "random" even though the process is strictly deterministic. It is useful to discuss chaos, partly to know what it is, but also partly because some of the tools used in chaos theory are useful for helping to understand nonlinear time series. Furthermore, you will see that nonlinear difference equations can have very different implications than the limited possibilities with linear difference equations.

Chaos theory is the analysis of certain nonlinear deterministic equations. *Chaos* is defined by two characteristics of some nonlinear difference equations and their solutions:

1. Sensitive dependence on initial conditions.
2. Seemingly “random” behavior in the sense of having a continuous distribution of the values produced.

As these characteristics suggest, nonlinear difference equations have quite different properties than linear ones.

2.1 Linear Difference Equations

A linear difference equation is rather simple. Consider the simple first-order difference equation

$$x_t = \alpha x_{t-1}. \tag{1}$$

The properties of this equation are well known. The behavior is defined by three regions. If $0 < |\alpha| < 1$, then the equation is *stable*: any deviation from the steady-state value is followed by return to that steady-state value.¹ If $\alpha = 1$, the equation is *metastable* with the value of x always being whatever it happens to be. For example, if $x_0 = 5$, then $x_1 = 5$, $x_2 = 5, \dots$. This is singularly uninteresting in a deterministic context. While it is dubious whether a univariate difference equation should ever be regarded as an “explanation” of anything, just saying that “something is whatever it is” certainly is uninformative. Stated differently, equation (1) with $\alpha = 1$ says that any deviation will persist forever. This can in fact be interesting behavior in a stochastic (or random) context.² From the standpoint

¹The steady-state value, of course, is zero. If you don’t see this right away, then you may want to review difference equations before proceeding further.

²Such behavior was thought to be the defining characteristic of an “efficient market” at one time.

of economic and financial applications, the form of metastability if $\alpha = -1$ is less interesting: in this case, $x_0 = 5$ is followed by $x_1 = -5, x_2 = 5, x_3 = -5, \dots$. If $|\alpha| > 1$, then the equation is *explosive* and any deviation from the steady-state value is followed by increasing divergence. More formally, if $\alpha > 1$, then $x_0 > 0$ implies that $x_t \rightarrow \infty$ and $x_0 < 0$ implies that $x_t \rightarrow -\infty$. If $\alpha < -1$, then x_t also diverges from the steady-state value but the value of x alternates between increasingly large positive and negative values. This strictly explosive behavior is not very interesting for saying something about financial markets. This uniform divergence suggests focusing on $|\alpha| \leq 1$. This leaves us with either uniform convergence or “something is whatever it is”.

Adding more lags to equation (1) introduces possibilities of cycles but nothing else.

In what follows, I probably would run out of symbols if I tried to use unique symbols for every different function. At the least, I would end up using some unusual letters. Instead, I will use ε, η and ζ to represent innovations and I will use other Greek letters to represent the coefficients. The coefficients are not necessarily related across the functions. Roman-alphabet letters are variables.

2.2 Nonlinear Difference Equations

Nonlinear difference equations are far more complex. Consider the seemingly simple nonlinear difference equation

$$x_t = \alpha x_{t-1} - \alpha x_{t-1}^2, \tag{2}$$

which does not seem all that different from equation (1).³ It is, however, quite different, a fact that may be more slightly more apparent if it is written

$$x_t = \alpha x_{t-1}(1 - x_{t-1}), \quad (3)$$

a version which suggests some possible special significance of the value unity for x_{t-1} . In fact, the range of equation (3) is not the same for values more or less than unity, which is quite different than for the linear difference equation (1). Suppose that $\alpha > 0$ and that we consider only starting values of x between zero and one. As for the linear difference equation (1), the behavior of x depends on the value of α , but the dependence is much more complicated for equation (3) than for equation (1).

Suppose that the value of α is 0.7, which would be a parameter value associated with convergence to zero for the linear difference equation. It is easy to compute that an initial value of x equal to 0.6 generates the sequence

0.6000
0.1680
0.0978
0.0618
0.0406
0.0273
0.0186
0.0127
0.0088
0.0061.

The values are converging to zero uniformly. This general behavior is not dif-

³This equation often is called the logistic difference equation.

ferent than the linear difference equation's behavior.

How about a value of α equal to unity, which would be a metastable linear difference equation. What happens now? With an initial value of $x = 0.6$ and $\alpha = 1$, the sequence is

0.6000
0.2400
0.1824
0.1491
0.1269
0.1108
0.0985
0.0888
0.0809
0.0744

This appears to be converging to zero. How can this be? The steady-state value can be determined by solving the quadratic equation

$$\bar{x} = \alpha\bar{x}(1 - \bar{x}), \quad (4)$$

where \bar{x} is the steady-state value by setting α equal to unity. This equation is

$$\bar{x} = \bar{x}(1 - \bar{x}),$$

which can also be written $\bar{x}^2 = 0$, which evidently can be true only if $\bar{x} = 0$. This is not always a good way to find steady-state values, but it is easy to see that zero is the only one that will work in equation (3) with α equal to unity. Convergence to zero for $\alpha = 1$ is an interesting difference from the linear difference equation,

but it is not exactly exciting.

Suppose, though, the value of α is 3.1 (not arbitrarily chosen). Now what happens? The first ten values of the sequence of x 's is

0.6000
0.7440
0.5904
0.7496
0.5818
0.7543
0.5746
0.7578
0.5691
0.7602

There is no tendency for the values to settle down to zero. In fact, it looks more like the values are converging to a two-period cycle with alternating values of about 0.57 and 0.76. Are they? Actually, iteration suggests that they will settle down to alternating values of about 0.5580 and 0.7646.⁴ What happens for other initial starting values? Iterating with a parameter value of $\alpha = 3.1$ and starting from any value of x between zero and one other than exactly $21/31$ (a set of measure zero), the resulting values of x are given by the phase diagram in Figure 1. The X 's mark the eventual two-period cycle.

