A Regression Test of Semiparametric Index Model Specifications

by

Diego Rodriguez and Thomas M. Stoker

MIT-CEEPR 93-011WP July 1993

A REGRESSION TEST OF SEMIPARAMETRIC INDEX MODEL SPECIFICATION

Diego Rodriguez and Thomas M. Stoker*

February 1992

Revised July 1993

* Department of Economics, MIT and Sloan School of Management, MIT, respectively. This research was funded by a grant from the MIT Center for Energy and Environmental Policy Research. The authors thank Gary Chamberlain, Joel Horowitz, Whitney Newey, Jim Powell, Hal White and Jeff Wooldridge for helpful comments.

ABSTRACT

This paper presents a straightforward regression test of parametric and semiparametric index models against more general semiparametric and nonparametric alternative models. The test is based on the regression coefficient of the restricted model residuals on the fitted values of the more general model. A goodness-of-fit interpretation is shown for the regression coefficient, and the test is based on the squared "t-statistic" of the coefficient estimate, where the variance of the coefficient has been adjusted for the use of nonparametric estimators. An asymptotic theory is given for the situation where kernel estimators are used to estimate unknown regression functions, and the variance adjustment terms are given for this case. The methods are applied to the empirical problem of characterizing environmental effects on housing prices in the Boston Housing Data, where a partial index model is found to be preferable to a standard log-linear equation, yet not rejected against general nonparametric regression. Various issues in the asymptotic theory and other features of the test are discussed.

A REGRESSION TEST OF SEMIPARAMETRIC INDEX MODEL SPECIFICATION

by Diego Rodriguez and Thomas M. Stoker

1. Introduction

The purpose of this paper is to propose and illustrate a straightforward specification test for index models. The test can be used to judge the empirical adequacy of parametric index models; such as a linear model or a probit model, against more general semiparametric or nonparametric models. Alternatively, the test can be used to judge the restrictions of a semiparametric partial index model, against more general semiparametric or nonparametric alternatives. As such, the test is intended as a diagnostic tool to be used in conjunction with empirical estimation of index models. We illustrate the methodology in a study of the index structure of pollution effects in the Boston Housing Data.

The test is based on the bivariate OLS coefficient of the residuals from the restricted model regressed on the fitted values from the general model. The test statistic is the square of the "t-statistic", or the ratio of the slope coefficient to its estimated standard error; which is compared to a $\chi^2(1)$ critical value. The value of the coefficient has a "goodness-of-fit" interpretation, as the percentage of variation of the general model that is not accounted for by the restricted model; and the restricted model is rejected when the coefficient is significantly different from zero.
Our proposal amounts to a completely standard test of a zero coefficient value, with one important proviso. In particular, the estimate of the standard error of the coefficient must be adjusted for the use of estimates (of parameters and functions) in both the restricted model and the general model.

The test is similar in spirit to the test of a linear model against nonparametric alternatives proposed by Wooldridge (1992) and Yatchew (1992), and related work by Hong and White (1991, 1993), Ellison and Fisher-Ellison (1992), Horowitz and Härdle (1992) and Eubank and Spiegelman (1990), among others. In line with the discussion of Hong and White (1991), this work is related to tests of moment restrictions as in Bierens (1990) and Lewbel (1991).

Our approach differs from the earlier proposals in that a wide range of restricted and general models are allowed, and that our test is based on an adjustment of the familiar "t-statistic." Our development of the limiting statistical theory of the test is based on index models, although similar tests could be devised for situations where the restricted and general models are nested in a particular fashion that we discuss below. We give the standard error adjustment terms that arise when kernel estimation methods are used to estimate unknown functions and index model coefficients. From Newey (1991), it is natural to conjecture that the asymptotic theory for the test will be the same when other kinds of nonparametric estimators are used, but the relevant adjustments would need to be derived.[1]

The exposition proceeds as follows. We begin with a brief layout of the models and the test in Section 2. Section 3 applies the test in an analysis of pollution effects on housing prices using the Boston Housing Data of Harrison and Rubinfeld (1978a, 1978b). Section 4 gives asymptotic approximation theory for the test statistic, as well as discussing implementation details and technical issues. Section 5 contains some concluding remarks.

2. Basic Layout

2.1 Basic Framework and Index Models

The empirical setting we consider is an analysis of data (y_i, x_i), $i = 1, \ldots, N$, which is assumed to be an i.i.d. random sample, where y_i is a response of interest and x_i is a k-vector of predictor variables. For the statistical theory of Section 4, we assume that x is continuously distributed with density f(x), where f(x) vanishes on the boundary of x values, and is also first differentiable. We assume that the mean of y exists, and denote the mean regression of y on x as $m(x) = E(y|x)$.

Our interest is in testing index model restrictions on the structure of m(x). To begin, m(x) is a single index model if there is a coefficient vector β and a univariate function G such that

$$m(x) = G(x^T\beta) \quad . \tag{2.1}$$

Familiar parametric models that are single index models include the standard linear model; $y = \alpha + x^T\beta + \epsilon$ with $E(\epsilon|x) = 0$; giving

$$m(x) = \alpha + x^T\beta \quad . \tag{2.2}$$

Likewise included is the standard probit model for analyzing binary responses; $y = 1[\epsilon < \alpha + x^T\beta]$ with $\epsilon \sim \mathcal{N}(0,1)$; giving

$$m(x) = \Phi(\alpha + x^T\beta) \tag{2.3}$$

with $\Phi(.)$ the cumulative normal distribution function.

A semiparametric single index model is written as

$$m(x) = G_1(x^T\beta) \quad , \tag{2.4}$$

which is in form (2.1), but G_1 is treated as an unknown, smooth univariate function. Estimation of (2.4) involves estimation of the coefficients β and the univariate function $G_1(.)$. A semiparametric partial index model is based on

$$(2.5) \qquad m(x) = G_2(x_1^T\beta_1, x_2)$$

where $x = (x_1, x_2)$ is a partition of x into a $k\text{-}k_2$ vector x_1, and a k_2 vector x_2, and G_2 is an unknown, smooth function of $k_2 + 1$ arguments. This form is useful if the impacts of some variables, here x_2, require more flexible treatment than permitted by the single index model (2.4). Our test is applicable to testing a restricted index model (for instance (2.4)), against a more general index model (for instance (2.5)).[2]

At the extreme end of generality, we consider the nonparametric regression model

$$(2.6) \qquad m(x) = g(x)$$

where g(x) is an unknown smooth function of k arguments. Failure to reject a proper index model against the general nonparametric regression constitutes practical acceptance of the index model restrictions. Likewise, failure to reject a parametric index model against the nonparametric regression constitutes practical acceptance of the parametric regression restrictions.

Our empirical and theoretical analysis employs kernel estimators for unknown functions in semiparametric and nonparametric regression models, and (kernel) average derivative estimators for index model coefficients.[3] The latter refer to using an instrumental variables estimator of the vector $\delta = E(m')$, where $m' = \partial m/\partial x$. For model (2.4), the coefficients β are proportional to δ, so we normalize the model by replacing β by δ, as in

$$(2.7) \qquad m(x) = G_1(x^T\delta) \quad ,$$

redefining G_1 to reflect the scale normalization. As such, δ is estimated directly (giving $\hat{\delta}$, say), and then G_1 is estimated by kernel regression of y on $x^T\hat{\delta}$.

Likewise, for the partial index model (2.5), we have that β_1 is proportional to the $k - k_2$ subvector δ_1 of δ (those components associated with x_1), and so we normalize (2.5) as

$$m(x) = G_2({x_1}^T\delta_1, x_2) \quad . \tag{2.8}$$

We denote estimators using hats; $\hat{\delta}$, $\hat{G}_1$, $\hat{G}_2$, $\hat{g}$, etc. One attractive feature of the index model framework is that a single consistent estimate of the average derivative vector δ can be used to estimate the relevant coefficients in single and partial index models as in (2.7) and (2.8).

We give the formulae for the kernel estimators used in Section 4.1. For clarity of the main theme, we now give a quick introduction to the ideas of the specification test, and follow it with an empirical application. We set aside many of the required technical details, deferring them until Section 4.

2.2 Quick Start: The Test and Its Motivation

We introduce the test by considering the problem of testing a (semiparametric) single index model against general (nonparametric) regression structure. In particular, the null hypothesis is that the true regression takes the restricted form

$$m(x) = G_1(x^T\delta) \quad . \tag{2.9}$$

The alternative is represented by

$$m(x) = g(x) \tag{2.10}$$

where g(x) obeys the smoothness conditions given in Section 4.2. The methods for applying the test with other restricted and alternative models will be clear from considering this case. Using the data (y_i, x_i), $i = 1,\ldots,N$, we assume that an estimator $\hat{\delta}$ of δ is computed, that G_1 is estimated by the kernel regression $\hat{G}_1$ of y_i on $x_i^T\hat{\delta}$, and that g is estimated by the kernel regression $\hat{g}$ of y_i on x_i. Following the results of Härdle and Stoker (1989), these procedures imply the $\hat{G}_1(x^T\hat{\delta})$ is a consistent (nonparametric) estimator of $E(y|x^T\delta)$ in general (i.e. with model (2.10)), so that when (2.9) is valid, $\hat{G}_1(x^T\hat{\delta})$ is a consistent nonparametric estimator of $G_1(x^T\delta)$.

The test statistic is computed as follows: for each observation i, form the residual from the restricted model $y_i - \hat{G}(x_i^T\hat{\delta})$ and the fitted value from the general model $\hat{g}(x_i)$, and then perform the bivariate OLS regression

$$(2.11) \qquad y_i - \hat{G}(x_i^T\hat{\delta}) = \hat{\alpha} + \hat{\gamma}\,\hat{g}(x_i) + \hat{u}_i \quad , \; i = 1,\ldots,N.$$

The test is based on the value of $\hat{\gamma}$; if large (indicating a significant difference from zero), we reject the single index model against the general regression; otherwise, we fail to reject.[4] In particular, if an estimate of the asymptotic variance of $\hat{\gamma}$ is denoted $\hat{\sigma}_\gamma$, then the appropriate "t value" is found as

$$(2.12) \qquad t = \sqrt{N}\,\hat{\gamma} \;/\; \sqrt{\hat{\sigma}_\gamma} \quad .$$

Our test compares t^2 to a $\chi^2(1)$ critical value. We discuss the estimate $\hat{\sigma}_\gamma$ below, following some basic motivation.

