Comput Stat (2011) 26:77–93
DOI 10.1007/s00180-010-0211-7

ORIGINAL PAPER

Detecting and modeling nonlinearity in the gas furnace data

Houston H. Stokes · Melvin Hinich

Received: 23 May 2007 / Accepted: 16 July 2010 / Published online: 29 July 2010

Abstract Using a modification of the Hinich, J Time Ser Anal 3(3):169–176, (1982) bispectrum test for nonlinearity and Gaussianity, the residuals of the Tiao and Box, J Am Stat Assoc 76:802–816, (1981) constrained and unconstrained VAR models for the gas furnace data reject the assumption of Gaussianity and linearity over a grid of bandwidths for estimating the bispectrum. These findings call into question the specification of the linear VAR and VARMA models assumed by Tiao and Box, J Am Stat Assoc 76:802–816, (1981). Utilizing the alternative Hinich J Nonparametr Stat 6:205–221, (1996) nonlinearity test, the residuals of the VAR model were shown to exhibit episodic nonlinearity. The sensitivity of the findings to outliers is investigated by estimating and testing the residuals of L1 and MINIMAX models from 1–6 lags. Building on the linear dynamic specification, a multivariate adaptive regression splines (MARS) model is estimated, using two software implementations, and shown to remove the nonlinearity in the residuals. Leverage plots were used to illustrate the "cost" of imposing a linearity assumption. Out-of-sample forecasting tests from 1–6 periods ahead found that using the sum-of-squared errors criteria, the MARS model out performed ACE, GAM and projection pursuit models.

Keywords Bispectrum · VAR · Nonlinear · Episodic nonlinearity · Gaussian · Skewness function · MARS · L1 estimation · MINIMAX estimation · Leverage plot · ACE · GAM · Projection pursuit

H. H. Stokes (✉)
Department of Economics, University of Illinois at Chicago,
601 South Morgan Street, Chicago, IL 60680, USA
e-mail: hhstokes@uic.edu

M. Hinich
Applied Research Laboratories, The University of Texas at Austin,
Austin, TX 78713, USA
e-mail: hinich@austin.utexas.edu

1 Introduction

Extensions of the Hinich (1982, 1996) nonlinearity tests are used to test the residuals of the classic gas furnace data studied most recently by Box et al. (2008). This data set had been selected by Tiao and Box (1981) when they illustrated a multistep identification strategy for identifying a vector autoregression moving average (VARMA) model from an assumed class of linear models. Stokes (1997) reported that the residuals of the constrained and unconstrained VAR model for the gas data suggested by Tiao and Box (1981) failed the Hinich (1982) Gaussianity and linearity tests. Although multiple nonlinear models that could be linearized were tried, the nonlinearity persisted in the residuals. The present paper first shows that the nonlinearity in the residuals is episodic, as measured by the Hinich (1996) test. The multivariate adaptive regression splines (MARS) model, first proposed by Friedman (1991), is shown to remove the nonlinearities in the gas furnace data model.

After a brief discussion of the Hinich bispectrum test, the gas furnace residuals are tested. Next, the model is estimated, using L1 and MINIMAX procedures to determine how sensitive the results are to outliers. An alternative Hinich (1996) nonlinearity test is used to determine if the measured nonlinearities are episodic. A MARS model was estimated for both the gas input series (GASIN) and the gas output series (GASOUT) that produce residuals that pass the Hinich (1982) nonlinearity test. Leverage plots were used to illustrate the "cost" of imposing a linearity assumption. Validation tests using out-of-sample forecasting tests from 1–6 periods ahead found that using the sum-of-squared errors criteria, the MARS model out performed ACE, GAM and projection pursuit models.

2 VAR model setup and overview of the Hinich bispectrum test

In an influential paper on the identification of VAR and VARMA models, Tiao and Box (1981) used the gas furnace data as an example. Their model relating the gas input (GASIN) and CO_2 concentration (GASOUT) was

$$A(B)Z_t = e_t \tag{1}$$

where Z_t' is a row vector of the tth observation of the two series, e_t' is a row vector for period t of the estimated error vector of the model for the two series, and $A(B)$ is the k by k autoregressive VAR matrix. Each element in $A(B)$ is a polynomial in the lag operator B, which maps Z_t into Z_{t-1}. Successful estimation of Eq. (1) assumes that the roots of the determinate $|A(B)|$ are outside the unit circle (the invertibility condition), the expected value of the error vectors is zero and the error vectors are pure white noise. Tiao and Box (1981) only tested the error terms for significant autocorrelations and cross correlations. No attempt was made to test whether the linear specification was appropriate.

Assuming linearity, Tiao and Box (1981) determined that an unconstrained model of order 6 would clean the residuals of any measurable autocorrelation. Next, they removed statistically insignificant VAR coefficients and estimated the constrained model, using conditional least squares. We have replicated both their constrained and unconstrained models and their results are reported in Table 1. We next applied the

Table 1 Estimated coefficients for gas furnace data

Lag	Unconstrained model				Constrained model			
	$A_{1,1}$	$A_{1,2}$	$A_{2,1}$	$A_{2,2}$	$A_{1,1}$	$A_{1,2}$	$A_{2,1}$	$A_{2,2}$
1	1.93 (0.0581)	−0.0508 (0.0457)	0.0632 (0.0743)	1.55 (0.0585)	1.982 (0.055)			1.522 (0.0571)
2	−1.20 (0.126)	0.0999 (0.0843)	−0.133 (0.161)	−0.593 (0.108)	−1.387 (0.0998)			−0.568 (0.1063)
3	0.17 (0.144)	−0.0796 (0.0881)	−0.441 (0.184)	−0.171 (0.113)	0.349 (0.0551)		−0.530 (0.0741)	−0.159 (0.0997)
4	−0.16 (0.145)	0.0269 (0.0877)	0.152 (0.186)	0.132 (0.112)			0.1180 (0.1631)	0.1312 (0.0431)
5	0.38 (0.137)	−0.0414 (0.0771)	−0.120 (0.175)	0.0569 (0.0985)			−0.0451 (0.1734)	
6	−0.214 (0.0839)	0.0305 (0.0328)	0.249 (0.107)	−0.0421 (0.0419)			0.2091 (0.1056)	
S	0.03408				0.03593			
	0.00229	0.0557			−0.00290	0.056143		

Standard errors are listed parenthetically. Constants were estimated for the constrained problem and were −0.004138 and 3.9992, respectively, with standard errors 0.01115 and 0.8335 for Eq. 1 and 2, respectively. The constrained and unconstrained AR models were selected by Tiao and Box (1981). S = residual error covariance matrix

Hinich (1982) test to the residuals of these equations to test for nonlinearity. The Hinich (1982) test uses an estimate of the bispectrum as an average value of a square of M^2 points. For a further discussion of properties of the bispectrum, see Hinich and Clay (1968); Subba Rao and Gabr (1984); Priestley (1988) and Hinich and Messer (1995). The larger (smaller) M, the smaller (larger) the finite sample variance and the larger (smaller) the sample bias. Because of this tradeoff, there is no one unique M that is appropriate to use for performing nonlinearity and Gaussianity tests. Hinich (1982) has suggested that a good value for M is $\sqrt{N}$. A lower M would be $\sqrt{N/3}$.