Suppose that the value of α is 4. With $\alpha = 4$, there is no tendency for the values to return to the same values **ever**. Figure 2 shows the sets of values, marked by much smaller empty circles than in Figure 1. The sequence of values fills up the space between zero and one in the same way that a continuous distribution

⁴You can check that these values are correct to the precision presented by substituting these values into the quadratic difference equation with $\alpha = 3.1$.

function would. Hence, it can be said that the sequence of values is “random” in the sense of being consistent with a probability distribution function even though the values are determined by the simple difference equation (3). This explains where the random characteristic of chaos comes from.

The sensitivity to initial conditions is illustrated in Figure 3. This figure shows the sequence of values from equation (3) with $\alpha = 4$ starting from initial values of 0.6, 0.600001 and 0.61. It would seem that 0.6 and 0.600001 are pretty close together. Yet, by the time that the iterations have taken about 80 steps, the values are quite different. The values starting from 0.6 sit close to zero for about 10 periods from 78 to period 87 and the values starting from 0.600001 do no such thing. This extreme dependence on the starting value is what is meant by “sensitive dependence on initial conditions.”

Sensitive dependence on initial conditions is not the same as appearing random. It is possible to get either without the other. The sensitive dependence on initial conditions is more interesting in terms of economic and financial applications and is not hard to mention in a little more detail.

The Lyapunov exponent λ_L determines whether or not an equation has such dependence. The Lyapunov exponent can be interpreted as being related to the eigenvalue of the system and depends on the behavior of the system with $\lambda_L > 0$ implying sensitive dependence on initial conditions. In financial data, we take randomness for granted. There are a lot of different factors at work and the ones that we ignore – the *imponderables* – are assumed to be a large number of small influences. Statistics developed in no small part because of economic and financial

data. While error terms may not be unimportant in some experiments with inanimate objects, imponderable influences are very important in financial data. This means that equations which determine the evolution of, say, stock prices deterministically are of little interest in terms of characterizing the price of IBM's stock price or the S&P 500.

What about sensitiveness to initial conditions? This may be important sometimes, although most current theory talks about convergence to a steady state or just assumes that the system's equilibrium is a steady state. Arbitrarily wandering around forever is not an important part of most financial theories and isn't likely to be part of them soon.

3 Nonlinear Time Series

Given the importance of generating randomness and sensitiveness to initial conditions in chaos theory and their possible unimportance in financial data, why bother with this nonlinear stuff at all?

3.1 Example

Let's look at some simple data on the deviations of stock and futures prices from their fair values determined by the cost of carry. Figure 4 shows the behavior of the adjusted basis between the futures price of the S&P 500 futures and cash on the New York Stock Exchange on February 13, 1989. The *basis* is the difference between the futures and the value of the S&P 500 index adjusted for the cost of

carry (i.e. the cost of borrowing the funds to hold a cash position and receipt of dividends). The logarithm of the basis is

$$b_t = {}_t f_T - p_t - a_t \quad (5)$$

where ${}_t f_T$ is the futures price at t for a contract that expires at T , p_t is the cash price at t and a_t is a term representing dividends and the cost of carry.⁵ An estimate of the value of a_t has been subtracted from the deviations of the futures and cash prices in Figure 4. Hence, in Figure 4, the term a_t in equation (5) equals zero when the futures price equals the cash price plus carrying cost. A simple deterministic equilibrium model implies that the equilibrium basis shown in Figure 4 always is zero. Clearly this is not true in Figure 4.

In a stochastic model, the implied equilibrium behavior of the basis might be given by a simple autoregression such as

$$b_t = \beta b_{t-1} + \varepsilon_t \quad (6)$$

$$E \varepsilon_t = 0, \quad E \varepsilon_t^2 = \sigma^2, \quad E \varepsilon_t \varepsilon_s = 0 \quad \forall t \neq s.$$

This linear autoregression is quite restrictive. If $|\beta| < 1$, it implies that the basis always is predicted to converge to the mean of zero at the same rate β . Or if $|\beta| = 1$, it never converges. Or if $|\beta| > 1$, it diverges forever. Uniform convergence may or may not be correct. $|\beta| \geq 1$ is inconsistent with the equilibrium in the determinis-

⁵If you are unfamiliar with this type of model, it would take me too far afield to explain more. Dwyer, Locke and Yu [1996] explain more about this application and how the basis is computed.

tic theory ever holding. Hence, among these choices, if the deterministic theory is a guide to what will happen in a stochastic model, $|\beta| < 1$ and the basis converges uniformly at the same rate for all deviations from the steady-state value of zero.

In fact, though, once the basis is outside bounds determined by transactions costs, it can pay to arbitrage.⁶ For example, if the basis is positive and above the transactions-cost bound, this means that it can pay to buy the cash (the cheaper) and sell the futures contract (the more expensive). When the basis returns to zero, which must be the value of the basis at the expiration of the contract, then the position can be unwound for a profit. Conversely if the basis is negative and below the transactions-cost bound. In short, in the figure, if the basis goes outside the bounds, arbitrage will be profitable. Otherwise not.

This profitability of index arbitrage at some times and not other times suggests that prices may behave differently inside and outside the transactions-cost bounds. When arbitrage is profitable, the basis may converge due to this arbitrage. When arbitrage is not profitable, the basis on a given day such as February 13, 1989 may not tend to converge at all.⁷ Then again, it may converge but not at the same rate as when arbitrage is profitable.