On broad grounds, basing a test on $\hat{\gamma}$ is sensible because if (2.9) is the true model, $y - G_1(x^T\delta)$ is uncorrelated with any function of x. Provided that $\hat{G}_1(x^T\hat{\delta})$ is an accurate estimator of $G_1(x^T\delta)$, then $y - \hat{G}_1(x^T\hat{\delta})$ should be approximately uncorrelated with $\hat{g}(x)$, which is what is being checked. More formally, suppose $G(x^T\delta) = E(y|x^T\delta)$ denotes the consistent limit of $\hat{G}_1(x^T\hat{\delta})$.

Consider the linear regression equation that holds if the true functions G and g were known:

$$(2.13) \qquad y - G(x^T\delta) = \alpha + \gamma\, g(x) + u$$

where the parameter γ is defined via OLS projection, as

$$(2.14) \qquad \gamma = \frac{E([g(x)-E(g)][y - G(x^T\delta)])}{E[g(x)-E(g)]^2} \quad .$$

Here u is uncorrelated with g(x) by definition. Equation (2.11) is just the sample analogue of the equation (2.13). Obviously, $\gamma = 0$ when $g(x) = G(x^T\delta)$, reflecting the lack of correlation.

The value of γ is also easy to characterize under the alternative, when $g(x) \neq G(x^T\delta)$. In particular, from the law of iterated expectations, we have that

$$(2.15) \qquad G(x^T\delta) = E[y|x^T\delta] = E[g(x)|x^T\delta].$$

Consequently,

$$(2.16) \qquad \begin{aligned} g(x) &= E[g(x)|x^T\delta] + (g(x) - E[g(x)|x^T\delta]) \\ &= G(x^T\delta) + U(x) \end{aligned}$$

where $U(x) = g(x) - E[g(x)|x^T\delta]$ has mean 0 conditional on $x^T\delta$. Therefore

$$(2.17) \qquad \gamma = \frac{E[U(x)^2]}{E[g(x)-E(g)]^2} > 0$$

when g(x) differs from $G(x^T\delta)$ on a set of positive probability. Therefore, γ is the percentage of (structural) variance of the true regression not accounted for by the restricted model. The statistic $\hat{\gamma}$ is an empirical measure of this "goodness of fit" value. The key feature of this motivation is that the restricted regression is the expectation of the general regression, conditional on the index argument(s) of the restricted model. This "nesting" is easily verified for comparing semiparametric index models (any coefficients in the general model must also be coefficients of the restricted model), and is assured by using kernel estimators for unknown functions and average derivative estimators for coefficients.[5]

We now describe how we measure the variance of $\hat{\gamma}$. If the parameters δ and the functions G and g were known, then the variance of $\hat{\gamma}$ would be consistently measured by the standard (White) heteroskedasticity consistent variance estimator. Our approach is to add adjustments to the standard term, to account for the presence of the estimates $\hat{\delta}$, $\hat{G}$ and $\hat{g}$. In particular, $\hat{\sigma}_\gamma$ is the sample variance of

$$(2.18) \qquad s_{\hat{g}}^{-1}\{[\hat{g}(x_i) - \bar{g}]\,\hat{u}_i + ra_i - la_i\}$$

where $\bar{g}$ and $s_{\hat{g}}$ are the sample average and sample variance of $\hat{g}(x_i)$ respectively, and $\hat{u}_i = y_i - \hat{G}_1(x_i^T\hat{\delta}) - \hat{\gamma}\,[\hat{g}(x_i) - \bar{g}]$ is the estimated residual. The term ra_i is the adjustment for the estimation of $g(x_i)$ (the "right-hand" function), and the term la_i is the adjustment for the estimation of $G(x_i^T\delta)$ (the "left-hand" function). These terms are spelled out in Section 4 and Appendix 2. It should be noted that the standard (White) variance statistic is constructed from (2.18) with $ra(x_i)$ and $la(x_i)$ omitted.

With this motivation, we now turn to an empirical example.

3. Index Structure of the Boston Housing Data

We illustrate the test by studying the index structure of the Boston Housing Data of Harrison and Rubinfeld (1978a,b). The focus of that study is on measuring environmental effects on housing prices, for the purpose of measuring the dollar-value benefits of lower air pollution levels, and their method of analysis is to estimate a standard log-linear hedonic price equation. All our nonparametric estimation uses kernel regression estimators, and testing is performed on a "trimmed" sample, that omits the 5% of the observations that displayed smallest estimated density values.

This data and the log-linear price equation has been extensively studied elsewhere, for instance, in the work of Belsley, Kuh and Welsch (1980) on regression diagnostics, among others. There is no particularly persuasive theoretical reason for choosing the log-linear form for the housing price equation; however, the amount of previous study of this equation makes it a good base case.[6] Our initial expectation was that our study of the index structure of the data would give some confirmation to the log-linear model.

We adopt the definitions of the observed variables in Harrison and Rubinfeld (1978a, 1978b). For notation, y_i denotes the log of the price of house i, and x_i denotes the vector of nine predictor variables that Harrison and Rubinfeld found to be statistically significant in their analysis. The data consists of 506 observations on the variables depicted in Table 3.1. As mentioned above, the earlier work produced a linear equation between y and x; of the form

$$y = \ln p = \alpha + x^T\beta + \epsilon \qquad . \tag{3.1}$$

The coefficients β summarize the proportional impacts of changes in x on housing prices. Table 3.2 contains the OLS estimates of these coefficients.

Our interest is in studying whether the log-linear model, or a more general index model, is a statistically adequate representation of the true regression $m(x) = E(y|x)$ of log-prices on the predictor variables.[7] We begin this by looking at a direct estimate of the average proportional impacts of changes in x on housing prices, or the average derivative $\delta = E[m'(x)]$. When the true model is linear as in (3.1), then $m(x) = \alpha + x^T\beta$, with $\delta = \beta$.

Moreover, as discussed above, (the appropriate components of) the average derivative δ represent the coefficients in semiparametric index and partial index models, so that our estimates can be used for coefficients of all such index specifications. In any case, we can regard the vector δ as giving generalized values of typical effects of the predictors on log housing prices. Our estimates are given in Table 3.2.[8]

We see that the basic difference between the OLS coefficient estimates $\hat{\beta}$ and the average derivative estimates $\hat{\delta}$ are minor. The Wald test that the differences are zero is based on the statistic

$$W = N\,(\hat{\delta} - \hat{\beta})^T V_{\hat{\delta}-\hat{\beta}}^{-1}(\hat{\delta} - \hat{\beta}) \tag{3.2}$$

where $V_{\hat{\delta}-\hat{\beta}}$ is the consistent estimator of the asymptotic variance of $\hat{\delta} - \hat{\beta}$ given by the sample variance of its influence representation. Here $W = 13.44$, which fails to reject for significance levels less that 15%.[9]

The largest qualitative difference in the coefficient estimates occurs for the coefficient of B, or the race effect. Because of the quadratic construction of B, for communities with small percentages of black residents, the OLS coefficient indicates a substantive negative race effect. The average derivative estimates indicate negligible race effects. From the consistency of average derivative estimates for coefficients of the single index model

$$m(x) = G_1(x^T\delta) \quad , \tag{3.3}$$

the difference in the B coefficient could arise because of nonlinearity in the function G_1 (which is not permitted by the log-linear model that motivates the OLS coefficients). We investigate this by computing and plotting the estimate of G_1 obtained by nonparametric regression of y_i on $x_i^T\hat{\delta}$, shown in Figure 3.1. This function appears as two lines with a shift (flat) in the center.

It seems that the negative race effect evidenced by the OLS coefficients may have resulted from forcing these two line segments together, as done by assuming that the overall model is linear.

To see whether this difference is statistically important, we apply our regression test to the linear model versus the single index model. All of our testing results are summarized in Table 3.3. Both the estimate $\hat{\gamma}$ and the "t-statistic" for testing the linear model against the single index model are quite small, so the linear model is not rejected. Therefore, the linear model (with the large race effect) and the single index model (with the negligible race effect but nonlinear function G_1) are statistically equivalent descriptions. Choice between these models rests on which has the more sensible interpretation; we would be inclined to use the single index model, but this is a purely subjective choice.

To see whether the linear model and/or the single index model stand up to further generalization, we compute the nonparametric regression of y on x, fitting the "model"

$$m(x) = g(x) \quad . \tag{3.4}$$

The nine-dimensional curve $\hat{g}(x)$ is difficult to plot and interpret, and so we mainly use it as the base case for the specification testing.

Again from Table 3.3, we see that the regression test rejects both the linear model and the single index model against the general regression. The estimates $\hat{\gamma}$ of the percentage of variance not accounted for by these models relative to general regression are 17.1% and 23.1%, which are each significantly different from zero.[10] Therefore, the restrictions of the single index model are too strong, and we must look further for a model that adequately captures the systematic variation between log price y and predictors x.

Our approach for this is to consider partial index models of increasing generality. In particular, we begin by estimating partial index models with one variable excluded from the index, so that the impact of the excluded variable is treated flexibly. This is computationally simple, since the average derivative estimates can be used as the coefficients for the variables remaining in the index. At any rate, the best model emerging from "one variable unconstrained" estimation is

$$(3.5) \qquad E(y|x) = G_2(x_1,\ x_{-1}'\delta_{-1})$$

where $x_{-1} = (x_2,\ldots,x_9)$ is the vector of all characteristics except for x_1 = NOXSQ, the pollution variable, and $\delta_{-1} = (\delta_2,\ldots,\delta_9)$ is the vector of average derivatives of the characteristics in the index. The function G_2 is a two dimensional function, and permits a general impact of the pollution variable x_1. In Table 3.3, we refer to this model as PARTIAL1.

We see that the single index model is rejected against model PARTIAL1. The graph of the function $\hat{G}_2$ in Figure 3.2 reveals some variation in the pollution effect, that is not consistent with the single index model (the "slices" of $\hat{G}_2$ for different values of x_1 have varying shapes). The model PARTIAL1 is rejected against the general regression, failing to account for an estimated 7.2% of the variation of the general regression. As such, we proceed to a next level of generalization, namely dropping two variables from the index.

Here, we find that the best model treating two variables flexibly is

$$(3.6) \qquad E(y|x) = G_3(y|x_1,x_9,x_{-19}{}^T\hat{\delta}_{-19}) \ ,$$

which permits flexible effects of the pollution variable x_1 = NOXSQ and the "lower status" variable x_9 = LSTAT.[11] The function G_3 is a three dimensional function, with the estimated model is referred to as PARTIAL2 in in Table 3.3

From Table 3.3, we see that the model PARTIAL2 gives a fairly parsimonious statistical depiction of the data. In particular, the estimate $\hat{\gamma}$ of the variation of the general regression not accounted for by PARTIAL2 is 1.16%, which is not significantly different from 0 at levels of significance lower than 3%. We likewise note that each more restricted index model we consider is rejected against PARTIAL2.

