# The efficacy of nonlinear models in earnings-returns relation of large firms

**Messod D. Beneish**
*Indiana University*

**Campbell R. Harvey**
*Duke University and NBER*

# The efficacy of nonlinear models in earnings-returns relation of large firms

There is a long history of research which examines the relation between unexpected earnings and unexpected returns on common stock. Early literature used simple linear regression models to describe this relation. Recently, a number of authors have proposed nonlinear models. These authors find that the earnings-returns relation is approximately linear for small changes but is 'S'-shaped globally. However, unexpected earnings are generated by the sum of a measurement error and a true earnings innovation, so the apparent nonlinearity could be an artifact of nonlinearity in the measurement errors. Using a research design that minimizes the presence of measurement errors, we provide evidence consistent with the hypothesis that measurement errors contribute to the nonlinearities in the earnings-returns relation. While we are not suggesting that the earnings-returns relation is linear, our evidence suggests that there is no advantage to using a nonlinear model for large firms that are widely followed by analysts.

# 1. Introduction

The widely used linear model of the earnings-return relation has recently come under criticism. Beneish and Harvey (1990) and Cheng, Hopwood, and McKeown (1992) offer evidence that linear models of the earnings-return relation are mispecified. Freeman and Tse (1992) and Das and Lev (1994) provide evidence of the superior performance of an S-shaped relation, that allows the stock price response to earnings announcements to vary according to the magnitude of the earnings surprise. Our study investigates the impact of measurement error on the estimation of the earnings-returns relation in cross-section and in time-series.

Previous evidence on this important issue is based on cross-sectional estimation. In contrast, our study provides firm-by-firm time-series analysis. This is of interest because changes over time in firm risk, dividend payout, and earnings persistence may imply departures from linearity. As Cheng, Hopwood and McKeown (1992) suggest, longitudinal models may mitigate some of these specification problems.

We estimate univariate and multivariate versions of various models in time-series and, to compare our findings to prior research, we also estimate the model in cross-section. We augment the usual specification with ex ante proxies for the expected discount rate, risk changes, and the accuracy of forecasts. These variables are included to capture variation over time. We also pool our data and estimate both the arctan and linear models as in Freeman and Tse (1992), and Das and Lev (1994). Finally, we introduce a new test to compare performance across models. One of the contributions of our study is to provide a test methodology to assess which model best fits the data.

We find that the linear model's performance is not distinguishable from that of alternative models. Reproducing Freeman and Tse's (1992) comparison of the arctan and linear models across classes of earnings surprise magnitude, we find that there is no difference between the arctan and the linear model when these models are estimated with 99.3% of the sample observations.

The message from this paper is that the linear model provides a simple, useful representation of the earnings-returns relation. Our evidence needs to be interpreted in light of our sample consisting of large firms. Our results provide support for the extensive use of a linear representation of the earnings return relation in prior work (e.g., see Brown (1993) for a review of this litterature). That is, prior studies also relied on large firms with a long time-series of earnings. More important, prior studies recognized the issue

1

of measurement error in expected earnings and typically truncated percentage errors greater than 100% from their samples. Applying such criterion we find no difference between the linear and all alternative models.

Our findings indicate that using a sample of large, well followed firms (where measurement error is minimized), there is little advantage to using a nonlinear alternative. We are not suggesting that the earnings-returns relation is linear, but we caution researchers to investigate alternative models only in contexts where measurement error is less likely to drive large earnings surprises. For example, nonlinear and nonparametric models may be appropriate in studies of growth firms or firms in financial distress where unanticipated structural changes rather than measurement error are driving large earnings surprises. In out-of-sample forecasting exercises by investors or managers seeking to assess the stock price reaction to an impending earnings announcement, the simplest and best forecast is given by the linear model.

The paper is organized as follows. The first section provides a discussion of the reasons one might expect a nonlinear relation between adjusted returns and unexpected earnings and why a measured nonlinearity could be spurious. The second section describes the data used in our study. The third section presents our evidence. Some concluding remarks are offered in the final section.

## 2. The Earnings-Returns Relation

Theory suggests that unexpected earnings should impact the firm's security price. In the most general form, the expected market-adjusted returns conditioned on an earnings surprise can be written as:

$$\mathrm{E}_t[AR_{it}|UE_{it}, z_{ijt}] = f_{it}(UE_{it}, z_{ijt}), \tag{1}$$

where $AR_{it}$ represents firm $i$'s stock return at time $t$ adjusted for market risk, $UE_{it}$ is the earnings surprise, $z_{ijt}$ represents $j = 1, ...k$, firm-specific and economy-wide variates that influence earnings responses and the return process, $\mathrm{E}_t$ is the conditional expectation operator and $f_{it}$ is an (unknown and potentially time-varying) firm-specific function. A linear model, linking earnings surprises to market-adjusted returns is, of course, just a special case of (1).

## 2.1 Nonlinear Earnings-Returns Relations

There are a number of reasons to expect departures from a simple linear function in (1): (i) the stock price responses may vary according to the magnitude of the earnings surprise, (ii) the response for positive and negative surprises may be asymmetric, and (iii) the response might differ over time and across firms.

Freeman and Tse (1992) and Das and Lev (1994) find that the stock price response to earnings surprises declines as the magnitude of these surprises increases. They suggest that large earnings forecast errors are more likely to result from large transitory components in actual earnings; these are more heavily discounted and result in a lower stock price response.

Abdel-khalik (1990) suggests asymmetry in the measured earnings returns relation. Measurement error in the proxy for unexpected earnings may be greater in cases of bad news if management has incentives to more quickly release information about the firms' good fortunes. If positive surprises are more likely occurrences than negative surprises and thus positive surprises are likely to be less noisy.[2]

There is also empirical evidence suggesting that the earnings-returns relation is not constant. For example, research studying the timeliness of earnings announcements (Chambers and Penman (1984), Hughes and Ricks (1987)) provides evidence suggesting that the stock price reaction is greater for early versus late disclosures. Hence, the estimated earnings-returns relation will be a function of the timeliness of announcements. Lang (1991) studies the earnings-returns relation in a sample of initial public offerings and suggests that the earnings returns relation changes through time as investors learn about the earnings process of firms.

## 2.2 Parametric and Nonparametric Models

The most frequently studied model of the earnings-returns relation is:

$$AR_{it} = \alpha_1 + \beta_1 UE_{it} + \tilde{\epsilon}_{1it}, \tag{2}$$

where $\alpha_1$ is the intercept, $\beta_1$ is the slope or earnings response coefficient (ERC). This

---

[2] Earnings surprises (U) are assumed to have zero mean (E(U)=0). If positive surprises are more frequent $(prob(U > 0) > .5)$ then $pE(U|U > 0) + (1-p)E(U|U < 0) = 0$ implies that $E(U|U > 0) < -E(U|U < 0)$, e.g., that positive surprises are less noisy.

specification is typically estimated in pooled time-series cross-sectional regression, forcing the relation to be constant both across firms and over time.

More recent studies have allowed the relation to vary across firms by estimating an additive or interactive variant of (2):

$$AR_{it} = \alpha_1 + \beta_1 UE_{it} + \sum_{j=1}^{k} \gamma_j z_{ijt} + \tilde{\nu}_{1it},$$

$$AR_{it} = \alpha_1 + \beta_1 UE_{it} + \sum_{j=1}^{k} \gamma_j z_{ijt} UE_{it} + \tilde{\nu}_{2it}. \tag{2a}$$

While these cross-sectional studies have documented which firm characteristics and economy wide variates ($z_{ij}$'s) influence the earnings-returns relation, interpreting their evidence depends on the assumption that the functional form of the earnings-returns relation, $f_i$, is identical for all firms. If this assumption is violated, inferences from these studies are unreliable.

We investigate a range of nonlinear models in both time-series and cross-sectional settings. The firm-by-firm time-series analysis is particularly important given evidence in Cheng, Hopwood and McKeown (1992) that nonlinearity, as well as omitted variables, is associated with systematic variation across firms in the earnings coefficient ($\beta_1$).

We use the arctan and modified quadratic models because they let the stock price response vary as a function of the magnitude of the earnings surprise. The arctan model allows the abnormal return return to be a function of the arctan of the earnings surprise. This delivers an 'S'-type nonlinear relation which damps the impact of large absolute earnings surprises. The modified quadratic model allows the abnormal return to be quadratic in the earnings surprise. A dummy variable is added to allow for different signs (but the same coefficient) on the quadratic term depending on whether the surprise is positive or negative. Importantly, these models treat positive and negative surprises symmetrically.

We also estimate three other simple nonlinear models. The piecewise linear model adds a slope dummy variable to the linear model that takes on the value of one if the earnings surprise is negative. We also estimate a plain quadratic model. Finally, we estimate a Fourier flexible model. This general nonlinear model augments the quadratic model with a trigonometic expansion.

We investigate two nonparametric kernel models. The nonparametric models are attractive because they do not require the prespecification of a functional form. They also

can capture simultaneously the large forecast error effect and the possible asymmetry between positive and negative errors. The technique relies on the data to inform us about the structure of the relation.

The two nonparametric models are called the fixed kernel and the adaptive kernel. With both of these methods, a curve (with a general shape) is fit through the data. With the fixed kernel, the degree of smoothing is fixed throughout the data. As a result, the fixed kernel is apt to overfit the data. That is, in areas of sparse data, the curve will be drawn towards influential observations. Indeed, the nonparametric fixed kernel regression estimates could be biased. We address both the sensitivity to influential observations and the bias problem by using the adaptive kernel methodology. With this method, the degree of smoothing increases in areas of sparse data. The regression curve is no longer unduly drawn towards extreme observations [see Härdle (1990) for a detailed comparison of various kernel models]. The details of the models are described in the Appendix.

*2.3 Criteria and Tests for Model Comparison*

One of the contributions of our study is to provide a test methodology to assess which model best fits the data. This is difficult because most of the models are nonnested. We assess the fit of the models both in and out of sample. First, we provide information of a standard set of criteria: $R^2$, adjusted-$R^2$, mean squared error, mean absolute error, Akaike's information criterion and Schwarz's criterion for each model. Second, we focus on a method to compare errors across different models.