In the models fit to the gas furnace data, where $N = 296$ for the complete sample, we have reported M values from 9 to 18, and averaged the result to insure that our findings are not sensitive to the M value selected. To test the null hypothesis that the error term is Gaussian, Hinich suggests the G statistic, which is normally distributed. To test whether the series is linear, Hinich suggests the normally distributed L statistic. For details of this test, see Hinich (1982); Hinich and Patterson (1985) and Ashley et al. (1986).

In this paper we report normal approximations of the Gaussianity and linearity tests. Mean values for G and L over all values of M are also reported. The Hinich critical values have been used to detect the presence of nonlinearity, although Lee (2001) in research involving simulations suggests that these values are overly conservative. Hinich et al. (2005) as well as Patterson and Ashley (2000) confirmed this finding. Hence, if the mean G and L values fail the Hinich test using the original Hinich critical values, it is more likely that the process is nonlinear.

Ashley et al. (1986, p.,174) presented an equivalence theorem which proved that the Hinich bispectral linearity test statistic is invariant to linear filtering of the data. This important result proves that the linearity test can be either applied to the raw series or to the residuals of a linear model. An additional important implication of the theorem is that if $X(t)$ is found to be nonlinear, then the residuals of a linear model of the form $y_t = f(X(t))$ will be nonlinear, since the nonlinearity in $X(t)$ will pass through any linear filter.

The above paper also reported tables on the power of the Hinich linearity test for detecting violations of the linearity assumption for a variety of common nonlinear models appearing in the literature and a table of the power of the linearity and Gaussianity tests for a number of sample sizes and M values. Their findings indicate substantial power for both tests, even when N is a small value, such as 256, if the value of M used is between 12 and 17. For this sample size, as M increases, the power of the test falls off. This is later illustrated in our test results.

Hinich (1996) proposed another testing strategy that could be applied to two series within the sample based on the sample cross-correlation at lags r and s, $C_{xy}(r,s)$ and the sample cross bicorrelation, $C_{xxy}(r,s)$. This test required whitened series and is suitable for the residuals of a model but not the raw series. Define $m = \max(r,s)$

$$C_{xxy}(r,s) = \frac{\sum_{t=1}^{N-m}(x_t x_{t+r} y_{t+s})}{N-m} \tag{1}$$

Let $L = N^c$, where $0 < c < 0.5$. Test statistics for non-zero cross correlations and cross bicorrelations are

$$H_{xy}(N) = \sum_{r=1}^{L} (N - r) C_{xy}^2(r) \tag{2}$$

$$H_{xxy}(N) = \sum_{s=-L}^{L} \sum_{r=1}^{L} (N - m) C_{xxy}^2(r, s), \tag{3}$$

where $s \neq 0$. $H_{xy}(N)$ and $H_{xxy}(N)$ are asymptotically Chi-squared with $L(2L + 1)$ degrees of freedom but for the purposes of this paper have been transformed to $U(0,1)$ under the null. The Hinich (1996) test has a number of advantages that include being able to test for nonlinearity mapping from one series to another and being relatively quick to calculate. This latter advantage allows the test to be performed within the sample to test for episodic nonlinearity. In the empirical section the probabilities of $H_{xy}(N)$ and $H_{xxy}(N)$ are given. For testing one series, say x, Hinich (1996, eq.,3.1) recommends using a variant of (1).

Define $G(r, s)$ as the r,s sample bicorrelation multiplied by $\frac{1}{\sqrt{(N-s)}}$ to standardize its variance. The statistic H_x, which is defined as $H_x = N^{-c} \sum_{s=2}^{L} \sum_{r=1}^{s-1} [G^2(r, s) - 1]$ where $0 < c < 0.5$ is $N(0,1)$, can be converted to a probability of rejection of the assumption of linearity.

3 Results using VAR models

Table 1 replicates the estimated coefficients of the Tiao and Box (1981) gas furnace data example. In Table 2 the Hinich bispectrum tests are performed on a grid of M values, going from 9 to 18, to test the residuals from the two equations implicit in both the unconstrained and constrained VAR(6) models listed in Table 1 for both the Gaussianity (G test) and the linearity (L test). The G test values are all above 4.99 for GASIN and 10.92 for GASOUT, indicating that both residual series, for both constrained and unconstrained models, reject the assumption of Gaussianity (G test) at a very high level of significance. For virtually all values of M, the assumption of linearity is also rejected by the L test. The lower L scores were found only with the higher M scores (17 and 18), which have a large bandwidth. For purposes of comparison, $e'e$ for the GASIN and GASOUT residuals was 9.8847 and 16.139, respectively.

As was mentioned earlier, Ashley et al. (1986, Tables 1 and 2) investigated the size (number of observations) needed for the Hinich linearity and Gaussianity tests and the power of such tests for various values of M and number of observations. Their findings indicate that for this example both G and L tests will give satisfactory convergence and that both tests detect nonlinearity with considerable frequency, even in cases in which $N = 256$. These simulation results suggest that it is appropriate to use the Hinich tests in the present case in which $N = 290$.