The three dimensional estimated function $\hat{G}_3$ of PARTIAL2 is somewhat more difficult to depict than $\hat{G}_1$ and $\hat{G}_2$ of the more restricted index models. Partial depictions are given in Figure 3.3, by plotting $\hat{G}_3$ holding x_9 constant at its mean, the lower status variable, (Figure 3.3a), and by plotting $\hat{G}_3$ holding the partial index $x_{-19}\hat{\delta}_{-19}$ constant at its mean (Figure 3.3b). The clearest difference between this model and the more restricted ones is the strong nonlinearity in the effect of x_1 the pollution variable, over ranges of x_9, the lower status variable. In particular, the marginal pollution effect is flat or slightly positive for high status communities (low "lower status" values), and strongly negative for low status communities (high "lower status" values. One interpretation of our testing results is that this nonlinearity is sufficiently strong to dictate a completely flexible treatment of both pollution and lower status effects on housing prices.

We close out this discussion by pointing our the effects of the nonparametric adjustments on the variances of the test coefficient $\hat{\gamma}$. In Table 3.4, we include different estimates of the variance of $\hat{\gamma}$ for the tests summarized in Table 3.3. The first column gives the standard OLS variance estimates, which neglect heteroskedasticity as well as the fact that estimated parameters and functions are used. The second column gives the (White) heteroskedasticity-consistent estimates, which likewise neglect that estimated functions are employed. Finally, the third column gives the variance estimates adjusted for the presence of estimated parameters and functions. Except for the test of PARTIAL2 against general regression, the adjustments for heteroskedasticity increase the variance estimates. In all cases, the adjustment for the use of estimated coefficients and functions increase the variance values. We will make reference to this feature when discussing issues with the limiting distributional theory below.

4. Technical Analysis of the Test Statistics

In this section, we give the explicit formulation of the estimators and test statistics, and summarize the theoretical results we have been able to obtain. We focus on the cases where the restricted and general models involve nonparametric estimation, and where kernel estimators are used for unknown regression functions. In Section 4.3A, we indicate how it is easy to incorporate cases where the restricted model is parametric.

We have introduced and applied our testing methodology in Sections 2 and 3, in order to motivate the value of a specification test as a tool for studying the appropriate index model structure in empirical data. The adjustments for nonparametric estimation that we have employed are based on familiar logic of the "delta method" as applied to the structure of our test statistic. We raise this now because our asymptotic theory for the test is incomplete, in a fashion that is not immediate from the derivations themselves - the standard asymptotic theory exhibits a singularity familiar from the discussion of Wooldridge (1992) and Yatchew (1992). The practical significance of this issue is that our critical values may not be tight enough - rejections of restricted models are unaffected, but a failure to reject could arise from our method overestimating the standard error of $\hat{\gamma}$ (i.e. for the Boston Housing Data, the model PARTIAL2 might still miss significant structure of the general regression model). We discuss this feature in detail after presenting the technical results.

Because the notation and the results become quite daunting in a hurry, it is useful to describe what we show. Standard semiparametric and nonparametric theory involves showing consistency and asymptotic normality under optimal approximation conditions - namely as sample size increases, nonparametric approximation parameters (here bandwidths) are adjusted optimally with sample size. This theory addresses the approximation capability of the statistic under study. We give a set of conditions under which $\sqrt{N}\,(\hat{\gamma} - \gamma)$ has a limiting normal distribution under those guidelines. As with most asymptotic theory, the derivation is based on isolating the leading terms of an asymptotic expansion, to which central limit theory is applied to show normality. Moreover, our method of estimating the variance of $\hat{\gamma}$ consistently estimates the variance of the relevant leading terms.

As it stands, this would give a traditional justification to our test. The singularity problem arises from the behavior of the test statistic in the limit of bandwidths approaching zero, when the restricted model equals the general model. In this case, the leading terms that our variance estimator is based on actually vanish, or that $\hat{\gamma}$ converges to $\gamma = 0$ at a rate faster than $\sqrt{N}$, with the limiting distribution of $\hat{\gamma}$ based on the next higher order terms in the expansion. At this time, we have not characterized these terms, to see whether a higher order analysis produces a better variance estimate. We will note how our results for the Boston Housing Data do not exhibit qualitative features consistent with the singularity, and as such we have proposed the method as it stands. At any rate, this issue appears to arise in most specification tests involving nonparametric alternatives, and we will discuss how other researchers have addressed the issue in the remarks following our results.

4.1 Estimation Formulae

Each of our comparisons involve nested index models, for which we enhance our notation as follows. Suppose that vector x of predictors is partitioned into $x = (x_{01}, x_{02}, x_1)$. In line with our treatment above, the symbol G is associated with the restricted model, and the symbol g is associated with the general model, as follows. The restricted model states that the regression $m(x) = E(y|x)$ is determined by d_1 arguments $z_1 = (x_{01}^T\delta_{01}+x_{02}^T\delta_{02}, x_1) = (x_0^T\delta_0, x_1)$, namely that $E(y|x) = E(y|z_1) = G(z_1)$. The general model states that the regression m(x) is determined by d_0 arguments $z_0 = (x_{01}^T\delta_{01}, x_{02}, x_1)$, $d_0 > d_1$, namely that $E(y|x) = E(y|z_0) = g(z_0)$. In the following, the notation g' refers to the partial derivative of $g(x_{01}^T\delta_{01}, x_{02}, x_1)$ with respect to its index argument $x_{01}^T\delta_{01}$, and G' is likewise the partial derivative of $G(x_0^T\delta_0, x_1)$ with regard to its index argument $x_0^T\delta_0$.

For estimating the density f(x) of x, we use the kernel density estimator

$$(4.1) \qquad \hat{f}(x) = N^{-1}h_f^{-k} \sum_{j=1}^{N} \mathcal{K}_f\left(\frac{x - x_j}{h_f}\right) \quad ,$$

where h_f is the bandwidth value and $\mathcal{K}_f$ is the kernel density that gives weights for local averages. One use of this estimator is to trim the sample for analysis, whereby we drop the observations with low estimated density. In particular, we drop observations with $\hat{I}_i = 1[\hat{f}(x_i) > b] = 0$, where b is a constant. The results of Section 3 had b set so that $I_i = 0$ for 5% of the observations. Our asymptotic results likewise take b as a fixed constant.

To measure the average derivatives (and therefore all index model coefficients), we use the "indirect IV" estimator of Stoker (1991,1992). This estimator is based on the density estimator $\hat{f}(x)$ of (4.1) as follows. Form the estimated "translation score" $\hat{\ell}(x_i) = -\hat{f}'(x_i)/\hat{f}(x_i)$ for each observation x_i. Take $\hat{\delta}$ as the instrumental variables estimator of the coefficients of y_i

regressed on x_i, using $\hat{\ell}(x_i)\hat{I}_i$ as the instrumental variable. Specifically, set

$$(4.2) \qquad \hat{\delta} = [\sum_i \hat{\ell}(x_i)\hat{I}_i\ (x_i - \bar{x})^T]^{-1}[\sum_i \hat{\ell}(x_i)\hat{I}_i\ (y_i - \bar{y})].$$

See Stoker (1992), among others, for explanation and motivation of this estimator.

The asymptotic results only require that we have an estimator $\hat{\delta}_0 = (\hat{\delta}_{01},\hat{\delta}_{02})$ of the coefficients that obeys

$$(4.3) \qquad \sqrt{N}(\hat{\delta}_0 - \delta_0) = N^{-1/2} \sum r_{\delta 0}(y_i,x_i) + o_p(1)$$

and therefore is $\sqrt{N}$ asymptotically normal. Denote the subvector of $r_{\delta 0}$ corresponding to $\hat{\delta}_{01} - \delta_{01}$ as $r_{\delta 01}$. The components of the estimator (4.2) have $r_{\delta 0}(y,x) = m_0'(x) - \delta_0 + [y - m(x)]\ell_0(x)$, where $m_0' = \partial m/\partial x_0$, and $\ell_0(x) = -\partial \ln f/\partial x_0$, as derived in Härdle and Stoker (1989) and Stoker (1991).

Nonparametric estimators of unknown regression functions are summarized as follows. The function G of the restricted model is estimated by $\hat{G}$, the d_1 dimensional kernel regression of y on $\hat{z}_1 = (x_0{}^T\hat{\delta}_0,x_1)$, using kernel function $\mathcal{K}_1$ and bandwidth h_1, or

$$(4.4) \qquad \hat{G}(z) = \hat{F}_1(z)^{-1}\left(N^{-1}h_1{}^{-d_1} \sum_{j=1}^{N} \mathcal{K}_1\left(\frac{z - \hat{z}_{1j}}{h_1}\right) y_j \right),$$

where

$$(4.5) \qquad \hat{F}_1(z) = \left(N^{-1}h_1{}^{d_1} \sum_{j=1}^{N} \mathcal{K}_1\left(\frac{z - \hat{z}_{1j}}{h_1}\right)\right) .$$

The function g of the general model is estimated by $\hat{g}$, the d_0 dimensional kernel regression of y on $\hat{z}_0 = (x_{01}{}^T\hat{\delta}_{01},x_{02},x_1)$, using kernel function $\mathcal{K}_0$ and bandwidth h_0, or

$$(4.6) \qquad \hat{g}(z) = \hat{F}_0(z)^{-1}\left(N^{-1}h_0^{-d_0} \sum_{j=1}^{N} \mathcal{K}_0\left(\frac{z - \hat{z}_{0j}}{h_0}\right) y_j\right),$$

where

$$(4.7) \qquad \hat{F}_0(z) = \left(N^{-1}h_0^{-d_0} \sum_{j=1}^{N} \mathcal{K}_0\left(\frac{z - \hat{z}_{0j}}{h_0}\right)\right) .$$

While these formulae are daunting, they are directly computed from the data, given bandwidth values and specifications of the kernel functions.[12] The same is true of the adjustment terms required for the variance of our t-statistic. Because of their size, we give the formulae for these adjustment terms in Appendix 2.

