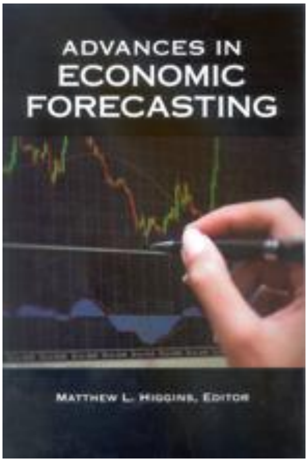
W.E. UPJOHN INSTITUTE
FOR EMPLOYMENT RESEARCH

Upjohn Institute Press

Forecasting Asset Prices Using Nonlinear Models



Michael D. Bradley
George Washington University

Dennis W. Jansen
Texas A&M University

Chapter 5 (pp. 65-104) in:
Advances in Economic Forecasting
Matthew L. Higgins, ed.
Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, 2011
DOI: 10.17848/9780880993937.ch5

Advances in Economic Forecasting

Matthew L. Higgins
Editor

2011

W.E. Upjohn Institute for Employment Research
Kalamazoo, Michigan

Library of Congress Cataloging-in-Publication Data

Advances in economic forecasting / Matthew L. Higgins, editor.
p. cm.
The papers in this volume were presented at the 2009–2010 Werner Sichel Lecture-seminar Series held at Western Michigan University.
Includes bibliographical references and index.
ISBN-13: 978-0-88099-383-8 (pbk. : alk. paper)
ISBN-10: 0-88099-383-9 (pbk. : alk. paper)
ISBN-13: 978-0-88099-384-5 (hardcover : alk. paper)
ISBN-10: 0-88099-384-7 (hardcover : alk. paper)
1. Economic forecasting. I. Higgins, Matthew L.
HB3730.A484 2011
330.01'12—dc23
2011044001

W.E. Upjohn Institute for Employment Research
300 S. Westnedge Avenue
Kalamazoo, Michigan 49007-4686

Cover design by Alcorn Publication Design.
Index prepared by Diane Worden.
Printed in the United States of America.
Printed on recycled paper.

5
Forecasting Asset Prices Using Nonlinear Models

Michael D. Bradley
George Washington University

Dennis W. Jansen
Texas A&M University

Over the past 25 years, a substantial body of research has produced evidence indicating the presence of nonlinearities in the behavior of both financial and real variables. Nonlinearity can arise for a variety of reasons. First, frictions and transaction costs can exceed gains from arbitrage when market deviations are small. Thus, the dynamic reaction to disequilibria may be dependent upon the size of the price change required to restore equilibrium. In other words, transaction costs may be large enough to preclude a complete price response to a small shock but not to large shocks, making the size of the reaction state-dependent.

Another source of nonlinearity is asymmetric dynamics, in which a variable's generating process following declines in its value may differ from the process following increases in its value. For example, the effects of positive shocks may be more persistent than the effects of negative shocks, which may be more rapidly offset. Similarly, herd behavior may cause market participants to overreact during periods of market stress, generating movements in asset prices that exceed normal dynamics. This is another reason a variable's dynamics would be state-dependent.

Still another source of nonlinearity is a variable's volatility. On one level, volatility may be a state variable, with a variable's dynamics changing depending on the state of volatility. Alternatively, volatility itself may be state-dependent, changing because of changes in a state variable such as the state of the economy (expansion, recession) or the state of the financial market (bear market, bull market). These examples show that nonlinearity can arise in the variance or in the mean of the variable. Nonlinearity in the variance typically arises because the variance is time-varying, such as in a generalized autoregressive conditional heteroskedasticity (GARCH) model, whereas nonlinearity in the mean arises because the equation generating the evolution of the mean is nonlinear.

There are a variety of approaches and specific classes of econometric models that have been developed to capture nonlinearities in economic relationships. These include the threshold models (e.g., threshold autoregressive (TAR) models—see Chan and Tong [1986], Tong [1990], and Tsay [1989]); smoothed versions of threshold models (e.g., smooth threshold autoregressive (STAR) models—see Granger and Teräsvirta [1993] and Teräsvirta [1994]); linear models with nonlinear appendages (e.g., the current depth of the recession (CDR) model—see Beaudry and Koop [1993]); the Markov switching model (see Hamilton [1989]); various artificial neural network models (see Cheng and Titterington [1994]); and various nonparametric models in general (see Li and Racine [2007]). These various models have all been employed in estimating nonlinear economic relationships, and most have seen some success as forecasting models.

Many of these models have been applied to business cycles. Neftçi (1984) and Falk (1986) ask whether business cycle dynamics are asymmetric. Teräsvirta and Anderson (1992) and Granger, Teräsvirta, and Anderson (1993) apply smooth transition models to capture business cycle nonlinearities including asymmetries. Van Dijk and Franses (1999) add multiple-regime smooth transition models. Beaudry and Koop (1993) take a different approach, which Potter (1995), Pesaran and Potter (1997), Jansen and Oh (1999), and Bradley and Jansen (1997, 2000) follow up on. However, few of these papers look in-depth at forecasting issues.

Even though nonlinear models have been successfully applied to model a wide variety of financial and macroeconomic variables, the results from using those models to forecast has been mixed. Nonlinear models generally improve upon linear models in terms of in-sample forecasting, but they often show little improvement in terms of out-of-sample forecasting. This somewhat disappointing performance has been ascribed to a number of causes. First, the nonlinearity may not occur in the forecast period. The forecasting advantage of a nonlinear model could arise from its ability to accurately capture the dynamics of a series during periods of time when it exhibits nonlinear behavior. If the out-of-sample forecasting period does not include any such periods, there will be no forecasting advantage for the nonlinear model.

This characteristic of nonlinear models to capture periods of time with "normal" dynamics as well as with exceptional dynamics can lead to other forecasting issues. One such issue is the need to forecast the regime switch or structural break in which the dynamics change. In an out-of-sample forecast, the moment of switch is unknown and may in itself be difficult to forecast, thus reducing the utility of the nonlinear model. Similarly, nonlinear forecasts may be state-dependent, meaning that an accurate out-of-sample forecast will require accurate forecasts of the state of nature over the period of the forecast. For linear models the impact of a shock is the same regardless of the state of the world in which the shock occurs. For nonlinear models this is not true—a disturbance or shock will have different impacts depending on the state of the world in which the shock occurs. Finally, if periods with exceptional dynamics are relatively rare but empirically significant, there may be a tendency for nonlinear models to overfit the sample, reducing their value in an out-of-sample forecast.

A last challenge for out-of-sample forecasting comes from the difficulties in using nonlinear models in multistep forecasts. Linear models can be solved recursively, making the calculation of multistep forecasts relatively straightforward. This is not true for nonlinear models. The nonlinearity makes multiple-step-ahead forecasting intrinsically more difficult. We will outline some of the difficulties with multiple-step-ahead forecasts later in this paper.

These problems notwithstanding, we believe it is important to continue the research into nonlinear forecasts so we can make better use of our ability to model the nonlinear aspects of the economy. We find the problems not to be drawbacks of nonlinear models so much as challenges that must be overcome to improve the accuracy of forecasts. The problems lead to inaccuracy in both linear and nonlinear forecasts, and the task is to better understand nonlinear forecasting in order to overcome these obstacles.

We thus examine an ongoing research question as to whether financial-sector variables help forecast real-sector variables, or real-sector variables help forecast financial sector variables—or both. We use nonlinear models to investigate this question. We do so because this allows us to investigate the out-of-sample forecasting ability of these nonlinear models in a multivariate context.

THEORETICAL PRELIMINARIES

The specific relationship that we investigate is between a key cyclical real-sector variable (industrial production) and two financial-sector variables, the 10-year Treasury bond rate and excess returns for Standard and Poor's 500. We model that relationship using nonlinear approaches and then examine the out-of-sample forecasting properties of those models. Before estimating the models and evaluating the forecasts we present three important definitions and a description of the data we use. What follows here is just the briefest of introductions to ideas of financial economics that provide the backdrop to any financial forecasting exercise.

The first definition is the asset pricing equation. A typical first-order condition from an asset allocation problem (i.e., a typical Euler equation for an asset pricing problem) is found in Equation (5.1):

$$(5.1) \quad u'(c_t) = \beta E_t \left[u'(c_{t+1}) R_{t+1} \right] ,$$

where $u(c_t)$ is a utility function, $u'(c_t)$ indicates the derivative of the utility function, R_{t+1} is real gross return on stocks purchased at time t and held until time $t + 1$, and c_t is real consumption at time t. Typically, optimization requires equating the marginal utility of current consumption (the left-hand side of Equation [5.1]) to the marginal utility of deferring consumption to the next period (the right-hand side of Equation [5.1]). The marginal utility of current consumption is straightforward and is written as $u'(c_t)$. The marginal utility of deferring consumption to the next period is calculated as the product of three terms: 1) the rate of return on a unit of deferred consumption, R_{t+1}; 2) the marginal utility of consumption that is deferred to the next period, $u'(c_{t+1})$; and 3) the discount factor β (used to calculate the present value of this additional future marginal utility).

Equation (5.1) is potentially highly nonlinear, and any variable affecting consumption can potentially affect forecasts of stock returns. This asset pricing condition provides a theoretical basis for nonlinear econometric modeling.

The second definition is of excess returns. Common models of equity returns focus on modeling excess returns, which are returns over and above a risk premium. Often the risk premium is a government bond yield, in which case excess returns are returns on equity over and above returns on the government bond. We thus define excess returns as

$$(5.2)\quad ER_{t+1} = \frac{S_{t+1} + D_{t+1}}{S_t} - I_t = R_{t+1} - I_t \ .$$

Here, ER_{t+1} is the excess return on stocks purchased at time t and held until time $t + 1$, and R_{t+1} is the gross return on stocks purchased at time t and held until time $t + 1$. The gross return on bonds purchased at time t and held until time $t + 1$ is I_t.