This different behavior depending on transactions costs is likely to be inconsis-

⁶This use of the term arbitrage is similar to its use in financial markets and not the same as in asset-pricing models.

⁷The basis must converge to zero at expiration of the contract. The data, though, are prices every 15 seconds during one particular day.

As it turns out in this particular application, the basis tends to zero inside the bounds but not as fast as when outside the bounds. Arguably, this occurs because some investors will buy and sell the S&P 500 the cheapest way – either with the cash or with the futures. This is sufficient to induce convergence within the bounds. This does not affect the basic point: the basis behaves differently outside and inside the bounds because index arbitrage occurs outside the bounds and not inside the bounds.

tent with a simple autoregression such as equation (6). The time-series behavior of the basis may be different inside and outside the transaction cost bounds. Outside the bounds, the basis will tend to converge *toward* zero, i.e. $|\beta| < 1$. Inside the bounds, the basis could be a random walk, i.e. $\beta = 1$. Alternatively, inside the transaction cost bounds, the basis may converge to zero but at a slower rate than when arbitrage is profitable. A model that can represent such behavior of the basis is

$$\begin{aligned}
 b_t &= \beta^u b_{t-1} + \varepsilon_t & \text{if } c < b_{t-d} \\
 b_t &= \beta^c b_{t-1} + \varepsilon_t & \text{if } -c < b_{t-d} < c \\
 b_t &= \beta^l b_{t-1} + \varepsilon_t & \text{if } b_{t-d} < -c
 \end{aligned} \tag{7}$$

where the parameter β differs depending on the value of the basis (i.e. $\beta^u \neq \beta^c \neq \beta^l$) and therefore on whether arbitrage is occurring. As equation (7) indicates, the return to zero may well differ depending on whether basis is above or below the bounds, i.e. β^u may not be the same as β^l .⁸ As equation (7) also suggests, the basis in the past may trigger arbitrage, and the basis may trigger arbitrage with a delay d that be more than one period. As a result, the determinant of arbitrage may be b_{t-d} and not just the basis last period, i.e. b_{t-1} .

Equation (7) is by no means the only possible model. We have used theoretical considerations to generate the representation (7) but not an explicit theory. This representation may be too simple in a variety of directions. On the other

⁸For that matter, the properties of the innovations ε need not be the same. This doesn't matter for the basic point, that something can be different from a linear equation.

hand, it may be too complicated: one nonlinear equation may capture the behavior embodied in the set of three equations in (7). The problem of picking *the correct* nonlinear model to fit is a tough one and has not been solved.⁹

Before jumping into nonlinear time series though, let's go back and review linear time series analysis briefly. The point of this review of linear time series is to highlight some basis aspects of nonlinear time series.

3.2 Linear Time Series Basics

What do we mean by *linear* functions anyway? We can define a linear function (or map) as a map $f(\cdot)$ by two characteristics:

$$\begin{aligned} f(x + y) &= f(x) + f(y) \\ f(\beta x) &= \beta f(x). \end{aligned} \tag{8}$$

Notice that, on this definition, the map $f(x) = \alpha + \beta x$ does not appear to be a linear map because $f(x + y) = \alpha + \beta(x + y)$ and $f(x) + f(y) = 2\alpha + \beta(x + y)$. This is trivial to fix though by a redefinition of the the map to $g(x) = \beta x$ and adding the constant α to the variable produced by the map.¹⁰ This definition works for deterministic functions and stochastic functions. We can say more about linear stochastic difference equations.

⁹Nor is it likely to be in general. The number of candidate functions is arbitrarily large and the the number of data points is anything but arbitrarily large. No matter how much data you have, there are many, many more functions.

¹⁰The constant can be removed from the function by a redefinition of the variable produced. Define $g(x) = f(x) - \alpha = \beta x$. Define the variable produced by the function $z = f(x)$ as $z^* = z - \alpha = g(x)$. In short, this is not a big deal.

The Wold representation provides a solid linear representation of a time series. Wold's Theorem [Anderson 1971, pp. 420-24] says that, if the series x is stationary, then there exists a representation of x such that

$$x_t = \delta_t + \sum_{i=0}^{\infty} w_i \varepsilon_{t-i} \quad (9)$$

It is standard in much of the time-series literature to use w_i to represent the moving-average coefficients and I follow that tradition even though it is inconsistent with the notational convention that coefficients are Greek letters. Lest there be any confusion, the w_i are coefficients. The first term has the property that

$$E[\delta_{t+j} | \delta_t, \delta_{t-1}, \dots] = \delta_{t+j},$$

which means that δ is perfectly predictable, i.e. deterministic, such as a constant term, a trend term or a set of seasonal factors. The second term has the properties that

$$\begin{aligned} \sum_{i=0}^{\infty} w_i^2 &< \infty \\ E[\varepsilon_t] &= 0 \\ E[\varepsilon_t \varepsilon_s] &= \begin{cases} \sigma^2 & \text{if } t = s \\ 0 & \text{if } t \neq s \end{cases} \end{aligned} \quad (10)$$

For simplicity, we will set the deterministic part δ_t to zero. It is useful to normalize w_0 to unity, i.e. $w_0 = 1$.¹¹ The representation using polynomials in the lag operator

¹¹This normalization of w_0 imposes no restriction on x or ε . It defines the relationship between

L such that $L^i \varepsilon_t = \varepsilon_{t-i}$ is

$$x_t = w(L)\varepsilon_t. \quad (11)$$

If the polynomial in the lag operator $w(L)$ is invertible, this moving-average representation can be written ¹²

$$w^{-1}(L)x_t = \varepsilon_t$$

or more familiarly as the linear autoregression

$$x_t = \pi(L)x_{t-1} + \varepsilon_t. \quad (12)$$

Linear autoregressions seem pretty general and are likely to capture a lot of behavior. And in fact, linear autoregressions are capable of characterizing much of the variation of time series that we observe.

