

Forecasting Stock Prices With Linear And Nonlinear Settings: A Comparison

Dimitri Tsoukalas, Purdue University-Calumet, USA

ABSTRACT

This paper is devoted to the application and comparison of linear (VAR) and nonlinear Multiple Adaptive Regression Splines (MARS) forecasting models, in estimating, evaluating, and selecting among linear and non-linear forecasting models for economic and financial time series. We argue that although the evidence in favor of constructing forecasts using non-linear models is rather sparse, there is reason to be optimistic. Nonlinear models reduce nonlinearity and Gaussianity in the residuals of the linear models. Linear models, however, demonstrate better forecasts than nonlinear. However, much remains to be done. Finally, we outline a variety of topics for future research, and discuss a number of areas which have received considerable attention in the recent literature, but where many questions remain.

Keywords: Forecasting; Gaussianity; Linear and Nonlinear Models; VAR; Multiple Adaptive Regression Splines (MARS)

1. INTRODUCTION AND LITERATURE REVIEW

This article examines the forecasting ability of the dividend growth rate and dividend yield for stock returns, in linear and non-linear settings. Previous work has typically only considered predictability in a linear framework, ignoring the substantial evidence of non-linear dynamics in stock returns, dividend growth and dividend yield. Predictability of returns has been extensively studied with arguments for and against it. Among the plethora of research papers, in support of linear models, are Campbell and Shiller (1988a,b) and Fama and French (1988). Campbell and Shiller (2001) and Campbell and Yogo (2006) have shown and supported that returns are predictable using financial variables such as the dividend yield. Cochrane (2008) presents additional evidence in support of the predictability of the dividend yield and dividend growth.

One of the primary uses of nonlinear models, is for their forecasting ability and it is in terms of their forecasting performance that are usually judged. The overall conclusion of the existing literature suggests, that the forecasting performance and accuracy of such models, over rival linear models, is not particularly superior. De Gooijer and Kumar (1992), M.P. Clements et al (2004) concluded that there is no clear evidence in favor of non-linear over linear forecasting performance. A possible problem may be that nonlinear models do not capture reality better than linear. Although, non-linear models do not perform better than linear (overall), we insist using them because we believe that linear models cannot capture certain types of economic behavior (such as expansions and recessions), or economic performance at certain times, or periods of high and low volatility, or self-exciting or catastrophic behavior in financial markets (Hamilton 1989; Sichel 1994). Diebold and Nason (1990) give a number of reasons why non-linear may fail to outperform linear models. Some of the most commonly used non-linear models are the ARCH type models (Engle 1982, Bollerslev 1986, Nelson 1991, Bollerslev, Engle, and Nelson 1994). M. Islam, L. E. Ali, N. Afroz (2012) used ARCH-type models to forecast volatility in the Bangladesh stock market index, show that linear dominate non-linear models. D. McMillan and M. E. Wohar (2008) used the ESTR introduced by Kapetanios et al (2006) model to predict stock returns. Results support the asymmetric ESTR model in forecasting. S. Corlay (2013) presents a new method the B-Spline technique to model volatility, using simulated data.

In this paper we use the MARS (Multiple Adaptive Recursive Splines), a non-linear technique (Breiman (1991), Stokes (1996)) to forecast stock returns in the NIKKEI-225 stock market index from 1/1/1955-12/30/2005. The overall results favor linear models.

2. METHODOLOGY, DATA ANALYSIS, VARIABLES IN THE SYSTEM, DEFINITIONS, AND HYPOTHESES TO BE INVESTIGATED

To analyze data with the Multiple Adaptive Splines (MARS) method someone has to develop a technique to start with. Since nonlinear is some combination of linear, we start with a linear model. MARS can be considered as being a generalization of VAR models. In our technique we will start with a Vector Autoregressive (VAR) model such as proposed by Tiao and Box (1981). The Hinich tests will follow then, to check for nonlinearity, and last if the VAR fails Gaussianity and Linearity we will proceed with the MARS methodology.

2.1 Vector Autoregressive (VAR) Method

In time series analysis econometricians assume, without testing, that the selection of the appropriate lag length removed all autocorrelation in the residuals of each VAR model. Although, linear models are considered a good estimation technique to study time series processes, it is important to test the non-linearity and non-gaussianity of the process. Non-linearities hidden in the data may contain useful information that may be exploited. Neuburger and Stokes (1991), and Stokes and Neuburger (1996) have shown that linearity can be rejected for monthly stock returns. Therefore, further testing is needed to determine whether non-linearity remains in the residuals of the VAR models. If the residuals of the VAR model of stock returns show non-linearity, we might extend our research to nonlinear models to reduce the residual sum of squares and the non-linearity in stock returns. The linear econometric method we employ, the (VAR), estimates unrestricted reduced form equations that have uniform sets of the lagged variables of every equation as regressors and are free of prior restrictions on the structure of relationships.

Assume that the unrestricted VAR model is:

$$X_t = \sum_{i=1}^n A_i X_{t-i} + u_t \quad (1)$$

where $X = (RS, D/P)'$: the vector of variables used in the model, RS = real stock returns, D/P = the dividend price ratio, $u = (u^{RS}_t, u^{D/P}_t)$ are serially and cross-equation uncorrelated.