4.2 Summary of the Test and Asymptotic Results

We now formally introduce the test, in order to present the asymptotic results as well as the ideas on which precision measurement is based. To keep the presentation compact, subscript "i" denotes evaluation of relevant terms at $(y,x) = (y_i,x_i)$; for instance, g_i denotes g evaluated at z_{0i}, $\hat{G}_i$ denotes $\hat{G}$ evaluated at $\hat{z}_{1i}$, and $\hat{I}_i$ is the trim indicator that is 1 if $\hat{f}(x_i) > b$, and 0 otherwise, as above.

With trimming incorporated, our test is based on the coefficient $\hat{\gamma}$ of the regression

$$(4.8) \qquad (y_i - \hat{G}_i)\hat{I}_i = \hat{\alpha}\,\hat{I}_i + \hat{\gamma}\,\hat{g}_i\hat{I}_i + \hat{u}_i$$

Letting

(4.9) $$S_{\hat{g}} = N^{-1} \sum (\hat{g}_i - \check{g})^2 \hat{I}_i ; \qquad \check{g} = N^{-1} \sum \hat{g}_i \hat{I}_i$$

denote the sample variance and mean of $\hat{g}_i \hat{I}_i$, we have that the coefficient $\hat{\gamma}$ is

(4.10) $$\hat{\gamma} = \frac{1}{S_{\hat{g}}} N^{-1} \sum (\hat{g}_i - \check{g})(y_i - \hat{G}_i)\hat{I}_i$$

In line with of the discussion of Section 2, this regression procedure amounts to fitting a sample analog of the equation

(4.11) $$(y_i - G_i)I_i = \gamma\, [g_i - E(gI)]I_i + u_i$$

where the parameter γ is defined via OLS projection as

(4.12) $$\gamma = \frac{E\{[g-E(gI)][y - G]I\}}{E\{[g-E(gI)]^2 I\}}$$

Consequently, γ is the percentage of variation of g not accounted for by G, over the untrimmed part of the population. Moreover, $\gamma = 0$ if and only if $g = G$ a.s. for x such that $f(x) > b$.

We require the following basic assumptions

<u>Assumption 1</u>: The fourth moments of (y,x) exist.

<u>Assumption 2R</u>: For F_0 the density of z_0, we have that $E(y^4|z_0)F_0(z_0)$ and F_0 are bounded, $(g - G)I$ is continuously bounded a.e., and $[g - G]F_0$ and F_0 are continuously differentiable of order $P_0 > d_0$.

<u>Assumption 2L</u>: For F_1 the density of z_1, we have that $E(y^4|z_1)F_1(z_1)$ and F_1 are bounded, GI is continuously bounded a.e., and GF_1 and F_1 are continuously differentiable of order $P_1 > d_1$.

<u>Assumption</u> <u>3R</u>: The kernel $\mathcal{K}_0$ has bounded support, is Lipschitz, $\int \mathcal{K}_0(u)\, du = 1$, and is of order $P_0 > d_0$.[13]

<u>Assumption</u> <u>3L</u>: The kernel $\mathcal{K}_1$ has bounded support, is Lipschitz, $\int \mathcal{K}_1(u)\, du = 1$, and is of order $P_1 > d_1$.

<u>Assumption</u> <u>4</u>: For f the density of x, fI is continuously bounded a.e., f is continuously differentiable of order $P_f > k$. The kernel $\mathcal{K}_f$ has bounded support, $\int \mathcal{K}_f(u)\, du = 1$, and is of order $P_f > k$.

Our approach to characterizing the limiting distribution of $\hat{\gamma}$ is to establish the following decomposition:

$$\sqrt{N}(\hat{\gamma} - \gamma) = \sqrt{N}(\tilde{\gamma} - \gamma) + RA_N - LA_N + o_p(1) \tag{4.13}$$

where $\tilde{\gamma}$ is the "estimator" based on known functions;

$$\tilde{\gamma} = \frac{1}{S_g} N^{-1} \sum [g_i - E(gI)](y_i - G_i)I_i \tag{4.14}$$

with

$$S_g = N^{-1} \sum [g_i - E(gI)]^2 I_i \quad . \tag{4.15}$$

an estimator of the (trimmed) variance $\sigma_g = E\{[g - E(gI)]I\}^2$. The remaining terms are the adjustments for using estimates on both sides of the regression equation: first,

$$(4.16) \qquad RA_N = \frac{1}{\sigma_g} N^{-1/2} \sum (\hat{g}_i - g_i)(y_i - G_i)I_i$$

is the adjustment for nonparametric estimation of the "right hand side", or predictor variable, and second,

$$(4.17) \qquad LA_N = \frac{1}{\sigma_g} N^{-1/2} \sum (\hat{G}_i - G_i)[g_i - E(gI)]I_i$$

is the adjustment for nonparametric estimation of the "Left-hand-side", or dependent variable, of the original regression. Standard limit theory applies to the "estimator" $\tilde{\gamma}$ of (4.14); with $u = (y - G)I - \gamma[g - E(gI)]$, we have that

$$(4.18) \qquad \sqrt{N}(\tilde{\gamma} - \gamma) = \frac{1}{\sigma_g} N^{-1/2} \sum [g_i - E(g)]u_i I_i + o_p(1) \ ,$$

so our conditions imply that $\tilde{\gamma}$ is asymptotically normal.

Therefore, the characterization of the limiting distribution of $\hat{\gamma}$ requires studying the adjustment terms, and establishing (4.13). The adjustment terms are characterized in two lemmas that follow from results of Newey (1992):

<u>Lemma R</u>: Given Assumptions 1, 2R and 3R suppose (a) $N \to \infty$, $h_0 \to 0$; (b) $Nh_0^{2d_0}/(\ln N) \to \infty$ and (c) $Nh_0^{2P_0} \to 0$. Let

$$r_{gi} = [g_i - G_i](y_i - g_i)I_i$$

$$r_{Ri} = r_{gi} + B_0\, r_{\delta 0}(y_i, x_i)I_i$$

where $B_0 = [B_{01}, 0]$ and

$B_{01} = E(g'[E[(y-G)x_0|z_0]I - (g-G)IE[x_0|z_1] + (g'-G')I[E[yx_0|z_1] - gIE[x_0|z_1]])$.

Then we have that

$$RA_N = \frac{1}{\sigma_g} N^{-1/2} \sum r_{Ri} + o_p(1) \quad .$$

(In the case where $d_0 = k$, where $g(x) = E(y|x)$ involves no estimated coefficients, we set $B_0 = 0$.)

<u>Lemma L</u>: Given Assumptions 1, 2L and 3L, suppose (a) $N \to \infty$, $h_0 \to 0$; (b) $Nh_1^{2d_1}/(\ln N) \to \infty$ and (c) $Nh_1^{2P_1} \to 0$. Let

$$r_{Gi} = [G_i - E(y)](y_i - G_i)I_i$$

$$r_{Li} = r_{Gi} + B_1\, r_{\delta 0}(y_i, x_i)I_i$$

where $B_1 = E(G' \; [E[(y+g)x_0|z_1] - 2\; G\; E[x_0|z_1]])$. Then we have that

$$LA_N = \frac{1}{\sigma_g} N^{-1/2} \sum r_{Li} + o_p(1) \quad .$$

The relation (4.13) is then shown as part of the proof of the Theorem 1.

<u>Theorem 1</u>: Suppose that Assumptions 1, 2R, 2L, 3R, 3L and 4 are valid, and assume the bandwidth conditions of Lemmae R and L. Suppose further that (a) $N \to \infty$, $h_f \to 0$; $Nh_f^{2k}/(\ln N) \to \infty$ and $Nh_f^{2P_f} \to 0$, (b) $Nh_f^k h_0^{d_0} h_1^{d_1}/(\ln N)^3 \to \infty$, (c) $Nh_f^k h_0^{2d_0}/(\ln N)^3 \to \infty$, (d) $Nh_f^k h_1^{2d_1}/(\ln N)^3 \to \infty$ and (d) $Nh_0^{d_0} h_1^{2d_1}/(\ln N)^2 \to \infty$. Define

$$r_{\gamma i} = [g_i - E(gI)]u_i I_i + r_{Ri} - r_{Li}$$

$$= [g_i - E(gI)]u_i I_i + [g_i - G_i](y_i - g_i) - [G_i - E(y)](y_i - G_i)$$

$$+ [B_0 - B_1] r_{\delta 0}(y_i, x_i)$$

We then have that

$$\sqrt{N}(\hat{\gamma} - \gamma) = \frac{1}{\sigma_g} N^{-1/2} \sum r_{\gamma i} + o_p(1)$$

so that $\sqrt{N}(\hat{\gamma} - \gamma) \rightarrow \mathcal{N}(0, \sigma_\gamma)$, where $\sigma_\gamma = \sigma_g^{-2} \operatorname{Var}\{r_{\gamma i}\}$. Further, the estimator $\hat{\sigma}_\gamma$ given in Appendix 2 is a consistent estimator of σ_γ.

Consequently, Theorem 1 gives conditions under which $\hat{\gamma}$ is asymptotically normal, with the squared "t-statistic" having a limiting $\chi^2(1)$ distribution.

4.3 Related Remarks

A. Testing Parametric Regression Models

When the restricted regression model is parametric, as with our tests of the linear model in Section 3, the test is modified in a straightforward way. In particular, suppose that the restricted model is $m(x) = \Gamma(x,\beta)$, and that we wish to test it against a general nonparametric regression, $m(x) = g(x)$ above. Suppose further that we have a $\sqrt{N}$ asymptotically normal estimator $\hat{\beta}$ of the parameters of the restricted model, wherein

$$\text{(4.19)} \qquad \sqrt{N}(\hat{\beta} - \beta) = N^{-1/2} \sum r_\beta(y_i, x_i) + o_p(1)$$

(where $\beta = \operatorname{plim} \hat{\beta}$ if the restricted model is not true).