It is important to realize that the Wold representation is not completely general. Wold's Theorem holds quite generally and the moving-average representation (9) with the restrictions (10) exists under these general conditions. In short, the existence of a linear representation with a constant variance and serially uncorrelated innovations is quite general. This does not mean that the representation is complete, or even adequate for many possible uses of the series x . The ε 's that generate x are guaranteed to be serially uncorrelated, but the higher moments of ε and x are not characterized at all. This representation is a partial characterization of the

the two as $\text{Var}[x_t - E[x_t | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots]] = \text{Var}[\varepsilon_t] = \sigma^2$.

¹²The assumption of invertibility is a trivial restriction. The function is invertible unless there is a unit root in the moving-average polynomial, in which case the unit root can be removed by summing (or integrating) the series.

series. The Wold representation has serially correlated innovations, ε 's, but they need not be independent.

A definition of a linear time series that is complete in terms of characterizing a times series includes independence of the innovations:

$$\begin{aligned}
 x_t &= \delta_t + \sum_{i=0}^{\infty} w_i \eta_{t-i} \\
 \sum_{i=0}^{\infty} w_i^2 &< \infty \\
 \eta_t &\sim \text{IID}.
 \end{aligned} \tag{13}$$

Requiring that the innovations be independent means that higher moments are characterized. The requirement that the innovations be independent is much more restrictive than that they be serially uncorrelated with a constant variance. Independence implies that third and higher-order non-contemporaneous moments are zero.¹³ In summary, we define a *linear time series* as one that has a linear moving-average representation with independent innovations (13).¹⁴

3.3 Nonlinear Time Series Basics

A general representation of a nonlinear time series follows from this discussion of a linear time series. For η independent as in (10), the Volterra expansion provides a general representation. This representation exists under general conditions and

¹³In short, $E \varepsilon_t \varepsilon_{t-i} \varepsilon_{t-j} = 0 \forall i \neq 0$ or $j \neq 0$ or both, and similarly for fourth and higher-order moments.

¹⁴The representation generally also will have a deterministic component.

is

$$\begin{aligned}
x_t = & \delta_t + \sum_{i=0}^{\infty} w_i \eta_{t-i} + \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} w_{ij} \eta_{t-i} \eta_{t-j} + \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} w_{ijk} \eta_{t-i} \eta_{t-j} \eta_{t-k} \\
& + \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} w_{ijkl} \eta_{t-i} \eta_{t-j} \eta_{t-k} \eta_{t-l} + \dots
\end{aligned} \tag{14}$$

$$\eta_t \sim \text{IID}.$$

A nonlinear time series is one that is not linear, and the equation is not linear if it has nonzero coefficients w_{ij} , w_{ijk} , w_{ijkl} , ... on the higher-order terms. The implications of stationarity of x for the sets of coefficients w in equation (14) are not easy to characterize. The general relationship between a linear time series and a nonlinear time series is easy to see: the nonlinear equation has a lot of cross-product terms. The implications of the additional terms are not so obvious. It makes no basic difference to the rest of this discussion, so I suppress the deterministic part of equation (14).

A couple of things follow from this discussion of the Wold Representation and the Volterra series expansion. First, it is not possible to look at means, variances and covariances of x to determine whether a series is linear. If a series x is linear, these first and second moments are the functions of the data that can be used to characterize the series x . It is necessary to look at higher-order moments to determine whether a series is nonlinear. Second, because higher-order moments require more data in order to be estimated adequately, a nonlinear model requires more data than a linear one.

As with the Wold representation, it is useful to write the Volterra series expansion in an autoregressive form. I know of no general characterization of the conditions under which the Volterra expansion (14) can be transformed into an autoregressive representation such as

$$x_t = f(x_{t-1}, x_{t-2}, \dots) + \eta_t \quad (15)$$

where $f(x_{t-1}, x_{t-2}, \dots)$ is some nonlinear function of the past values of x . This representation need not exist in general but can be a useful starting point for empirical analysis. How might this representation follow from the Volterra series expansion? As for the Wold representation, suppress the deterministic term d_t for simplicity. The Volterra expansion can be written in general form as

$$x_t = f^V(\eta_t, \eta_{t-1}, \eta_{t-2}, \dots)\eta_t$$

and it obviously is no restriction at all to write this as

$$x_t = f^x(x_{t-1}, x_{t-2}, \dots) + f^V(\eta_t, \eta_{t-1}, \eta_{t-2}, \dots)\eta_t.$$

The restrictiveness of the nonlinear autoregressive representation with an additive innovation (15) is due to the implicit assumption that, loosely speaking, the nonlinear function of lagged values of x $f^x(x_{t-1}, x_{t-2}, \dots)$ are able to sufficiently characterize the behavior of x that $f^V(\eta_t, \eta_{t-1}, \eta_{t-2}, \dots)$ is redundant. This need not be true always, or even possibly in general.