Our test to compare the different models is general in that errors can be nonnormal, heteroskedastic and even serially correlated. Let $\epsilon_{1i}$ and $\epsilon_{2i}$ be forecast errors from two models. The following equality holds:

$$E[(\epsilon_{1i} + \epsilon_{2i})(\epsilon_{1i} - \epsilon_{2i})] = \sigma_{1i}^2 - \sigma_{2i}^2, \tag{3}$$

where $E[\cdot]$ denotes the expectation operator, and $\sigma_{1i}^2$, $\sigma_{2i}^2$ are the residual variance of models 1 and 2 for firm $i$. This equality says that the differences between two error variances can be measured by the covariance between the sum of the errors and the difference of the errors. The null hypothesis is that the error variances (and thus the MSEs) are not different.[4] The tests are interpreted as follows: a significantly positive (negative) covari-

---

[4] Most of these tests rely on either the squared error loss function or the absolute error. In principal, residuals and forecasts can be evaluated with other loss functions.

ance indicates that model 2 (model 1) dominates in that it has a higher $R^2$ and a lower error variance. If the covariance is zero, then the models exhibit similar performance. The covariance can be estimated following White (1980) and Hansen (1982), with estimators consistent with the presence of heteroskedasticity and nonnormality. A sufficient condition in Hansen (1982) is that the data are stationary and ergodic. The standard error of the covariance is determined as follows. Let $u_t = \epsilon_{1t} - \epsilon_{2t}$ and $v_t = \epsilon_{1t} + \epsilon_{2t}$. Define $s(j)_t = 1/n \sum_{t=1}^{N} u_t v_t u_{t-j} v_{t-j}$, where $N$ is the number of observations.

We get the standard error of the covariance from the fact that:

$$\sqrt{N} \frac{1/N \sum_{t=1}^{N} u_t v_t}{\left[ \sum_{j=-K}^{K} (1 - \frac{|j|}{K+1}) s(j) \right]^{1/2}} \sim N(0,1) \tag{3a}$$

where $K$ is the order of the moving-average process in the errors. In our analysis, $K = 0$.

Equation (3) can easily be modified to allow for a test of differences in $R^2$. Since the denominator in the $R^2$, total sum of squares (TSS), is identical across models with a common dependent variable, multiplying $\epsilon_{1i}$ and $\epsilon_{2i}$ by $\sqrt{(N-1)/TSS_i}$, yields $\epsilon_{1i}^*$ and $\epsilon_{2i}^*$ where $N$ is the number of observations, and (3) becomes,

$$E[(\epsilon_{1i}^* + \epsilon_{2i}^*)(\epsilon_{1i}^* - \epsilon_{2i}^*)] = R_{2i}^2 - R_{1i}^2, \tag{4}$$

Tests (3) and (4) differ only by scale so the model with the larger error variance will have a smaller $R^2$ and, in fact, the MSE and $R^2$ tests yield identical results.[6]

We further modify (3) to test for the equality of adjusted-$R^2$s. To do this, $\epsilon_{1i}$ and $\epsilon_{2i}$ are multiplied by $\sqrt{(N-1)/TSS_i(N-k_1)}$ and $\sqrt{(N-1)/TSS_i(N-k_2)}$, yielding $\epsilon_{1i}^{**}$ and $\epsilon_{2i}^{**}$ where $N$ is the number of observations and $k_1$ and $k_2$ are the number of parameters estimated in models 1 and 2.

$$E[(\epsilon_{1i}^{**} + \epsilon_{2i}^{**})(\epsilon_{1i}^{**} - \epsilon_{2i}^{**})] = adj R_{2i}^2 - adj R_{1i}^2. \tag{4a}$$

---

[6] Each term in (3) is divided by $TSS_i$ which we assume is known. We were concerned that there is a Jensen's inequality problem going from (3) to (4). If the problem exists, it is very small because we executed tests based on both (3) and (4) and obtained the same results.

# 3. Data and Method

## 3.1 Sample Selection

Our sample is designed to minimize measurement error in expected earnings and consists of 40 large firms followed by Value Line over the period 1968-1988. The forty firms included in our sample met the following criteria:

(i) Daily stock returns are available from 1967 to 1988 on the CRSP daily returns tape.

(ii) Quarterly earnings announcement dates are published in *The Wall Street Journal Index* from 1968 to 1988.

(iii) The *Value Line Investment Survey* provided earnings per share forecasts for the 79 quarters between 1969:1 and 1988:3.

Criteria (i) and (ii) are imposed to obtain the returns data necessary to estimate the stock price reaction to quarterly earnings announcements. Criterion (iii) is imposed to minimize measurement error in the computation of unexpected earnings (see 3.2 below). Out of a random sample of 100 firms meeting criterion (i), we find 40 firms followed by Value Line in the period 1969-1988. Out of 3160 (40 firms × 79 quarters) observations, we eliminate 28 for which no announcement date is available in *The Wall Street Journal* and 203 for which we could either not find a specific issue of the *Value Line Investment Survey*, or the timing of publication was such that no one quarter ahead forecast was available. Our analysis is conducted on 2929 firm-quarter observations.

Our criteria result in a sample of large surviving firms. The mean (median) market value of equity in December 1978 is $3,377.52 ($842.0) million; all our sample firms are larger than the median NYSE firm in December 1978; 60% of our sample firms are in the top NYSE size quintile. There are three reasons for deliberately selecting a sample of large firms. Large firms are more widely followed, increasing the likelihood that analysts' forecasts approximate market expectations of earnings. Large firms have lower disagreement among analysts (Barron and Stuerke (1997)), and Abarbanell, La-nen, and Verrechia (1995) show that measurement error increases with the disagreement in analysts' forecasts. Large firms are also less subject to post–announcement drift, which Bernard and Thomas (1989) attribute to incomplete price responses to earnings

announcements in the case of extreme surprises (upper and lower deciles).

### 3.2 Measuring Unexpected Earnings and Adjusted Returns

Earnings forecast errors are calculated as the difference between Value Line actual earnings per share at quarter $t$ and the corresponding one-quarter ahead Value Line analysts' forecast. Unexpected earnings are equal to the ratio of earnings forecast errors to the price of a firm's common share in day -2 relative to the day of announcement of earnings (Christie (1987)).

Computing forecast errors as the difference between Value Line actual earnings and Value Line earnings forecast has two advantages. First, it has the advantage of reducing measurement error. Philbrick and Ricks (1991) provide evidence that the Value Line actual/forecast pairing yields smaller forecast errors than pairings based on IBES, S&P's Earnings Forecaster, Compustat or Zacks Investment Research. This is significant since some recent studies attribute nonlinearity in the earnings-returns relation to the magnitude of forecast errors without regard to measurement error in unexpected earnings. Second, there is no significant serial correlation in Value Line earnings forecast errors (Brown et al. (1987)). The absence of serial correlation is also consistent with the notion that the Value Line forecasts contain minimal measurement errors.[8]

The mean absolute (standard deviation) and median absolute forecast errors in our sample equal 0.0051 (0.0042) and 0.0039. In Philbrick and Ricks (1991, p.405), corresponding statistics are 0.0077 (0.0075) and 0.0026. Mean and median are closer in our sample than in theirs suggesting that our average error is less affected by large errors than Philbrick and Ricks' (1991). Our comparison to Freeman and Tse (1992) is more limited because they only report their mean absolute forecast error (0.0159) based on IBES forecasts. Their average is over three times larger than the average in our sample. Overall, the smaller forecast errors in our sample are consistent with our sample being composed of large, widely followed firms – a sample that is likely to contain less measurement error in the earnings forecasts.

Market-adjusted returns at the time of quarterly earnings announcements are com-

---

[8] For our sample, the average first to fourth order serial correlation coefficients are: 0.21, 0.11, 0.07 and 0.08. Given the sample sizes of 62 to 78, we cannot reject the null hypothesis that these coefficients are zero.

puted as follows:

$$AR_{it} = \sum_{t=-1}^{0} (R_{it} - \hat{\alpha}_i - \hat{\beta}_i R_{mt}) \tag{5}$$

where:[9]

$AR_{it}$ = market-adjusted return from day -1 to 0 relative to the day of earnings announcement (day 0).

$R_{it}$ = daily return, firm $i$, day $t$.

$R_{mt}$ = equally weighted NYSE and AMEX index, day $t$.

$\hat{\alpha}_i, \hat{\beta}_i$ = market model parameters estimated over 300 trading days from day -360 to day -61.

We choose a two-day window to capture the effect of earnings announcements since earnings are likely to be released the day prior to publication in *The Wall Street Journal*. The choice of a short window reduces error in measuring the stock price response relative to longer periods where non-earnings information released can confound the experiment. This increases the likelihood that the price reaction observed is due to the earnings surprise.

## 4. Empirical Results

### 4.1 In-Sample Estimation

Table 1 summarizes the results of estimating the time-series relation between market-adjusted returns and unexpected earnings for 40 firms over 79 quarters from 1969–1988. Panel A reports averages of standard goodness-of-fit statistics: adjusted $R^2$, mean absolute error (MAE) and root mean squared error (RMSE).

Table 1 shows that the linear model has the smallest $R^2$. The average adjusted $R^2$ of 5.4% is comparable to previous research which has focussed on cross-sectional analysis. All of the alternative models yield higher $R^2$s. For example, the explanatory power is more than doubled when the fixed kernel is applied to the same data. However, an analysis of the MAE and the RMSE statistics reveals little apparent difference among

---

[9] The three trading days of October 16, 19 and 20, 1987 corresponding to the October 1987 stock market crash are excluded from the estimation period. No announcement of quarterly earnings occurs in this 3-day period.

9

the models.[11] For example, the average RMSE of the fixed kernel model is 0.031. The application of the linear model increases the RMSE only slightly to 0.032 suggesting little gain in insample fit.

Panel B provides statistical tests of model performance across models. Using Hansen's (1982) generalized method of moments to estimate the covariance,[12] we categorize the proportion of cases where: the linear model has a significantly higher forecast error variance, the alternative model has significantly higher forecast error variance, and the models' forecast errors are not significantly different. Our analysis is focused on the linear model versus the other nonlinear models because the linear is the most parsimonious and the most prevalent in the literature.