Table 2 documents that our findings of nonlinearity are invariant as to whether the estimated form of Eq. (1) is unconstrained or constrained. In results not reported, we experimented with increasing the lag length of the unconstrained VAR model from 6 to 12. The findings were similar. We conclude that even though the distribution of the Hinich tests is known only asymptotically, the magnitude of the Z scores at a high

Table 2 Z scores for Gaussianity and linearity tests for unconstrained and constrained 6th order VAR model of the GAS data

Equation 1					Equation 2			
Unconstrained			Constrained		Unconstrained		Constrained	
M	G	L	G	L	G	L	G	L
9	10.86	5.80	11.76	7.75	11.27	5.53	11.64	5.81
10	12.07	6.30	12.53	5.91	11.86	4.52	12.05	6.34
11	7.05	6.77	8.62	7.30	11.24	5.29	11.35	6.00
12	12.75	3.08	12.77	7.01	12.22	6.27	12.21	4.87
13	5.99	2.63	6.94	2.74	11.19	4.00	11.41	5.81
14	7.51	1.45	8.07	1.37	10.92	8.45	11.36	4.75
15	4.99	4.04	4.98	1.98	11.45	3.27	11.58	6.21
16	6.47	3.40	6.95	8.19	12.91	3.46	12.91	2.49
17	7.63	1.11	9.30	6.59	12.05	0.17	12.39	0.99
18	6.48	.60	6.90	0.63	12.46	4.23	12.67	4.06
Mean	8.18	3.52	8.88	4.95	11.76	4.52	11.96	4.73

G = Z score for normal approximation for Gaussianity test. L= Z score for linearity test. M = Square root of the number of terms used to estimate the bispectrum at the center of the square. The number of residuals was 290. Equation 1 is for the gas furnace input data. Equation 2 is for the gas furnace output data. Coefficients for the unconstrained (Model 1) and constrained (Model 2) are given in Table 1. Estimated coefficients are consistent with those of Tiao and Box (1981). For further detail on Tables 1 and 2, see Stokes (1997, Chapter 8)

confidence level indicates that both the input series and the output series fail the null hypotheses of Gaussianity and linearity. In an attempt to remove the indications of nonlinearity in the test statistics, an exhaustive search of alternative nonlinear models that could be linearized was attempted without success. To investigate the possibility that outliers were giving the illusion of nonlinearity, L1 and MINIMAX models were tried. L1 models minimize $\sum |e_t|$ and are less sensitive to outliers than OLS models. MINIMAX models minimize max $|e_t|$ and are more sensitive to outliers than OLS models. Using all three estimation methods, six alternative models of GASOUT of the form

$$GASOUT_t = \alpha + \beta_0 + \sum_{i=0}^{k} GASIN_i + \sum_{j=1}^{k} GASOUT_j \tag{4}$$

were tried for lags of $k = \{1, \ldots, 6\}$ and the results are reported in Table 3. Note that in this model $GASIN_t$ was included. Gaussianity is rejected for lags 1–6 for all estimators. Linearity is rejected for lags 2–6 for L1 and OLS. The MINIMAX model rejects linearity for models with 3 lags and 4 lags. In the MINIMAX model, the cost of reducing the maximum $|e_t|$ is reflected by $e'e$ values that are two or more times bigger than their OLS and L1 counterparts. MINIMAX models with 1, 2, 5 and 6 lags have no indication of nonlinearity, but have $e'e$ values that are in general more than two times their OLS counterparts, (32.36 verses 16.44 for lag 5 and 30.41 verses 16.09

Table 3 Tests on the residual of alternative estimators of a modified linear model

Lag		L1			OLS			MINIMAX	
		G	L		G	L		G	L
1		17.11	−0.57		17.19	−0.1		8.16	−3.11
2		21.63	10.11		17.15	8.89		18.22	1.56
3		26.95	7.31		16.15	4.26		36.48	2.33
4		16.41	6.61		13.61	5.17		7.96	2.89
5		15.68	7.35		11.54	4.69		9.92	0.28
6		14.56	5.10		11.51	4.39		17.03	0.13
	$e'e$	max|e|	Σ|e|	$e'e$	max|e|	Σ|e|	$e'e$	max|e|	Σ|e|
1	76.84	2.49	111.90	76.51	2.47	112.3	160.3	1.73	178.5
2	20.25	1.53	51.49	190.60	1.54	52.55	52.69	0.91	105.3
3	19.43	1.47	50.23	18.02	1.50	51.32	48.59	0.84	104.3
4	17.26	1.54	47.83	16.71	1.49	48.76	37.61	0.79	85.09
5	17.57	1.55	47.53	16.44	1.40	48.89	32.36	0.73	77.66
6	16.82	1.53	47.06	16.09	1.41	48.28	30.41	0.73	74.78

for lag 6). Since L1 models are less sensitive to outliers than OLS models, the poor performance of L1 in removing the nonlinearity suggests that outliers are not tripping the Hinich (1982) test.

The next experiment was to test whether the measured nonlinearity varies over time. Table 4 uses the Hinich (1996) test to investigate the within-sample properties of the residuals of the VAR(6) model of the gas furnace data series given in Table 1. Two window sizes of 20 and 30 were used. For notational simplicity, define $x =$ the residuals of GASIN and $y =$ the residuals of GASOUT. H_x and H_y measure the probability of nonlinearity remaining in the residuals of the GASIN and GASOUT models, respectively. H_{xy} and H_{yx} measure the probability of there being a nonlinear relationship between the residuals of GASIN to the residuals of GASOUT or the residuals of GASOUT to the residuals of GASIN, respectively. P_x, P_y and P_{xy} measure the probability of autocorrelation in the GASIN residuals, the GASOUT residuals and between the GASIN and GASOUT residuals, respectively.

Inspection of H_x and H_y for both windows shows periods of nonlinearity in each series. Using a window size of 20, H_x was 0.92 and 0.90 for windows 1 and 2, respectively, which included observations 1–40. Using a window size of 30, H_x was 0.94 in window 1, covering observations 1–30. For the 20-observations window, H_x and H_y were, 0.94 and 0.91 respectively in window 6 that covered observations 101–120. For the same window H_{xy} and H_{yx} were 0.9995 and 0.9837, respectively. Using the 30-observations window, H_x and H_{xy} had values of 0.98 and 0.94, respectively, for window 4, covering observation 91–120, indicating that there was a relationship between the nonlinearity in the GASIN series residual and GASOUT series residual. The autocorrelations and cross correlations of the residual series for the complete sample are flat.