An example of a function that is not consistent with the nonlinear autoregressive representation with an additive innovation (15) is an ARCH (autoregressive conditionally heteroskedastic) model. An example of an ARCH model is

$$\begin{aligned}
 x_t &= \sum_{i=1}^{\infty} \beta_i x_{t-i} + \varepsilon_t, & \varepsilon &= h_t \zeta_t \\
 h_t^2 &= \gamma + \sum_{j=1}^{\infty} \gamma_j \varepsilon_{t-j}^2 & & (16) \\
 E \zeta_t &= 0, & E \zeta_t^2 &= 1.
 \end{aligned}$$

Such a model can be called *linear in mean* [Brock, Hsieh and LeBaron 1989] because the conditional expected value is linear in the observations, i.e.,

$$E[x_t | x_{t-1}, x_{t-2}, \dots] = \sum_{i=1}^{\infty} \beta_i x_{t-i}. \quad (17)$$

It is easy to see that the conditional expected value (17) is a linear function of the data in the sense of the definition of linear functions (8).¹⁵ ARCH models and their numerous generalization are very useful, as the huge literature on them testifies.¹⁶ ARCH models and their elaborations help to characterize the time-series behavior of the volatility of a series and are not linear in the sense that they have nonzero terms on higher-order moving average coefficients in the Volterra expansion.

¹⁵An elaborated ARCH model that is not linear in the mean is the MARCH model in which the conditional standard deviation is included in the right-hand-side of the equation determining the conditional expected value.

¹⁶Enders [1995, Ch.3], Tsay [2002, Ch. 3] and Gouriéroux and Jasiak [2002, Ch. 6] summarize the models.

3.4 Nonlinear Functions

The set of nonlinear functions is arbitrarily large and, at this level of generality, there is no obvious reason to limit ourselves to any particular functions. From this point of view, linear models have the definite advantage that, once we know that we are going to estimate a linear autoregression, we know that we are going to estimate the coefficients in equation (12). Once we decide to estimate a nonlinear autoregression, we have the task of deciding which of an arbitrary large number of functions to estimate. In practice, a few nonlinear functions have received most of the attention.

The ARCH model (16) in the last section and elaborations are well known models that are used often in various parts of finance. In addition, various stochastic volatility models are additional often used models of asset prices [Tsay 2002, Chs. 3, 10] These models differ from the ARCH models by including stochastic variation in the volatility (h_t in the set of equations (16)). While not exactly in the immediate spirit of the single-shock Volterra expansion, these equations are no less nonlinear than the ARCH models themselves.

In the rest of this section, I focus on some basic models that are nonlinear in mean. For all of the functions,

$$E \varepsilon_t = 0, E \varepsilon_t^2 = \sigma^2, E \varepsilon_t \varepsilon_s = 0 \quad \forall t \neq s.$$

For the functions to be complete representations, it would have to be the case that

$$\varepsilon_t \sim \text{IID}.$$

Constants are suppressed in all equations; they merely would add notational complexity. I use just one lag where feasible to simplify the notation. The generalizations to multiple lags are relatively obvious.

In the finance literature, the nonlinear autoregressions that have received the most attention are threshold autoregressions, discussed briefly above. These can be written most simply as

$$x_t = \beta^u x_{t-1} + \varepsilon_t^u \text{ if } x_{t-d} \geq c \quad (18)$$

$$x_t = \beta^l x_{t-1} + \varepsilon_t^l \text{ if } x_{t-d} < c \quad (19)$$

This threshold autoregression essentially allows the behavior of x to evolve differently depending on whether the value of x_t at d periods before t is above or below a threshold value c . They can have more than two states, as they did in the arbitrage example above. As the notation here suggests, the innovations in the series can have different variances in the different states. A more general representation would have a data value other than x determining the state. An important question is whether the values of the parameters are consistent with transitions back and forth between the states, so that x does not get stuck in one of the two states.¹⁷ These piece-wise linear functions are far from the only possible nonlinear

¹⁷This is essentially a question of whether the function is ergodic.

functions.

A nonlinear model suggested early in the literature is the bilinear function

$$x_t = \beta x_{t-1} + \gamma x_{t-1} \varepsilon_{t-1} + \varepsilon_t.$$

This model essentially allows for some of the interaction in the Volterra expansion. It has been used relatively little, perhaps because estimation involves the cross-product of x_{t-1} and ε_{t-1} . Estimating a moving-average representation in general is harder than estimating an autoregressive representation, which no doubt has a lot to do with the greater popularity in finance and economics of autoregressive representations than moving-average or autoregressive moving-average representations.

Another nonlinear model is the exponential autoregression

$$x_t = \beta_t x_{t-1} + \varepsilon_t \tag{20}$$

$$\beta_t = \phi_0 + \phi_1 \exp(-\gamma x_{t-1}^2). \tag{21}$$

In this model, the autoregressive coefficient β_t depends on x_{t-1} . The limiting behavior of β_t is

$$\lim_{x_t \rightarrow 0} \beta_t = \phi_0 + \phi_1$$

$$\lim_{x_t \rightarrow \infty} \beta_t = \phi_0.$$

The squared value of x_{t-1} in $\exp(-\gamma x_{t-1}^2)$ imposes symmetry on deviations from

zero (or the mean of the series if the series is in terms of deviations from its mean).

The smooth transition autoregression function

$$x_t = \beta x_{t-1} + \gamma F(x_{t-1}) x_{t-1} + \varepsilon_t$$

where $F(x_{t-1})$ is a continuous function of x_{t-1} . There are many different possible functions $F(\cdot)$ that can be used. One is the logistic function

$$F(x_{t-1}) = \frac{1}{1 + \exp(-x_{t-1}^2)}.$$

Others are the exponential

$$F(x_{t-1}) = \exp(-\delta x_{t-1}^2)$$

and functions based on probability distribution functions such as the normal cumulative distribution function or the normal probability distribution function.

There are many other nonlinear models. Models of autoregressive conditional duration have recently been used to address the timing of trades in financial markets [Engle and Russell 1998]. A model that is popular in macroeconomics but does not appear much in the finance literature is Hamilton's Markov state-transition model [1989].