The results are striking. Using the degrees of freedom adjustment, the linear model is never dominated by the piecewise, quadratic, arctan or Fourier flexible. Even when we drop the degrees of freedom adjustment which favors the high parameterization models such as the Fourier flexible, there is only one firm where the alternative model statistically beats the linear model. In four firms, the nonparametric model beats the linear model. However, for 36 firms or 90% of the sample, there is no significant difference between the nonparametric and the linear in-sample predictions.

To illustrate the fit of the models, we provide plots of the fitted and actual abnormal returns versus unexpected earnings for two sample firms. The two firms selected are the first and fourth quartile midpoints of the difference between the fixed kernel and linear models' adjusted $R^2$s. A plot of the first quartile midpoint (the series for the firm Caterpillar) in figure 1 shows that the fitted values from all the specifications are similar to the linear model. In this case, no obvious nonlinearities are detected and the linear model is a good approximation of the earnings-return relation.

Plots of the fourth quartile midpoint, Eagle Picher, are presented in figure 2. The fitted values suggest that the fit of the nonlinear models and the fixed kernel are affected

---

[11]   The inconsistencies between the results obtained for $R^2$ and RMSE, are due to the fact that RMSEs are unscaled and that we average these measures across firms. Two firms could have identical RMSEs but their $R^2$s could be different because their total sum of squares is different.

[12]   The tests relate to MSE, $R^2$ and adjusted $R^2$ criteria. For the MAE, we compared model performance using a Wilcoxon rank-sign test. We also tested for serial correlation in the forecast errors of the linear model and found only two cases of significant autocorrelation. As a result, we did not modify the standard error of the covariance to account for a first-order moving average process.

by extreme observations in the low and high unexpected earnings range. The fit of the adaptive kernel is less susceptible to the influence of extreme observations because a larger degree of smoothing is applied in areas of sparse data. The graphical analysis suggests that some of the evidence of nonlinearity in the data is determined by the extreme earnings forecast errors.[13]

Since previous research has estimated the earnings-returns relation cross-sectionally, we replicate the analysis in a cross-sectional context. Table 2 summarizes 79 cross-sectional regressions (one per quarter) from 1969–1988.

The in-sample cross-sectional results are broadly similar to the time-series results. The average adjusted $R^2$ from the linear models is 5.3% (compared to 5.4% in the time-series models). According to the $R^2$ criterion, the linear model appears to be the least desirable. However, the linear model is not statistically dominated by the piecewise, quadratic, arctan or Fourier flexible in any of the 79 cross-sectional regressions. Moreover, the nonparametric models dominate the linear model in only three quarters.

What do we learn from the in-sample analysis? Numerically higher goodness-of-fit measures, such as adjusted $R^2$, obtain for nonlinear models. However, the tests indicate that these nonlinear models do not statistically outperform the simple linear model. Furthermore, the plots suggest that the nonlinearity is being driven by a few extreme observations.

### 4.2 Comparison to Previous Research

Our results provide support for the extensive use of a linear representation of the earnings-return relation in prior work. However, our evidence appears to conflict with recent studies by Freeman and Tse (1992) and Das and Lev (1994). We thus investigate whether the differences occur because of the relative accuracy of earnings forecasts and/or because of influential observations. As in recent studies, we estimate pooled time-series cross-sectional regressions with all 3,000 observations and in subsamples based on the magnitude of unexpected earnings.

---

[13] We recalculated Table 1, Panel A when observations more than two standard errors from the mean unexpected earnings were stripped. The results are similar to the ones reported in table 1. We also recalculated Panel A using seasonal random walk forecast errors. These errors are much noisier with mean error (standard deviation) of 0.0138 (0.0104) compared to Value Line's 0.0051 (0.0042). The fit of all the models deteriorates making it even more difficult to discriminate across models.

Consider the results in Freeman and Tse's (1992, p.195), Table 1, Panel B. They report dramatic differences in adjusted-$R^2$ between the linear and arctan models when price deflated unexpected earnings are in absolute value ($|UE|$) greater or equal to 0.05. This occurs for 701 observations or 5.7% of their sample of 12,381 observations.[14] To compare our results to those in Freeman and Tse's, we replicate their analysis with our sample in Table 3, panel B. As mentioned earlier, the forecasts are more accurate in our sample. We find only 20 observations out of 2929 or 0.7% of our sample with price deflated $|UE| > .05$. We further note that the adjusted-$R^2$s are equal (3.4%) for both the linear and arctan models when 2909 observations with $|UE|$ less or equal to 0.05. However, when the 20 observations with $|UE| > 0.05$ are included in the estimation, the adjusted-$R^2$ for the arctan increases to 5.6% and that of the linear model decreases to 3.0%.

Our analysis suggests that the success of the nonlinear model is due to a small number of extreme observations. The performance of the arctan is nearly doubled as a result of less than one percent of our sample. Indeed, Freeman and Tse (1992) conjecture that the linear model may be appropriate if large unexpected earnings are discarded. Is there a basis for discarding these observations?

Most prior research treated observations with percentage earnings forecast errors (as opposed to price-deflated errors) greater than 100% in absolute value as unreliable and either winsorized or excluded these observations from their analysis (see for example Foster (1977), Hagerman, Zmijewski and Shah (1984), Brown et al. (1987), and Easton and Zmijewski (1989)). There is no way to distinguish whether these are influential

---

[14] Freeman and Tse (1992, p. 208) conclude that "the linear model appears to be well specified when the earnings surprise is no greater than 0.5% of firm value, which includes approximately 60% of our sample." This conclusion is based on observations with $|UE| <= .005$ (7,361/12,381=59.5%, Table 1, p. 195). However, Freeman and Tse do not provide any statistical test to compare adjusted-R$^2$s across models. Thus, examining observations with $|UE| <= .01$ we observe that the linear and arctan models adjusted-R$^2$s are not that different (.049 vs. .053) and believe that the linear model does well as the arctan model for at least 9,363/12,381 or 75.6% of their observations. Furthermore, in the absence of tests comparing fit, we do not know whether there is a significant difference between explaining 5.2% vs. 7.5% of the variation. As such, the linear model could be as good as the arctan model for 11,680 out of 12,381 or 94.3% of their sample. What is clear, however, is that the arctan model is superior when the 346 observations with largest $|UE|$ are also included in the estimation. This is because, with extreme observations, the linear model collapses–it has a .4% adjusted-R$^2$ and the regression is not likely to be significant.

observations or data errors. Assume that percentage absolute errors greater that 100% are based on forecasts with large measurement error and can be truncated. To the extent that our sample of large, widely followed firms are less likely to have surprises driven by unanticipated structural change, surprises larger than 100% are less likely to be true innovations. Applying such criterion to our sample eliminates 27 observation and we find no difference between the linear and all alternative models. Applying the same criterion to Freeman and Tse's sample would also eliminate most of the observations which result in a higher $R^2$ for the arctan model.[15]

In addition to the arctan model, Das and Lev (1994) evaluate a quadratic model and the locally weighted regression of Cleveland and Devlin (1988). Evidence in tables 1 and 2 indicates that the linear model is superior to the quadratic model. The linear model dominates the quadratic model in 24 out of the 40 firms in time-series analysis, and 25 out of 79 quarters in cross-section. Also the linear model is never dominated by the quadratic. We believe that these results obtain because the quadratic model yields unreasonable predictions. That is, if the quadratic coefficient is positive (negative), the earnings response function indicates large rewards (penalties) to both large positive and negative surprises.

We conducted another experiment with our pooled time-series cross-section. The cross-sectional model assumes that coefficients are the same for each firm. It seems plausible that firms with high variance of unexpected earnings would have lower response coefficients. When the coefficients are constrained to be equal in the cross-section, those firms with high unexpected earnings variances might produce extreme observations. That is, if we group firms by variance, we might expect different response coefficients. The constraining coefficients to be identical might induce nonlinearity.

To address this possibility, we segmented the sample into two equal sized groups: low unexpected earnings variance and high unexpected earnings variance. We re-estimated

---

[15] We are not suggesting that this is a valid cut-off. Since prior research used percentage errors and Freeman and Tse use price deflated errors, we convert the latter using a price-earnings ratio [forecast error/earnings=(forecast error/price)×(price/earnings)]. The mean and median P/E ratios for their data are 16.72 and 16.77 (see their Table 4, p. 201) indicating that $|UE| > 1/16.72$ or $>1/16.77$, that is, $|UE|$ greater than about .06 would have been deleted from the sample using the 100% error criterion. In our sample the mean P/E is 18.33 suggesting that 18 observations with $|UE| > 0.055$ would be deleted. Computation of the actual percentage error with our data suggests that 27 firm quarter observations, nine more than with the previous criterion, have greater than 100% percentage forecast errors and would be deleted.

the quadratic and modified quadratic models in these two samples. The results are available on request. For the quadratic model, the quadratic term does not enter the low variance sample but is strongly significant in the high variance sample. In the modified quadratic model, the quadratic term enters both samples.

This exercise suggests that nonlinearity is not an artifact of contraining cross-sectional coefficients to be identical across firms. Even within our sample of large, well-followed firms, the high variance sample is the group with extreme observations and possible measurement error. Our sub-sampling by variance is consistent with our hypothesis that measurement error plays a role in the evidence of nonlinearity.

*4.3 Out-of-Sample Prediction*

Part of the interest in the earnings-returns relation is its use in out-of-sample forecasting. Note, however, that (2) describes market-adjusted returns at time $t$ as a function of the earnings surprise which is observed at time $t$. Assume that the investor has a better forecast than the market, which is represented by Value Line in our case. Then with knowledge of the earnings response function, a stock market reaction can be estimated ex ante and a trading strategy implemented.

Out-of-sample forecasts are calculated for all seven models by holding out the last 12 quarters of data for each firm. The other quarters are used to estimate the models' parameters. For the linear, piecewise, quadratic, modified quadratic, arctan and Fourier models, the parameters of the returns-earnings model are estimated with data up to quarter $t$ and are combined with the actual earnings surprise in quarter $t + 1$, to forecast the market-adjusted return in quarter $t + 1$. This procedure, sometimes called "conditional prediction," is repeated with new parameter estimates through quarter $t + 12$. A similar method is used to generate nonparametric forecasts.