Table 4 Episodic nonlinearity tests on residuals of VAR(6) model of the gas furnace data

Window	Obs begin	Obs end	H_x	H_y	H_{xy}	H_{yx}	P_x	P_y	P_{xy}
20-Observations window									
1	1.000	20.00	0.9161	0.2585	0.8313	0.1657	0.9319	0.0298	0.8673
2	21.00	40.00	0.9023	0.2420	0.9977	0.2821	0.6265	0.5210	0.7833
3	41.00	60.00	0.5499	0.1330	0.9882	0.9550	0.4752	0.0829	0.7788
4	61.00	80.00	0.3074	0.3405	0.4314	0.7664	0.0388	0.9000	0.7628
5	81.00	100.0	0.6689	0.5759	0.1725	0.8941	0.5412	0.9743	0.4024
6	101.0	120.0	0.9445	0.9145	0.9995	0.9837	0.4257	0.4402	0.9271
7	121.0	140.0	0.1372	0.1906	0.03957	0.0240	0.6553	0.9899	0.1687
8	141.0	160.0	0.7008	0.5053	0.4304	0.0192	0.5055	0.7427	0.0988
9	161.0	180.0	0.5193	0.8839	0.4963	0.8676	0.9271	0.9729	0.7095
10	181.0	200.0	0.7077	0.9568	0.1443	0.8591	0.8089	0.7413	0.7274
11	201.0	220.0	0.2225	0.3257	0.5370	0.9212	0.8503	0.9087	0.6343
12	221.0	240.0	0.0353	0.4755	0.0503	0.2564	0.4339	0.3199	0.9025
13	241.0	260.0	0.0461	0.1492	0.7665	0.8273	0.9695	0.1152	0.9379
14	261.0	290.0	0.3780	0.4675	0.0046	0.1469	0.1768	0.9986	0.5782
30-Observations window									
1	1.000	30.00	0.9401	0.5035	0.5274	0.5016	0.8710	0.1031	0.9428
2	31.00	60.00	0.1143	0.2178	0.1715	0.8537	0.8200	0.1457	0.9589
3	61.00	90.00	0.6934	0.5751	0.3824	0.7185	0.1962	0.9526	0.1077
4	91.00	120.0	0.9829	0.8760	0.9427	0.7067	0.2149	0.7155	0.8948
5	121.0	150.0	0.4982	0.3916	0.2873	0.5187	0.6239	0.9851	0.1880
6	151.0	180.0	0.1118	0.9450	0.5312	0.7611	0.8083	0.9856	0.4810
7	181.0	210.0	0.9247	0.9925	0.4945	0.9901	0.8557	0.8896	0.8737
8	211.0	240.0	0.2800	0.5237	0.2165	0.2548	0.4533	0.4997	0.9695
9	241.0	290.0	0.6338	0.8891	0.9999	1.000	0.9579	0.8227	0.9774

H_x and H_y measure the probability on nonlinearity in x and y, respectively. H_{xy} measures the probability of nonlinearity in x being reflected in y, while H_{yx} measures the probability of nonlinearity in y being reflected in x. P_x and P_y measure the probability of autocorrelation in x and y,respectively, while P_{xy} measures the probability of cross correlation between x and y

For the subsamples, significant values show up. For the 20-observations window, P_x is significant for windows 1 (0.93), 9 (0.93) and 13 (0.97). P_y is significant for windows 5 (0.97), 7 (0.99) and 14 (0.99). For the 30-observations window, P_x was significant for window 9 (0.96) and P_y was significant for window 3 (0.95), 5 (0.99), and 6 (0.99). P_{xy} was significant for windows 2 (0.96), 8 (0.97) and 9 (0.98). The results of the sub-sample estimation suggest that there are episodic periods of both nonlinearity and linear memory in the model. These results suggest that it might be promising to attempt a modeling strategy that includes a quite general class of threshold models to remove the nonlinearity. The technique chosen was MARS, a general data-driven, nonparametric approach, which has had success in other recent applications. These results are discussed next.

4 The MARS approach

The results in the preceding section indicate that the residuals in both the constrained and unconstrained models of the gas furnace data suggested by Tiao and Box (1981) fail the Hinich (1982) tests for nonlinearity and Gaussianity. The residuals of the unconstrained VAR model for GASIN and GASOUT are graphed in the tops of Figs. 1 and 2, respectively. Looking first at the GASOUT residual series, shown on the top of Fig. 2, note the relatively homogenous pattern for the first 60% of the series, with larger spikes at observation 193 of 0.975, observation 230 of −0.7526 and observation 259 of 1.431. The GASIN residual series graphed in the top of Fig. 1 shows outliers at observation 37 of 1.016 and observation 49 of −0.8851. The task is to select an estimation method having VAR as a special case, but allowing possible level-specific function changes, including interactions, that might be able both to reduce the variance of the residual and model some of these outliers while preserving the property that the sum of the residuals is constrained to be 0.0.

The MARS approach, first proposed by Friedman (1991) and used successfully by Lewis and Stevens (1991) in a time series context was shown later by Lewis and Ray (1997, p., 883) to generalize the Tong (1990) threshold autoregression (TR) model to fit nonlinear threshold models that are continuous in the domain of the predictor variables and allow for interactions among lagged predictor variables. Using the MARS method of estimation, it was thus possible to have lagged predictor variable thresholds, thus overcoming the limitations of Tong's approach. Lewis and Ray (1997) called