The problem of choosing a particular nonlinear model is tough. A solution that seems to work is to ask what behavior is being explained and then limit the choices based on what seem like plausible implications of that behavior. I sometimes think

of this as: “A little theory goes a long way.”¹⁸ This does not mean that one has ruled out all alternatives functions. Ruling out all the alternatives is an impossible task anyway because the set of all alternatives always is an arbitrarily large set. It does mean that one has focused on a function that is likely to characterize the behavior under examination.

While it would be limiting to look at functions using only one time series at a time, there is no necessary reason that the x 's above generally can't be treated as vectors. For practical reasons, the size of nonlinear systems of equations are limited, but nonlinear time series would be pretty limited if it couldn't say anything about how time series are related.

4 Conclusion

I have barely scratched the surface of nonlinear time series, but I hope that I have left you with two impressions.

First, linear time series analysis is a very restrictive way to look at the world. Essentially, linear time series analysis can be a thorough characterization of a time series if the series has constant means, variances and covariances and nonzero higher moments only contemporaneously. It will imply that a series converges uniformly, has no tendency to converge, or possibly has cycles. A non-normal distribution of the innovations in a linear representation can help to characterize

¹⁸Bendat [1990, 1998] presents analyses using implications of various functions for statistics, but I think that he has data with less noise and looks at a more restricted range of functions. Nonetheless, the basic idea that I am suggesting is similar even if it does include as many diagnostics of whether the function is consistent with a statistical summary of the data.

the series better than a normal distribution, but this fix will not work if the non-normality of the series varies over time. The problem of selecting a particular non-normal distribution to estimate, which is similar to the problem of selecting a particular nonlinear model to estimate, is itself a very tough problem.¹⁹

Second, nonlinear time series analysis is likely to be very useful for analyzing some aspects of financial data. While I have discussed things that may be new to you (I hope so, or I've been wasting your time and mine), nonlinear time series is a natural extension of linear time series. It's not trivial to delve into this material, but much of it is based on things that you already know. While there are many different nonlinear models, the set of such models that are useful for a particular problem is smaller. It is not necessary to be a nonlinear-time-series guru to find nonlinear time series useful and informative.

References

- [1] This is a short list of references, not a bibliography. A bibliography would be a very large undertaking. There are some additional references in these references. Ramsey [1988] is an informative discussion aimed at economists. For detailed statistical surveys, I suggest Priestley [1988], Tong [1990] and, with a slant to smooth-transition autoregressive models, Granger and Teräsvirta [1993]. Tsay [2002] includes material on nonlinear models in financial applications throughout his book but especially in chapters 3 and 4.
- [] Anderson, T. W. 1971. *The Statistical Analysis of Time Series*. New York: John Wiley & Sons, Inc.

¹⁹And I think that economics is likely to be less helpful for limiting the range of interesting alternative distributions than for limiting the range of interesting alternative mean functions. After all, error terms represent ignorance almost by definition.

- [] Bendat, Julius S. 1990. *Nonlinear System Analysis and Identification from Random Data*. New York: John Wiley & Sons.
- [] Bendat, Julius S. 1998. *Nonlinear System Techniques and Applications*. New York: John Wiley & Sons, Inc.
- [] Brock William A., David A. Hsieh and Blake LeBaron. 1991. *Nonlinear Dynamics, Chaos, and Instability*. Cambridge, Massachusetts: The MIT Press.
- [] Dwyer, Jr., Gerald P., Peter Locke and Wei Yu. 1996. "Index Arbitrage and Nonlinear Dynamics between the Futures and Cash S&P 500." *Review of Financial Studies*. 9 (Spring), 301-32.
- [] Enders, Walter. 1995. *Applied Economic Time Series*. New York: John Wiley & Sons, Inc.
- [] Engle, Robert F., and Jeffrey R. Russell. 1998. "Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data." *Econometrica* 66 (September), 1127-62.
- [] Gourieroux, Christian, and Joann Jasiak. 2001. *Financial Econometrics*. Princeton: Princeton University Press.
- [] Granger, Clive W. J., and Timo Teräsvirta. 1993. *Modelling Non-Linear Economic Relationships*. Cambridge: Cambridge University Press.
- [] Hamilton, James D. 1989. "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle." *Econometrica* 57 (March), 357-84.
- [] Hsieh, David A. 1991. "Chaos and Nonlinear Dynamics: Applications to Financial Markets." *Journal of Finance* 46 (December), 1839-77.
- [] Priestley, M. B. 1988. *Non-linear and Non-stationary Time Series Analysis*. London: Academic Press.
- [] Ramsey, James B. 1990. "Economic and Financial Data as Nonlinear Processes," in *The Stock Market: Bubbles, Volatility, and Chaos*, edited by Gerald P. Dwyer, Jr. and R. W. Hafer, pp. 81-134. Boston: Kluwer Academic Publishers.
- [] Tong, Howell. 1990. *Non-linear Time Series: A Dynamical Systems Approach*. Oxford: Clarendon Press.

- [] Tsay, Ruey. 2002. *Analysis of Financial Time Series*. New York: John Wiley & Sons, Inc.