The specification test is applicable as above, namely by computing the OLS regression coefficient $\hat{\gamma}$ of

$$(4.20)\qquad y_i - \Gamma(x_i,\hat{\beta}) = \hat{\alpha} + \hat{\gamma}\,\hat{g}(x_i) + \hat{u}_i \quad , \; i = 1,\ldots,N.$$

Testing is based on whether $\gamma = 0$, which is likewise tested by the square of the "t-statistic." The only complication (actually simplification) is that the asymptotic variance of $\hat{\gamma}$ must reflect the fact that the estimator $\hat{\beta}$ is used. The only change to the above development is that the "left" adjustment only contains the influence of $\hat{\beta}$, with the "right" adjustment left unaffected. In particular, here we have

$$(4.21)\qquad LA_N = \frac{1}{\sigma_g} N^{-1/2} \sum [\Gamma(x_i,\hat{\beta}) - \Gamma(x_i,\beta)][g_i - E(gI)]I_i .$$

This term is analyzed in an entirely standard fashion, namely we have

$$(4.22)\qquad LA_N = \frac{1}{\sigma_g} E\{\partial\Gamma(x_i,\beta)/\partial\beta\ [g_i - E(gI)]I_i\} \sqrt{N}(\hat{\beta} - \beta) + o_p(1) \ .$$

If $\hat{r}_\beta(y_i,x_i)$ is a (uniformly) consistent estimator of the influence $r_\beta(y_i,x_i)$, then the relevant estimate for the influence term of the left hand adjustment is

$$(4.23)\qquad la_i = \{N^{-1} \sum \partial\Gamma(x_i,\hat{\beta})/\partial\beta\ [\hat{g}_i - \hat{\bar{g}}]\hat{I}_i\}\ \hat{r}_\beta(y_i,x_i) \quad .$$

We then estimate the asymptotic variance of $\hat{\gamma}$ by the sample variance of (2.18). This method was applied for the test statistics involving the linear model of Section 3.

B. Issues of Practical Implementation

As is now standard, our asymptotic results above have assumed the use of higher order kernels for nonparametric estimation. It is also well known that such kernels, with giving positive and negative local weighting, do not often give good estimator performance in small samples. Consequently, for our estimation of Section 3, we have used positive kernels throughout. In particular, each kernel function is the product of biweight kernels: for estimation of a d dimensional function, we used

$$(4.24) \qquad \mathcal{K}(u_1,\ldots,u_d) = \Pi\, k(u_j)$$

where $k(u_j)$ is given as

$$(4.25) \qquad k(u) = \left(\frac{15}{16}\right) (1 - u^2)^2 \, 1[|u| \leq 1]$$

We have likewise used these kernel functions in the variance adjustment formulae.

Since there is no developed theory for optimal bandwidth choice for the purpose of our specification test, we chose bandwidth values using Generalized Cross Validation (GCV) of Craven and Wahba (1979). For instance, to estimate the general regression m(x), let Y denote the vector of observations $\{y_i\}$ and M_h denote the vector of values $\{\hat{m}(x_i)\}$ computed with bandwidth h. Consider the weight matrix W_h defined from

$$(4.26) \qquad M_h = W_h Y$$

The GCV bandwidth is the value of h that minimizes

$$(4.27) \qquad \frac{N^{-1}|(I - W_h)Y|^2}{[N^{-1}\,\mathrm{Tr}(I - W_h)]^2}$$

We also standardized the predictor data for the nonparametric estimation.[14]

This method of bandwidth choice was used for simplicity. However, it is unlikely that this method applied in increasingly large samples will give the bandwidth conditions of Theorem 1 above. In particular, those conditions require pointwise bias to vanish faster than pointwise variance, which is not implied by GCV bandwidths chosen for each sample size.

As indicated above, we have incorporated the trimming indicator, dropping the 5% of data values with lowest estimated density values. In practical terms, this drops observations with isolated predictor values, such as remote outliers. Moreover, since the regression estimators involve dividing by estimated density, dropping observations with small estimated density likely avoids erratic behavior in the nonparametric estimates.

4.4 The Singularity Issue

The singularity issue that we alluded to at the beginning of this section is evident from Theorem 1. In particular, under the limiting bandwidth conditions of the theorem and under the null hypothesis that $G = g$ a.e., we have that the influence function $r_{\gamma i} = 0$ for all i. Under those provisos, with $\gamma = 0$ in that case, we have that $\sqrt{N}(\hat{\gamma} - \gamma) = o_p(1)$, of that $\hat{\gamma}$ converges to the true value $\gamma = 0$ at a rate faster than $\sqrt{N}$. This is not true under any circumstance where $G \neq g$ for a set of x's of positive probability, nor is it true if the limiting theory did not take the bandwidths closer to 0 in an optimal fashion with increases in sample size N.

Since we have departed from the conditions of Theorem 1 for our application, by using positive kernels (positive local weighting), one might wonder how relevant the singular y problem is. We can look at one feature of our results, by noting how the singularity arises. In particular, the leading (White) terms of the influence given in Theorem 1 (or their sample analogue, the first term of (2.18)) have variance, and the singularity means that the nonparametric adjustments actually cancel this variation, so that the overall influence terms vanish. In Table 3.4, we see that the adjustments for nonparametric estimation actually increase the estimated variances, which is the opposite of what we would expect (under the null) if our variance estimator were estimating 0. For instance, this is the case with our test of PARTIAL2 versus general regression, where if the singularity were important, our estimate of $\hat{\gamma} = .0116$ would be more precise (higher t-statistic) than we have displayed. While these observations don't prove anything, they suggest that the standard leading terms of $\hat{\gamma} - \gamma$ may not have negligible variance in our application.

Another way of looking at the singularity is to consider varying the conditions of Theorem 1 in a way that might lead to a more accurate distributional approximation for $\hat{\gamma}$. Since we have used positive kernels, and higher order kernels are a technical device for increasing the order of bias, we might ask how the approximate distribution of $\hat{\gamma}$ would look if the bias were explicitly recognized. One of the authors has recently studied finite sample bias issues on the basis of approximation for large N but where the bandwidth parameters are not decreased with sample size (Stoker 1993a, 1993b). This kind of approximation is not without controversy (N treated as large but bandwidths as fixed), but it does shed light on our method of estimating the variance of $\hat{\gamma}$, so we discuss it for a moment. In particular, we formulate the nonparametric adjustment terms on the basis of the variation of the (U statistic) structure of $\hat{\gamma}$, and nothing in this demonstration uses higher order kernels. Suppose that we are testing a single index model $E(y|x) = G_1(x^T\delta)$ versus general regression $E(y|x) = g(x)$. Suppose that conditional on the values of the bandwidths used, the limits of $\hat{\delta}$, $\hat{G}_1$, $\hat{g}$ and $\hat{I}$ are denoted with overbars as $\bar{\delta}$, $\bar{G}_1$, $\bar{g}$ and $\bar{I}$. Further define

$$\bar{\gamma} = \frac{E\{[\bar{g}-E(\bar{g}\bar{I})][y - \bar{G}_1(x^T\bar{\delta})]\bar{I}\}}{E\{[\bar{g}-E(\bar{g}\bar{I})]^2\bar{I}\}} \tag{4.28}$$

While we have not provided a formal proof, the proof of Theorem 1 suggests that we can conjecture with positive kernels that 1) $\sqrt{N}\,(\hat{\gamma} - \bar{\gamma}) \to \mathcal{N}(0,\bar{\sigma}_\gamma)$; where $\bar{\sigma}_\gamma > 0$ when bandwidths are nonzero, and 2) $\hat{\sigma}_\gamma$ is a consistent estimator of $\bar{\sigma}_\gamma$. Given that this conjecture is verified, it would give some validation for our method of estimating variance and setting critical values. However, it also shows the difficulty of including bias - namely $\bar{\gamma}$ is the regression coefficient of the test using the (potentially biased) functions $\bar{g}$ and $\bar{G}$. There is a practical sense in which one must take the nonparametric estimates of unknown functions as the best representation of the true regressions, but if the biases are severe and systematic in an unfortunate way, we could be verifying $\bar{g}(x) = \bar{G}_1(x^T\bar{\delta})$ a.e. without the corresponding equality among the true functions. Even under the controversial position that a better distributional approximation might result from holding bandwidths fixed in the theory, we do not have a totally satisfactory answer.

Other authors, notably Wooldridge (1992) and Yatchew (1992), have made different proposals in light of the singularity. Wooldridge develops a theory involving "asymptotic poor fitting" of the general nonparametric regression, whose analogue here would be to use a bandwidth sequence for the general regression that shrinks more slowly than under the conditions of Theorem 1. Yatchew proposes sample splitting, wherein nonparametric estimation is done with half (say) of the sample and specification testing is done using the other half. Other artificial methods could be proposed, such as adding random noise to the y data (after estimating computing the nonparametric estimates), but these artificial methods may avoid but do not address the basic issue.

Finally, within the context of the standard theory, one could do estimation with the proper higher order kernels, and characterize the higher order terms that represent the variation in the test statistic under the nonparametric approximation theory. Some recent unpublished work has made progress in this direction. Horowitz and Härdle (1993) consider testing a parametric model against a nonparametric alternative using an approach similar to ours, focusing on a part of a regression statistic for which the higher order terms can be successfully characterized. Hong and White (1993) develop some general theory toward characterizing the higher order terms, although we have not succeeded in verifying their convergence conditions for our test. Such verification would permit solving for the higher order terms, and our method of estimating variance could be assessed relative to alternatives.

5. Conclusion

In this paper we have presented a simple specification test for assessing the appropriate index model in an empirical application. The index model framework gives a generalization of linear models that may be informative for applications where there are no theoretical reasons for specifying a particular functional form. Our application to measuring environmental effects from housing prices had this feature, and we have tried to illustrate how index models can give an enhanced depiction of the data relationships relative to standard linear modeling. We have used our test to check to the adequacy of a parametric (linear) model versus nonparametric regression, and it seems natural that the testing approach will be useful for other (nested) testing problems.

We have focused on the use of nonparametric kernel estimators. While the adjustment terms listed in Section 2 involve large formulae, they are computed directly from the data and do not involve more complicated computation than required for the kernel estimators themselves. We also have developed a standard asymptotic theory for using kernel estimators; but from the results of Newey (1991), it is natural to conjecture that the same distributional results would be obtained when other nonparametric estimators are used, such as truncated polynomials or other series expansions. We have raised the singularity issue for tests using nonparametric estimators, and discuss various ways our basic method might be further studied or justified.

We do want to stress one feature of our method that we find appealing relative to alternative testing proposals. In particular, focusing on the single coefficient $\hat{\gamma}$ is valuable because of its goodness of fit interpretation. The cost of this was a fairly complicated technical analysis, such as the formulation of the adjustments required to account for nonparametric estimation. But in our view, the value of focusing on an interpretable statistic is the immediate practical sense it gives for which models "fit" the data and which do not. For instance, the model PARTIAL2 accounts for an estimated $1-\hat{\gamma} = .9884$ of the variation of the general regression, which is strong support for the notion that the model PARTIAL2 captures the systematic features of the the log housing price regression in the Boston Housing Data, especially relative to the more restricted models. As such, we feel our method is more appealing on practical grounds than specification tests that just take on an uninterpretable "accept or reject" posture, without giving further information.