The first panel of Table 4 presents goodness-of-fit statistics for the stepahead predictions. The Fourier flexible fares, by far, the worst. It has the lowest RMSE in-sample (0.022) and the largest RMSE out-of-sample (0.043). This model has the largest number of parameters and is the most susceptible to overfitting. For none of the 40 firms in the sample does the piecewise, quadratic, arctan, or Fourier flexible statistically dominate the linear model. In only one instance are the nonparametric models' forecasts better than the linear model.

14

Table 5 presents the out-of-sample cross-sectional analysis. Parameters estimated from the previous quarter are used to forecast the abnormal returns in the next quarter. There are 78 quarters of out-of-sample predictions. Consistent with the in-sample analysis, no model has lower MAE than the linear model. The Fourier model (which had the second best in-sample performance) clearly fails when taken out of sample. The Fourier forecast errors in Panel A are an order of magnitude higher than any other models' forecast errors. The adaptive kernel model has a slightly lower RMSE than the linear model (0.033 versus 0.034). However, the tests in panel B reveal that the adaptive kernel provides a statistically superior fit in only 3 of 78 quarters.

Both the in-sample and out-of-sample analysis provide evidence that some of the apparent nonlinearities in the earnings-returns relation is likely due to overfitting. None of the alternative nonlinear models results in significantly lower errors than the traditional linear model. In the next section, we extend our analysis to a multivariate setting and examine the role of other conditioning information. This is important because, for example, the univariate time-series models do not control for sources of potential variation.

*4.4 The Role of Conditioning Information*

In this section, we focus on other conditioning information, $z_j$, that could influence the relation between market-adjusted returns and unexpected earnings. We identify proxies for expected discount rates, measurement error in expected earnings and firms' growth prospects. We follow Collins and Kothari's (1989) suggestion that using proxies for *expected* discount rates at which unexpected earnings are capitalized enhances the specification of the earnings-returns relation.

We consider two firm-specific variables and one economy-wide variable. The first variable we consider is a lagged firm-specific return, $R_{i,-60:-2}$. Easton and Zmijewski (1989) suggest that this variable mitigates the measurement error problem in expected earnings. The second firm-specific variable is the lagged variance of returns, $\sigma^2_{i,-60:-2}$. This variable proxies for firm risk characteristics not captured in the market model beta (see, for example, Fama and MacBeth (1973)). Finally, we consider an economy wide variable, $SPREAD_t$. This variable is defined as the difference in yields between Moody's Baa and Aaa rated bonds. Keim and Stambaugh (1986) show that this ex ante variable has the ability to forecast market returns. Fama and French (1989) demonstrate that

the SPREAD variable has distinct business cycle patterns and, as such, it may proxy for the expected risk premium.

Both $\sigma^2_{i,-60:-2}$ and SPREAD are linked to the fundamental valuation model that transforms unexpected earnings into price changes. In most applications, both the risk and the market premium are assumed to be constant. The variable $\sigma^2_{i,-60:-2}$ is designed to capture the effects of both shifts in risk through time and risk that is not priced in the capital asset pricing model framework. The SPREAD variable is included to capture time-variation in the market premium. Both of these variables could affect the discount rate and hence affect the earnings-returns relation.

We consider two general specifications:[16]

$$AR_{it} = a_0 + a_1 UE_{it} + a_2 R_{i,-60:-2} + a_3 \sigma^2_{i,-60:-2} + a_4 \text{SPREAD}_t + \eta_{it} \qquad (6)$$

and

$$AR_{it} = b_0 + b_1 UE_{it} + b_2(UE_{it} \times R_{i,-60:-2}) + b_3(UE_{it} \times \sigma^2_{i,-60:-2})$$
$$+ b_4(UE_{it} \times \text{SPREAD}_t) + \nu_{it}. \qquad (7)$$

In the first specification, the other conditioning information enters additively. This specification tests whether these other conditioning variables linearly influence market-adjusted returns holding the level of the unexpected earnings constant. The second model is an interactive specification. This specification allows for a simple type of nonlinearity. The conditioning information directly affects how market-adjusted returns respond to unexpected earnings. While collinearity is more likely to be a problem in this model, a

---

[16] We also examined but did not include in our final analysis three other variables. First, we included the ratio of market to book value to proxy for the firms' growth prospects. Collins and Kothari (1989) use this variable as a cross-sectional proxy for earnings persistence. Fama and French (1992) find that this variable is related to differences in cross-sectional expected returns. Second, given the evidence in Chambers and Penman (1984) and Hughes and Ricks (1987), we used two variables that proxy for the timeliness of earnings announcements. For each firm, we defined two dummy variables EARLY and LATE taking values of 1 (0 otherwise) when the number of days between a fiscal year end and the date of announcement was 2 standard errors each way of the mean number of days to announcement. For 34 of the 40 firms, both dummies were constants equal to zero. Finally, drawing on the evidence of Cornell and Landsman (1989) that the fourth quarter announcements provide more information to the market, we included quarter indicator variables to capture differential seasonal responses. None of these variables enhanced the time-series specifications and hence are not reported. However, our sample design has large, small earnings surprise firms. It is possible that some of these other variables might be important in a different sample.

diagnostic check of the instrument matrix reveals that the square root of the ratios of the largest eigenvalue to the individual eigenvalues lie between 6.7 and 23.2. Belsley, Kuh and Welsch (1980) show that collinearity is not likely to be a problem if this condition index is less than 30.[17]

Table 6 presents a summary of the firm by firm time-series regressions that include the augmented conditioning information. In these regressions, firm-specific coefficients are allowed in (6) and (7). In panel A, the additive specification shows a mean adjusted $R^2$ of 8% which is higher than the 5.4% average $R^2$ reported in Table 1 for the linear specification without the conditioning information. However, the importance of the conditioning variables depends on the firm. In 9 of 40 firms, an F-test provides evidence against the hypothesis that the coefficients on the additional conditioning information are zero. For five firms, the lagged return is significantly different from zero. The SPREAD variable also enters significantly for five of 40 firm regressions.

The interactive specification does not improve the fit of the time-series regressions. The average adjusted $R^2$ is 7.3%. While this is higher than the linear model without the conditioning information, it is lower than the additive specification. The interaction with the lagged return is significant in 6 firms' regressions. There is also evidence in 4 firms that the lagged volatility influences the relation between returns and unexpected earnings.

Panel B presents the results using the nonlinear models. Consistent with the results of the linear specification, the extra conditioning information improves the fit of the time series regressions on average. The piecewise specification exhibits an average adjusted $R^2$ of 8.4% in the additive specification compared to 6.1% reported in Table 1. The quadratic and arctan mean adjusted $R^2$s increase from 6.5% and 6.4% in Table 1 to means of 8.7% and 8.1% when the extra conditioning information is included in Table 6. Finally, the Fourier flexible form explains, on average, 11.1% of the variation compared to 8.2% in Table 1.

A similar but slightly better fit is found when the additional conditioning information enters the nonlinear models interactively. The average adjusted $R^2$s increase on average for the piecewise, quadratic and modified quadratic specifications. The Fourier flexible model is not estimated in this specification because it requires 20 regressors which is

---

[17] Another specification is to augment (6) with the interaction terms. The diagnostics revealed high correlation between some of the the additive and interactive variables. As a result, we do not report the augmented additive specification.

probably not appropriate when there are less than 70 time-series observations.[18]

Although the average adjusted $R^2$s increase in panel B, the coefficients on the extra conditioning variables are generally not significant. Table 6 also reports exclusion tests on all variables related to the augmented information. In the interactive specification, we cannot reject the null hypothesis that the coefficients on the extra conditioning information are zero in 6 cases for the piecewise model, 4 cases for the quadratic model and 9 cases for the modified quadratic model. Similar results obtain for the additive specification. Hence, there are more than 30 different firms where the extra conditioning information does not increase the ability to explain the market-adjusted returns. Furthermore, a comparison of the adjusted-$R^2$s of the augmented linear versus alternative models indicates that the linear model is only dominated by one of the alternative models for, at most, two firms, that is for less than 5% of the sample. This evidence, based on our time-series estimation, is important as it suggests a minor role for these variables which attempt to capture time-variation.

*4.5 Comparison to Previous Cross-Sectional Research using Other Information*

In Table 7, we pool our 40 firm-specific time-series observations into a constrained cross-sectional regression. We provide regression estimates based on 2929 observations. This allows us to assess the joint significance of the extra variables.

Two features of the results in Table 7 are noteworthy. First, the coefficients on the other conditioning information attain significance at standard levels of confidence.[19] While this is consistent with previous work (Collins and Kothari, 1989), some caution should be exercised in interpreting the statistical results. Given the evidence on individual firms, it is possible that 'significance' is attained due to the dramatic increase in sample size.

---

[18] For the same reason, the multivariate nonparametric density is not reported. It would not be meaningful to approximate a five-dimensional probability density with less than 70 data points.

[19] We also estimated the regression with proxies for persistence (market to book value ratio) and size to capture cross-sectional differences. While we obtain similar results, the estimates on these two proxies for cross-sectional variation are not significant.

# 5. Conclusions

There are a number of theoretical reasons for believing that the relation between market-adjusted stock returns and unexpected earnings is not linear. Indeed, a number of recent empirical papers have pursued various nonlinear functional form.

Our paper evaluates a broad range of nonlinear models as alternatives to a linear relation. We conduct in-sample and out-of-sample analyses. We consider other variables that might affect the earnings returns relation and conduct our tests both in time series and in cross section. Our tests for model comparison are robust to nonnormalities, heteroskedasticity and serial correlation in the models' errors.

Our paper constructs a sample of firms from the largest NYSE quintile and uses the most reliable earnings expectations. This sample is designed to minimize measurement error. Our in-sample and out-of-sample analysis turns up no convincing evidence that the simple linear model is dominated by any of the nonlinear models.

While previous research has documented nonlinearity in the earnings returns relations, these samples contained both large and small firms. The nonlinearities could be driven by characteristics of these firms and/or measurement error in the expectations data. Our evidence is consistent with the measurement error explanation.

In the end, the message of our paper is simple. If one is studying larger firms, a simple linear model relating unexpected returns to unexpected earnings is hard to beat.