Figure 1

Phase Diagram for Logistic Equation

Parameter of 3.1 and Starting Value of 0.6

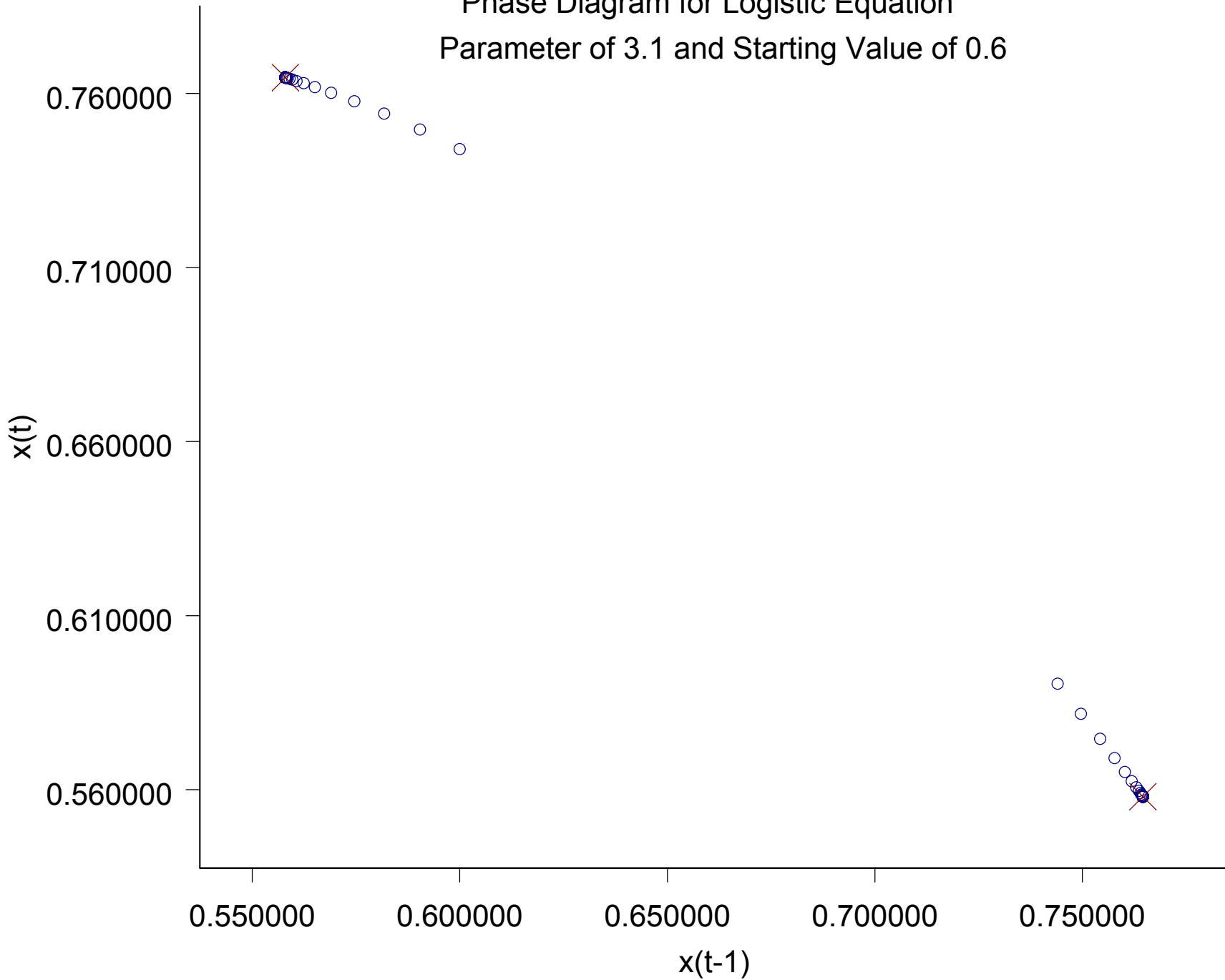


Figure 2

Phase Diagram for Logistic Equation
Parameter of 4.0 and Starting Value of 0.6