The third and last definition is that of the information set used for forecasting. It is important to think carefully about what should be included in the information set when investigating the forecasting performance of various models. The excess return formula indicates that, at time t, an investor knows the nominal return on bonds between t and $t + 1$. If one were to buy a bond at time t, one would know what its interest payments were between t and $t + 1$. But when one buys equity, one won't know the equity's return between t and $t + 1$ until time $t + 1$ occurs and one can observe the price of stock at time $t + 1$. With this definition of the information set, Ω, we can define the forecast for an excess return as

$$(5.3)\quad E(ER_{t+1} \mid \Omega_t) = E\left(\frac{S_{t+1} + D_{t+1}}{S_t} \mid \Omega_t \right) - I_t \ .$$

DATA DESCRIPTION

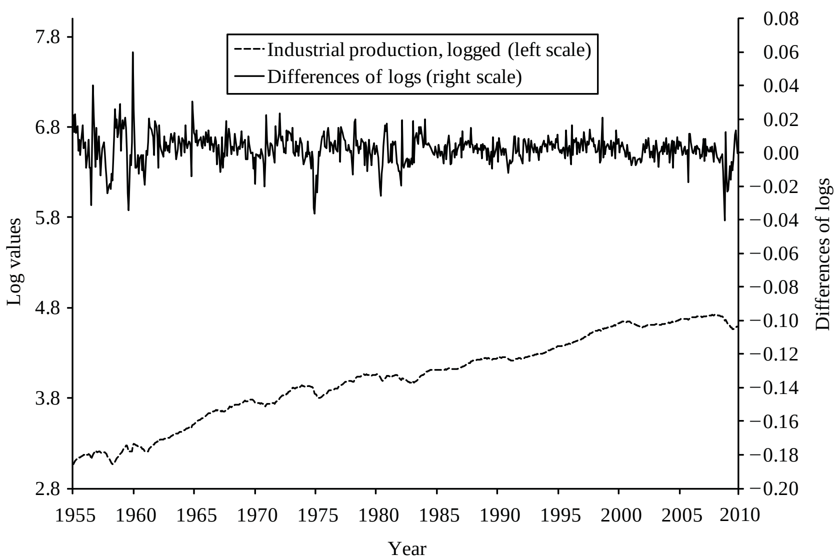
We estimate models of excess returns on equities and bond interest rates using monthly data for the United States. The two sources for our data are 1) Shiller's monthly data set on stocks and associated variables and 2) Federal Reserve System data on industrial production. We also employ a measure of the general price level, the Consumer Price Index calculated by the Bureau of Labor Statistics.

Our data include the value of the Standard and Poor's (S&P) 500 index at the end of each month, calculated as the monthly average of daily closing prices. We represent this variable as the stock index value S_t. Dividends are represented by the symbol D_t, and according to Shiller

(2011) are "computed from the S&P four-quarter tools for the quarter since 1926, with linear interpolation to monthly figures." The price index, P_t, is the CPI-U series. Industrial production is y_t. Finally, we use Shiller's 10-year government security rate (GS10) as our measure of the gross long-term nominal interest rate, I_t.

Below are plots of the key variables. Figure 5.1 graphs the log of industrial production (left scale) and the growth rate, calculated as changes in the log of industrial production (right scale). The general upward trend in industrial production is clearly visible, as are various periods when industrial production was declining. These periods are typically recessions, such as the period around 1975 and the period at the end of our sample, 2009. The plot of the growth rate indicates periods of greater volatility, especially at the beginning of our sample, 1955–1960, and again at the end of our sample. The impact of the recession in 1974–1975 is clear. The long period of relatively low volatility in the growth rate of industrial production from the mid-1980s until 2005 is also apparent.

Figure 5.1 U.S. Industrial Production, January 1955–December 2009

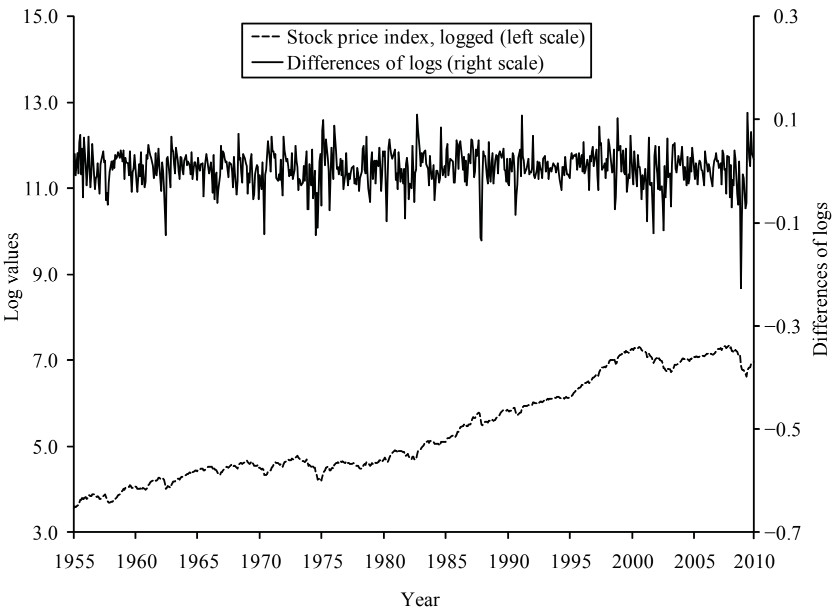


SOURCE: Authors' calculations.

Figure 5.2 graphs the log of the stock price index, both its level (left-hand scale) and its growth rate (right-hand scale). Again the general upward trend in the stock price index is clear, as are periods of declining stock prices after 2000 and again at the end of our sample. The growth rate of stock prices shows considerably more volatility than the growth rate of industrial production. For stock prices, monthly changes of plus or minus 0.1 (10 percent) occur at times, whereas we do not see such large movements in industrial production.

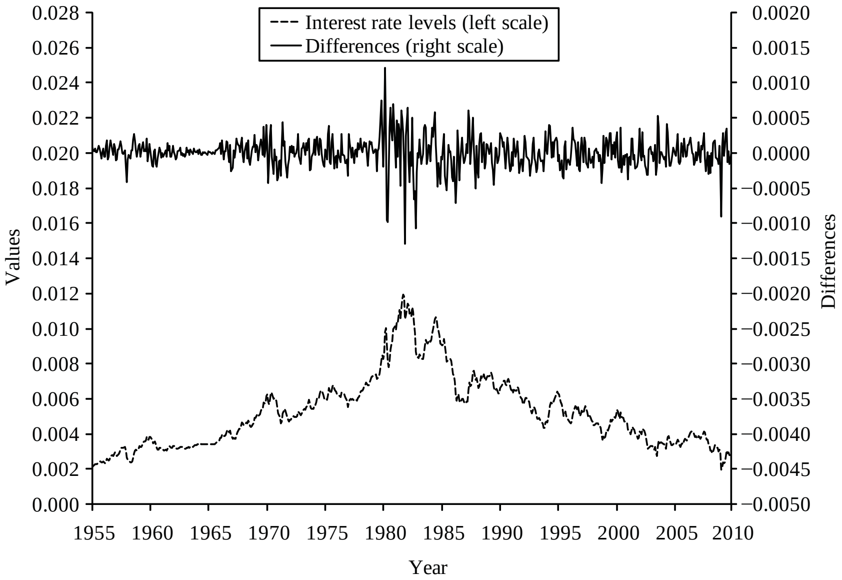
Figure 5.3 graphs our interest rate data, where we have converted this series to monthly net interest rates. Interest rate levels are shown on the left-hand scale, while changes in the interest rate levels, calculated as simple differences, are shown on the right-hand scale. Most apparent is the secular increase in interest rates from the beginning of our sample until the very early 1980s, and the secular decline from the early 1980s until the end of our sample. Interest rates begin our sample at about 0.2 percent per month (roughly 2.4 percent per year), increase to a rate of

Figure 5.2 U.S. Stock Price Index, January 1955–December 2009



SOURCE: Authors’ calculations.

Figure 5.3 U.S. Interest Rate, January 1955–December 2009



SOURCE: Authors' calculations.

almost 1.2 percent per month (roughly 14.4 percent per year), and then return by the end of our sample to a rate near 0.2 percent per month. This run-up and subsequent decline in rates is most often blamed on inflation rates, which increased from the mid-1960s through the 1970s, peaking in the early 1980s before declining gradually throughout the next several decades. Of course, the large secular movements contain many shorter periods of ups and downs in interest rates, as the graph of interest rate changes makes clear. Also apparent in the graph of interest rate changes is the high volatility from the late 1970s through the early 1980s.

ESTIMATING THE NONLINEAR MODELS

In this section we describe how we estimate the nonlinear models that we use for producing our out-of-sample forecasts. Our estimation proceeds in three steps. First we test for linearity in our three variables. The null hypothesis will be that the modeled relationship is linear, and a rejection leads to continued estimation of a nonlinear model. The idea is not to estimate a nonlinear model if that is unnecessary. Second, when nonlinearity is detected, we will estimate a threshold model. Third, we will also estimate a second nonlinear model, which we will call a "current depth of recession" model.

Threshold models capture the possibility that the dynamics of a variable may be state-dependent. They allow the data-generating process to vary across two or more states of nature. For this reason, they are also often called "regime-switching" models. Seasonal models for industrial production or "day of week" models for stock returns are examples of *deterministic* threshold models. For these models the occurrence of a regime switch is known with certainty. However, many interesting cases involve *stochastic* threshold models, in which the regime switch is unknown. An example is given by a model in which stock market returns are driven by a different dynamic process after large declines in stock prices. Bradley and Jansen (2004) provide one attempt at forecasting stock returns in a nonlinear framework.

Threshold models are an example of a model in which the state variable is observable. In a deterministic threshold model it is clear that we observe the day of the week and that we can allow our model to behave differently on different days of the week. In a stochastic threshold model we can observe that there has been a large decline in stock prices, and then we can allow our model to behave differently after such a large decline. The key feature is that the state variable, either the day of the week or the decline in stock prices, is observable. This stands in contrast to models with unobserved state variables, such as the various Markov switching models. In those models the state variable is an unobserved variable, and changes in the underlying hidden state variable lead to changes in the behavior of the variable we are modeling. Thus a key modeling decision is between using a nonlinear model with observable state variables and using a nonlinear model with hidden or unobserved state variables. The choice depends in part on whether there are observed variables that can adequately indicate or represent the state of the world. Here we estimate and forecast with observable state variables.

We investigate two types of threshold models, the threshold autoregressive (TAR) model and the smooth transition autoregressive (STAR) model. A TAR model specifies (at least) two sets of dynamics for the variable of interest, y, with the regime switch dependent upon the value of a "transition" variable, labeled z. The threshold model can be written as follows:

$$y_t = a_0 + \sum_{i=1}^{p} \alpha_i y_{t-i} + \left(\sum_{i=1}^{p} \beta_i y_{t-i} \right) \delta_t + \varepsilon_t ,$$

where

$$(5.4) \qquad \delta_t = \begin{cases} 0 \; if \; z_{t-d} \leq c \\ 1 \; if \; z_{t-d} > c \end{cases} .$$

In this threshold model, the behavior of the variable y_t is a p^{th} order autoregressive model governed by the coefficients α_i when the transition variable z_{t-d} is below the threshold value, c. When the transition variable is greater than the threshold value, the behavior of the variable y_t is an autoregressive model governed by the coefficients $\alpha_i + \beta_i$. Thus the variable y_t changes behavior depending on the relationship between the transition variable and the threshold.