Appendix 1: Proofs of Results

The structure of the terms that adjust for estimated functions and parameters are quite similar, so we present generic results which specialize to Lemmas L and R above. For this section, refine the notation slightly for any partial index model: suppose that x is partitioned as $x = (x_0, x_1)$, with x_1 a d-1 vector, $d \leq k$, and z denotes the the d vector of predictors for a partial index model, namely $z = (x_0^T\delta_0, x_1)$. Thus, the notation can range from the case of a single index model, where d = 1 and $z = x^T\delta$, to the general regression case where d = k, where without loss of generality we set z = x (and ignore the adjustment term for the estimation of δ_0 below).

Further, let x_2 denote a k - d subvector of x_0, where the remaining component of x_0 has a positive coefficient δ_1. The transformation

$$(z, x_2) = \tau(x)$$

is linear and nonsingular with (constant) Jacobian δ_1, so that the Jacobian of τ^{-1} is $1/\delta_1$. Below, we need to consider several functions of x as functions of (z, x_2). To keep this compact, we use a "*" to signify this simply: for a function a(y,x), we have

$$a^*(y, z, x_2) = a(y, \tau^{-1}(z, x_2))$$

We will mention this explicitly when necessary for clarity.

We will focus on adjustment terms that arise from the estimation of the regression function $M(z) = E(y|z)$. Recall that the marginal density of x is f(x), and the joint density of y and x is $q(y|x)f(x)$. The regression of y on z is written explicitly as

$$M(z) = C(z)/F(z)$$

where C(z) is

$$C(z) = \delta_1^{-1} \int yq^*(z,x_2)f^*(z,x_2)\, dx_2 = \delta_1^{-1} \int m^*(z,x_2)f^*(z,x_2)\, dx_2$$

and F(z) is the marginal distribution of z; namely

$$F(z) = \delta_1^{-1} \int f^*(z,x_2)\, dx_2 \ .$$

All the adjustment terms that we consider are based on kernel estimation of $M(z) = E(y|z)$. Let $\hat{z}_i = (x_{0i}^T\hat{\delta}_0, x_{1i})$ denote the i^{th} observation of the predictor based on estimated index coefficients, and $z_i = (x_{0i}^T\delta_0, x_{1i})$ denote the analogous vector based on the true coefficient values. The kernel estimator used in estimation is $\hat{M}(z)$, computed using $\hat{z}_i$, y_i, namely

$$\hat{M}(z) = \hat{C}(z)/\hat{F}(z)$$

where

$$\hat{C}(z) = N^{-1}h^{-d} \sum_{j=1}^{N} \mathcal{K}\left(\frac{z - \hat{z}_j}{h}\right) y_j$$

where $\mathcal{K}$ is a kernel function, h is a bandwidth parameter that must be set for estimation, and

$$\hat{F}(z) = N^{-1}h^{-d} \sum_{j=1}^{N} \mathcal{K}\left(\frac{z - \hat{z}_j}{h}\right)$$

Finally we will need to make reference to the kernel estimator that would be computed if the coefficients δ_0 were known, namely

$$\tilde{M}(z) = \tilde{C}(z)/\tilde{F}(z)$$

where

$$\tilde{C}(z) = N^{-1}h^{-d} \sum_{j=1}^{N} \mathcal{K}\left(\frac{z - z_j}{h}\right) y_j$$

and

$$\tilde{F}(z) = N^{-1}h^{-d} \sum_{j=1}^{N} \mathcal{K}\left(\frac{z - z_j}{h}\right)$$

Each adjustment term takes the following form:

$$\text{(A.1)} \qquad A = N^{-1/2} \sum [\hat{M}(\hat{z}_i) - M(z_i)]\ a(y_i,x_i)\ I_i$$

where a(y,x) has mean 0 and finite variance. We first split this into variation due to the estimation of δ_0, and due to the estimation of M:

$$A = A_\delta + A_M$$

where

$$A_\delta = N^{-1/2} \sum [\hat{M}(\hat{z}_i) - \tilde{M}(z_i)]\ a(y_i,x_i)\ I_i$$

$$A_M = N^{-1/2} \sum [\tilde{M}(z_i) - M(z_i)]\ a(y_i,x_i)\ I_i$$

Again, recall that for k = d, we set $A_\delta = 0$.

First, consider the adjustment for nonparametric estimation, or A_M. This is analyzed by linearizing $\tilde{M}$ in terms of its numerator and denominator, analyzing its U-statistic structure to show asymptotic normality, and analyzing its bias separately, along the lines of Härdle and Stoker (1989). Fortunately, some recent unifying theory is applicable. Let

$$\mathcal{A}(z) = E[a(y,x)I|z] \quad .$$

Begin with the following generic assumption:

<u>Assumption</u> <u>A1</u>: We assume that

1) $E(y^4) < \infty$,
2) $E(y^4|z)F(z)$ and $F(z)$ are bounded,
3) $E[a(y,x)^2 I] < \infty$
4) The kernel $\mathcal{K}$ has bounded support, is Lipschitz, $\int \mathcal{K}(u)\, du = 1$, and is of order $P > d$.
5) $\mathcal{A}(z)F(z)$ and $F(z)$ are continuously differentiable of order at least P,
6) There exists a compact set $\mathfrak{A}$ such that $\mathcal{A}(z) = 0$ for $z \in \mathbb{R}^d/\mathfrak{A}$
7) $\mathcal{A}(z)$ is continuously bounded a.e.

The adjustment for nonparametric estimation, A_M, is characterized by applying Theorem 3.4 of Newey (1992).

<u>Lemma</u> <u>1</u>: Given Assumption A1, if $Nh^{2d}/(\ln N) \to \infty$ and $N\, h^{2P} \to 0$, then

$$A_M = N^{-1/2} \sum r_{AMi} + o_p(1),$$

where $r_{AMi} = \mathcal{A}(z_i)\ [y_i - M(z_i)]$, and $A_M \to \mathcal{N}[0,\ E(r_{AMi} r_{AMi}^T)]$.

The adjustment for using estimated coefficients is characterized directly as follows. Recall that

$$\sqrt{N}(\hat{\delta}_0 - \delta_0) = N^{-1/2} \sum r_{\delta 0}(y_i, x_i) + o_p(1)$$

Let

$$\mathcal{B} = E\{ \partial M/\partial z_1(z)[E(ax_0|z) - \mathcal{A}(z)E(x_0|z)] + \partial\mathcal{A}/\partial z_1(z)[E(yx_0|z) - M(z)E(x_0|z)]\}$$

then we have

Lemma 2: Given Assumption A1, if $Nh^{d+2}/(\ln N) \to \infty$ and $h \to 0$, then

$$A_\delta = \mathcal{B}\sqrt{N}(\hat{\delta}_0 - \delta_0) + o_p(1),$$

$$= N^{-1/2}\sum \mathcal{B}\, r_{\delta 0}(y_i,x_i) + o_p(1) \quad .$$

Proof: Denote the kernel regression as a function of x_i and δ as

$$M^+(x_i;\delta) = \left[\sum_{j=1}^{N} \mathcal{K}\left(\frac{(x_{0i}-x_{0j})^T\delta}{h};\ \frac{x_{1i}-x_{1j}}{h}\right)\right]^{-1} \sum_{j=1}^{N} \mathcal{K}\left(\frac{(x_{0i}-x_{0j})^T\delta}{h};\ \frac{x_{1i}-x_{1j}}{h}\right) y_j$$

$$= \quad s_{\mathcal{K}}(x_i;\delta)^{-1} \sum_{j=1}^{N} \mathcal{K}\left(\frac{(x_{0i}-x_{0j})^T\delta}{h};\ \frac{x_{1i}-x_{1j}}{h}\right) y_j$$

By the Mean Value Theorem, we have that

$$A_\delta = \{N^{-1}\sum [\partial M^+(x_i;\tilde{\delta}_i)/\partial\delta]\ a(y_i,x_i)\ I_i \sqrt{N}(\hat{\delta}_0 - \delta_0)$$

where $\tilde{\delta}_i$, i=1,,...,N lies on the line segment between $\hat{\delta}_0$ and δ_0. Therefore, if

$$\mathcal{B}_N = N^{-1}\sum [\partial M^+(x_i;\tilde{\delta}_i)/\partial\delta]\ a(y_i,x_i)\ I_i$$

and we can characterize plim $\mathcal{B}_N = \mathcal{B}$, then we will have

$$A_\delta = \mathcal{B}\sqrt{N}(\hat{\delta}_0 - \delta_0) + o_p(1)$$

We have

$$\partial M^+(x_i;\delta)/\partial\delta_0 = S_{\mathcal{K}}(x_i;\delta)^{-1} \sum_{j=1}^{N} \frac{x_{0i}-x_{0j}}{h} \mathcal{K}'\left(\frac{(x_{0i}-x_{0j})^T\delta}{h}; \frac{x_{1i}-x_{1j}}{h}\right) y_j$$

$$- M^+(x_i;\delta)\ S_{\mathcal{K}}(x_i;\delta)^{-1} \sum_{j=1}^{N} \frac{x_{0i}-x_{0j}}{h} \mathcal{K}'\left(\frac{(x_{0i}-x_{0j})^T\delta}{h}; \frac{x_{1i}-x_{1j}}{h}\right)$$

$$= x_{0i}\ (M^+)_1'(x_i;\delta)$$

$$- S_{\mathcal{K}}(x_i;\delta)^{-1} \sum_{j=1}^{N} \frac{1}{h} \mathcal{K}_1'\left(\frac{(x_{0i}-x_{0j})^T\delta}{h}; \frac{x_{1i}-x_{1j}}{h}\right) x_{0j}y_j$$

$$+ M^+(x_i;\delta)\ S_{\mathcal{K}}(x_i;\delta)^{-1} \sum_{j=1}^{N} \frac{1}{h} \mathcal{K}_1'\left(\frac{(x_{0i}-x_{0j})^T\delta}{h}; \frac{x_{1i}-x_{1j}}{h}\right) x_{0j}$$

where " $_1'$ " denotes differentiation with regard to the index, or first argument. Under our conditions, as $h \to 0$ and $\delta \to \delta_0$, these terms estimate

$$\partial M(x_i;\delta)/\partial\delta_0 = x_{0i}\ M_1'(z_i) - [F(z_i)]^{-1}[\partial[E(x_0y|z_i)F(z_i)]/\partial z_1]$$
$$+ M(z_i)[F(z_i)]^{-1}[\partial[E(x_0|z_i)F(z_i)]/\partial z_1]$$
$$= x_{0i}\ M_1'(z_i) - [E_1'(x_0y|z_i) - M(z_i)E_1'(x_0|z_i)]$$
$$- [E(x_0y|z_i) - M(z_i)E(x_0|z_i)][F(z_i)]^{-1}F_1'(z_i)$$