We consider four alternative nonlinear parametric models given the evidence in Abdel-khalik (1990), Beneish and Harvey (1990) and Freeman and Tse (1992). The first is a piecewise linear model:

$$AR_{it} = \alpha_{2i} + \beta_{2i} UE_{it} + \gamma_{2i} D_{it} UE_{it} + \tilde{\epsilon}_{2it} \qquad (A.1)$$

where $D_{it}$ is an indicator variable taking on the value of one if the earnings surprise is negative and zero if it is positive. A quadratic specification is:

$$AR_{it} = \alpha_{3i} + \beta_{3i} UE_{it} + \gamma_{3i} UE_{it}^2 + \tilde{\epsilon}_{3it} \qquad (A.2)$$

For a given earnings surprise, $\frac{\partial AR_{it}}{\partial UE_{it}}$, the response is $\beta_{3i} + 2\gamma_{3i} UE_i$. However, if the quadratic coefficient, $\gamma_{3i}$, is positive, the earnings response function indicates large rewards to both large positive and negative surprises, which seems unreasonable. A specification explored by Freeman and Tse (1992) in their cross-sectional work is:

$$AR_{it} = \alpha_{4i}^* + \beta_{4i}^* UE_{it} + \gamma_{4i}^* D_{4t} UE_{it}^2 + \tilde{\epsilon}_{4it}^* \qquad (A.3)$$

where $D_4$ is an indicator variable taking on the value of 1 when the earnings surprise is positive and -1 when the surprise is negative. The form delivers a convex-concave function. Freeman and Tse predict that $\gamma_{4i}^*$ is negative. Their intuition is that investors' perceptions of the persistence of the surprise is negatively correlated with its magnitude.

Freeman and Tse (1992) and Das and Lev (1994) also pursue a model where the stock return is a function of the arctan of the earnings surprise:

$$AR_{it} = \alpha_{4i} + \beta_{4i} \arctan(\gamma_{4i} UE_{it}) + \tilde{\epsilon}_{4it} \qquad (A.4)$$

For a given earnings surprise, the response function is $\beta_{4i}\gamma_{4i}/(1 + \gamma_{4i}^2 UE_{it}^2)$. This form also provides a convex-concave function.

These models can be extended to incorporate additional terms by a a Taylor series. However, Gallant (1981) argues that the Taylor series expansions are often undesirable because they only apply locally.

Gallant (1981) proposes a Fourier series approximation:

$$
\begin{aligned}
AR_{it} = \alpha_{5i} + \beta_{5i} UE_{it} + \gamma_{5i} UE_{it}^2 &+ \gamma_{6i} sin(UE_{it}^*) + \gamma_{7i} cos(UE_{it}^*) \\
&+ \gamma_{8i} sin(2UE_{it}^*) + \gamma_{9i} cos(2UE_{it}^*) + \tilde{\epsilon}_{5i}
\end{aligned}
\qquad (A.5)
$$

where $UE^* = 2\pi(UE - min(UE))/(max(UE) - min(UE))$ which scales the unexpected earnings to fall in the range of $(0, 2\pi)$. The minimum and maximum are taken over the whole sample. As the number of *sine* and *cosine* terms increases, the in-sample prediction error approaches zero. Following Pagan and Schwert's (1990) application, we truncate the expansion of the Fourier flexible form to two terms.

Nonparametric regression [see Silverman (1986), Härdle (1990)] provides a way to estimate the expected value in (1) without knowing the (parametric) functional form. The nonparametric regression uses the method of kernels to empirically estimate the densities and then uses these densities to obtain fitted values.

The regression function for the conditional mean can be written (by Bayes' theorem) as:

$$E[Y|X = x] = \frac{\int yg(x,y)dy}{\int g(x,y)dy} \qquad (A.6)$$

where $g(x, y)$ is the joint density of **x** and $y$. We assume that there are $k$ variables in **x**.

The discrete-time regression function can be written:

$$\hat{E}[y_t|x_t] = \sum_{j=1}^{T} w_{jt}y_j, \qquad (A.7)$$

where the expected value of $y_t$ is just a weighted sum of all the $y$s and the weights sum to unity. Typically, the weights are defined by least squares:

$$w_{jt} = x_t(x'x)^{-1}x_j. \qquad (A.8)$$

This just follows from the definition of the slope coefficients.

To implement this approach, define a simple kernel function, $K(x)$:

$$K(x) = \begin{cases} \frac{1}{2}, & \text{if } |x| < 1 \\ 0, & \text{otherwise.} \end{cases} \qquad (A.9)$$

The histogram or probability density takes the form:

$$\hat{g}(x) = \frac{1}{hT} \sum_{j=1}^{T} K\left[\frac{x - x_j}{h}\right]. \qquad (A.10)$$

The parameter $h$ is called the bandwidth.

Although the estimator $\hat{g}(x)$ is a density function (nonnegative and integration over the whole space equal to 1), it is not a continuous function because it jumps at points $x_j \pm h$. Notice that $K(x)$ in (A.9) is also a density function called a *kernel* and, if it is replaced by any other density function, the resulting estimator $\hat{g}(x)$ given by (A.10) will be a density function as well. So, if we choose a variety of densities as $K(x)$, we get a variety of *kernel estimators* $\hat{g}(x)$ defined by (A.10). If $K(x)$ is continuous, then so is $\hat{g}(x)$. In fact, $\hat{g}(x)$ preserves all the smooth properties of $K(x)$. Generally, if we choose the kernel $K(x)$ such that

$$\int K(x)dx = 1, \quad \int xK(x)dx = 0, \quad \int x^2K(x)dx < \infty,$$

then, the estimator $\hat{g}(x)$ given by (A.10) will converge to the true density with probability one. Cacoullos (1966) shows that these assumptions are sufficient to guarantee uniform consistency of the estimator. Parzen (1962) proves that the estimator of the density is asymptotically normal.

The choice of $h$ will affect the prediction of the model. As $h$ gets small, the estimated density is less smooth, but the prediction fits the actual data closely. An extreme case is when $h$ is very small, then $w_{jt} = 1$ when $j = t$ and zero elsewhere. This is the case of perfect overfitting. Obviously, this model will fare poorly in any validation exercise.

Besides the bandwidth, the researcher must choose the kernel function. In our analysis, we choose the multivariate Epanechnikov kernel. This kernel is considered 'optimal' in the sense of minimizing the smallest mean squared error achievable [see Silverberg (1988) and Härdle (1990)].

The second kernel we refer to as the *adaptive kernel* because the bandwidth 'adapts' to the local nature of the data. This is particularly useful if there are regions of the data with few observations. Intuition would suggest that we would like to increase the smoothing in areas of sparse data. If a fixed bandwidth is used, some observations could be very influential in determining the fit in these regions of low density. This is exactly analogous to adding variables to a multivariate regression model that (over)fit the extreme observations in the data.

# References

Abarbanell, J. S., W. N. Lanen and R. E. Verrechia, 1996, Analysts' forecasts as proxies for investor beliefs in empirical research. Journal of Accounting and Economics 20, 31–60.

Abdel-khalik, A. R., 1990, Specification problems with information content of earnings: revisions and rationality of expectations, and self-selection bias, Contemporary Accounting Research 7, 142–172.

Ball, R., S.P. Kothari and R.L. Watts, 1993, Economic determinants of the relation between earnings changes and stock returns, The Accounting Review 68, 622–638.

Barron, O. E. and P. S. Stuerke, 1997, Dispersion in analysts earnings forecasts as a measure of uncertainty, unpublished working paper, Indiana University.

Beneish, M. D., 1991, Stock prices and the dissemination of analysts' recommendations, Journal of Business 64, 393–416.

Beneish, M. D. and C. R. Harvey, 1990, A new approach to estimating the earnings response function, Working paper, Duke University.

Bernard, V.L. and J. K. Thomas, 1989, Post-announcement drift: Delayed price response or risk premium? Journal of Accounting Research, 27, 1–36.

Brown, L. D., "Earnings forecasting research: Its implications for capital markets research," International Journal of Forecasting, 1993, Vol 9: 295-320.

Brown, L., P. Griffin, R. Hagerman, and M. Zmijewski, 1987, Security analyst superiority relative to univariate time series models in forecasting quarterly earnings, Journal of Accounting and Economics 9, 61–87.

Cacoullos, T., 1966, Estimation of a multivariate density, Annals of the Institute of Mathematical Statistics 18, 176–189.

Chambers, A.E. and Penman, S.H., 1984, Timeliness of reporting and the stock price reaction to earnings announcements, Journal of Accounting Research 22, 21–47.

Cheng, C.S.A., W.S. Hopwood and J.C. McKeown, 1992, Nonlinearity and specification problems in unexpected earnings response regression model, The Accounting Review 67, 379–398.

Cleveland, W. S. and S. J. Devlin, 1988, Locally weighted regression: An approach to regression analysis by local fitting, Journal of the American Statistical Assocation 83, 596–609.

Collins, D.W. and S.P. Kothari, 1989, An analysis of intertemporal and cross-sectional determinants of earnings response coefficients, Journal of Accounting and Economics 11, 143–181.

Cornell, B., and Landsman, W.R., 1989, Security price response to quarterly earnings announcements and analysts' forecast revisions, The Accounting Review 64, 680–692.

Christie, A.A., 1987, On cross-sectional analysis in accounting research, Journal of Accounting and Economics 9, 231–258.

Das, S. and B. Lev, 1994, Nonlinearity in the returns-earnings relation: Tests of alternative specification and explanations, Contemporary Accounting Research 11, 353–379.

Easton, P. and M. Zmijewski, 1989, Cross-sectional variation in the stock market response to accounting earnings announcements, Journal of Accounting and Economics 11, 117–141.

Fama, E. F. and K. R. French, 1989, Business conditions and expected returns on stocks and bonds, Journal of Financial Economics 25, 23–50.

Fama, E. F. and K. R. French, 1992, The cross section of expected returns, Journal of Finance 42, 427–466.

Fama, E. F. and J. MacBeth, 1973, Risk, return and equilibrium: Empirical tests, Journal of Political Economy 71, 607-636.

Freeman, R. N. and S. Tse, 1992, A nonlinear model of security price responses to unexpected earnings, Journal of Accounting Research 30, 185–209.

Fukunaga, K., 1972, Introduction to statistical pattern recognition, New York: Academic Press.