The transition variable is the observable state variable we mentioned above. Changes in this variable lead to changes in the behavior of y_t. The transition variable has a subscript $t - d$ to indicate that it is a lagged value, and d is an integer value of 1 or higher. The parameter d is known as the "delay" and indicates the delay between changes in the transition variable and changes in the behavior of y_t.

The TAR model seems simple, with an indicator variable δ switching from zero to one as the transition variable crosses a threshold value. This is a step function, with δ equaling zero when z_{t-d} is on one side of the threshold and one when z_{t-d} is on the other side of the threshold. Yet, despite this simplicity, the TAR model has proven useful as a model to capture nonlinear behavior.

The STAR model generalizes the threshold approach by allowing a smooth transition between the two regimes. This transition is governed by a function of the threshold variable, z, and the transition function is usually specified as being either logistic (LSTAR) or exponential (ESTAR). The shape of the transition function governs the nature of the movement from one regime to another. The main difference is that logistic specification is one-sided, in the sense that there are alternative dynamics for either large or small values of the transition function, and an intermediate range where the dynamics are a combination of the dynamics at the two extremes. The exponential specification is two-sided, in the sense that there is a set of dynamics for both large and small values, and a different set of dynamics for intermediate values.

The structure of the STAR model is given by

$$y_t = a_0 + \sum_{i=1}^{p} \alpha_i y_{t-i} + \left(\sum_{i=1}^{p} \beta_i y_{t-i} \right) F(z_{t-d}) + \varepsilon_t \ ,$$

where

$$(5.5) \qquad F(z_{t-d}) = \left[1 + e^{-\gamma(z_{t-d} - c)}\right]^{-1} \qquad \text{(LSTAR)}$$

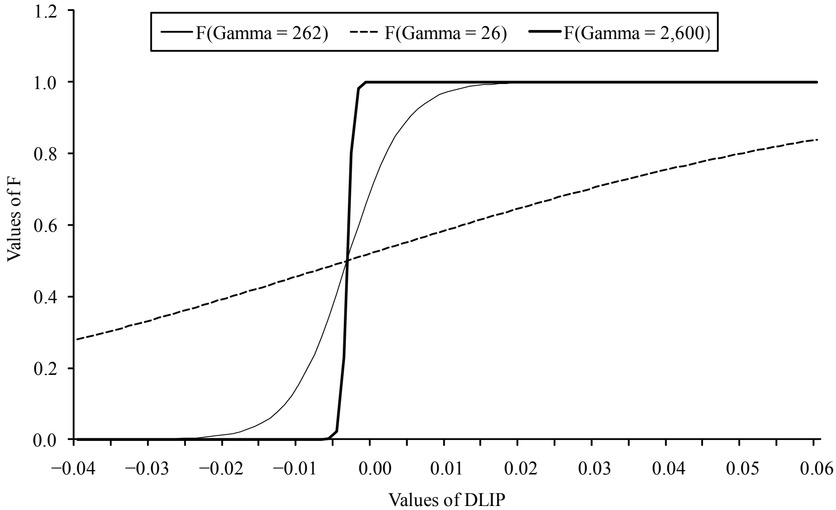
or

$$F(z_{t-d}) = \left[1 - e^{-\gamma(z_{t-d} - c)^{\wedge 2}}\right], \gamma > 0 \qquad \text{(ESTAR).}$$

We illustrate the nature of a STAR model transition function in Figure 5.4. In a TAR model there is a discrete switch between the two regimes, as illustrated by the line that goes almost straight up. The transition function takes a value of zero before the period of the switch and a value of one afterward. In a STAR model the switch between the regimes is more gradual, with the degree of smoothness depending upon the size of the transition parameter, γ. When gamma takes a small value the transition is very gradual, as illustrated by the dotted line. As gamma gets larger, the STAR model begins to approximate the discrete switch of the TAR model, as illustrated by the curved lighter line. Thus, one advantage of the STAR model is that it permits, but does not require, a relatively abrupt switch between regimes.

There are four steps involved in identifying and estimating a STAR model. The first step is the identification and estimation of a linear auto-

Figure 5.4 STAR Transition Functions—Role of Gamma



SOURCE: Authors' calculations.

regressive model. The primary purpose of this step is to determine the lag lengths that will be used for linearity testing and, if nonlinearity is detected, for estimating the STAR model. Step two is to test for linearity, to make sure that a nonlinear model is needed. If linearity is not rejected, there is often no need to continue the process of estimating a nonlinear model. If linearity *is* rejected, the third step is to identify the STAR model specification. Here identification is used in the time series sense and is meant to specify the various features of the model, such as lag lengths. It is in this third step that one determines whether an exponential or a logistical star model is appropriate. The last step is the actual estimation of the specified STAR model. This can be done with various nonlinear optimization procedures, and we use nonlinear least squares.

In this exercise we will identify and estimate three models: one for industrial production, one for a long-term bond rate, and one for excess equity returns. We begin the estimation of the linear models with stationarity testing. We perform both the Augmented Dickey-Fuller (ADF) and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests, and the results are presented in Table 5.1. Note that the ADF test has a null hypothesis of nonstationarity or "integrated of order one," written as I(1). The KPSS test has a null hypothesis of stationarity, or I(0).

Our ADF tests fail to reject the null of I(1) for the levels of the log industrial production [Log(IP)] and the bond interest rate (INT), but reject that null for the level of excess returns (ER). This indicates that log industrial production and the bond interest rate should be differenced, while excess returns are stationary as calculated. We note that the ADF test fails to reject the null of I(1) for the S&P 500, which is consistent with the result that excess returns, calculated in part from differencing the log S&P 500 index [Log(S)], are I(0). The ADF results are corroborated by the KPSS tests, which reject the null of I(0) for the levels of the log of industrial production and the bond rate, but fail to reject that null for excess returns.

Table 5.1 Testing for Stationarity

Variable	ADF test[a] series in levels	ADF test[a] series in first differences	KPSS test[b] series in levels	KPSS test[b] series in first differences
Log(IP)	−2.55	−12.25	0.42	0.14
	(p = 0.31)	(0.00)	(5% CV = 0.15)	(5% CV = 0.46)
	(2 lags, trend)	(1 lag)	(trend)	
	Fail to reject	Reject	Reject	Fail to reject
Log(S)	−1.71	−19.29	0.63	0.10
	(p = 0.75)	(p = 0.00)	(5% CV = 0.15)	(5% CV = 0.46)
	(1 lag, trend)	(0 lags)	(trend)	
	Fail to reject	Reject	Reject	Fail to reject
INT	−1.73	−18.08	0.96	0.27
	(p = 0.42)	(0.00)	(5% CV = 0.15)	(5% CV = 0.46)
	(2 lags)	(1 lag)		
	Fail to reject	Reject	Reject	Fail to reject
ER	−19.20		0.09	
	(p = 0.00)		(5% CV = 0.15)	
	(0 lags)			
	Reject		Fail to reject	

NOTE: Blank = not applicable.
[a] ADF test: Null hypothesis is I(1).
[b] KPSS test: Null hypothesis is I(0).
SOURCE: Authors' calculations.

Based upon these two tests, we estimate linear models in the first differences of the log of industrial production (DLIP) and the bond interest rate (DINT), and in the level of excess returns (ER). We proceed to determine the appropriate lag length for these linear models, and here we select the lag lengths using a standard goodness of fit criterion, the Schwartz Information Criterion (SIC). Using the SIC to pick lag lengths involves searching over a range of possible lag lengths selected a priori and finding the lag length within that set that will minimize the SIC. Here we searched over a range from 1 lag to 12 lags.

Our models will possibly contain multiple right-hand-side variables at various lags, and not just lags of the dependent variable. This raises a few issues with lag selection. One procedure is to search over the entire lag space, with 1 through 12 lags of each variable. If there are three right-hand-side variables, as there are in some of our models, this involves 12 cubed regressions. An alternative is to use an iterative procedure, first picking the lags of the dependent variable and then proceeding with the other explanatory variables. We follow this latter approach. We first selected the best univariate model, then the best lags of the second variable, holding constant the lags specified for the dependent variable in the univariate specification, and then the best lags of the third variable given the lags of the first two variables. The results are provided in Table 5.2.

The linear model for the change in the log of industrial production contains two lags of the change in the log of industrial production, two lags of excess stock returns, and one lag of the change in the interest rate. The linear models for the asset returns are more parsimonious. Neither one contains any lags of change of the log of industrial production, so the real sector variable will not be included in the models for the financial variables.

Table 5.2 Determining the Lag Length for the Linear Models

	DLIP	ER	DINT
Univariate	AR(2); −6.74	AR(1); −3.93	AR(2); −14.14
Bivariate	2 lags ER; −6.75	0 lags DLIP; −3.93	2 lags ER; −14.15
Trivariate	1 lag DINT; −6.75	1 lag DINT; −3.94	0 lags DLIP; −14.15

SOURCE: Authors' calculations.

The estimated linear models are presented in Table 5.3. Estimation was by ordinary least squares. Excess returns have a positive effect on current growth in industrial production. Interestingly, increases in the interest rate also have a positive effect on current growth in industrial production. As for excess returns, increases in the interest rate have a decidedly negative impact on excess returns. Finally, changes in the interest rate are affected positively by excess returns.

The next step is to test for linearity. We use the approach derived by Luukkonen, Saikkonen, and Teräsvirta (1988) and Teräsvirta and

Table 5.3 Linear Model Estimates

	DLIP	ER	DINT
Constant	0.0015	4.43E-04	1.20E-06
	(0.0004)	(1.36E-03)	(8.18E-06)
	p = 0.0000	p = 0.7440	p = 0.8829
DLIP(−1)	0.3098	—	—
	(0.0404)	—	—
	p = 0.0000	—	—
DLIP(−2)	0.1028	—	—
	(0.0399)	—	—
	p = 0.0103	—	—
ER(−1)	0.0231	0.2086	6.27E-04
	(0.0100)	(0.0397)	(2.50E-4)
	p = 0.0211	p = 0.0000	p = 0.0122
ER(−2)	0.0308	—	6.31E-04
	(0.0099)	—	(2.50E-04)
	p = 0.0020	—	p = 0.0119
DINT(−1)	4.6626	−26.5785	0.3863
	(1.5652)	(6.2857)	(0.0404)
	p = 0.0030	p = 0.0000	p = 0.0000
DINT(−2)	—	—	−0.1925
	—	—	(0.0406)
	—	—	p = 0.0122
R^2	0.1951	0.0842	0.1720
Std. error	0.0081	0.0332	0.0002
SIC	−6.7509	−3.9441	−14.1487

NOTE: — = data not available.
SOURCE: Authors' calculations.