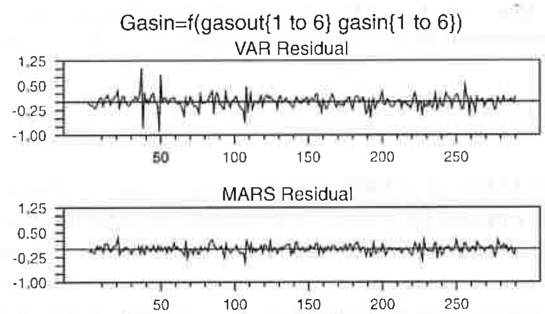

Fig. 1 GASIN residuals for VAR and MARS models

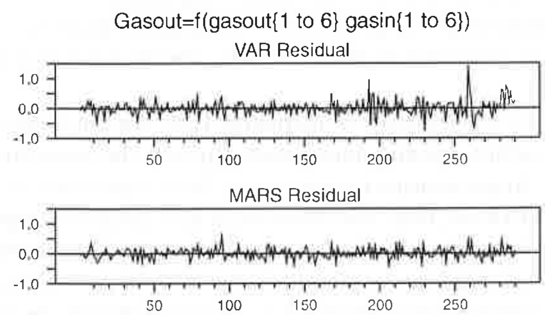

Fig. 2 GASOUT residuals for VAR and MARS models

Table 5 MARS and MARSPLINE models of GASOUT in the gas furnace data

Model	G	L	$e'e$
OLS 6th order VAR	11.756	4.518	16.14
Max interaction 1			
MARS model	12.000	4.463	15.00 15.73
MARSPLINE Model $d = 2$	8.15	5.43	12.28
MARSPLINE Model $d = 3$	7.66	4.20	14.09
Max interaction 2			
MARS model	7.812	3.925	11.31 13.76
MARSPLINE Model $d = 2$	0.93	0.72	6.35
MARSPLINE Model $d = 3$	−0.05	0.22	9.14
Max interaction 3			
MARS model	−0.736	−0.209	7.11 11.61
MARSPLINE Model $d = 2$	−0.86	−0.31	5.38
MARSPLINE Model $d = 3$	−0.50	0.12	6.02
Max interaction 4			
MARS model	−0.637	−1.04	7.14 12.66
MARSPLINE Model $d = 2$	1.56	1.18	5.00
MARSPLINE Model $d = 3$	2.27	1.84	5.90

For MARS™ models the first $e'e$ is the piecewise-linear approximation and the second $e'e$ is the piecewise-cubic approximation. For MARSPLINE models the penalty is set as 2 and 3, respectively. For all models the upper limit on the knots is 80. The GASOUT series is the dependent variable in all cases. All models were estimated with independent variables GASIN_{t-1}, GASIN_{t-2}, GASIN_{t-3}, GASIN_{t-4}, GASIN_{t-5}, GASIN_{t-6}, GASOUT_{t-1}, GASOUT_{t-2}, GASOUT_{t-3}, GASOUT_{t-4}, GASOUT_{t-5} and GASOUT_{t-6}

their approach TSMARS, or MARS related to time series. In related work, Chen and Tsay (1993a) used arranged local regression models to model the chickenpox data, the sunspot data and a simulated series. Chen and Tsay (1993b) used a nonlinear additive autoregressive model with exogenous variables for nonlinear time series model fitting.

The MARS approach can be thought of as a generalization of this approach in that higher order interaction terms are allowed. Some years after 1991 Friedman trademarked his implementation of the MARS algorithm: the resulting program is distributed by Salford and sold as MARS™. Estimation using the 1991 3.5 version of this code was first reported in Stokes (1997), who had obtained the Fortran from Friedman in 1991.

Hastie and Tibshirani released an alternative GPL Fortran implementation of the MARS technique with some differences regarding how the knots are handled, which is contained in R. Improvements to this version have been made by Stokes and will be referred to as MARSPLINE. The differences between the two programs suggest reporting both to validate the calculations as will be done in Table 5. As a preview, note the GASIN and GASOUT residual plots in the bottom of Figs. 1 and 2, which show the magnitude of the residual reduction that was obtained when the linearity assumption was dropped and a MARS model was estimated using MARSPLINE software.

The MARS technique assumes a nonlinear model of the form

$$y = f(x_1, \ldots, x_m) + e, \tag{5}$$

involving N observations on m right-hand-side variables, $x_1, \ldots, x_m$, which are column vectors in the N by m matrix X. The function $f(X)$ is approximated by

$$\hat{f}(X) = \sum_{j=1}^{s} c_j K_j(X), \tag{6}$$

where $\hat{f}(X)$ is an additive function of the product basis functions $\{K_j(X)\}_{j=1}^{s}$ associated with the s disjoint sub-regions $\{R_j\}_{j=1}^{s}$ of D $(D = \chi_{j=1}^{s} R_j)$ and c_j is the coefficient for the jth product basis function. An OLS model is a special case of a MARS model, if all sub-regions include the complete range of each of the right-hand-side variables. In this situation, the coefficients $\{c_j\}_{j=1}^{s}$ can be interpreted as OLS coefficients of the right-hand-side variables. The MARS approach identifies the sub-regions under which the coefficients are stable and detects any possible interactions up to a maximum number of possible interactions controllable by the user.

In contrast to other spline approaches that require the user to specify the knots τ^*, the MARS algorithm produces an estimate of the knot. If all knots are found to be at the minimum of the x variable, then the MARS algorithm has signaled that OLS is the correct estimation procedure. In this case all variables are figuring in the calculation of $\hat{y}$, no matter what their level. The VAR model maintains this assumption and can be thought of as a special case of MARS. The derivative of the spline function is not defined for values of x at the knot value. Friedman (1991) suggests using either a linear or cubic approximation to determine the exact y value and has implemented this in his code MARS™. In contrast, the Hastie and Tibshirani (1990) MARSPLINE program does not make this adjustment.

Models in this paper have been estimated with the original MARS™ Fortran code and with the newer Hastie and Tibshirani (1990) MARSPLINE Fortran code. In the results reported later in Table 5, both of the Friedman evaluation techniques to calculate $e'e$ have been reported and the one with the lowest sum of squares of the residual has been selected. In setting up a MARS estimation, the user selects the maximum number of knots to consider and the maximum order of interaction to investigate. It will be shown that the order of interaction makes a difference in removing the nonlinearity. As an aid in determining the degree of model complexity, Friedman (1991) suggests using a modified form of the generalized cross validation criterion (*MGCV*).

$$MGCV = \frac{(1/N)\sum_{i=1}^{N}(y_i - \hat{f}(X_i))^2}{1 - [C(M)^*/N]^2} \tag{7}$$

where there are N observations, $\hat{y}_i = \hat{f}(X_i)$ and $C(M)^*$ is a complexity penalty. The default is to set $C(M)^*$ equal to a function of the effective number of parameters. The formula used is

$$C(M)^* = C(M) + dM \tag{8}$$

The parameter d, which is user-controlled, has been set to the default value of 3 as suggested by Friedman (1991) for MARS™ estimation and 2–3 for MARS-PLINE estimation, since Hastie et al. (2009) suggest $d = 2$ if the number of interactions is 1 and $d = 3$ for higher-order interactions. As will be illustrated, $e'e$ will in general be larger when $d = 3$, since from Eqs. (7) and (8) in most cases simpler models will be selected. $C(M)$ is the number of parameters being fit and M the number of nonconstant basis functions in the model. The MARS approach starts by choosing where to place the knots for a non-interaction model. Next, more complex interactions are chosen up to a user-controlled maximum number of interactions and maximum number of parameters in the model. Once the forward selection is completed, the *MGCV* statistic is used to eliminate parameters that improve the model only slightly.