Since x can be regarded as bounded because of trimming on small positive density, then uniform convergence follows as in Newey (1992), since $Nh^{d+2}/(\ln N) \to \infty$ as $h \to 0$ and $\hat{\delta}_0 - \delta_0 = o_p(1)$. Therefore

$$\mathcal{B} = E\{a(y,x)[x_0\ M_1'(z) - [E_1'(x_0y|z) - M(z)E_1'(x_0|z)]$$
$$- [E(x_0y|z) - M(z)E(x_0|z)][F(z)]^{-1}F_1'(z)]\}$$
$$= E\{\ M_1'(z)[E(ax_0|z) - \mathcal{A}(z)E(x_0|z)]$$

$$+ \mathcal{A}_1'(z)[E(yx_0|z) - M(z)E(x_0|z)]\}$$

giving the characterization of A_δ above. QED

Consequently, we conclude that

Lemma 3: If $Nh^{2d}/(\ln N) \to \infty$ and $N h^{2P} \to 0$, then

$$A = \mathcal{B}\sqrt{N}(\hat{\delta}_0 - \delta_0) + N^{-1/2}\sum r_{AMi} + o_p(1),$$

$$= N^{-1/2}\sum B\, r_\delta(y_i,x_i) + N^{-1/2}\sum r_{AMi} + o_p(1)$$

$$= N^{-1/2}\sum r_{Ai} + o_p(1)$$

where $r_{Ai} = \mathcal{B}\, r_\delta(y_i,x_i) + r_{AMi}$

Applying Theorem 1 to RA_N and LA_N yields Lemmae R and L.

Estimation of asymptotic variance is accomplished by using an estimate of the influence terms for the adjustment factors, with the consistency of this procedure verified by an argument similar to that in Härdle and Stoker (1989). With regard to the generic adjustment term (A.1), the matrix $\mathcal{B}$ is consistently estimated by evaluating the expression for $\mathcal{B}_N$ above at $\hat{\delta}$ and the bandwidth used for estimation. The influence term r_{AMi} is estimated from the U-statistic structure of A_M, which would be used in a direct proof of Lemma 1 above. In particular, we have that

$$A_M = N^{1/2}[U_1 - U_2] + o_p(1)$$

where

$$U_1 = \binom{N}{2}^{-1} \sum_{i=1}^{N} \sum_{j=i+1}^{N} p_{1ij}$$

with

$$p_{1ij} = 1/2\ h^{-d}\ \mathcal{K}\left(\frac{z_i - z_j}{h}\right) \left(\frac{a(y_i, x_i) y_j I_i}{F(z_i)} + \frac{a(y_j, x_j) y_i I_j}{F(z_j)} \right)$$

and

$$U_2 = \binom{N}{2}^{-1} \sum_{i=1}^{N} \sum_{j=i+1}^{N} p_{2ij}$$

where

$$p_{2ij} = 1/2\ h^{-d}\ \mathcal{K}\left(\frac{z_i - z_j}{h}\right) \left(\frac{a(y_i, x_i) M(z_i) I_i}{F(z_i)} + \frac{a(y_j, x_j) M(z_j) I_j}{F(z_j)} \right)$$

If $\hat{p}_{1ij}$ and $\hat{p}_{2ij}$ denote the above expressions evaluated at $\hat{\delta}$, $\hat{M}$, $\hat{F}$, $\hat{I}$ and the bandwidth used for estimation, then the influence term r_{AMi} is estimated by $\hat{r}_{AMi} = N^{-1} \sum_j (\hat{p}_{1ij} - \hat{p}_{2ij}) \hat{I}_i$. Carrying out these manipulations for the "right" adjustment RA_N and and the "left" adjustment LA_N give the estimators presented in Appendix 2.

Therefore, the remainder of the proof of Theorem 1 rests on the validity of

$$\sqrt{N}\ (\hat{\gamma} - \bar{\gamma}) = RA_N - LA_N + o_p(1)$$

This equation is demonstrated by verifying two features: namely that trimming with regard to the estimated density gives the same results as trimming with regard to the true density; and that the equation can be linearized into the adjustment terms above.

The first piece requires showing that the estimated trimming index $\hat{I}_i$ =

$1[\hat{f}(x_i) > b]$ can be replaced by $I_i = 1[f(x_i) > b]$ in the terms

$$N^{-1/2} \sum (\hat{g}_i - \check{g})(y_i - \hat{G}_i)\hat{I}_i$$

$$N^{-1} \sum (\hat{g}_i - \check{g})^2 \hat{I}_i$$

that comprise $\hat{\gamma}$, without affecting their asymptotic distribution. This feature follows from a term-by-term analysis which we highlight below. In particular, we have that

$$N^{-1/2} \sum (\hat{g}_i - \check{g})(y_i - \hat{G}_i)(\hat{I}_i - I_i) = N^{-1/2} \sum (\hat{g}_i - g_i)(y_i - G_i)(\hat{I}_i - I_i)$$

$$- N^{-1/2} \sum (\hat{g}_i - g_i)(\hat{G}_i - G_i)(\hat{I}_i - I_i) - N^{-1/2} \sum (g_i - E(gI))(\hat{G}_i - G_i)(\hat{I}_i - I_i)$$

$$+ N^{-1/2} \sum (g_i - E(gI))(y_i - G_i)(\hat{I}_i - I_i) - N^{-1/2} \sum (\check{g} - E(gI))(y_i - G_i)(\hat{I}_i - I_i)$$

$$+ N^{-1/2} \sum (\check{g} - E(gI))(\hat{G}_i - G_i)(\hat{I}_i - I_i)$$

and

$$N^{-1} \sum (\hat{g}_i - \check{g})^2 (\hat{I}_i - I_i) = N^{-1} \sum (\hat{g}_i - g_i)^2 (\hat{I}_i - I_i) + N^{-1} \sum (\check{g} - E(gI))^2 (\hat{I}_i - I_i)$$

$$+ N^{-1} \sum (g_i - E(gI))^2 (\hat{I}_i - I_i) - 2 N^{-1} \sum (\check{g} - E(gI))(g_i - E(gI))(\hat{I}_i - I_i)$$

$$- 2N^{-1} \sum (\hat{g}_i - g_i)(g_i - E(gI))(\hat{I}_i - I_i) + 2 N^{-1} \sum (\hat{g}_i - g_i)(\check{g} - E(gI))(\hat{I}_i - I_i)$$

Each of the terms in these expression can be shown to be $o_p(1)$ by a similar method, which we outline as follows. Begin by noting that that our assumptions implies uniform convergence of $\hat{f}(x)$ to $f(x)$ (when $f(x) > \epsilon > 0$), so that with high probability

$$f(x) - c_N < \hat{f}(x) < f(x) + c_N$$

where $c_N = \psi\, [(Nh_f^k/\ln N)^{-1/2}]$, ψ a constant. If $\tilde{I} = 1[b-c_N < f(x) \leq b+c_N]$, note that

$$\mathrm{Prob}(\ \hat{I} - I \text{ nonzero}) = E[(\hat{I} - I)^2] \leq E(\tilde{I}) = \Psi\ [(Nh_f^{\ k}/\ln N)^{-1/2}].$$

Further, let $N_I = \sum (\hat{I}_i - I_i)^2$ denote the number of nonzero terms in each of the terms above.

To illustrate how the terms are analyzed, consider the first term of the first expression, for which we have

$$N^{-1}E[\sum (\hat{g}_i - g_i)(y_i - G_i)(\hat{I}_i - I_i)]^2$$

$$\leq N\ [\mathrm{Prob}(\hat{I}_i - I_i \text{ nonzero})]\ (\sup |\hat{g}_i - g_i|)^2\ [\sum |(y_i - G_i)(\hat{I}_i - I_i)|/N_I]^2 + o_p(1)$$

$$\leq O[\ N\ (Nh_f^{\ k}/\ln N)^{-1/2}\ (Nh_0^{\ d_0}/\ln N)^{-1}] = O([N\ h_f^{\ k} h_0^{\ 2d_0}]^{\ -1/2}\ (\ln N)^{3/2}]$$

$$= o(1)$$

given our bandwidth conditions. Similarly, the third term of the second expression is

$$N^{-2}\ E[\sum(g_i - E(gI))^2(\hat{I}_i - I_i)]^2 = (N_I/N)^2\ E[\sum(g_i - E(gI))^2(\hat{I}_i - I_i)/N_I]^2$$

$$= O[(Nh_f^{\ k}/\ln N)^{-1}] = o(1)$$

and so forth. All the other terms are treated similarly.

Finally, with trimming based on the true density, the linearization is shown by uniformity arguments analogous to those used above. Denote the sample variance based on trimming with the true density as $S_{\hat{g}I} = N^{-1} \sum (\hat{g}_i - \check{g})^2 I_i$. It is easy to show that $\mathrm{plim}\ S_{\hat{g}I} = \sigma_g$, so

$$\sqrt{N}\,(\hat{\gamma} - \tilde{\gamma}) = \frac{1}{S_{\hat{g}I}} N^{-1/2} \left(\sum (\hat{g}_i - \check{g})(y_i - \hat{G}_i)I_i - \sum [g_i - E(g)](y_i - G_i)I_i\right)$$

$$+ \frac{S_g - S_{\hat{g}I}}{S_g S_{\hat{g}I}} \left(N^{-1/2} \sum [g_i - E(g)](y_i - G_i)I_i\right)$$

$$= \frac{1}{\sigma_g} N^{-1/2} \left(\sum (\hat{g}_i - \check{g})(y_i - \hat{G}_i)I_i - \sum [g_i - E(g)](y_i - G_i)I_i\right)$$

$$+ o_p(1)$$

so we focus on the overall adjustment term

$$\mathrm{ADJ}_N = N^{-1/2} \left(\sum (\hat{g}_i - \check{g})(y_i - \hat{G}_i)I_i - \sum [g_i - E(g)](y_i - G_i)I_i\right)$$

Some tedious arithmetic gives that

$$\mathrm{ADJ}_N = N^{-1/2} \sum (\hat{g}_i - g_i)(y_i - G_i)I_i - N^{-1/2} \sum [g_i - E(g)](\hat{G}_i - G_i)I_i)$$

$$- T_{1N} + T_{2n} + T_{3N}$$

where

$$T_{1N} = [\check{g} - E(g)]\; N^{-1/2} \sum (y_i - G_i)I_i$$

$$T_{2N} = N^{-1/2} [\check{g} - E(g)]\; (\check{G} - \bar{G}) \sum I_i$$

$$T_{3N} = N^{-1/2} \sum (\hat{g}_i - g_i)(G_i - \hat{G}_i)I_i$$

Moreover, by the methods used above, if is easy to verify that each T is $o_p(1)$. For instance, for T_{3N}, we have

$$|T_{3N}| = N^{1/2} \sup(|(\hat{g}_i - g_i)I_i|) \sup(|(G_i - \hat{G}_i)I_i|)$$

$$= O_p[N^{-1/2} h_0^{-d_0/2} h_1^{-d_1/2} (\ln N)] = o_p(1)$$

since $Nh_0^{d_0} h_1^{d_1}/(\ln N)^2 \to \infty$. The other terms follow similarly. Thus, we have that

$$ADJ_N = RA_N - LA_N + o_p(1)$$

which completes the proof of the Theorem. QED

Appendix 2: Variance Adjustment Terms

Recall that we use subscript "i" to compactly denote evaluation of relevant terms at $(y,x) = (y_i,x_i)$; for instance, g_i denotes g evaluated at $\hat{z}_{0i}$, $\hat{G}_i$ denotes $\hat{G}$ evaluated at $\hat{z}_{1i}$, and $\hat{I}_i$ is the trim indicator that is 1 if $\hat{f}(x_i) > b$, and 0 otherwise, as above.