Anderson (1992), in which linearity is tested with the "approximating equation." The advantage of this approach is that it simultaneously tests for linearity and provides guidance about the specification of the nonlinear model. The approximating equation is given as follows:

$$y_t = \beta_0 + \sum_{i=1}^{p} \beta_{1,i} y_{t-i} + \sum_{i=1}^{p} \beta_{2,i} y_{t-i} z_{t-d} + \sum_{i=1}^{p} \beta_{3,i} y_{t-i} z_{t-d}^2 + \sum_{i=1}^{p} \beta_{4,i} y_{t-i} z_{t-d}^3 + \varepsilon_t \quad (5.6)$$

where y_t is the variable being modeled, z_{t-d} is the transition variable, and d is the delay between when the transition variable crosses the threshold value and the variable of interest's alternative dynamics become active. The null hypothesis of linearity is a test of $\beta_{2i} = \beta_{3i} = \beta_{4i} = 0$.

We take a general approach to testing for linearity by using five possible transition variables and up to a three-period delay. We start by using the dependent variable as the transition variable to see if the variable's own values indicate the source of nonlinearity. This would mean, for example, that the change in the log of the industrial production would be the transition variable for itself. We then look at the possibility that one of the other two variables being modeled could be the transition variable. For the change in the log of industrial production, this means testing whether the change in the bond rate or excess returns is the transition variable. Finally, we consider two external variables. The first, called current depth of the recession, or CDR, is the distance from the past peak in industrial production and its current value. This variable would capture a situation in which recessions and expansions had alternative dynamics. A variable similarly defined for the stock index, CDB (current depth of stocks), measures the difference between the previous peak in the S&P 500 index and index's current value. This variable would capture a situation in which rising and falling S&P 500 index values generated different dynamics.

The results of estimating the approximating equation and testing for linearity are given in Table 5.4. First, there is no evidence suggesting rejection of linearity for excess returns. No tests for any threshold variable for any delay are close to suggesting a rejection of linearity. We conclude that excess returns are best modeled here as a linear process. There is abundant evidence, however, to support rejecting

Table 5.4 Testing for Linearity

	Dependent variables for linearity test					
Threshold	DLIP		ER		DINT	
variables	Chi-sq.	*P*-value	Chi-sq.	*P*-value	Chi-sq.	*P*-value
CDB(−1)	35.59	0.0020	5.03	0.5402	18.27	0.1076
CDR(−1)	30.63	0.0098	4.34	0.6302	28.47	0.0047
ER(−1)	22.08	0.1056	2.43	0.8764	52.13**	0.0000
DINT(−1)	16.59	0.3442	8.68	0.1925	51.96*	0.0000
DLIP(−1)	52.25**	0.0001	8.77	0.1871	25.80	0.0114
CDB(−2)	41.87*	0.0002	2.73	0.8424	20.88	0.0522
CDR(−2)	28.53	0.0185	7.91	0.2445	30.83	0.0021
ER(−2)	29.70	0.0131	5.36	0.4985	24.30	0.0185
DINT(−2)	17.23	0.3052	6.83	0.3369	43.54	0.0000
DLIP(−2)	25.84	0.0398	5.03	0.5400	19.52	0.0768
CDB(−3)	29.04	0.0159	1.68	0.9467	15.93	0.1944
CDR(−3)	14.45	0.4917	5.75	0.4515	35.18	0.0004
ER(−3)	19.50	0.1921	5.20	0.5182	39.93	0.0001
DINT(−3)	31.66	0.0072	3.48	0.7469	40.05	0.0001
DLIP(−3)	17.54	0.2878	6.50	0.3692	9.66	0.6459

NOTE: * significant at the 0.10 level (two-tailed test); ** significant at the 0.05 level (two-tailed test).

SOURCE: Authors' calculations.

linearity for both the change in the log of industrial production and the change in the bond rate. Of the 15 different tests for linearity, 9 of them support rejection, suggesting that a finding of nonlinearity is not dependent upon a very specific combination of threshold variable and delay. A review of all instances that show rejection indicates that a single lag of the change in the log of industrial production should be chosen as the threshold variable. Similar results are found for the change in the bond rate, in that 11 of the 15 tests produce evidence indicating rejection of linearity. A review of those tests indicates that a single lag of excess returns should be chosen as the threshold variable for the change in the bond rate.

Given that we find evidence rejecting linearity for two of the variables, the next step is to identify which STAR model is appropriate for each, the LSTAR or the ESTAR model. This is done through a series of hypothesis tests on the coefficients in the approximating equation. The null hypothesis of linearity is tested through setting to zero all of the estimated coefficients on the threshold variable. The identification of the model specification looks at similar tests for subsets of the coefficients. Teräsvirta and Anderson (1992) specify a set of three hypotheses:

(5.7) $H_1 : \beta_{4,i} = 0, \forall i$,
$H_2 : \beta_{3,i} = 0 \mid \beta_{4,i} = 0, \forall i$, and
$H_3 : \beta_{2,i} = 0 \mid \beta_{3,i} = \beta_{4,i} = 0, \forall i$.

Rejection of H_1 indicates that an LSTAR model is appropriate. Failure to reject H_1 but rejection of H_2 indicates that an ESTAR model is appropriate. Finally, failure to reject H_1 or H_2 but rejection of H_3 indicates that an LSTAR model is appropriate.

The results of testing theses hypotheses for our models are presented in Table 5.5. The table shows that H_1 is rejected for both variables, indicating an LSTAR specification is appropriate for both the change in the log of industrial production and the change in the bond rate.

At this point we estimated the LSTAR model for changes in the log of industrial production by nonlinear least squares, with results reported in Table 5.6. There are some important issues in estimating TAR and STAR models that have to do with discontinuities in the likelihood function, and these have been documented and discussed in Hansen (1997). Our solution is to conduct a grid search for various values of the threshold value in the TAR model, and the parameter estimates of the TAR model are used as starting values for the STAR estimation.

In Table 5.6 we will first examine the transition variable and its role in our model. The transition variable is one lag of the change in the log of industrial production. Here the estimated value for the threshold is −0.0035, which suggests that, roughly, the dynamics of the growth in industrial production will differ when that growth is positive (specifically, above −0.0035) as compared to when it is negative (specifically, below −0.0035).

Table 5.5 Identifying the STAR Model Specifications

Hypothesis tests	Dependent variable: DLIP Chi-sq.	Prob.	Dependent variable: DINT Chi-sq.	Prob.
	Threshold variable: DLIP(−1)		Threshold variable: ER(−1)	
$H_0 : \beta_{2,i} = \beta_{3,i} = \beta_{4,i} = 0, \forall i$	52.25	0.0000	52.13	0.0000
$H_1 : \beta_{4,i} = 0, \forall i$	16.90	0.0047	13.85	0.0078
$H_2 : \beta_{3,i} = 0 \mid \beta_{4,i} = 0, \forall i$	14.01	0.0155	3.14	0.5340
$H_3 : \beta_{2,i} = 0 \mid \beta_{3,i} = \beta_{4,i} = 0, \forall i$	21.34	0.0007	35.13	0.0000
	Threshold variable: CDB(−2)		Threshold variable: DINT(−1)	
$H_0 : \beta_{2,i} = \beta_{3,i} = \beta_{4,i} = 0, \forall i$	41.87	0.0002	51.96	0.0000
$H_1 : \beta_{4,i} = 0, \forall i$	12.79	0.0254	9.91	0.0420
$H_2 : \beta_{3,i} = 0 \mid \beta_{4,i} = 0, \forall i$	2.60	0.7611	17.62	0.0015
$H_3 : \beta_{2,i} = 0 \mid \beta_{3,i} = \beta_{4,i} = 0, \forall i$	26.48	0.0001	24.43	0.0001

SOURCE: Authors' calculations.

One important issue to examine is whether one regime of the STAR model is just being estimated on a single or a very few data points, so that the model is really just showing that a few data points are a special case. To examine this issue we use the following histogram, Figure 5.5, which shows that the estimated threshold does not simply identify a few observations at the extreme tail of the distribution. Instead, we see that, over the history of the variable, many observations occur above, and below, the threshold. Thus both regimes occurred with some regularity.

The estimated value of the transition parameter, γ, is relatively large at 262.04, suggesting a relatively sharp transition between the regimes, as DLIP(−1) varies around −0.0035. We thus also estimate a TAR model for the change in the log of industrial production to serve as a

Table 5.6 LSTAR for DLIP, January 1955–December 2004

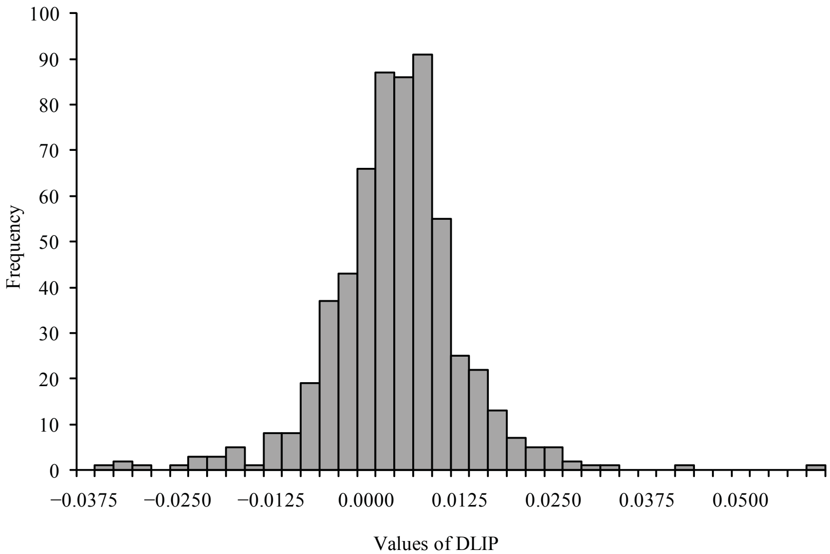
Constant	0.0018	(0.0004);	$p = 0.0001$
DLIP(−1)	0.1795	(0.1034);	$p = 0.0830$
DLIP(−2)	0.3527	(0.1265);	$p = 0.0055$
ER(−1)	−0.0116	(0.0313);	$p = 0.7115$
ER(−2)	0.1109	(0.0381);	$p = 0.0037$
DINT(−1)	9.0912	(3.4673);	$p = 0.0090$
DLIP(−1) × F[DLIP(−1)]	0.1581	(0.1325);	$p = 0.2331$
DLIP(−2) × F[DLIP(−1)]	−0.3357	(0.1508);	$p = 0.0263$
ER(−1) × F[DLIP(−1)]	0.0502	(0.0407);	$p = 0.2173$
ER(−2) × F[DLIP(−1)]	−0.1145	(0.0486);	$p = 0.0189$
DINT(−1) × F[DLIP(−1)]	−7.3670	(4.9938);	$p = 0.1407$
Gamma	262.04	(187.44);	$p = 0.1626$
Threshold for DLIP(−1)	−0.0035	(0.0032);	$p = 0.2758$
R^2		0.2283	
Std. error		0.0079	
SIC		−6.7184	
Log likelihood		2,057.110	

SOURCE: Authors' calculations.

basis for comparison with the STAR model. The estimated value for the threshold for the TAR model is reported in Table 5.7 and is very close to that of the STAR model. Figure 5.6 graphs the transition functions for the TAR and STAR model. Of course this is a step function for the TAR model, and, as the graph makes clear for the STAR model, there is a large range where the two extreme regimes are "smoothly" combined to generate the dynamics that we observe. As the following graph shows, the switch between regimes takes place just below zero.