From (1) we get the reduced form:

$$X_t = \sum_{i=1}^n G_i X_{t-i} + e_t \quad (2)$$

or

$$RS_t = d_{11} RS_{t-1} + d_{12} D/P_{t-1} + e_{1t} \quad (2.a)$$

$$D/P_t = d_{21} RS_{t-1} + d_{22} D/P_{t-1} + e_{2t} \quad (2.b)$$

In the above model the returns equation is given by (2.a).

2.2 The Hinich Test

To detect non-linearity, we applied a statistical technique involving the estimation of the bispectrum of the observed time series, the Hinich (1982) test. Linear modeling methods such as VAR or OLS require that the values of the error term be normally distributed. The basis for this assumption is that statistical errors are the sum of a number of independent effects. If the normality assumption is violated, problems in model specification are often present. There is a need therefore, when we use linear models such as VAR or OLS, to verify that the behavior of the residual is linear, or close to linear rather than assuming it. The Hinich procedure uses the bispectrum to test a series for three way systematic relationships. The autocorrelation tests whether x_t and x_{t-k} are related, the Hinich test however, examines whether x_t , x_{t-k} and x_{t-j} are jointly related for $k \neq j$. Let $X(n)$ be a stationary time series with mean zero, $m_x = E[X(n)] = 0$, let the autocovariance of $X(n)$ be $C_x(n) = E[X(t+n)X(t)]$ and let $C_{xx}(n,m) = E[X(t+n)X(t+m)X(t)]$ be the third order moments of $\{X(N)\}$, which is also called bicovariance signal. To decide whether a series is white noise researchers examine the covariance of the series $C_x(n)$. If $C_x(n) = 0$ for all n not zero,

the series is white noise. By assuming that the errors of their models are Gaussian in testing for white noise, researchers ignore information regarding possible nonlinear relationships that are found in the third order moments, (Stokes 1991). Hinich and Patterson (1985), have argued that although a series may be white noise if it is not Gaussian, $X(n)$ and $X(m)$ cannot be independent. This can create biased parameter estimates in statistical analysis. If the distribution of $X(n)$ is multivariate normal then the series in addition to being white noise is also Gaussian. Stokes (1991), defines a white noise series as a pure white noise if in addition to being white noise is also Gaussian, pre-assuming that $X(n_1), \dots, X(n_t)$ are independent and identically distributed random variables for all values of n_1, \dots, n_t . All pure white noise series are white, but not all white noise series are pure white unless they are also Gaussian. Brocket, Hinich and Patterson (1988), argue that it is possible for a series to be linear without being Gaussian, but all stationary Gaussian series are linear. The Hinich procedure uses the bispectrum of $X(n)$ which is defined as the (two-dimensional) Fourier Transform of the third-moment function of $X(n)$. If a series fails the Hinich linearity test then it indicates that a linear model cannot describe the series. The test generates two values: the Z value for linearity and the G value for Gaussianity. For values smaller than two the probability that the series is linear and Gaussian at the 95% confidence interval is very strong. See also Lim and Hinich (2005b) and Hinich and Patterson (2005).

2.3 Multiple Adaptive Splines (MARS) Method

Breiman (1991) introduced a methodology for nonlinear regression modeling, the Multivariate Adaptive Regression Splines (MARS), which fits splines into the model instead of simple functions. Assume a linear model: $Y = f(x_1, \dots, x_k) + \epsilon$ (4), where (x_1, \dots, x_k) are the right-hand side variables with N observations on each variable. ϵ is a random error variable with zero mean and variance σ^2 . The error variable reflects the dependence of Y on other quantities than x. The goal of the MARS methodology is to find an estimate \hat{f} to approximate the nonlinear function $f(x^1, \dots, x^k)$. The MARS estimate is $\hat{f}(x) = \sum_{j=1}^s c_j K_j(x)$. Here $K_j(x)$ is a basis function associated with s subregions $\{R\}_{j=1}^s$, c_j is the coefficient for the j^{th} product basis function. For a given set of product basis functions the coefficients $\{c_j\}_{j=1}^s$ can be determined by an OLS regression. The MARS methodology can identify the subregions where the coefficients are stable and detects interactions up to a maximum number of possible interactions (Neuberger and Stokes 1996). The MARS methodology makes an assumption on levels.

For example assume a model:

$$Y = \alpha + \beta_1 + \epsilon \text{ for } x > 100 \tag{5}$$

$$= \alpha + \beta_2 + \epsilon \text{ for } x < 100$$

MARS model will be:

$$Y = \alpha + \beta_1(x-\tau^*)_+ - \beta_2(x-\tau^*)_+ + \epsilon \tag{6}$$

where $(\)_+$ is the right (+) truncated spline function. The MARS model uses a knot which is a function of the right truncated spline function $(x-\tau^*)_+$. In the above equation the first knot is $K_1(X) = (x-\tau^*)_+$ and the second $K_2(X) = (\tau^*-x)_+$. If the expression in the parenthesis is greater than zero then it takes its real value, if it is smaller than zero then it is equal to zero. At $x = \tau^*$ there is no solution. Assume $\tau^* = 100$, then the MARS model will be: $Y = \alpha + \beta_1(100 - x)_+ - \beta_2(x - 100)_+ + \epsilon$. To determine the exact value of Y Friedman (1991) uses a linear or cubic approximation. Both estimation techniques are estimated and the one with the lowest sum of squares is selected. With the MARS models we can get very complex interactions. Example of an interaction model for $Y = f(x, z)$:

$$Y = \alpha + \beta_1(x-\tau_1^*) + \beta_2(\tau_1^* - x)_+ + \beta_3(x-\tau_1^*)_+(z-\tau^*_{z2})_+ + \epsilon \tag{7}$$

which implies:

$$Y = \alpha + \beta_1x - \beta_1\tau_1^* + \epsilon \quad \text{for } x > \tau_1^* \text{ and } z < \tau_2^* \tag{7a}$$

$$= \alpha + \beta_2x + \beta_2\tau_1^* + \epsilon \quad \text{for } x < \tau_1^* \tag{4.8.4b}$$

$$= \alpha + \beta_1 x - \beta_1 \tau_1^* + \beta_3 (xz - \tau_1^* z - \tau_2^* x + \tau_1^* \tau_2^*) + \dots \quad \text{for } x > \tau_1^* \text{ and } z > \tau_2^* \quad (7b)$$

To find the degree of complexity of the model, we use a modified generalized cross validation criterion (MGCV) Friedman (1991).

$$MGCV = [(1/N) \sum_{i=1}^N (Y - f(x))^2] / [1 - [C(M)^*] / N]^2$$

where N is the number of observations, $f^{\wedge}(x_i) = Y_i^{\wedge}$ and $C(M)^*$ is a complexity penalty. $C(M)^*$ can be expressed as: $C(M)^* = C(M) + d.M$, where the parameter $d = 3$ by default, $C(M)$ is the number of parameters being fit and M is the number of nonconstant basis functions in the MARS model. The MGCV statistic is used to eliminate parameters that do not improve the model. The MARS model is powerful for low order interactions and large number of right-hand variables.

3. EMPIRICAL RESULTS

3.1 Causality and Hinich Test Results for VAR, MARS

In this section first, the results of the Hinich test for the MARS model along with the R^2 are compared with those of the VAR model. Second the Granger F-statistic is obtained for the MARS model, using the same procedure as in the VAR model, which enhances the predictability of stock returns we found in the linear models. The results of the Hinich test are reported in Table 1.

These test statistics are normally distributed, if G or L is greater than or equal to 2, we can reject the assumption of normality or linearity at the 95% level. For the VAR models, the R^2 values are quite low; the G and L scores are well above 2 indicating that the residuals of the VAR models have failed the non-Gaussianity and linearity tests in every market. The MARS model however, performs better increasing significantly the R^2 in every model. The linearity and non-Gaussianity scores are improved compared to VAR. For example, in model (RJP,D/P)', the G and L scores were decreased from 6.051 and 4.37 of the VAR model to 3.364 and 1.93 respectively. The R^2 increased from .2445 to .5038.

The Granger F value is obtained by just comparing the residual sum of squares for the constrained model: $RS_t = a_0 + a_1 RS_{t-1} + \dots + RS_{t-k} + \epsilon_t$, and unconstrained regression model $RS_t = a_0 + a_1 RS_{t-1} + \dots + RS_{t-k} + b_1 (D/P)_{t-1} + \dots + (D/P)_{t-k} + \epsilon_t$. The appropriate test statistic is: $F_{M,T,K-1} = [(RSS_R - RSS_U) / M] / [(RSS_U / T-K-1)$ or $F_{M,T,K-1} = [(R^2_U - R^2_R) / M] / [(1 - R^2_R) / (T-K-1)]$. Where: M is the number of the constrained independent variables indicating the degrees of freedom in the denominator, K is the total number of independent variables, and T is the total number of observations. (T-K-1) indicates the degrees of freedom in the denominator. We test whether the informational variables (D/P) predict stock returns, using the F distribution at the level of significance of 5% or 10%. Results in Table 2, show that the F values are more significant in the MARS (F = 45.670) than in the VAR (F = 11.515) model. (Note: this result enhances the relative importance of variables in the MARS model.) We also find that in all models the dividend growth rate predicts real stock returns (explanations of variables (RJP, DG) are available upon request.

Table 1: Comparison of Hinich Test Results for VAR and MARS Models

Variables	VAR Models	MARS Models
(RJP, DG)'	G = 8.032 L = 5.710 $R^2 = .3594$	G = 1.906 L = 1.789 $R^2 = .5144$
(RJP, DP=JPD/JPS)'	G = 6.051 L = 4.370 $R^2 = .2445$	G = 3.364 L = 1.930 $R^2 = .5038$

Note: The above table presents and compares the Hinich test results and the R^2 of two-variable VAR and MARS models.