Gallant, A.R., 1981, On the bias in flexible functional forms and an essentially unbiased form: The Fourier flexible form, Journal of Econometrics 15, 211-224.

Granger, C. W. J. and P. Newbold, 1986, Forecasting economic time series, New York: Academic Press.

Hansen, L. P., 1982, Large sample properties of generalized method of moment estimators, Econometrica 50, 1029–1054.

Härdle, W., 1990, Applied Nonparametric Regression. Cambridge University Press.

Hughes, J. S. and Ricks, W. E., 1987, Association between forecast errors and excess returns near to earnings announcements, The Accounting Review 62, 158–175.

Keim, D. B. and R. F. Stambaugh, 1986, Predicting returns in the bond and stock market, Journal of Financial Economics 17, 357–390.

Lang, M., 1991, Time-varying stock price responses to earnings induced by uncertainty about the time-series properties of earnings, Journal of Accounting Research 27, 229–257.

Lev, B., 1989, On the usefulness of earnings and earnings research: Lessons and directions from two decades of empirical research, Journal of Accounting Research Supplement 27, 153–192.

Mizrach, B., 1992a, Multivariate nearest-neighbour forecasts of EMS exchange rates, Journal of Applied Econometrics 7, S151–S163.

Mizrach, B., 1992b, Forecast comparison in L2, Working paper, Federal Reserve Bank of New York.

Pagan, A. and G.W. Schwert, 1990, Alternative models for conditional stock volatility, Journal of Econometrics 45, 267–290.

Parzen, E., 1962, On the estimation of probability density and mode, Annals of Mathematical Statistics 33, 1065–1076.

Philbrick, D. R. and W. E. Ricks, 1991, Using Value Line and IBES analysts forecasts in accounting research, Journal of Accounting Research 29, 397–415.

Schipper, K., 1991, Analysts' forecasts, Accounting Horizons, 5, 105-121.

Silverman, B. W., 1986, Density estimation for statistics and data analysis, London: Chapman and Hall.

White, H. , 1980, A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, Econometrica 48, 817–838.

# Fitted and Actual Returns: Caterpillar



A. Linear OLS

B. Piecewise Linear

C. Quadratic

D. Arctan

E. Fourier Flexible

F. Nonparametric
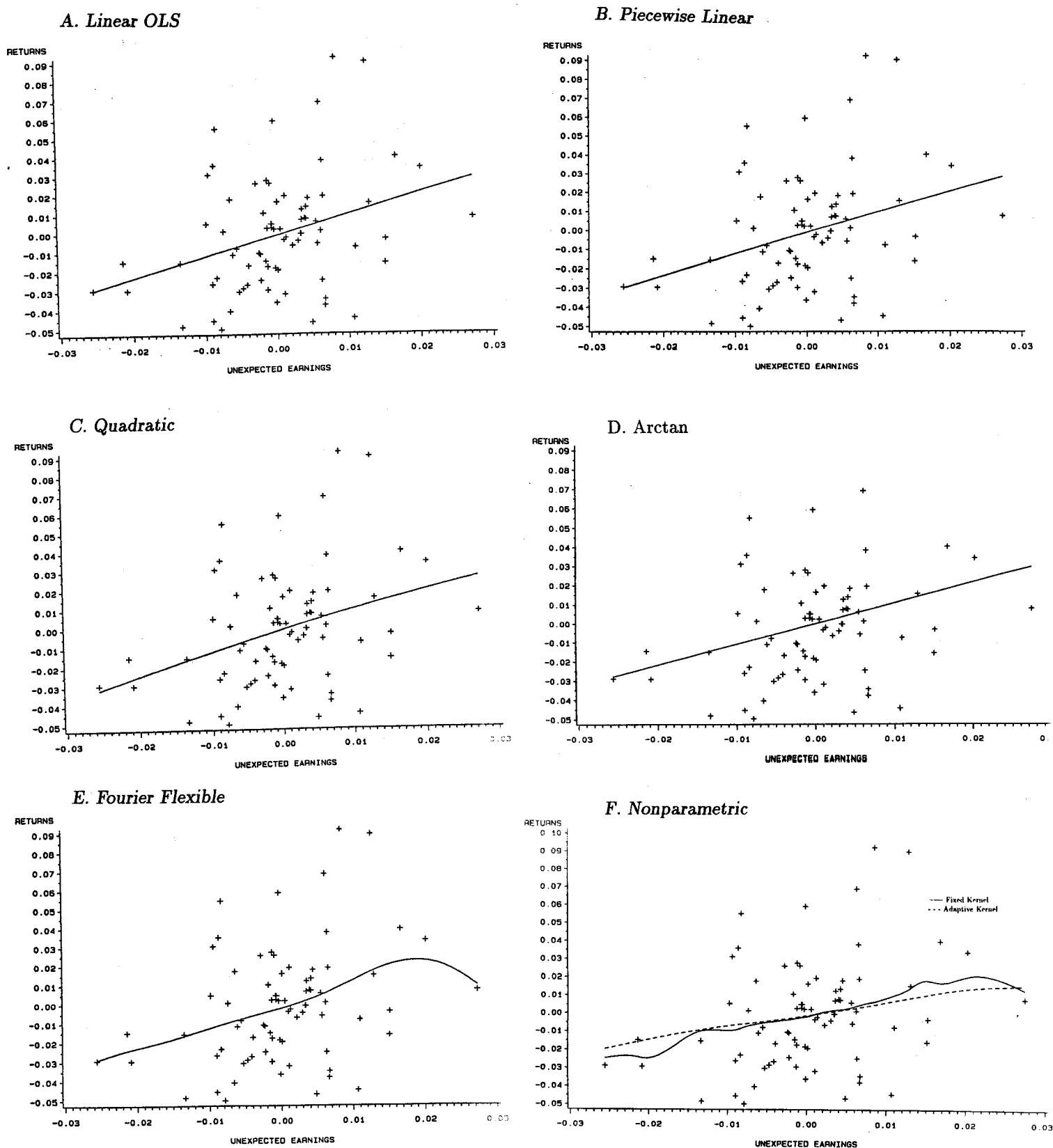
Fig. 1. A comparison of the fitted values from seven different models relating market-adjusted returns to unexpected earnings for Caterpillar.

# Fitted and Actual Returns: Eagle Picher
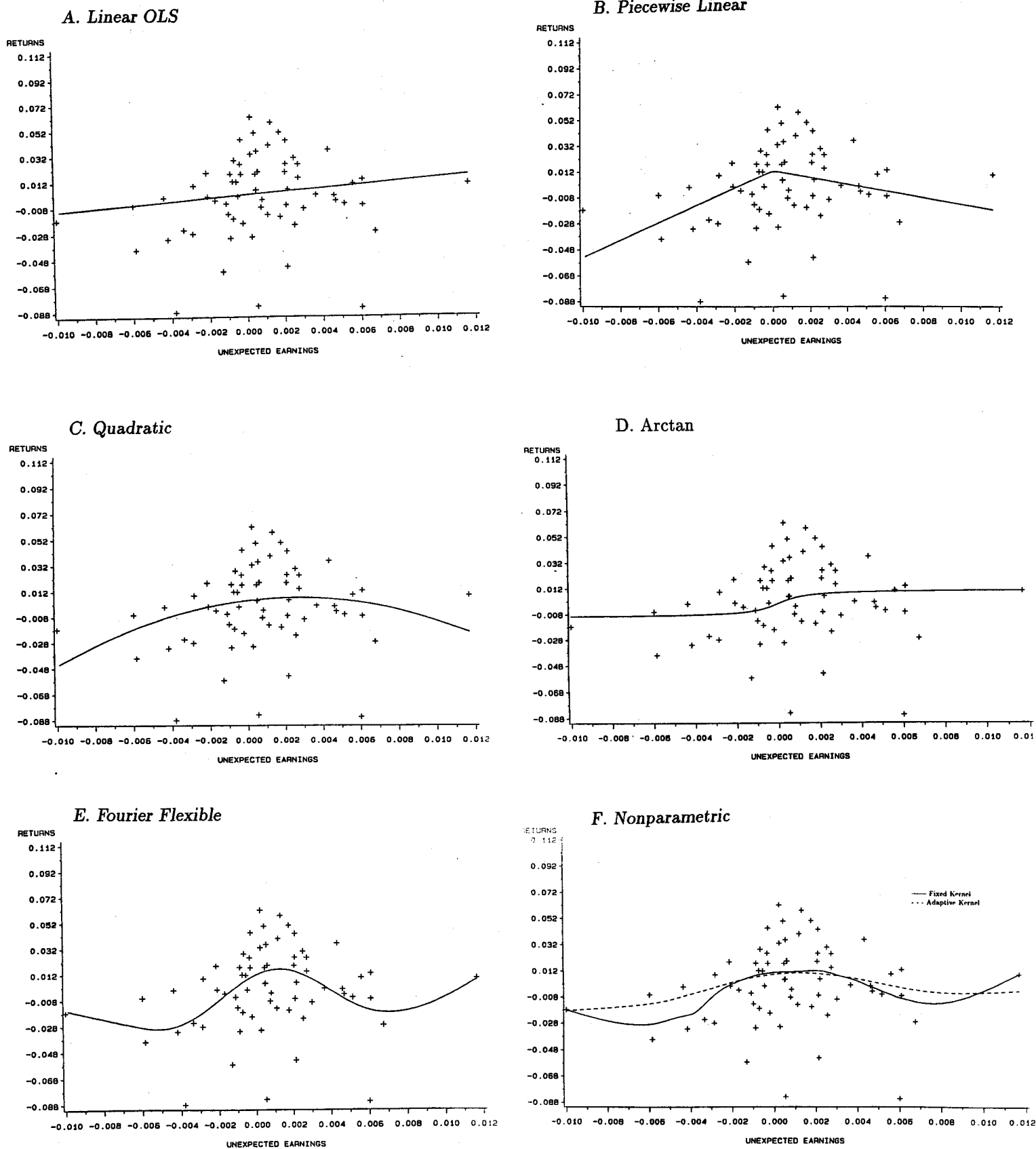


Fig. 2. A comparison of the fitted values from seven different models relating market-adjusted returns to unexpected earnings for Eagle Picher.