The estimated LSTAR model for the change in the bond rate is presented in Table 5.8. The coefficients on lags of the changes in the bond rate in the lower regime were statistically insignificant in initial estimates of both the STAR and TAR models, and their inclusion caused convergence difficulties for the STAR model, so we set these two coefficients to zero for the results reported in Table 5.8 (and for the TAR model reported in Table 5.9).

Figure 5.5 Histogram for DLIP



SOURCE: Authors' calculations.

For the bond rate the transition variable is one lag of the change in the bond rate, or DINT(−1). The estimated threshold value is −0.00048, a value that falls far to the left in the distribution of changes in the bond rate between positive and negative values, suggesting that changes in the bond rate are largely governed by one regime, with occasional large declines in the bond rate leading to alternative dynamics. Figure 5.7 presents the histogram of excess returns. We see that there are only a small, though not trivial, number of observations in the lower regime.

The estimated value for the transition parameter is also large for this model, 464.7, so we again estimate a TAR model for comparison, with results reported in Table 5.9. In the TAR model, the estimated threshold value is −0.00048, nearly the same as the STAR model. Given this and the size of the transition variable in the LSTAR model, the two models provide very similar results in terms of the region around the switch, as seen in Figure 5.8.

So far, we have estimated four nonlinear models to be used in out-of-sample forecasting, an LSTAR and a TAR for both the change in

Table 5.7 TAR For DLIP, January 1955–December 2004

Constant	0.0018	(0.0005);	$p = 0.0001$
DLIP(−1)	0.2199	(0.0818);	$p = 0.0074$
DLIP(−2)	0.2635	(0.0823);	$p = 0.0014$
ER(−1)	0.0013	(0.0190);	$p = 0.9461$
ER(−2)	0.0800	(0.0178);	$p = 0.0000$
DINT(−1)	9.3324	(2.6351);	$p = 0.0004$
DLIP(−1) × δ	0.0971	(0.1095);	$p = 0.3755$
DLIP(−2) × δ	−0.2099	(0.0932);	$p = 0.0246$
ER(−1) × δ	0.0303	(0.0222);	$p = 0.1727$
ER(−2) × δ	−0.0679	(0.0213);	$p = 0.0015$
DINT(−1) × δ	−7.1795	(3.2521);	$p = 0.0277$
Threshold for DLIP(−1)		−0.003	
R^2		0.2258	
Std. error		0.0079	
SIC		−6.7365	
Log likelihood		2,056.135	

SOURCE: Authors' calculations.

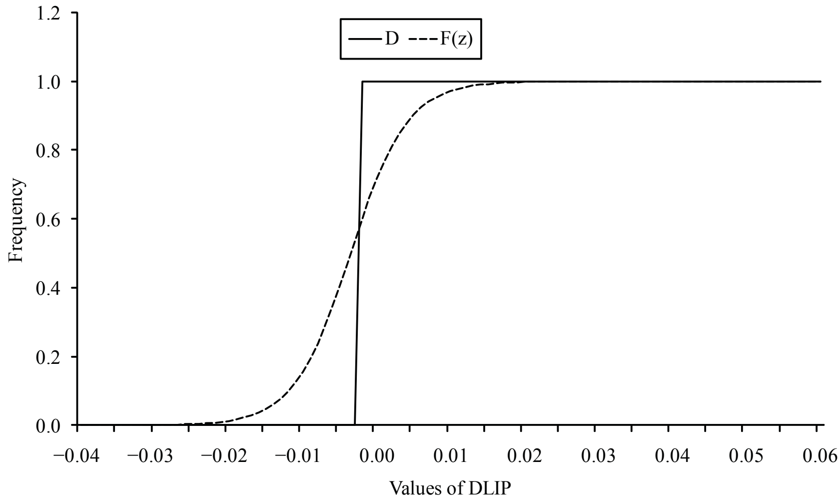
the log of industrial production and the change in the bond rate. We now supplement these models with an alternative approach to capturing nonlinearity among the real and financial variables. We specify and estimate CDR models for the change in the log of industrial production, for excess returns, and for the change in the bond rate.

Beaudry and Koop's (1993) CDR model is designed to capture the asymmetric dynamic caused by the fact that negative shocks to real and financial variables tend to have temporary effects but positive shocks tend to have permanent effects. This asymmetry is embodied in the model through the inclusion of a CDR term, which measures the distance from the previous peak of the variable to the current value. This term is positive when the current value is below the previous peak:

$$(5.8) \quad CDR_t = max(Y_{t-j})_{j \geq 0} - Y_t \ .$$

Inclusion of this CDR term converts an otherwise linear model (like an AR model) into a nonlinear model:

Figure 5.6 DLIP Models: Transition Functions TAR and STAR



SOURCE: Authors' calculations.

(5.9) $\Theta(L)_t \Delta Y_t = \delta + [\Omega(L) - 1]\text{CDR}_t + \varepsilon_t$.

Here $\Theta(L)$ and $\Omega(L)$ represent polynomials in the lag operator L, a convenient way to represent that there are lags of ΔY and lags of CDR in the equation. If the coefficient on the CDR term is positive, then ΔY grows faster when CDR increases—that is, when the recession is deeper. In other words, ΔY grows fasters after a negative shock has placed the economy in a deep recession. When the economy recovers and is growing above its previous peak, this extra growth in ΔY is eliminated. In this case, positive shocks will have longer-lasting positive effects on ΔY than negative shocks. Of course, if the coefficient on the CDR term is < 0, the opposite case holds: a negative shock leads to more persistent performance below the previous peak.

To allow for possible nonlinear effects from the financial markets and the real sector, we investigate two versions of a CDR-type model, one for industrial production and one for stock prices. The CDR term has a positive value when industrial production is below its previous peak, and the CDB term has a positive value when the S&P 500 is below its

Table 5.8 LSTAR for DINT, January 1955–December 2004

Constant	1.95E-06	(8.07E-06);	$p = 0.8090$
DINT(−1)	—	—	—
DINT(−2)	—	—	—
ER(−1)	−0.0118	(0.0028);	$p = 0.0000$
ER(−2)	0.0083	(0.0017);	$p = 0.000$
DINT(−1) × F	0.3995	(0.0461);	$p = 0.0000$
DINT(−2) × F	−0.2235	(0.0409);	$p = 0.0000$
ER(−1) × F	0.0129	(0.0029);	$p = 0.0000$
ER(−2) × F	−0.0078	(0.0170);	$p = 0.0000$
Gamma	464.68	(183.01);	$p = 0.0014$
Threshold for DINT(−1)	−4.81E-04	(4.78E-05);	$p = 0.0000$
R^2		0.2276	
Std. error		1.94E-04	
SIC		−14.1756	
Log likelihood		4,281.459	

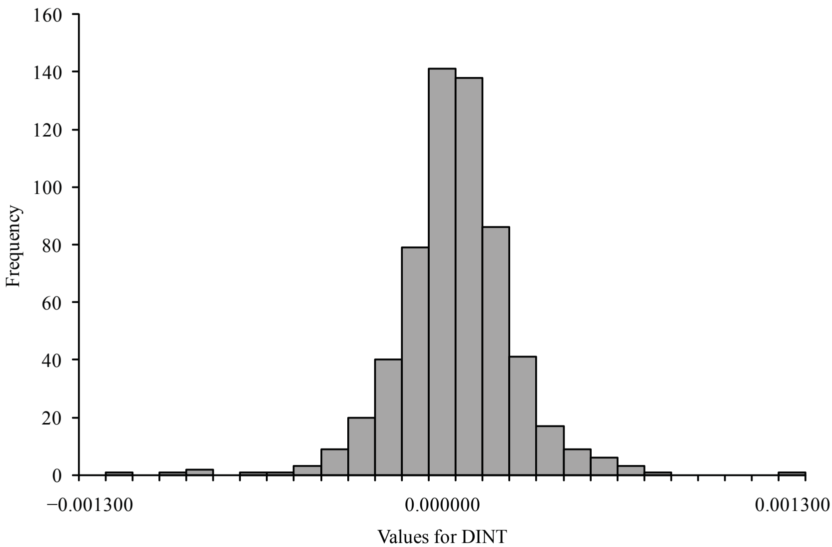
NOTE: — = data not available.
SOURCE: Authors' calculations.

Table 5.9 TAR for DINT, January 1955–December 2004

Constant	1.29E-06	(7.97E-06);	$p = 0.8710$
DINT(−1)	—	—	—
DINT(−2)	—	—	—
ER(−1)	−0.0099	(0.0014);	$p = 0.0000$
ER(−2)	0.0086	(0.0014);	$p = 0.0000$
DINT(−1) × δ	0.3869	(0.0439);	$p = 0.0000$
DINT(−2) × δ	−0.2159	(0.0401);	$p = 0.0000$
ER(−1) × δ	0.0108	(0.0014);	$p = 0.0000$
ER(−2) × δ	−0.0080	(0.0014);	$p = 0.0000$
Threshold for DINT(−1)		−0.00048	
R^2		0.2245	
Std. error		1.94E-04	
SIC		−14.1928	
Log likelihood		4,280.239	

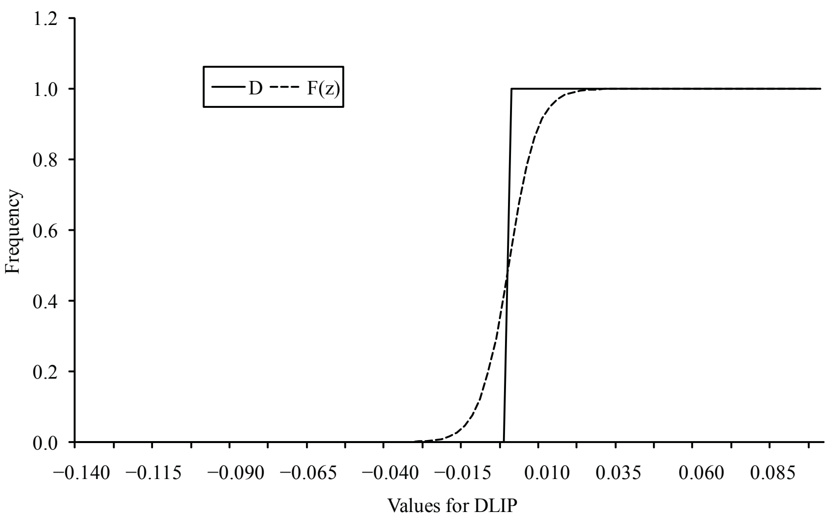
NOTE: — = data not available.
SOURCE: Authors' calculations.