The *MGCV* value controls how many parameters finally remain in the model and can be used to form an estimate of the relative importance of each x_i variable in the model. The MARS technique requires that the user select the variables $x_1, \ldots, x_m$ to use in (6). Since the gas furnace data model involves lags, an immediate concern is how to select the appropriate lags of GASIN and GASOUT to place in the $x_1, \ldots, x_m$ vector. The technique proposed in this paper is first to use the VAR model, such as proposed by Tiao and Box (1981), to determine the maximum number and placement of the lags of x and y to estimate a VAR model of the series in the linear domain. If the resulting residuals show evidence of nonlinearity, as measured by the Hinich (1982) test, then these would first be used on the right-hand side of the MARS model equation. It must be emphasized that such a procedure would not be strictly appropriate if evidence of feedback were found in the VAR or VARMA step of the model. In a VAR model, since contemporaneous effects (instantaneous causality) are seen in the off-diagonal elements of the covariance matrix, contemporaneous values of some X variables must be included on the right-hand side of the MARS equation if the VAR model has identified instantaneous causality.

A MARS model with up to 80 knots and from 1–4 interactions was estimated using the MARS and the MARSPLINE code for GASIN and GASOUT. Figures 1 and 2 show the MARSPLINE and VAR residuals for the interaction = 4, $d = 2$ model. Detailed inspection of the plots graphically shows the gains from dropping the linearity assumption implicit in the VAR model. Note that for both the GASIN and GASOUT residuals, the MARS residuals are within much tighter bounds as compared with the corresponding VAR residuals. Using the Hinich (1982), test note that for interactions 2–4 and $d = 2$, the MARSPLINE program removes the nonlinearity as shown by L values of 0.72, −0.31 and 1.18, respectively. For the MARS program for interactions 3–4 the nonlinearity was removed as shown by L values of −0.209 and −1.04, respectively. The differences may be due in part to the default penalty being set for MARS™ at 3.0 and MARSPLINE at 2.0. To test for this possible effect, MARSPLINE models were re-estimated for $d = 3$, where the Hinich L values were 0.22, 0.12 and 1.84, respectively, for interactions 2–4. Note that for every interaction setting, $e'e$ is lower for the MARSPLINE approach, no matter what the d setting. For purposes of comparison, the VAR (6) unconstrained G and L values from Table 2 are also shown. In comparison to the VAR results where $e'e$ was 16.14, the MARSPLINE results for interaction = 4, $d = 2$ had been reduced to 5.00.

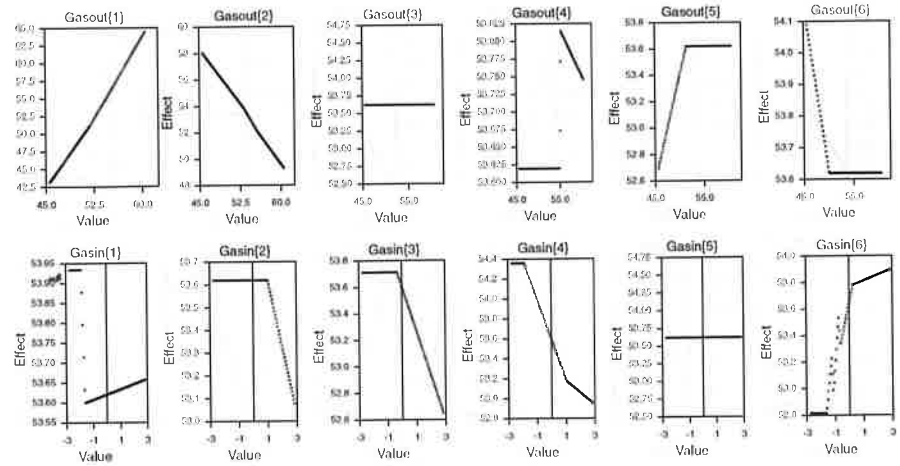

Fig. 3 Leverage plots for a MARSPLINE model of GASOUT-max interaction = 1

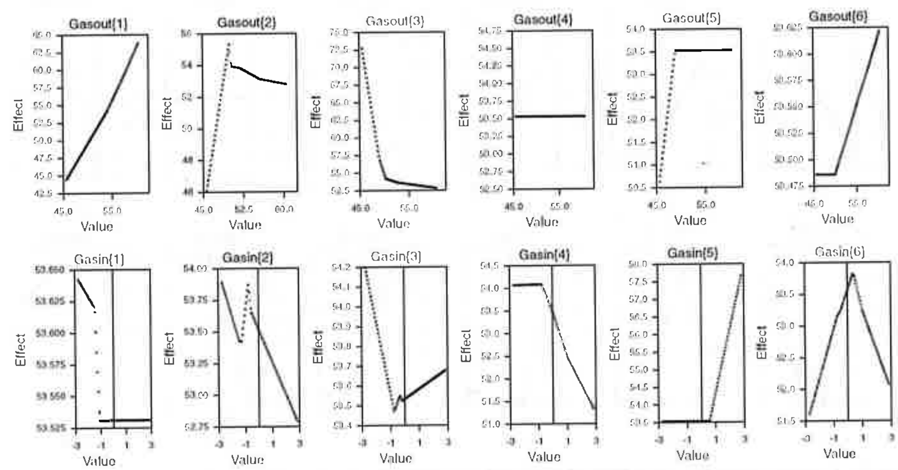

Fig. 4 Leverage plots for a MARSPLINE model of GASOUT-max interaction = 2

The MARSPLINE code allows calculation of standard errors, since after $k_j(X)$ has been determined, OLS is used to obtain the estimated coefficients c_j and $\hat{y}_t$ using the transformed data vectors that contain knots and interaction terms. The MARSPLINE interaction = 4, $d = 2$ model contained 48 estimated coefficients, while the more restricted $d = 3$ model contained 40. The MARS model had 28 coefficients, making the results difficult to summarize without recourse to leverage plots. Define a leverage plot as a graph of the forecasted left-hand-side variable against each right-hand-side variable over its range when all other variables are held at their means or expected value. Leverage plots have been calculated for MARSPLINE models for all right-hand-side variables for interactions 1–4 and are shown in Figs. 3, 4, 5 and 6.