Table 2: Comparison of Granger Causality Results for VAR and MARS

Variables	VAR Models	MARS Models
(RJP, DG)'	F = 15.972 (1.00)	F = 11.940 (.999)
(RJP, DP=JPD/JPS)'	F = 11.515 (1.00)	F = 45.670 (1.00)

The Granger F-test statistic is given by: $F_{M,T,K-1} = [(R^2_U - R^2_R) / M] / [(1 - R^2_R) / (T-K-1)]$ where, $R^2_U = 1 - RSS_U / TSS_U$ and $R^2_R = 1 - RSS_R / TSS_R$; M is the number of the constrained independent variables indicating the

degrees of freedom in the denominator, K is the total number of independent variables, and T is the total number of observations. (T-K-1) indicates the degrees of freedom in the denominator. R^2_U is the unrestricted R^2 , R^2_R is the restricted R^2 ; RSS_U is the unrestricted residual sum of squares TSS_U is unrestricted total sum of squares, RSS_R is the restricted residual sum of squares and TSS_R is the restricted total sum of squares.

3.2 Empirical Results (MARS)

3.2.1 Equation #1 (Estimation of MARS Model: (RJP_t vs D/P))

$$RJP_t = 0.04237 - 13.03(DP_{t-1}-0.0007)_+ + 9.068(0.0007-DP_{t-1})_- - 7.978(DP_{t-3}-0.2051E-20) + 19.46(0.2051E-20-DP_{t-3})_+ - 1.856(RJP_{t-1}-0.1210)_+ - 0.410(0.1210-RJP_{t-1})_- + 558.6(RJP_{t-4}-0.04438)_+(DP_{t-1}-0.0007)_+ + 256.3(0.04438-RJP_{t-4})_+(DP_{t-1}-0.0007)_+ + 69.26(DP_{t-5}-0.0002)_+(0.1210-RJP_{t-1})_+ + 451.1(0.04994-RJP_{t-1})_+(DP_{t-3}-0.2051E-20)_+ + 265.2(RJP_{t-3}+0.02343)_+(DP_{t-3}-0.2051E-20)_+ - 381.0(RJP_{t-3}+0.05326)_+(DP_{t-1}-0.0007)_+ - 287.9(0.05326-RJP_{t-3})_+(DP_{t-1}-0.0007)_+ - 166.9(RJP_{t-1}-0.1589)_+(0.2051E-20-DP_{t-3})_+ + 357.4(0.01589-RJP_{t-1})_+(0.2051E-20-DP_{t-3})_- - 177.1(0.07872-RJP_{t-5})_+(0.2051E-20-DP_{t-3})_+ + 174.4(0.09222-RJP_{t-1})_+(0.0007-DP_{t-1})_+ - 5.154(-0.01127-RJP_{t-4})_+(0.1210-RJP_{t-1})_+ + 2926(-0.3881-RJP_{t-1})_+(DP_{t-4}-0.0001)_+ + 84.38(RJP_{t-4}+0.04492)_+(DP_{t-4}-0.0001)_+ + 424.0(-0.04492-RJP_{t-4})_+(DP_{t-4}-0.0001)_+ + \epsilon_t$$

$R_2 = .5038$ MGCV = .001469 SSR = .5116

$G = 3.364$ L = 1.930 $F_{5,438} = 45.67$

Table 3: Relative Importance of Each Variable

#	Variable	RI	ΔMGCV	#	Variable	RI	ΔMGCV
1	RJP _{t-1}	93.22	.001746	6	DP _{t-1}	100.0	.001787
2	RJP _{t-2}	0.0	.001469	7	DP _{t-2}	0.0	.001469
3	RJP _{t-3}	45.45	.001534	8	DP _{t-3}	78.18	.001663
4	RJP _{t-4}	57.80	.001575	9	DP _{t-4}	47.21	.001540
5	RJP _{t-5}	28.46	.001494	10	DP _{t-5}	9.047	.001471

Table 4: ANOVA Decomposition for 11 Basis Functions of the Model

Function	S.D.	ΔMGCV	# Basis	Variables
1	.01809	.001559	1	DP _{t-1}
2	.02535	.001546	2	DP _{t-3}
3	.01965	.001688	1	RJP _{t-1}
4	.01656	.001525		RJP _{t-4} , DP _{t-1}
5	.004843	.001471	1	RJP _{t-1} , DP _{t-5}
6	.02069	.001572	2	RJP _{t-1} , DP _{t-5}
7	.01177	.001520	1	RJP _{t-3} , DP _{t-3}
8	.01186	.001472	2	RJP _{t-3} , DP _{t-1}
9	.01125	.001494	1	RJP _{t-5} , DP _{t-3}
10	.007506	.001495	1	RJP _{t-1} , DP _{t-1}
11	.009998	.001537	1	RJP _{t-1} , RJP _{t-4}
12	.008197	.001517	1	RJP _{t-1} , DP _{t-4}
13	.008481	.001500	1	RJP _{t-4} , DP _{t-4}

Note: The relative importance estimate (RI) measures the importance of each variable used in the model. ΔMGCV is the modified GCV score a specific variable or function were removed from the model. The column variables indicates the variables used in the model. ANOVA decomposition measures the effect on the MGCV if the indicated function were removed from the model. The S.D. gives a measure of the relative importance of the function to the model. Footnotes 2, 3