Figure 5.7 Histogram for Changes in the Interest Rate



SOURCE: Authors' calculations.

Figure 5.8 DINT Models: Transition Functions TAR and STAR

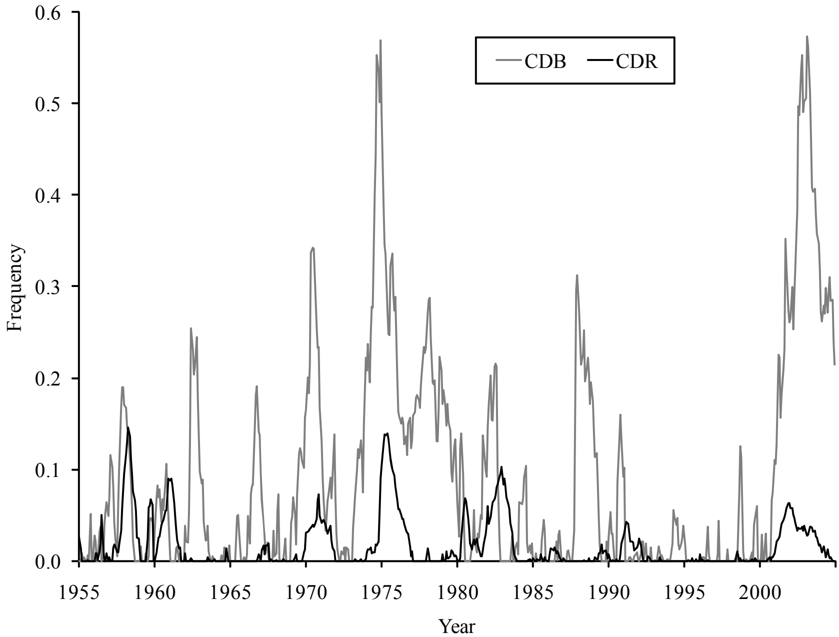


SOURCE: Authors' calculations.

previous peak. The two graphs in Figure 5.9 display the values for CDR and CDB and illustrate the differences in their time series histories.

To investigate this specification of nonlinearity, we took the linear models presented in Table 5.3 and augmented them with both terms—CDR and CDB. Neither the CDR term nor the CDB term was significant for the change in the bond interest rate, indicating that the CDR class of models was not appropriate for that variable. In contrast, both the CDR and the CDB terms were significant in the model for the change in the log of industrial production (see Table 5.10). The sum of the coefficients on the CDR term is positive, meaning that industrial production grows faster after a negative shock to industrial production. This means that negative real shocks have shorter lasting effects than positive real shocks and industrial production tends to grow relatively rapidly after

Figure 5.9 Current Depth of Recession and Current Depth of Stocks, January 1955–December 2004



NOTE: CDR stands for current depth of recession. CDB stands for current depth of stocks.
SOURCE: Authors' calculations.

Table 5.10 CDR Model for DLIP

Constant	0.0023	(0.0006);	$p = 0.0001$
DLIP(−1)	0.1296	(0.0872);	$p = 0.1378$
DLIP(−2)	0.1229	(0.0411);	$p = 0.0029$
ER(−1)	0.0770	(0.0258);	$p = 0.0030$
ER(−2)	—	—	—
DINT(−1)	4.9270	(1.5277);	$p = 0.0013$
CDR(−1)	−0.1840	(0.0997);	$p = 0.0655$
CDR(−2)	0.2241	(0.0958);	$p = 0.0197$
CDB(−1)	0.0611	(0.0286);	$p = 0.0335$
CDB(−2)	−0.0722	(0.0278);	$p = 0.0096$
R^2		0.2274	
Std. error		0.0079	
SIC		−6.7599	
Log likelihood		2,056.741	

NOTE: — = data not available.
SOURCE: Authors' calculations.

recessions. In contrast, the sign of the sum of the coefficients on the CDB term is negative. This suggests that industrial production grows more slowly after stock market declines.

Only the CDR term was significant in the excess returns equation, and the estimated model is presented in Table 5.11. The coefficient on the CDR term is positive, suggesting that excess returns grow faster

Table 5.11 CDR Model for ER

Constant	1.19E-03	(1.59E-03);	$p = 0.4529$
ER(−1)	0.1996	(0.0399);	$p = 0.0000$
DINT(−1)	−26.4039	(6.2711);	$p = 0.0000$
CDR(−1)	0.0960	(0.0486);	$p = 0.0486$
R^2		0.0902	
Std. error		0.0331	
SIC		−3.9400	

SOURCE: Authors' calculations.

when industrial production is in its “recovery” phase and expanding out of a recession.

FORECAST EVALUATION

The above models were all estimated over a sample period from January of 1955 to December of 2004 (1955.01–2004.12). We reserved the final 58 data points, 2005.01–2009.10, for an out-of-sample forecasting comparison. The idea is to estimate the model up to 2004.12, as if we are actually in 2004.12, and use that information and parameter estimates to forecast in 2005.01. Then we update the sample to 1955.01–2005.01 and use that information to forecast 2005.2. We continue this exercise through our last data point, forecasting 2009.10 using the sample 1955.01–2009.09. In this way our forecasts are all constructed using only information available at the time of the forecast.

The above description is an ideal, however, as data revisions occur after the fact, and we have used data available to us late in 2009. If data revisions occurred—and they certainly did to industrial production—then our entire sample in 2009 contains data different from what a forecaster would have available in real time. This is a topic of great interest in the current literature but not one we deal with in this study. Fortunately, financial series such as stock prices and returns are not typically subject to the data revision problem.

An important issue is how to judge forecasting performance. We can calculate how far off each individual forecast is for the various models, and average the forecast errors over our 58-data-point forecasting sample. More often, we calculate the average of the squared forecast errors, and still more often we calculate the square root of the average of the squared forecast errors, or the root mean square forecasting error (RMSFE). This is probably the most widely cited measure of forecast accuracy. Another widely used measure is the average of the absolute value of the forecasting errors, the mean absolute forecasting error (MAFE). Other loss functions are possible, including measures of turning points and loss functions based on utility or profit functions, but we will not pursue those alternatives here.

Table 5.12 provides the out-of-sample measures of the RMSFE and MAFE for our three variables (DINT, DLIP, ER) and our four models (linear, TAR, STAR, CDR). For excess returns we also provide a random walk model. We highlight the model that achieves the best (lowest) value for each variable. For the RMSFE criterion, the best DINT model is the linear model. The TAR model does a particularly poor job. For DLIP, the best RMSFE values are given by the STAR model, followed by the TAR model. The linear model only does better than the CDR model. Finally, for ER the best RMSFE value is provided by the CDR model, and both models beat a random walk.

For the MAFE criterion, we again find that the best forecasts of DINT are provided by the linear model, although here the STAR model appears to do almost as well as the linear model. For DLIP, the best forecasts are from the TAR model, followed by the STAR model, with the linear model third. Again the CDR model does the worst of the four. For ER, the best forecasts come from the CDR model, followed by the linear model, with the random walk bringing up the rear.

More insight into the relative forecast performance can be gleaned from examining the forecasts and forecast errors. Figure 5.10 plots the values for the changes in the bond rate, DINT, along with forecasts of DINT from the linear model and the TAR model. These are graphed on the left-hand scale. The actual values are represented by the line identi-

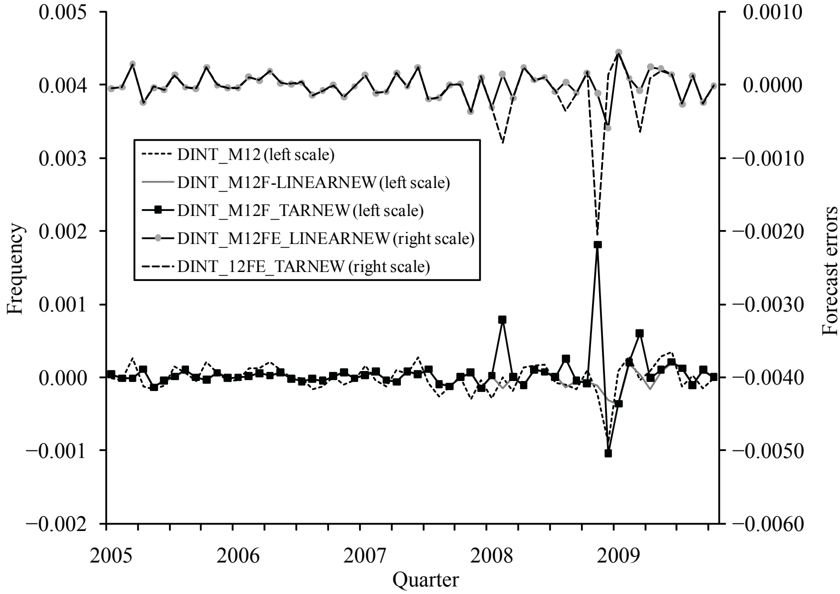
Table 5.12 Performance Measures for Out-of-Sample Forecasts, January 2005–October 2009

RMSE loss criterion	Linear	TAR	STAR	CDR	Random walk
DINT	1.76E-04	3.40E-04	1.92E-04		—
DLIP	8.28E-03	7.64E-03	7.54E-03	8.64E-03	—
ER	4.52E-02	—	—	4.46E-02	4.77E-02

MAE loss criterion	Linear	TAR	STAR	CDR	Random walk
DINT	1.33E-04	1.82E-04	1.35E-04		—
DLIP	5.67E-03	5.16E-03	5.28E-03	5.72E-03	—
ER	2.90E-02	—	—	2.84E-02	3.11E-02

NOTE: — = model not estimated; blank = not applicable.
SOURCE: Authors' calculations.