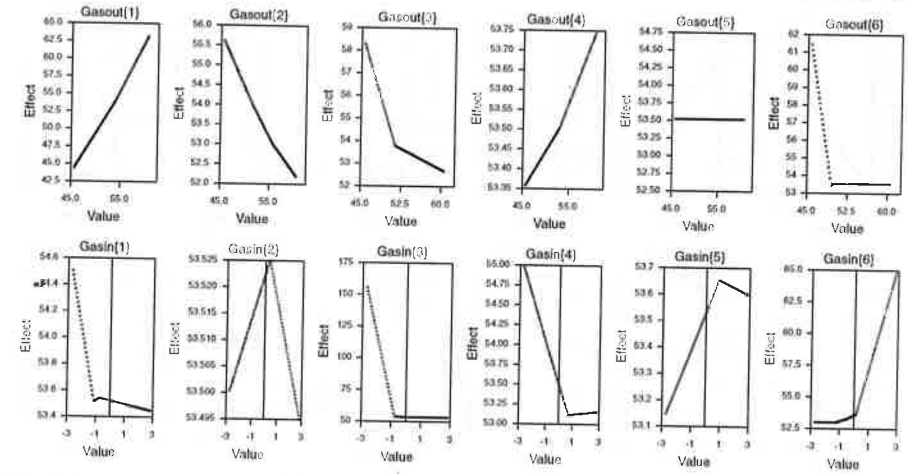

Fig. 5 Leverage plots for a MARSPLINE model of GASOUT-max interaction = 3

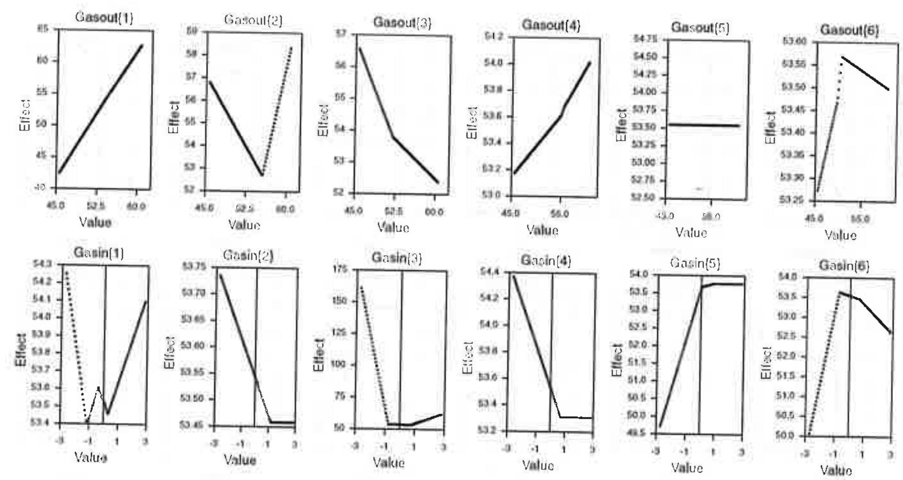

Fig. 6 Leverage plots for a MARSPLINE model of GASOUT-max interaction = 4

What is occurring can be illustrated by a number of examples. Consider GASOUT{1}($\equiv$ GASOUT$_{t-1}$) shown as upward sloping for all graphs. This suggests that this variable enters in a positive and linear fashion that is invariant to the number of interactions. GASOUT{3} illustrates nonlinear effects. For interaction = 1, the plot is flat, showing no effect on the left-hand-side variables, given all other variables were at their means. However, the plot changes for interactions 2–4 when we see a downward slope that flattens out one third of the way into its range. For a linear model estimated by the MARS approach, the leverage plots would be linear, either upward, downward or horizontal. The fact that kinks are shown illustrates the cost of imposing linearity that is documented in the higher $e'e$ values for the linear models and the Hinich (1982) test finding evidence of nonlinearity.

5 Alternative models and out-of-sample validation

To investigate if the estimates of nonlinearity in the residuals detected by the Hinich (1982) test were sensitive to OLS being used, the more robust estimator L1 estimator that minimizes the sum of the absolute values of the residual was tried on the exact VAR model estimated with OLS. For the GASIN equation the L statistic was 6.054 and for the GASOUT equation the L test was 4.71. Since these equations used the exact VAR specification, not a model with a contemporaneous GASIN series on the right as was reported in Table 3, these findings are consistent with the view that the estimated nonlinearity in the residuals found was not due to OLS, which is more sensitive to outliers, being used for estimation.

Nonlinearity tests on the raw series are reported to control for any possible effect of estimated models on the nonlinearity finding. Hinich (1982) tests on the raw GASIN and GASOUT series produce G and L values for GASIN of 2.95 and 0.0980, respectively, and test statistics for GASOUT of 8.214 and 4.691, respectively. This finding of nonlinearity in the GASOUT series and not in the GASIN series is consistent with the hypothesis that the nonlinearity found in the residuals of the GASOUT equation appears to be coming from the nonlinearity in the raw GASOUT series that passed through the linear VAR model.

The next task is to test out-of-sample performance of the MARS model against OLS and other nonlinear estimation alternatives. The results of this exercise are reported in Table 6, where from 1–6 out of sample forecasts were made for OLS, GAM, ACE, MARS and projection pursuit models at the end of the data period. These alternative techniques and the results will be briefly discussed next.

The GAM (general additive model) approach, initially developed by Hastie and Tibshirani (1990) and discussed more recently in Hastie et al. (2009) and Faraway (2006), estimates a model of k explanatory variables X_j of the form

$$y = \beta_0 + \sum_{j=1}^{k} f_j(X_j) \tag{9}$$

where $f_j(X_j)$ is the smoothed X_j series estimated with the use of an iterative backfitting approach. In the forecasts reported later, the smoothing was done with