3.2.2 Interpretation of Equation #1

Equation 1 shows the response structure of Japanese stock returns in relation to dividend yields. The MARS model improves non-Gaussianity and eliminates nonlinearity from 6.051 and 4.370 of the linear models to 3.364 and 1.930 respectively, shown on Table 1. The R^2 increases substantially from .2445 to .5038. The order of the importance, table 3, for variables in this model is DP_{t-1}(100.0), RJP_{t-1}(93.22), DP_{t-3}(78.18), RJP_{t-4}(57.80), DP_{t-4}(47.21), RJP_{t-3}(45.45), RJP_{t-5}(28.46), DP_{t-5}(9.407). Even though the most important variables are DP_{t-1} and RJP_{t-1}, variables DP_{t-5} and DP_{t-4}

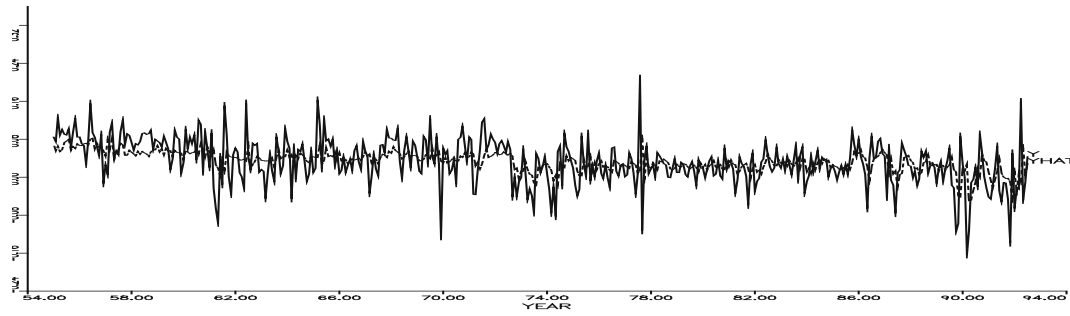
which are very important in the VAR model (not shown) are less important in the nonlinear model. This however, implies that the nonlinear model does not lose any information contained in the VAR model. The Δ MGCV suggests that if DP_{t-1} were removed from the model MGCV would increase to .001787 from .001494, if RJP_{t-1} were removed the MGCV would increase to .001746 from .001469, but if DP_{t-5} were removed the MGCV would increase to .001471 from .001540. This confirms the relative importance score of each variable. Variables that are more important weigh more on the nonlinear model.

Analyzing the MARS model we see more relationships between RJP_t and these variables, shown in table 4. The term $-13.03(DP_{t-1}-0.0007)_+$ shows that if $DP_{t-1} > 0.0007$ then RJP_t declines while, the term $+9.068(0.0007-DP_{t-1})_+$ indicates that if $DP_{t-1} < 0.0007$ then RJP_t increases. If these terms were removed the MGCV would increase to .001559. This suggests that for different threshold patterns DP_{t-1} affects RJP_t differently. The term $-7.978(DP_{t-3}-0.2051E-20)_+$ indicates that if $DP_{t-3} > 0.2051E-20$ RJP_t declines but, the term $19.46(0.2051E-20-DP_{t-3})_+$ shows that when $DP_{t-3} < 0.2051$ RJP_t rises. If the terms were removed the MGCV increases to .001546. This also implies that for certain thresholds the effect on RJP_t is different. In contrast to linear models which constrained these effects to be nonsignificant the MARS procedure allows us to detect these effects. The terms $558.6(RJP_{t-4}-0.04438)_+(DP_{t-1}-0.0007)_+$ and $256.3(0.04438-RJP_{t-4})_+(DP_{t-1}-0.0007)_+$ indicate that RJP_t continuously increase if DP_{t-1} remains above a threshold $DP_{t-1} > 0.0007$ irrelevant of whether RJP_{t-4} takes a value greater or smaller than 0.04438. If these terms were removed the MGCV would increase to .001656. This pattern explains a market overreaction (see Endnote 1) to good news about dividend yields (or a "price pressure effect") which can also explain temporary components in stock prices. Term $-381.0(RJP_{t-3}-0.05326)_+(DP_{t-1}-0.0007)_+$ and term $-287.9(0.05326-RJP_{t-3})_+(DP_{t-1}-0.0007)_+$ show a different effect. Here, we can identify an interaction effect between two related two way terms given the value of the one-month lag of dividend yields is greater than 0.0007. First, if the three-month lag of stock returns is greater than 0.0532 current stock returns (RJP_t) decline by -381.0 and if the four-month lag in stock returns is smaller than 0.0532 RJP_t declines by -287.9. ANOVA analysis indicates that if these terms were removed, the MGCV would increase to .001472. This result can be interpreted as showing that for certain threshold patterns, the market oversells bad news. The following three related two way interaction terms show a more complicated subtle effect. Term $2926(-0.03881-RJP_{t-1})_+(DP_{t-4}-0.0001)_+$ indicates that if $DP_{t-4} > 0.0001$ while $RJP_{t-1} < -0.03881$ current stock returns rise. Term $84.38(RJP_{t-4}+0.04492)_+(DP_{t-4}-0.0001)_+$ shows that if $DP_{t-4} > 0.0001$ and $RJP_{t-4} > -0.04492$ RJP_t increases more. Finally, term $424.0(-0.04492-RJP_{t-4})_+(DP_{t-4}-0.0001)_+$ shows that if $DP_{t-4} > 0.0001$ and $RJP_{t-4} < -0.04492$ RJP_t rises even more. If these terms were removed the MGCV would increase to .001525 for the first term or .001500 for the second and third. The economic interpretation of these terms is that the more constrained dividend yields are, the higher the valuation of firms by the market, therefore the higher the share prices.