Figure 5.10 Changes in the Bond Rate and Forecasts, January 2005–September 2009

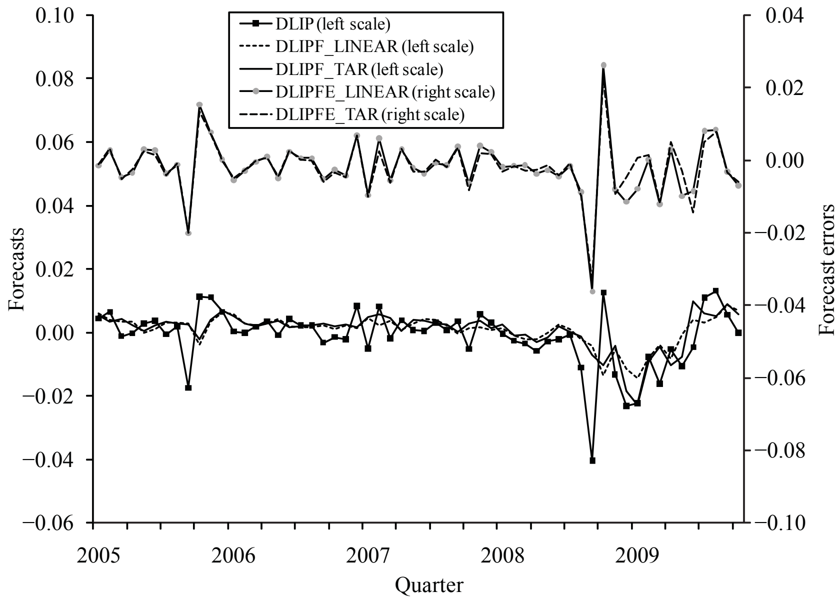


SOURCE: Authors' calculations.

fied in the legend as "DINT_M12," and the forecasts are the other lines. There is a big difference between the actual values and the forecast from the TAR model in the middle of 2008. The TAR model predicted a large value for DINT at this time, but the large value failed to materialize. The difference between the actual value of DINT and the forecast from the TAR model is about −0.002, a large value that led the TAR model to perform quite poorly based on the RMSFE. The forecast errors themselves are plotted at the top, and on the right-hand scale, of Figure 5.10. The large downward spike in mid-2008 is the TAR model forecast error we have just discussed.

Figure 5.11 plots the forecasts, actual value, and forecast errors for changes in the log of industrial production. The line with black squares in the lower part of the graph is the actual value of DLIP, which experienced large upward and downward moves in the latter half of 2008. These movements in DLIP were not forecast by either the linear or TAR

Figure 5.11 Forecasts, Actual Value, and Forecast Errors for Changes in the Log of Industrial Production, January 1955–September 2009

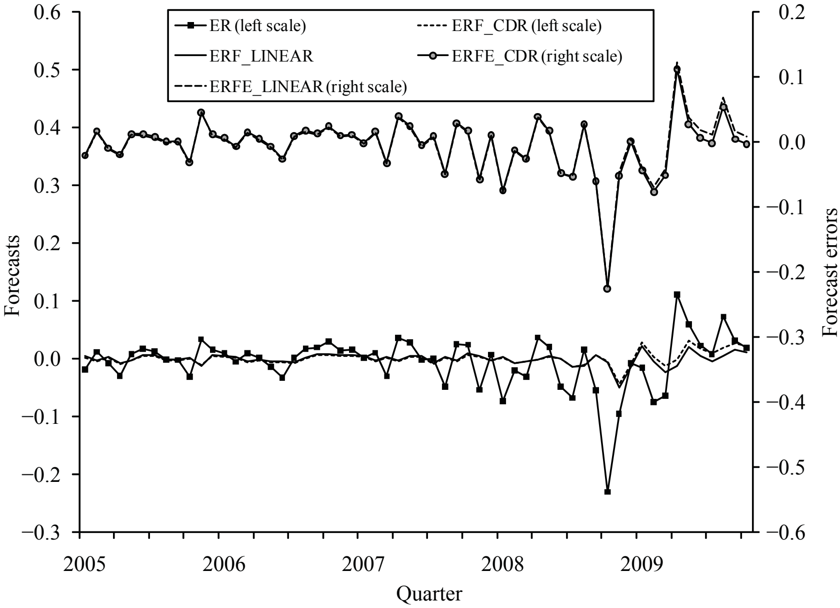


SOURCE: Authors' calculations.

models. Hence this movement generated forecast errors, which can be seen in the top of Figure 5.11 on lines graphed against the right-hand scale. The forecast errors from the linear and TAR models seem similar in Figure 5.11, although there is some discrepancy near the end of 2008 and at the beginning of 2009. During this period the TAR model does slightly better, and this leads to the TAR model having a lower RMSFE in Table 5.12.

Finally, Figure 5.12 plots linear and CDR forecasts of excess returns. Again we see there was a large downward spike in ER in the fourth quarter of 2008 that was not forecast by either the linear or the CDR models. Thus this spike shows up in the top of Figure 5.12 as forecast errors for both the linear and CDR models. As with industrial production in Figure 5.11, it is difficult to see much difference in the forecasts, or forecast errors, from the linear and CDR models. The main

Figure 5.12 Linear and CDR Forecasts of Excess Returns, January 1955–September 2009



SOURCE: Authors' calculations.

difference appears beginning in the middle of 2009, and this small difference leads to the CDR model having somewhat lower RMSFE values compared to the linear model.

One issue with results such as in Table 5.12 is the lack of a measure for saying just how much better one model's forecasts are over another's. We would like a way of answering this question. Usually this is phrased as an issue of statistical significance. We want to know whether, for the ER model, the CDR model forecasts are statistically significantly better than the forecasts of the linear model.

There are a variety of tests available for answering this question. A classic test of forecast accuracy is the Diebold and Mariano (1995) test, which has as its null hypothesis that two forecasts are equally accurate by the chosen criterion (say RMSFE), and the alternative that one of the two is better. Another test is called the encompassing test, which compares two forecasts and asks whether, given one forecast, there is additional useful information in the second forecast. If the answer is yes, then you might want to combine the two forecasts. If the answer is no, then you might want to use the best forecast and ignore the second forecast as containing no additional information once you have the first forecast. Encompassing tests have been in use for quite some time, and some early advocates include Chong and Hendry (1986) and Harvey, Leybourne, and Newbold (1989).

In conducting these tests, an important practical issue is whether or not your forecasting models are nested. The initial Diebold and Mariano test was designed for use with nonnested forecasting models, which are basically two unrelated forecasting models. Nested forecasting models, in contrast, are models where one model is a subset of another model. In our models, the linear model is nested inside the CDR model. If we just eliminate the CDR terms—say, by setting the coefficients on the CDR terms to zero—we get back the linear model. Similarly, our TAR and STAR models also nest the linear model. If we just set the terms multiplying the transition variable all to zero, then we have a one-regime linear model.

Statistical comparisons of nested models bring up complications relative to comparisons of nonnested models. This issue has been explored by a number of authors, including work by West (1996), Clark (1999), McCracken (2000), and Clark and McCracken (2001), and we refer the interested reader to those papers.

Giacomini and White (2006) suggest a new approach to statistical comparisons of forecasts from nested models. Basically they have a version of the Diebold and Mariano test that works for nested models, a model based on the idea of conditional forecast comparisons, and we use their approach here. In Table 5.13 we report tests of the RMSFE loss function for our various models. In these tests we select a baseline model and compare our other models to the baseline. For the DINT and DLIP forecasts we use the linear model as the baseline. For the ER forecasts we use the random walk model as the baseline.

For the DINT forecasts we see in Table 5.12 that the smallest RMSFE was for the linear model. Thus the test results in Table 5.13 are basically tests of whether the baseline linear model is statistically significantly better than the TAR or STAR models. The answer is that while the linear model has a lower RMSFE, it is not statistically sig-

Table 5.13 GW Version of DM Test (unconditional)—RMSE Loss Function

Variable to forecast	Random walk	Linear	TAR	STAR	CDR/CDB
DINT	—	Baseline	1.2871 (p = 0.203)	0.9974 (p = 0.323)	—
DLIP	—	Baseline	1.5092 (p = 0.131)	1.2880 (p = 0.198)	−0.7353 (p = 0.465)
ER	Baseline	1.5329 (p = 0.125)			0.5257 (p = 0.601)

NOTE: — = model not estimated; blank = not applicable.
SOURCE: Authors' calculations.

nificantly lower than either of the two nonlinear models. This is even true for the TAR model, which appeared to perform quite poorly in terms of RMSFE. Still, this is no victory for the nonlinear models. A linear model is much easier to estimate and to use for forecasting. If a linear model gives forecasts that are as good or better than the nonlinear model alternatives, then we would usually avoid going to the trouble of forecasting from a nonlinear model.

For the DLIP forecasts, we see in Table 5.12 that the lowest RMSFE values were generated by the two nonlinear models. In Table 5.13 we see that even though TAR and STAR both provided better RMSFE values, the improvement was not statistically significant. The TAR model has the best marginal probability value, 13.1 percent, but that means that at conventional significance levels of 5 or even 10 percent we would not reject the hypothesis of equal RMSFE for the linear and TAR forecasts.

For the ER forecasts, we compared both the linear model and the CDR model to a random walk baseline. In Table 5.12 we see that the linear model had better RMSFE values than the random walk model, and that the CDR model had better RMSFE values than the linear model, but in Table 5.13 we see that neither the linear nor the CDR model improves on the random walk model in a statistically significant amount.

The results for the DLIP and ER forecasts are disappointing for fans of the nonlinear model. In both cases a nonlinear model or models made improvements in terms of RMSFE values, but these improvements were not statistically significant.

In Table 5.14 we report a similar exercise using the MAFE criterion. Here we find a bit better news for the nonlinear models. For DINT we again find that the linear model is best, but for DLIP we find that the forecasts from the TAR model are statistically significantly better than forecasts from the linear model. The marginal probability value is 3.2 percent, indicating that at standard significance levels of 5 percent we would reject the hypothesis of equal forecasting accuracy—in terms of MAFE—of the linear model and the TAR model.

For the ER model the results in Table 5.14 indicate that forecasts from the linear model are statistically significantly better than forecasts from the random walk model. But we find the disappointing result that forecasts from the CDR model are statistically insignificantly different in accuracy from forecasts of the random walk model. Even though the CDR model generated a better MAFE value compared to the linear model, the variability of the forecasts from the CDR model means that the difference is judged to be statistically insignificant.

Overall, then, our forecast evaluation indicates only weak support for the superiority of forecasts of DLIP from a TAR model, and even weaker support for using a CDR model to forecast ER. We find no support for using anything other than a linear model for forecasting DINT. Basically we find the result, familiar to many in this literature, that nonlinear models appear to fit well in estimation samples but that these models don't fair nearly as well in out-of-sample forecasting exercises.

Table 5.14 GW Version of DM Test (unconditional)—MAE Loss Function

	Random walk	Linear	TAR	STAR	CDR/CDB
DINT	—	Baseline	1.3157	1.6432	—
			(p = 0.194)	(p = 0.106)	
DLIP	—	Baseline	2.1438	1.4329	−0.2010
			(p = 0.032)	(p = 0.152)	(p = 0.841)
ER	Baseline	2.0097			0.7158
		(p = 0.044)			(p = 0.477)

NOTE: — = model not estimated; blank = not applicable.
SOURCE: Authors' calculations.