## Table 1
### A Comparison of the In-Sample Performance of
### Seven Specifications of the Earnings-Returns Relation
### Individual Estimation for 40 firms over 79 quarters from 1969 to 1988

Panel A:   Goodness-of-fit Statistics

| Mean | Linear[a] | Piecewise | Quadratic | Arctan | Fourier Flexible | Nonparametric Kernel Fixed | Adaptive |
|---|---|---|---|---|---|---|---|
| $R^2$ | .067 | .087 | .091 | .078 | .158 | .133 | .096 |
| Adjusted $R^2$ | .054 | .061 | .065 | .064 | .082 | .121 | .083 |
| MAE[b] | .024 | .024 | .023 | .024 | .022 | .022 | .023 |
| RMSE[b] | .032 | .032 | .032 | .032 | .031 | .031 | .032 |

Panel B:  Model Comparisons:  In-Sample fit

MSE, $R^2$ and Adjusted $R^2$ Comparisons

Time series estimation for 40 firms
Frequencies where, at the 5% level, the:

| Comparison Model | Linear model is not distinguishable from[c] | Linear model dominates | Linear model is dominated by |
|---|---|---|---|
| Piecewise | | | |
| Unadjusted[c] | 100.0% | 0.0% | 0.0% |
| Adjusted[c] | 72.5% | 27.5% | 0.0% |
| Quadratic | | | |
| Unadjusted | 95.0% | 2.5% | 2.5% |
| Adjusted | 40.0% | 60.0% | 0.0% |
| Arctan | 100.0% | 0.0% | 0.0% |
| Fourier Flexible | | | |
| Unadjusted | 97.5% | 0.0% | 2.5% |
| Adjusted | 70.0% | 30.0% | 0.0% |
| Nonparametric (fixed kernel) | 90.0% | 0.0% | 10.0% |
| Nonparametric (adaptive kernel) | 90.0% | 0.0% | 10.0% |

a. **Linear Model:**  $AR_{it} = \alpha_1 + \beta_1 UE_{it} + \tilde{\varepsilon}_{1it}$

   **Piecewise Linear Model:**  $AR_{it} = \alpha_2 + \beta_2 UE_{it} + \gamma_2 D_1 UE_{it} + \tilde{\varepsilon}_{2it}$

   **Quadratic Model:**  $AR_{it} = \alpha_3 + \beta_3 UE_{it} + \gamma_3 UE_{it}^2 + \tilde{\varepsilon}_{3it}$

   **Arctan Model:**  $AR_{it} = \alpha_4 + \beta_4 \text{Arctan}(\gamma_4 UE_{it}) + \tilde{\varepsilon}_{4it}$

   **Fourier Flexible Model:**  $AR_{it} = \alpha_5 + \beta_5 UE_{it} + \gamma_5 UE_{it}^2 + \gamma_6 \sin(UE_{it}^*) + \gamma_7 \cos(UE_{it}^*) + \gamma_8 \sin(2UE_{it}^*) + \gamma_9 \cos(2UE_{it}^*) + \tilde{\varepsilon}_{5it}$

   **Nonparametric Model [Fixed Kernel]:**  $AR_{it} / \alpha_6 + \beta_6 f(UE_{it}) + \tilde{\varepsilon}_{6it}$

   **Nonparametric Model [Adaptive Kernel]:**  $AR_{it} = \alpha_7 + \beta_7 f(UE_{it}) + \tilde{\varepsilon}_{7it}$

**Variable definitions:**

$AR_{it}$ = Abnormal return in days -1 and 0 relative to the day of the earnings announcement in quarter t in the period 1969-1988. Market model parameters are estimated using days -360 to -61 relative to the earnings announcement.

$UE_{it}$ = Value Line Earnings forecast error for quarter t deflated by the price of a firm's common stock on day -2.

$D_1$ = Slope dummy taking values of 1 when $UE_{it}$ is negative and 0 otherwise.

Arctan is the inverse tangent; the model is estimated using the same method as Freeman and Tse (1992), that is, with the Gauss-Newton iterative method.

$UE_{it}^*$ = $UE_{it}$ scaled to lie in the interval $[0, 2\Pi]$. The scaling is given by:

$$UE_{it}^* = \frac{2\Pi(UE_{it} - \min(UE_{it}))}{Max(UE_{it} - Min(UE_{it}))}$$

b. Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) are computed for each firm and model as:

$$\frac{1}{N}\sum_{t=1}^{N} |AR_{it} - PRED_{it}| \quad \text{and} \quad \left[\frac{1}{N}\sum_{t-1}^{N}(AR_{it} - PRED_{it})^2\right]^{1/2}$$

where $AR_{it}$ is the actual abnormal return, $PRED_{it}$ denotes the fitted value for each model and N the number of observations.

c. The tests denoted unadjusted are based on equations (3) and (4) for MSE and $R^2$ comparisons. The tests denoted adjusted are based on equation (4a) for adjusted-$R^2$ comparisons. A significant positive (negative) covariance indicates that the linear model is dominated by (dominates) the comparison model. When the covariance is not distinguishable from zero at the 5% level, the linear model is not distinguishable from the comparison model.

## Table 2
### A Comparison of the In-Sample Performance of
### Seven Specifications of the Earnings-Returns Relation
### Cross-sectional Estimation for 79 quarters from 1969 to 1988

Panel A: Goodness-of-fit Statistics

| Mean | Linear[*] | Piecewise | Quadratic | Arctan | Fourier Flexible | Nonparametric Kernel Fixed | Adaptive |
|------|--------|-----------|-----------|--------|---------|--------------------|----------|
| $R^2$ | .079 | .115 | .119 | .127 | .252 | .161 | .119 |
| Adjusted $R^2$ | .053 | .062 | .066 | .075 | .099 | .136 | .093 |
| MAE | .024 | .023 | .023 | .023 | .020 | .022 | .023 |
| RMSE | .032 | .032 | .032 | .032 | .031 | .030 | .031 |

Panel B: Model Comparisons In-Sample fit

MSE, $R^2$ and Adjusted $R^2$ Comparisons

Cross-sectional estimation in 79 quarters
Frequencies where, at the 5% level, the:

| Comparison Model | Linear model is not distinguishable from | Linear model dominates | Linear model is dominated by |
|------------------|------------------------------------------|------------------------|------------------------------|
| Piecewise | | | |
| Unadjusted | 100.0% | 0.0% | 0.0% |
| Adjusted | 67.1% | 32.9% | 0.0% |
| Quadratic | | | |
| Unadjusted | 100.0% | 0.0% | 0.0% |
| Adjusted | 68.4% | 31.6% | 0.0% |
| Arctan | 100.0% | 0.0% | 0.0% |
| Fourier Flexible | | | |
| Unadjusted | 84.8% | 0.0% | 15.2% |
| Adjusted | 69.6% | 30.4% | 0.0% |
| Nonparametric (fixed kernel) | 96.2% | 0.0% | 3.8% |
| Nonparametric (adaptive kernel) | 96.2% | 0.0% | 3.8% |

[*] Model specifications, variable definitions and test explanations are in Table 1.

## Table 3

### A Reproduction of and Comparison to
### Freeman and Tse's (1992) tests

**Panel A:** A summary of Freeman and Tse's (1992), Table 1, Panel B, p. 195

| \|UE\| range[a] | N- | % | Linear Adj-$R^2$ | Arctan Adj-$R^2$ |
|---|---|---|---|---|
| \|UE\| < ∞ | 12,381 | 100.0% | .004 | .072 |
| \|UE\| < =.1 | 12,035 | 97.2% | .039 | .075 |
| \|UE\| < =.05 | 11,680 | 94.4% | .052 | .075 |
| \|UE\| < =.01 | 9,363 | 75.6% | .049 | .053 |
| \|UE\| < =.005 | 7,361 | 59.5% | .038 | .038 |
| \|UE\| < =.001 | 2,768 | 22.4% | .007 | .007 |

**Panel B:** A reproduction of their tests with our sample

| \|UE\| range[a] | N- | % | Linear Adj-$R^2$ | Arctan Adj-$R^2$ |
|---|---|---|---|---|
| \|UE\| < ∞ | 2,929 | 100.0% | .030 | .056[b] |
| \|UE\| < =.1 | 2,925 | 99.9% | .031 | .055[b] |
| \|UE\| < =.05 | 2,909 | 99.3% | .034 | .034 |
| \|UE\| < =.01 | 2,541 | 86.8% | .040 | .040 |
| \|UE\| < =.005 | 2,064 | 70.5% | .038 | .039 |
| \|UE\| < =.001 | 847 | 28.9% | .003 | .003 |

[a] The mean absolute earnings surprise in our sample (.0051) is smaller than in Freeman and Tse (1992) (.0159) a result that we attribute to our use of Value Line actual and forecasts (see Philbrick and Ricks (1991)) and to our sample consisting of large firms. The distributions in our respective samples according to \|UE\| range are consistent with our smaller forecast errors.

[b] While we cannot test for equality of adjusted-$R^2$s in Freeman and Tse's sample, we find, using the test described in Table 1, that the arctan model dominates the linear model when the 20 largest \|UE\| are included.

## Table 4
### A Comparison of the Out-of-Sample Predictive Ability of
### Seven Specifications of the Earnings-Returns Relation
### Estimated with Time-Series Data for each Firm
### Statistics Based on One-Step Ahead Forecasts for 12 Quarters

**Panel A: Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) Summary Statistics**

| Mean | Linear[a] | Piecewise | Quadratic | Modified Quadratic | Fourier Flexible | Nonparametric Kernel Fixed | Nonparametric Kernel Adaptive |
|------|-----------|-----------|-----------|--------------------|------------------|----------------------------|-------------------------------|
| MAE[b] | .027 | .027 | .027 | .027 | .031 | .027 | .027 |
| RMSE[b] | .034 | .035 | .035 | .035 | .043 | .035 | .034 |

**Panel B: Model Comparisons: Out of Sample**

MSE Comparisons

12 one-step ahead forecasts for 40 firms
Frequencies where, at the 5% level, the:

| Comparison Models | Linear model is not distinguishable from[c] | Linear model dominates | Linear model is dominated by |
|-------------------|---------------------------------------------|------------------------|------------------------------|
| Piecewise | 85.0% | 15.0% | 0.0% |
| Quadratic | 95.0% | 5.0% | 0.0% |
| Arctan | 95.0% | 5.0% | 0.0% |
| Fourier Flexible | 95.0% | 5.0% | 0.0% |
| Nonparametric (fixed kernel) | 97.5% | 0.0% | 2.5% |
| Nonparametric (adaptive kernel) | 97.5% | 0.0% | 2.5% |

a. The seven specifications are described in Table 1.

b. Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) are given by for each model and firm:

$$MAE_t = \frac{1}{12} \sum_{j=1}^{12} |AR_{it} - PRED_i| \quad and \quad RMSE = \left[ \frac{1}{12} \sum_{t=1}^{12} (AR_{it} - PRED_i)^2 \right]^{1/2}$$

where $AR_{it}$ is the abnormal return and $PRED_{it}$ the predicted abnormal return for each firm and model.

c. The tests are described in Table 1.