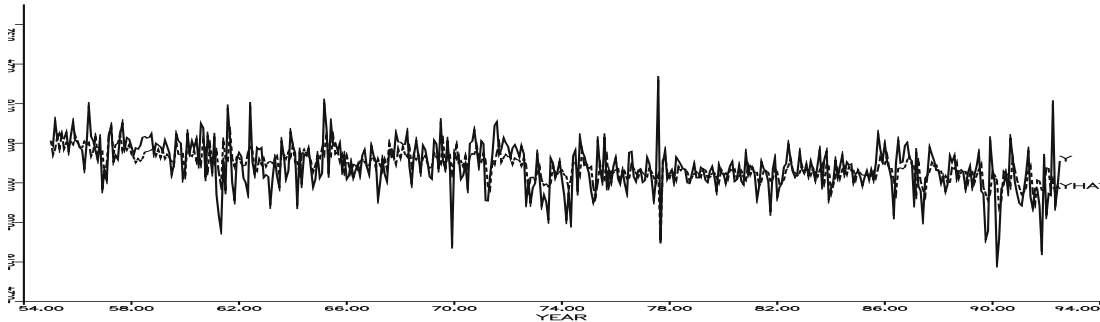
3.2.3 Comparison of the Residuals between VAR and MARS Models

The residuals of VAR and MARS models are plotted and compared in Figure 1, for model (RJP, D/P)' of the Nikkei-225, market. The figure consists of, and compares four graphs: Graph I, contains the plot of the actual (Y) and estimated real stock returns (YHAT) for an estimated VAR model. Graph II presents Y and YHAT for an estimated MARS model. In these graphs the vertical axis depicts the rate of stock return while the horizontal axis depicts time period. The dashed line is series Y, and the solid line is series YHAT. Graphs III and IV present the plots of residuals series, which is given as the difference between Y and YHAT, for the estimated VAR and MARS models. The vertical axis depicts the rate of returns for residuals, and the horizontal axis depicts time period.

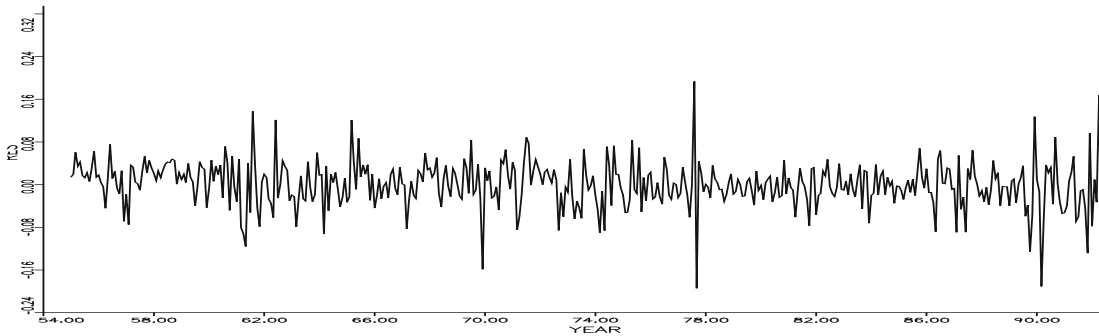
As shown, in Graphs I and II, the Y (dashed line) and YHAT (solid line) series for the VAR and MARS models fluctuate between -16 and 16, but the plot of YHAT series is smoother than the plot of Y. In Graphs III and IV the residuals of the VAR model fluctuate between -16 and 20 but, those of the MARS model are between -12 and 12. Therefore, the residuals of the MARS model are less fluctuant. In Graph II the series Y and YHAT of the model (RJ, D/P)' fluctuate within the range -40 to 55 with the YHAT series more fluctuate than the Y series. Note however, that in Graphs III and IV the residual series of the MARS model being between -17 to 16 is much less fluctuant than the residuals of the VAR model being in the range -20 to 25. Interpreting the remaining figures by the same way, we observe that the residual series for the MARS models are less fluctuant (smoother) than for those of the VAR models indicating that the MARS model improves the fitness of the data compared to VAR. The better performance of MARS versus VAR models may be supported by the success of the residuals of MARS models, to reduce substantially or eliminate nonlinearities.



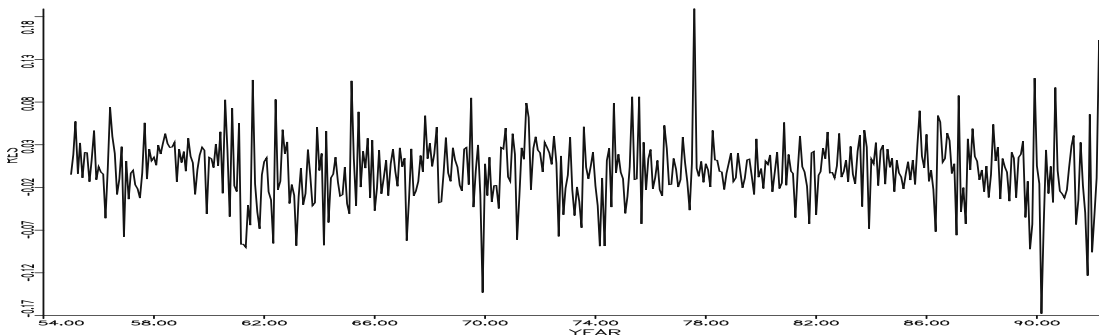
Graph I: Plot of Y and YHAT for VAR for Japanese Market