SOME COMMENTS ON MULTIPLE-STEP-AHEAD FORECASTING

In the analysis above we have investigated the ability of a set of nonlinear models to generate forecasts of two financial variables and industrial production that are better than forecasts from a linear model. This analysis has looked at one-step-ahead forecasts, or forecasts made at time t for the value of variables at time $t + 1$. It is also possible, of course, to construct multiple-step-ahead forecasts—forecasts made at time t for the value of variables at time $t + 2$ or later (better represented as $t + H$, where H stands for the horizon) and for multiple-step-ahead forecasts $H > 1$.

Nonlinear models present particular challenges when constructing multiple-step-ahead forecasts. For linear models the law of iterated expectations and the use of the linear expectation operator on linear equations makes multiple-step-ahead forecasting a straightforward extension of one-step-ahead forecasts. To see this, consider a simple AR(1) model:

$$(5.10) \quad y_t = \beta y_{t-1} + \varepsilon_t \, .$$

To calculate the one-step-ahead forecast we rewrite Equation (5.10) for time $t + 1$ and take expectations conditioned on knowledge of the value of y at time t:

$$(5.11) \quad y_{t+1} = \beta y_t + \varepsilon_{t+1} \, .$$

Then, taking expectations conditioned on knowledge of y_t, we have

$$(5.12) \quad E(y_{t+1} \mid y_t) = \beta y_t \, .$$

To think about a two-step-ahead forecast made at time t, rewrite Equation 5.10 for time $t + 2$ and iteratively substitute to write the result as a function of the value of y at time t:

$$(5.13) \quad y_{t+2} = \beta y_{t+1} + \varepsilon_{t+2} = \beta(\beta y_t + \varepsilon_{t+1}) + \varepsilon_{t+2} = \beta^2 y_t + \beta\varepsilon_{t+1} + \varepsilon_{t+2} \, .$$

Then, taking expectations conditioned on knowledge of y_t, we have

(5.14) $E(y_{t+2}|y_t) = \beta^2 y_t$.

Thus the two-step-ahead forecast in Equation 5.14 is a simple extension of the one-step-ahead forecast in Equation 5.12. While the above is a particularly simple model in terms of notation, the principle holds more generally in forecasts from linear models.

Consider now a nonlinear model. A simple TAR model would be

(5.15) $y_t = \beta y_{t-1} + \gamma y_{t-1} \times I(y_{t-1} < c) + \varepsilon_t$.

To calculate the one-step-ahead forecast, we rewrite Equation (5.15) for time $t + 1$ as

(5.16) $y_{t+1} = \beta y_t + \gamma y_t \times I(y_t < c) + \varepsilon_{t+1}$.

Then, taking expectations conditioned on knowledge of y_t, we have

(5.17) $E(y_{t+1}|y_t) = \beta y_t + \gamma y_t \times I(y_t < c)$.

So far this looks straightforward, much like the one-step-ahead forecast from the linear model. However, consider the two-step-ahead forecast made at time *t*. Rewrite Equation (5.15) for time $t + 2$ and iteratively substitute to write the result as a function of the value of *y* at time *t* :

(5.18) $y_{t+2} = \beta y_{t+1} + \gamma y_{t+1} \times I(y_{t+1} < c) + \varepsilon_{t+2}$

or

$$y_{t+2} = \beta[\beta y_t + \gamma y_t \times I(y_t < c) + \varepsilon_{t+1}] + \gamma[\beta y_t + \gamma y_t \times$$

$$I(y_t < c) + \varepsilon_{t+1}] \times I\{[\beta y_t + \gamma y_t \times I(y_t < c) + \varepsilon_{t+1}] < c\} + \varepsilon_{t+2} .$$

Then, taking expectations conditioned on knowledge of y_t, we have

$$(5.19)\quad E(y_{t+2}|y_t) = \beta[\beta y_t + \gamma y_t \times I(y_t < c)] + \gamma[\beta y_t + \gamma y_t \times I(y_t < c)]$$
$$\times E\Big(I\{[\beta y_t + \gamma y_t \times I(y_t < c) + \varepsilon_{t+1}] < c\}|y_t\Big) + E\{\varepsilon_{t+1}$$
$$\times I[\varepsilon_{t+1} < c - \beta y_t - \gamma y_t \times I(y_t < c)]|y_t\}.$$

Clearly, Equation (5.19) is not a straightforward extension of Equation (5.17). In fact, the last term in Equation (5.17) involves expectation of the disturbance term ε_{t+1} interacted with a function of the same disturbance term ε_{t+1}. It is evident, then, that the two-step-ahead forecast involves considerations of higher moments than the mean. To put this in practice requires distributional assumptions on the error term or else some sort of bootstrap procedure to calculate expectations from the empirically realized (i.e., estimated) disturbances. None of this makes multiple-step-ahead forecasts from nonlinear models impossible, but they are much more involved than such forecasts in a linear model, and as this chapter is already quite long we do not pursue such forecasts here.

CONCLUSION

Our study demonstrates once again how nonlinear models can fit very well in-sample and yet struggle to outperform linear models in out-of-sample forecasting. This finding is not unusual, but it is frustrating to proponents of nonlinear modeling. Nonlinear modelers usually exert care in trying to avoid overfitting within sample, and yet the out-of-sample performance difficulties point to overfitting as one possible source of the problem. The exact reason for these difficulties with forecasts from nonlinear models remains an open issue.

References

Beaudry, Paul, and Gary Koop. 1993. "Do Recessions Permanently Change Output?" *Journal of Monetary Economics* 31(2): 149–163.

Bradley, Michael D., and Dennis W. Jansen. 1997. "Nonlinear Business Cycle Dynamics: Cross-Country Evidence on the Persistence of Aggregate Shocks." *Economic Inquiry* 35(3): 495–509.

———. 2000. "Are Business Cycle Dynamics the Same across Countries? Testing Linearity around the Globe." *Studies in Nonlinear Dynamics and Econometrics* 4(2): 51–71.

———. 2004. "Forecasting with a Nonlinear Dynamic Model of Stock Returns and Industrial Production." *International Journal of Forecasting* 20(2): 321–342.

Chan, Kung-sik, and Howell Tong. 1986. "On Estimating Thresholds in Autoregressive Models." *Journal of Time Series Analysis* 7(3): 179–190.

Cheng, Bing, and D.M. Titterington. 1994. "Neural Networks: A Review from a Statistical Perspective." *Statistical Science* 9(1): 2–30.

Chong, Yock Y., and David F. Hendry. 1986. "Econometric Evaluation of Linear Macro-Economic Models." *Review of Economic Studies* 53(4): 671–690.

Clark, Todd E. 1999. "Finite-Sample Properties of Tests for Equal Forecast Accuracy." *Journal of Forecasting* 18(7): 489–504.

Clark, Todd E., and Michael W. McCracken. 2001. "Tests of Equal Forecast Accuracy and Encompassing for Nested Models." *Journal of Econometrics* 105(1): 85–110.

Diebold, Francis X., and Roberto S. Mariano. 1995. "Comparing Predictive Accuracy." *Journal of Business and Economic Statistics* 13(3): 253–263.

Falk, Barry. 1986. "Further Evidence on the Asymmetric Behavior of Economic Time Series over the Business Cycle." *Journal of Political Economy* 94(5): 1096–1109.

Giacomini, Raffaella, and Halbert White. 2006. "Tests of Conditional Predictive Ability." *Econometrica* 74(6): 1545–1578.

Granger, Clive W.J., and Timo Teräsvirta. 1993. *Modelling Dynamic Nonlinear Relationships*. Oxford: Oxford University Press.

Granger, Clive W.J., Timo Teräsvirta, and Heather M. Anderson. 1993. "Modeling Nonlinearity over the Business Cycle." In *New Research on Business Cycles, Indicators, and Forecasting*, James H. Stock and Mark W. Watson, eds. Chicago: University of Chicago Press, pp. 311–326.

Hamilton, James D. 1989. "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle." *Econometrica* 57(2): 357–384.

Hansen, Bruce E. 1997. "Inference in TAR Models." *Studies in Nonlinear Dynamics and Econometrics* 2(1): 1–14.

Harvey, David I., Stephen J. Leybourne, and Paul Newbold. 1989. "Tests for Forecast Encompassing." *Journal of Business and Economic Statistics* 16(2): 254–259.

Jansen, Dennis W., and Wankeun Oh. 1999. "Modeling Nonlinearity of Business Cycles: Choosing between the CDR and STAR Models." *Review of Economics and Statistics* 81(2): 344–349.

Li, Qi, and Jeffrey Scott Racine. 2007. *Nonparametric Econometrics: Theory and Practice*. Princeton, NJ: Princeton University Press.

Luukkonen, Ritva, Pentti Saikkonen, and Timo Teräsvirta. 1988. "Testing Linearity against Smooth Transition Autoregressive Models." *Biometrika* 75(3): 491–499.

McCracken, Michael W. 2000. "Robust Out-of-Sample Inference." *Journal of Econometrics* 99(2): 195–223.

Neftçi, Salih N. 1984. "Are Economic Time Series Asymmetric over the Business Cycle?" *Journal of Political Economy* 92(2): 307–328.

Pesaran, M. Hashem, and Simon M. Potter. 1997. "A Floor and Ceiling Model of U.S. Output." *Journal of Economic Dynamics and Control* 21(4–5): 661–696.

Potter, Simon M. 1995. "A Nonlinear Approach to U.S. GNP." *Journal of Applied Econometrics* 10(2): 109–125.

Shiller, Robert S. 2011. *Online Data*. Hew Haven, CT: Yale University, Department of Economics. http://www.econ.yale.edu/~shiller/data.htm (accessed April 6, 2011).

Teräsvirta, Timo. 1994. "Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models." *Journal of the American Statistical Association* 89(425): 208–218.

Teräsvirta, Timo, and Heather M. Anderson. 1992. "Characterizing Nonlinearities in Business Cycles Using Smooth Transition Autoregressive Models." *Journal of Applied Econometrics* 7(S1): S119–S136.

Tong, Howell. 1990. *Non-Linear Time Series: A Dynamical System Approach*. Oxford: Oxford University Press.

Tsay, Ruey S. 1989. "Testing and Modeling Threshold Autoregressive Processes." *Journal of the American Statistical Association* 84(405): 231–240.

van Dijk, Dick, and Philip Hans Franses. 1999. "Modeling Multiple Regimes in the Business Cycle." *Macroeconomic Dynamics* 3(3): 311–340.

West, Kenneth D. 1996. "Asymptotic Inference about Predictive Ability." *Econometrica* 64(5): 1067–1084.