Table 6 Out-of-Sample forecasting using MARS, PPREG, GAM, ACE and OLS models

Out of sample	OLS_ESS	ACE_ESS	MARS_ESS	GAM_ESS	PP_ESS
1	0.0860	0.0517	0.0106	0.1053	0.7425E-5
2	0.0719	0.0443	0.0649	0.0814	0.0614
3	0.0857	0.0495	0.0600	0.0811	0.0209
4	0.2270	0.1424	0.0589	0.1997	0.1120
5	0.2425	0.1780	0.0813	0.2074	0.0782
6	0.4119	0.3163	0.4581	0.3765	0.5457
Mean of col	0.1875	0.1304	0.1223	0.1752	0.1362

a third-degree polynomial. The ACE (alternating conditional expectations) approach, discussed in Faraway (2006), generalizes the GAM model to smooth both the left-hand and right- hand sides of the model to form

$$\theta(y) = \beta_0 + \sum_{j=1}^{k} f_j(X_j) \tag{10}$$

Imposing the restriction that the variance of $\theta(y) = 1$, the ACE model minimizes $\sum_{i=1}^{N} (\theta(y) - \sum_{j=1}^{k} f_j(x_{ij}) - \beta_0)^2$. The projection pursuit regression model estimation method of Friedman and Stuetzle (1981), discussed in some detail in Hastie et al. (2009), can be thought of as a generalization of the GAM specification, where for M trees $f(X) = \sum_{m=1}^{M} g_m(\omega_m' X)$. The M functions $g_m(\omega_m' X)$ are estimated along directions ω_m, using a flexible smoothing function. The idea is to form nonlinear functions of linear combinations of the k X variables. ω_m' is a unit k-vector of the unknown parameters. As noted by Friedman and Stuetzle (1981), if M is taken to be sufficiently large, the projection pursuit approach is a universal approximation in that any continuous function can be approximated arbitrarily well. Note that in the above discussion M has been redefined from its use in the discussion of the Hinich (1982) test.

The results of this exercise are listed in Table 6. Out-of-sample tests were performed by holding out from 1–6 observations, estimating a model and using this model to forecast ahead. The columns represent the average of out-of-sample error sum of squares for from 1–6 out-of-sample periods. Note that for the means of the six out-of-sample tests, the MARS_ESS value is less. This indicates that for the experiment, the MARS model out performed the other methods. Going from best to worst and using the average of the columns as a heuristic test statistic, the ranking was MARS, ACE, projection pursuit, GAM and OLS. All calculations in this paper have been done with version 8.11E of the B34S software built with the Lahey 7.2 Fortran compiler using optimization level 1. Graphs have been drawn using RATS version 7.30.

6 Conclusion

The Hinich (1982) test was used to test the adequacy of the linearity assumptions in the classic Box et al. (2008) gas furnace data. After finding evidence of nonlinearity, various linearizable, nonlinear models were tried without success. L1 and MINIMAX estimation models were used to determine if the measured nonlinearity of the exact model used to estimate the VAR was sensitive to outliers, where the L1 (MINIMAX) model is less (more) sensitive to outliers than OLS models. The L1 models of GASOUT were found not to remove the measured nonlinearity in the residuals. In four out of six MINIMAX models, measured nonlinearity was removed at the cost of relatively large $e'e$ values. The Hinich (1996) test, applied to subsamples of residuals, indicated that the nonlinearity was episodic. The MARS approach was shown to remove the measured nonlinearity in the model residuals for GASIN and GASOUT and produce a closer fit. Leverage plots were utilized to show the nonlinearity in the effects of the right-hand-side variables. Out-of-sample forecasts for from 1–6 periods were compared using OLS, GAM, ACE, projection pursuit and MARS models, with the latter technique giving the smallest error sum of squares.

Acknowledgments A number of comments of the referees of this paper are acknowledged. Their suggestions have greatly improved the final product. The authors thank Diana Stokes for editorial assistance. Any remaining errors are our responsibility.

References

Ashley R, Patterson DM, Hinich MJ (1986) A diagnostic test for nonlinear serial dependence in time series fitting errors. J Time Ser Anal 7(3):165–178
Box GEP, Jenkins GM, Reinsel G (2008) Time series analysis—forecasting and control. 4. Wiley, New Jersey
Chen R, Tsay R (1993) Functional-coefficient autoregressive models. J Am Stat Assoc 88(421):298–308
Chen R, Tsay R (1993) Nonlinear additive ARX models. J Am Stat Assoc 88(423):955–967
Faraway J (2006) Expanding the linear model: with R. Chapman and Hall, New York
Friedman J, Stuetzle W (1981) Projection pursuit regression. J Am Stat Assoc 76(376):817–823
Friedman J (1991) Multivariate adaptive regression splines. Ann Stat 19(1):1–67
Hastie T, Tibshirani R (1990) Generalized additive models. Chapman and Hall, New York
Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning: data mining and prediction. 2nd edn. Springer, New York
Hinich MJ (1982) Testing for gaussianity and linearity of a stationary time series. J Time Ser Anal 3(3):169–176
Hinich MJ (1996) Testing for dependence in the input to a linear time series model. J Nonparametr Stat 6:205–221
Hinich MJ, Clay C (1968) The application of the discrete Fourier transform in the estimation of power spectra, coherence and bispectra of geophysical data. Rev Geophys 6(3):347–363
Hinich MJ, Messer H (1995) On the principal domain of the discrete bispectrum of a stationary signal. IEEE Trans Signal Process 43(9):2130–2134
Hinich MJ, Mendes E, Stone L (2005) Detecting nonlinearity in time series: surrogate and bootstrap approaches. Stud Nonlinear Dyn 9(4):1–13
Hinich MJ, Patterson DM (1985) Evidence of nonlinearity in daily stock returns. J Bus Econ Stat 3(1):69–77
Lee J-M (2001) A Monte Carlo study of asymmetric time series. Unpublished Ph.D. Dissertation, University of Illinois at Chicago
Lewis PA, Stevens JG (1991) Nonlinear modeling of time series using multivariate adaptive regression splines (MARS). J Am Stat Assoc 86(416):864–877
Lewis PA, Ray B (1997) Modeling long-range dependence, nonlinearity, and periodic phenomena in sea surface temperature using TSMARS. J Am Stat Assoc 92(439):881–893
Patterson DM, Ashley R (2000) A nonlinear time series workshop: a toolkit for detecting and identifying nonlinear serial dependence. Kluwer Academic Publishers, Boston
Priestley MB (1988) Non-linear and non-stationary time series analysis. Academic Press, London
Subba Rao T, Gabr MM (1984) An introduction to bispectral analysis and bilinear time series models. In: Lecture notes in statistics, vol 24. Springer, New York
Stokes HH (1997) Specifying and diagnostically testing econometric models. 2nd edn. Quorum Books, New York
Tiao G, Box GEP (1981) Modeling multiple time series with applications. J Am Stat Assoc 76:802–816
Tong H (1990) Nonlinear time series. Oxford University Press, New York