Graph II: Plot Y and YHAT of MARS for Japanese Market



Graph III: Plot Residuals of VAR for Japanese Market



Graph IV: Plot of Residuals of MARS Model for Japanese Market

Figure 1: Comparison of VAR and MARS Models

3.2.4 *Over-Fitting Linear and Non-Linear Models*

Although the MARS model gives substantially better within sample fits than the VAR model (see the graphs of Y, YHAT and Residuals that compare the goodness of fit between VAR and MARS models) there is always concern of over-fitting these models. As Stokes and Neuburger (1996) have stated: “by over-fitting, we mean modeling noise as if it were part of the structure of the model”. To test the fitness of the models we generate out-of sample forecasts for both the VAR and MARS models. For the VAR models we generate 48 in and out of sample forecasts (see Endnotes 2, 3). Table 5 lists and compares the out-of-sample forecasts of the VAR and MARS models. As shown in the table the correlations between the forecasted and actual values of stock returns vary by models. For instance, in the case of model (RJP,D/P)' the correlation values of the VAR model is .1771 while the correlation of the MARS model for the same variables is .2378 with variance of forecast error .00157 and .00101, respectively. This indicates that the VAR significantly outperforms the MARS model. Examining Graphs III and IV of Figure 1 where the residuals of the VAR and MARS models are presented, we can see that the MARS model fits the data better than the VAR in the in sample forecasts. Looking at Graphs III and IV on page 7, we see that MARS outperforms the VAR in both in and out-of-sample tests. The superiority of the MARS model over the VAR for the in-sample forecasts is not confirmed by the out-of sample tests. This suggests that some of what appears to be structure is actually noise that has been included in the model as a result of over-fitting.

Table 5: Out-Of-Sample Forecasts of VAR and MARS Models

Validation Of Out-Of-Sample Forecasts of VAR and MARS Models Variables	Correlation (Between Forecasted And Actual Returns)		Variance of Forecast Error	
	VAR	MARS	VAR	MARS
(RJP,DG)'	.3374	.04548	.00399	.01142
(RJP,DP=JPD/JPS)'	.1771	.2378	.00157	.00101

Note: We examine the first equation of the two-variable (basic) VAR models for each country of the form: $RS_{it} = \alpha_0 + \sum_{j=1}^n a_j RS_{it-j} + \sum_{j=1}^n \beta_j X_{it-j} + \varepsilon_t$, where RS_{it} is real stock returns of each country at time t, and X_{it-j} is dividend growth rate (DG) or the dividend-price ratio (D/P, or LD/P) in each country at time t-j. "Significance" is the joint significance of all the coefficients other than the constant. Note: comments regarding model (RJP, DG)', are available upon request from the author.

4. THE ECONOMIC IMPORTANCE OF MARS MODELS' RESULTS

The application of the MARS to model predictability of returns is statistically significant, but what is its economic significance? Why do we observe this behavior in stock markets? We begin by referring to bubble models of Diba and Grossman (1987), (1988) where they suggest that their bubble tests should allow for nonlinearities. Or the rational bubble model of Blachard and Watson (1982) which allows the bubble to grow by the exact amount needed to compensate investors that the price will crash and the stock price will revert to the small initial bubble. The rational bubble model of Blachard and Watson allows for the unexpected price changes from two sources: a) unexpected changes in the bubble and b) unexpected changes in the fundamental value¹. The overreaction hypothesis (Endnote 1) can also be considered to explain the results of the nonlinear models. De Bondt and Thaler (1985) have shown that if prices systematically overshoot, then their reversal should be predictable based on past information alone. They observed that extreme stock price movements are followed by subsequent price movements in the opposite direction and that the more extreme the initial price movement, the greater will be the subsequent adjustment. Other causes of nonlinearities might include the possibility that the effects of shocks in financial and economic data (business cycles or periods of high and low volatility) accumulate until the process explodes (self-exciting or catastrophic behavior Hamilton (1989) and Clements (2004)). MARS analysis also shows that, stock market indexes are characterized by "runs" such as in McQueen and Thorley (1994) that assuming nonlinearities are predictable. Also "price pressure effects" may be able to cause the behavior of stock prices that we observed in the previous sections.

ENDNOTES

1. See McQueen and Thorley (1994) for more on speculative bubbles, predictability of stock returns and nonlinearities.

¹ The forecasts were generated by using the FORECAST comment of the B34S statistical package.

2. I only test a representative model for the market. Results and interpretation of results for models (RJP, DG and RJP, LD/P) are available upon request from the author of this paper.
3. The forecasts were generated by using the FORECAST comment of the B34S statistical package.