## Table 5
### A Comparison of the Predictive Ability of
### Seven Specifications of the Earnings-Returns Relation
### Estimated Cross Sectionally each Quarter
### Statistics Based on One-Step Ahead Forecasts for 78 Quarters

**Panel A: Mean Absolute Error (MAE) Summary Statistics**

| Mean | Linear* | Piecewise | Quadratic | Arctan | Fourier Flexible | Nonparametric Kernel Fixed | Adaptive |
|------|---------|-----------|-----------|--------|---------|-------|----------|
| MAE  | .025    | .026      | .027      | .027   | .563    | .025  | .025     |
| RMSE | .034    | .035      | .039      | .041   | .721    | .034  | .033     |

**Panel B: Model Comparisons: out of sample**

MSE Comparison

78 one quarter ahead forecasts
Frequencies where, at the 5% level, the:

| Comparison Models | Linear model is not distinguishable from | Linear model dominates | Linear model is dominated by |
|-------------------|------------------------------------------|------------------------|------------------------------|
| Piecewise | 94.9% | 3.8% | 1.3% |
| Quadratic | 92.4% | 3.8% | 3.8% |
| Arctan | 92.3% | 5.1% | 2.6% |
| Fourier Flexible | 41.0% | 59.0% | 0.0% |
| Nonparametric (fixed kernel) | 91.1% | 3.8% | 3.8% |
| Nonparametric (adaptive kernel) | 91.1% | 5.1% | 3.8% |

* Model specifications, variable definitions and test explanations are described in Table 1.

## Table 6
### Effect of Other Conditioning Information
### on the Earnings-Return Relation:
### Additive and Interactive Models[a]
### Estimated for Each of 40 Firms Followed by Value Line From 1969-1988

(1) Additive: $AR_{it}=a_o+a_1 UE_{it}+a_2 R_i(-60,-2)+a_3\sigma^2_{Rit}+a_4 SPREAD_t+\eta^b_{it}$

(2) Interactive: $AR_{it}=a_o+a_1 UE_{it}+a_2 UE^*_{it}R_i(-60,-2)+a_3 UE^*_{it}\sigma^2_{Rit}+a_4 UE^*_{it}SPREAD_t+\upsilon_{it}$

---

Panel A:     Linear Model

Additive Model
Descriptive Statistics for 40 Firms Estimation

|  | $a_1$ | $a_2$ | $a_3$ | $a_4$ | Adj.$R^2$ | DW | Exclusion[e] F |
|---|---|---|---|---|---|---|---|
| Mean | 1.78 | -.014 | .032 | -.002 | .080 | 1.97 | 1.71 |
| Median | 1.06 | -.008 | .024 | -.001 | .081 | 1.93 | 1.22 |
| # positive | 38 | 16 | 22 | 31 | 36 | --- | --- |
| # significant at 5% | 19 | 5 | 3 | 5 | --- | --- | 9 |

Interactive Model
Descriptive Statistics for 40 Firms Estimation

|  | $a_1$ | $a_2$ | $a_3$ | $a_4$ | Adj.$R^2$ | DW | Exclusion[d] F |
|---|---|---|---|---|---|---|---|
| Mean | 2.62 | -.011 | 17.69 | -.234 | .073 | 1.94 | 1.52 |
| Median | 2.18 | -.010 | 5.11 | -.071 | .064 | 1.97 | 1.44 |
| # positive | 32 | 17 | 21 | 11 | 37 | --- | --- |
| # significant at 5% | 8 | 6 | 4 | 1 | --- | --- | 5 |

**Panel B:**     Nonlinear Models Summary Statistics

| Additive | Piecewise | Quadratic | Arctan | Fourier Flexible |
|---|---|---|---|---|
| **Individual Firms (N = 40)** | | | | |
| Mean Adj $R^2$ | .084 | .087 | .081 | .111 |
| Median Adj $R^2$ | .085 | .086 | .080 | .110 |
| # significant Exclusion tests (5%) | 8 | 6 | 6 | 8 |
| Frequency where linear model is not distinguishable from:[d] | 100% | 97.5% | 100% | 95% |

| Interactive | Piecewise | Quadratic | Arctan | Fourier Flexible |
|---|---|---|---|---|
| **Individual Firms (N = 40)** | | | | |
| Mean Adj. $R^2$ | .088 | .093 | .088 | -[e] |
| Median Adj. $R^2$ | .076 | .077 | .089 | --- |
| % significant exclusion tests (5%) | 6 | 4 | 9 | --- |
| Frequency where linear model is not distinguishable from: | 97.5% | 95.0% | 95.0% | --- |

[a]    The interactive model is a test of the following general model $E_t\ (AR_{it}/UE_{it}, Z_t) = f_i(AR_{it}(Z_t), Z_t)$ where $Z^t$ is the information set at time t. In our specification, the interaction terms explain the effect on the earnings response coefficients of investors expectations about future discount rates and are expected to be negative. By contrast, the additive model is a test of the following general model $E_t(AR_{it}/UE_{it}, Z_t) = f_i(UE_{it}, Z_t)$. In this model, additional conditioning information captures whether abnormal returns are explained by these variables. Both models are presented to allow for two different general model forms. While collinearity is more likely a problem in the interactive model, the highest condition index ranges between 6.72 and 23.19 across functional specifications. According to Belsley et al. (1980) condition numbers lower than 30 are not indicative of collinearity effects.

[b]    Variable definitions.
$AR_{it}$ and $UE_{it}$ are as defined in Table 1 footnote a. $R_{it}$ (-60, -2) is firm i's return from day -60 to day -2 relative quarter t earnings announcement date, $\sigma^2_{Rit}$ the variance of returns for day -60 to day -2 relative to quarter t earnings announcement date and SPREAD$_t$ the difference in the yields between Moody's Baa and Aaa rate bonds in quarter t.

[c]    F tests that the coefficients on other conditioning variables are zero. If N is the number of observations in each estimation and number of regressors the statistic is distributed F(3, N-k-3).

[d]    The test is described in Table 1.

[e]    The fourier flexible is not estimated interactively firm by firm because it requires the estimation of 20 regressors with an average 70 observations. Similarly, the number of observations in each time-series prevents the application of the nonparametric density estimation.

## Table 7
### Effect of Cross-Sectional Aggregation:
### Additive and Interactive Models[*]
### of the Earnings-Returns Relation Estimated with 2929 Observations

(1)     Additive:     $AR_{it} = a_o + a_1 UE_{it} + a_2 R_i(-60, -2) + a_3 \sigma_{Rit}^2 + a_4 SPREAD_t + \eta_{it}^b$

(2)     Interactive:     $AR_{it} = a_o + a_1 UE_{it} + a_2 UE_{it}^* R_i(-60, -2) + a_3 UE_{it}^* \sigma_{Rit}^2 + a_4 UE_{it}^* SPREAD_t + \upsilon_{it}$

Panel A:     Linear Model

| | $a_1$ | $a_2$ | $a_3$ | $a_4$ | Adj.$R^2$ | Average Adj $R^2$ (individual firms) |
|---|---|---|---|---|---|---|
| Additive Model | .883 (4.65)[*] | -.008 (-1.67)[**] | 1.62 (1.19) | -.077 (-1.95)[**] | .031 | .080 |
| Interactive Model | .567 (9.12)[*] | -.008 (-1.60) | .021 (2.16)[*] | -.002 (-3.06)[*] | .034 | .073 |

Panel B:     Nonlinear Models

| | Additive Model | | Interactive Model | |
|---|---|---|---|---|
| | Aggregate Estimation Adj$R^2$ | Average Adj$R^2$ (individual firms) | Aggregate Estimation Adj$R^2$ | Average Adj$R^2$ (individual firms) |
| Piecewise | .038 | .084 | .033 | .088 |
| Quadratic | .038 | .087 | .034 | .093 |
| Arctan | .058 | .081 | .052 | .088 |
| Fourier Flexible | .041 | .111 | ---[c] | --- |

a. The interactive model is a test of the following general model $E_t (AR_{it}/UE_{it}, Z_t) = f_i(AR_{it}(Z_i), Z_t)$ where $Z_t$ is the information set at time t. In our specification, the interaction terms explain the effect on the earnings response coefficients of investors expectations about future discount rates and are expected to be negative. By contrast, the additive model is a test of the following general model $E_t(AR_{it}/UE_{it}, Z_t) = f_i(UE_{it}, Z_t)$. In this model, additional conditioning information captures whether abnormal returns are explained by these variables. Both models are presented to allow for two different general model forms. While collinearity is more likely a problem in the interactive model, the highest condition index ranges between 6.39 and 22.38 across functional specifications. According to Belsley et al. (1980) condition numbers lower than 30 are not indicative of collinearity effects.

b. Variable definitions.
$AR_{it}$ and $UE_{it}$ are as defined in Table 1 footnote a. $R_{it}$ (-60, -2) is firm i's return from day -60 to day -2 relative quarter t earnings announcement date, $\sigma^2_{R_{it}}$ the variance of returns for day -60 to day -2 relative to quarter t earnings announcement date and SPREAD$_t$ the difference in the yields between Moody's Baa and Aaa rate bonds in quarter t.

c. The fourier flexible is not estimated interactively firm by firm because it requires the estimation of 20 regressors with an average 70 observations. Similarly, the number of observations in each time-series prevents the application of the nonparametric density estimation.

* (**) significant at the 5% (10%) level (two-tailed test)