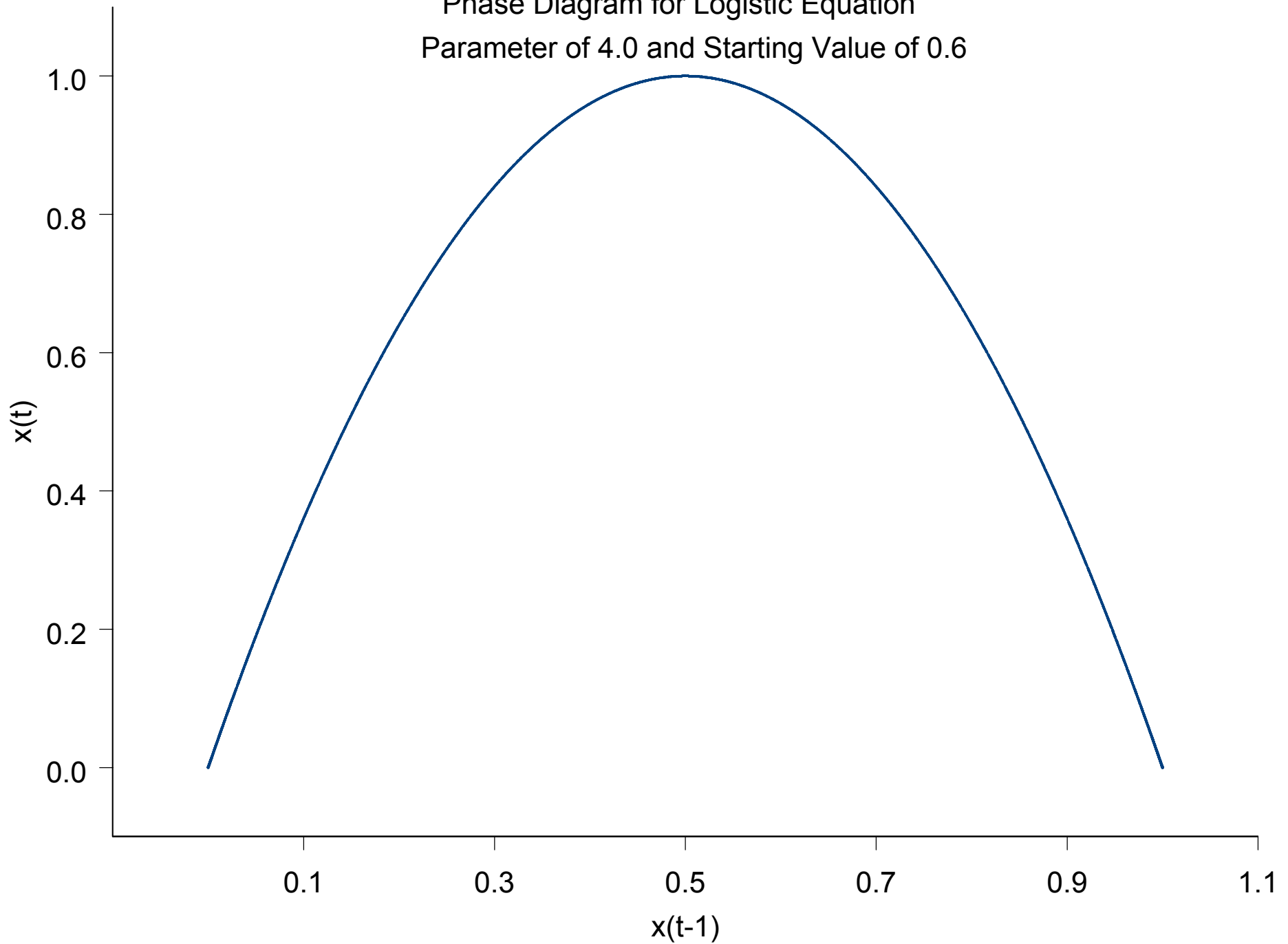


Figure 3

Sequences of Values From Different Starting Points

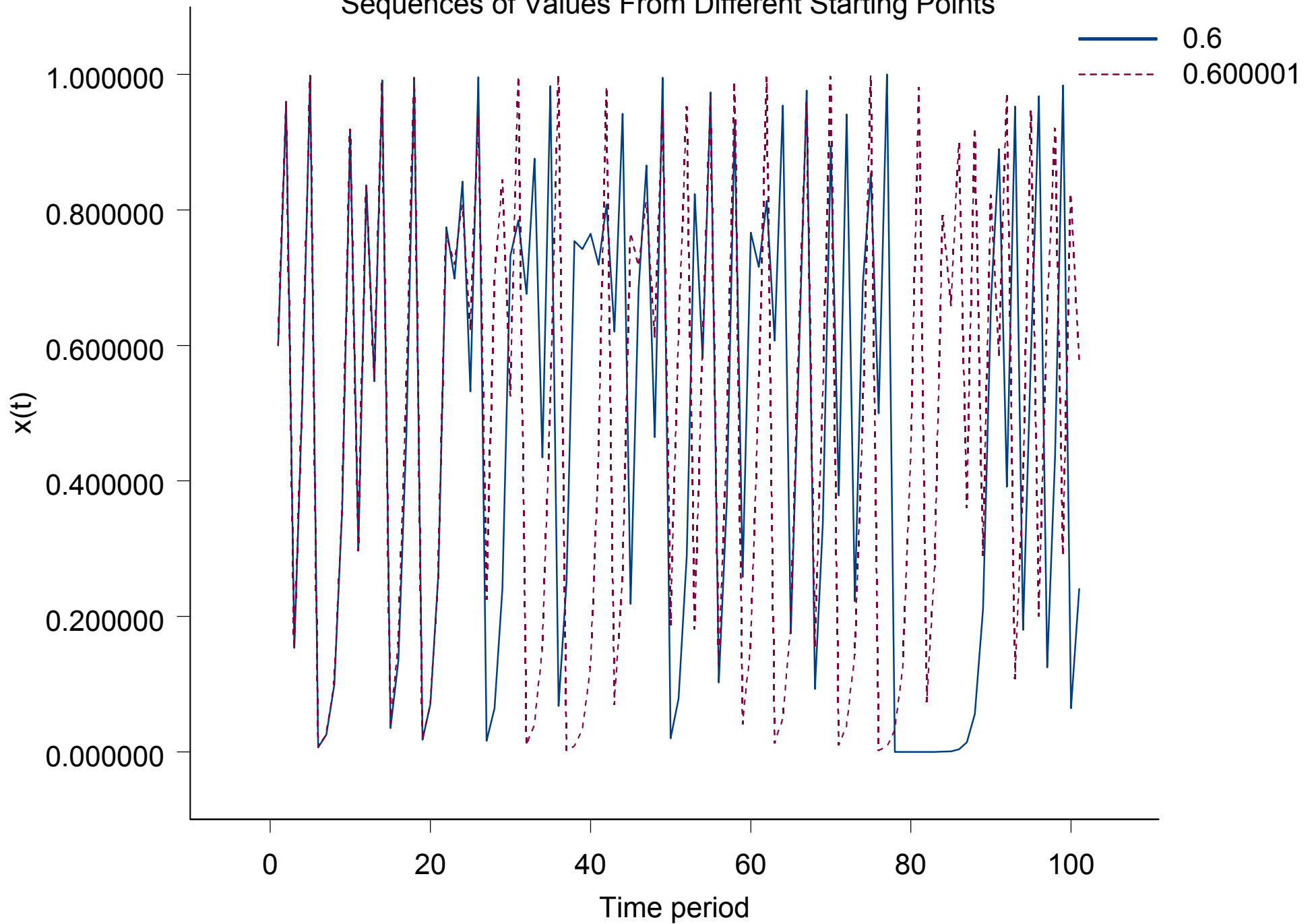


Figure 4

S&P 500 Futures and Cash and Rough Estimate of Transactions Costs
February 13, 1989