AUTHOR INFORMATION

Dimitrios Tsoukalas is a professor of Financial Economics/Time Series Analysis, at Purdue University-Calumet. He teaches graduate and undergraduate courses in Business Statistics, Econometrics and Money Banking and Financial Markets. Professor Tsoukalas has published numerous articles in refereed journals with high rate of acceptance. His interests are national and international stock and bonds markets, in developed and emerging economies. E-mail: tsoukala@purduecal.edu

REFERENCES

1. Abhyankar, Copeland, and Wong (1997), Uncovering Nonlinear Structure In Real Time Stock-Market Indexes: The S&P-500, the DAX, the NIKKEI-225, and the FTSE-100, *Journal of Business and Economic Statistics*, January 1997, Vol.15, No.1.
2. Arouri M., and Fouquau J. (2009), On the Short-term Influence of Oil Price Changes on Stock Markets in GCC Countries: Linear and Nonlinear Analyses, *Economics Bulletin*, Vol. 29 no.2 pp795-804.
3. Barnett, and Chen (1988), The Aggregation- Theoretic Monetary Aggregates Are Chaotic and Have Strange Attractors, in *Dynamic Econometric Modeling*, eds. W. Barnett, E. 711-728.Berndt, and H. White, Cambridge U.K.: University Press.
4. Blachard, and Watson (1982), Bubbles, Rational Expectations, and Financial Markets. In *Crisis in the Economic and Financial System*, P. Wachtel, ed. Lexington, M.A.: Lexington Books (1982).
5. Blank (1991), Chaos in Futures Markets? A Nonlinear Theoretic Monetary Aggregates Are Chaotic and Have Strange Dynamic Analysis, *The Journal of Futures Markets*, Vol. 11, No. 6.
6. Brieman (1991), Response. *Technometrics*, 33, no.2. (May 1991b): 155-160.
7. Brieman (1991), Discussion. *The Annals of Statistics*, 19, no. 1(1991c): 82-91.
8. Brock, and Potter (1993), Nonlinear Time Series and Macroeconometrics, in *Handbook of Statistics* (Vol.11), eds. G. S. Maddala, C. R. Rao, and H. D. Vinod, New York:North-Holland, pp 195-229.
9. Clements P. M., Hans Frances P., and Swanson R. N. (2004), Forecasting Economic and Financial Time-Series with Nonlinear Models, *International Journal of Forecasting* 20, (2004) pp. 169-183.
10. DeCoster, and Mitchell (1991), Nonlinear Monetary Dynamics, *Journal of Business and Economic Statistics*, 9, 455-461. Corlay Sylvian, "B-Spline techniques for Volatility Modeling", arXvi: 1306.0995v2 [q-fin.CP], 4 Jul 2013.
11. DeBondt, and Thaler (1985). Does the Stock Market Overreact? *Journal of Finance* 40: 793-805.
12. Diba, and Grossman (1987), On the Inception of Rational Bubbles. *Quarterly Journal of Economics*, 102, (Aug. 1987):697-700.
13. Granger (1969), Investigating casual relations by Econometric Models and Cross-Spectral Methods. *Econometrica* Vol. 37, No.3 (July 1969).
14. Hinich (1982), Testing for Gaussianity and Linearity of Stationary Time Series. *Journal of Time Series Analysis* 3 no.5 (1982): 169-176.
15. Hinich, and Patterson (1985), Evidence of Non-linearity in Daily Stock Returns. *Journal of Business and Economic Statistics* 3, vol.1 (January 1985):69-77.
16. Hsieh (1989), Testing for Nonlinear Dependence in Daily Foreign Exchange Rates, *Journal of Business*, 62, 339-368.
17. Hsieh (1991), Chaos and Nonlinear Dynamics: Application to Financial Markets, *Journal of Finance*, 46, 1839-1877.
18. Islam M., Lasker Ershad A., and Nahida A. (2012), Forecasting Volatility of Dhaka Stock Exchange: Linear Vs Non-linear Models. *International Journal of Science and Engineering*, Vol. 3(2): 4-8, October 2012, Masudul Islam et al.
19. Jaditz and Sayers (1998), Out -of-Sample Forecast Performance as a Test for Nonlinearity in Time Series, *Journal of Business and Economic Statistics*, Vol. 16, No. 1.

20. Just and Rauser (1981), Commodity Price Forecasting with Large Scale Econometric Models and the Futures Market, *American Journal of Agricultural Economics*, 63: 197-208.
21. McMilan D. G, and Wohar M. (2008), Stock Return Predictability and the Dividend-Price Ratio: A Nonlinear Approach. Draft (University of St. Andrews and University of Nebraska, September, 2008).
22. McQueen and Thorley (1994). Bubbles, Stock Returns, and Duration Dependence. *Journal of Financial and Quantitative Analysis* vol. 29, no.3, (September 1994):379-401.
23. Neuburger and Stokes. (1991), Testing the Appropriateness of Statistical Methods. *Financial Analysts Journal* (July/August 1991): 83-88.
24. Scheikman, and LeBaron (1989a), Nonlinear Dynamics and Stock Returns, *Journal of Business*, 62, 311-337.
25. Seyyed Ali Paytakhti Oskooe (2012), Emerging Markets Efficiency: Nonlinearity and Episodic Dependences Evidence from Iran Stock Market, *Journal Basic. Appl. Sci. Res*, 2(11) 11370-11380, 2012.