To account for the estimation of δ (or a subvector), we use the "slope" influence estimator discussed in Stoker (1992), namely

$$\hat{r}_\delta(y_i,x_i) = [N^{-1} \textstyle\sum_i \hat{\ell}_i \hat{I}_i \, (x_i - \bar{x})^T]^{-1}$$

$$\left(\hat{\ell}_i \hat{I}_i \hat{\upsilon}_i + N^{-1} h_f^{-k} \sum_{j=1}^{N} \left[h_f^{-1} \mathcal{K}_f{}' \left(\frac{x_i - x_j}{h_f} \right) - \mathcal{K}_f \left(\frac{x_i - x_j}{h_f} \right) \hat{\ell}_j \right] \frac{\hat{I}_j \hat{\upsilon}_j}{\hat{f}_j} \right)$$

where $\hat{\upsilon}_i = (y_i - \bar{y}) - (x_i - \bar{x})^T \hat{\delta}$ is an estimated residual. The asymptotic covariance matrix of $\hat{\delta}$ is estimated as the sample variance of $\hat{r}_\delta(y_i,x_i)$.

The adjustment terms are given as follows. The "right-hand" adjustment is

$$ra_i = \hat{r}_{gi} + \hat{B}_0 \, \hat{r}_{\delta 0}(y_i,x_i)$$

where $\hat{r}_{\delta 01}$ refers to the subvector of $\hat{r}_{\delta 0}$ corresponding to the coefficients of the more general (right hand) regression function, and where

$$\hat{r}_{gi} =$$

$$h_0^{-d_0} \sum_{j=1}^{N} \left(\mathcal{K}_0 \left(\frac{\hat{z}_{0i} - \hat{z}_{0j}}{h_0} \right) \left(\frac{(y_i - \hat{G}_i) y_j \hat{I}_i}{\hat{F}_{0i}} + \frac{(y_j - \hat{G}_j) y_i \hat{I}_j}{\hat{F}_{0j}} \right) - \right.$$

$$\left. \mathcal{K}_0 \left(\frac{\hat{z}_{0i} - \hat{z}_{0j}}{h_0} \right) \left(\frac{(y_i - \hat{G}_i) \hat{g}_i \hat{I}_i}{\hat{F}_{0i}} + \frac{(y_j - \hat{G}_j) \hat{g}_j \hat{I}_j}{\hat{F}_{0j}} \right) \right)$$

Recall $B_0 = 0$ if $\hat{m}$ does not have an index variable as an argument, otherwise

$$B_0 = N^{-1} \sum \hat{D}_0 (y_i - \hat{G}_i) \hat{I}_i$$

where $\mathcal{K}_0'$ denotes the derivative of $\mathcal{K}_0$ with regard to its index argument, and

$$\hat{D}_0 = s_{0\mathcal{K}}(x_i)^{-1} \sum_{j=1}^{N} \frac{x_{0i}-x_{0j}}{h_0} \mathcal{K}_0'\left(\frac{(x_{0i}-x_{0j})^T\hat{\delta}}{h_0}; \frac{x_{1i}-x_{1j}}{h_0}\right) y_j$$

$$- \hat{g}_i\, s_{0\mathcal{K}}(x_i)^{-1} \sum_{j=1}^{N} \frac{x_{0i}-x_{0j}}{h_0} \mathcal{K}_0'\left(\frac{(x_{0i}-x_{0j})^T\hat{\delta}}{h_0}; \frac{x_{1i}-x_{1j}}{h_0}\right)$$

$$s_{0\mathcal{K}}(x_i) = \left[\sum_{j=1}^{N} \mathcal{K}_0\left(\frac{(x_{0i}-x_{0j})^T\delta}{h_0}; \frac{x_{1i}-x_{1j}}{h_0}\right)\right].$$

Finally, the "left hand" adjustment is

$$la_i = \hat{r}_{Gi} + \hat{B}_1\, \hat{r}_{\delta 01}(y_i, x_i)$$

where

$$\hat{r}_{Gi} = h_1^{-d_1} \sum_{j=1}^{N} \left(\mathcal{K}_1\left(\frac{\hat{z}_i - \hat{z}_j}{h_1}\right) \left(\frac{(\hat{g}_i - \check{g})y_j\hat{I}_i}{\check{F}_{1i}} + \frac{(\hat{g}_j - \check{g})y_i\hat{I}_j}{\check{F}_{1j}} \right) - \right.$$

$$\left. \mathcal{K}_1\left(\frac{\hat{z}_i - \hat{z}_j}{h_1}\right) \left(\frac{(\hat{g}_i - \check{g})\hat{G}(\hat{z}_j)\hat{I}_i}{\check{F}_{1i}} + \frac{(\hat{g}_j - \check{g})\hat{G}(\hat{z}_j)\hat{I}_j}{\check{F}_{1j}}\right)\right)$$

$$B_1 = N^{-1} \sum \hat{D}_1 (\hat{g}_i - \check{g}) \hat{I}_i$$

and where $\mathcal{K}_1'$ denotes the derivative of $\mathcal{K}_1$ with regard to its index argument

$$\hat{D}_1 = S_{1\mathcal{K}}(x_i)^{-1} \sum_{j=1}^{N} \frac{x_{0i}-x_{0j}}{h_1} \mathcal{K}_1'\left(\frac{(x_{0i}-x_{0j})^T\hat{\delta}}{h_1}; \frac{x_{1i}-x_{1j}}{h_1}\right) y_j$$

$$- \hat{G}_i S_{\mathcal{K}}(x_i)^{-1} \sum_{j=1}^{N} \frac{x_{0i}-x_{0j}}{h_1} \mathcal{K}_1'\left(\frac{(x_{0i}-x_{0j})^T\hat{\delta}}{h_1}; \frac{x_{1i}-x_{1j}}{h_1}\right)$$

$$S_{1\mathcal{K}}(x_i) = \left[\sum_{j=1}^{N} \mathcal{K}_1\left(\frac{(x_{0i}-x_{0j})^T\delta}{h_1}; \frac{x_{1i}-x_{1j}}{h_1}\right)\right]$$

With these assignments, the asymptotic variance of $\hat{\gamma}$ is estimated as the sample covariance $\hat{\sigma}_\gamma$ of

$$\hat{r}_{\gamma i} = s_{\hat{g}}^{-1}([\hat{g}_i - \bar{g}]\, \hat{u}_i + ra_i - la_i)\hat{I}_i$$

and so the variance of $\hat{\gamma}$ is estimated by $\hat{\sigma}_\gamma/N$.

Notes

1 We focus on kernel estimators in line with our application, but not because their analysis is necessarily easier than with other estimators. In particular, Newey (1992) develops some prerequisite theory for polynomial estimators of index coefficients (average derivatives), which would provide the initial foundation for tests based on polynomial estimation of the restricted and general models.

2 We could likewise apply our test using other kinds of index models as either the restricted model (null) or the general model (alternative), such as the multiple index model $m(x) = G(x_1^T\beta_1, x_2^T\beta_2)$. The key requirement for our development is that the restricted model is nested in the more general model, as discussed later.

3 See Härdle (1991) for a thorough development of nonparametric regression estimation and Stoker (1992) for a discussion of average derivatives, kernel estimation and the connection to index models. This choice of estimation method has some attractive features, such as permitting two-stage estimation of index models (estimate coefficients, then estimate unknown functions) as described below. However, there are many alternatives methods, such as estimation of the coefficients and unknown functions simultaneously by least squares. Stoker (1992) gives references to many of the proposals for estimating index models.

4 We include the constant term to permit minor differences in the mean of the fitted values of the restricted and general models.

[5] This "goodness of fit" interpretation may not apply for parametric model - semiparametric model comparisons where different estimation methods are used for the restricted and unrestricted models. For example, when the null hypothesis is a linear model, the mean of y conditional on the index $x^T\beta$ will be nonlinear under general alternatives, so that the relevant analog of (2.15) will not hold. Also, for considering other semiparametric methods of estimating index models, the nesting implied in (2.15) is what is necessary for γ to have the goodness of fit interpretation; one could always base a test on $\gamma = 0$, but the nesting is required to assert that $\gamma \neq 0$ when the general model captures structure that the restricted model misses.

[6] We have written a companion paper, Rodriguez and Stoker (1993), that discusses various issues of using hedonic price equations, as well as gives a detailed graphical analysis of the partial index model PARTIAL2, the "final model" that results from the specification testing.

[7] We do not take account of the jointness of the hypotheses to be tested. It would be useful to develop Bonferoni critical values or a Scheffe S-method for the tests involved with characterizing index structure.

[8] These are "indirect IV" estimates in the terminology of Stoker (1992). Details on estimation are discussed in Section 4.

[9] Strictly speaking, this is a test of the equality of the average derivative $\delta = E(m')$ and the limit of the OLS coefficient $\beta = [Var(x)]^{-1}Cov(x,y)$, which must coincide when the model is linear.

[10] Two observations are warranted on the fact that the log-linear model appears to explain more structural variation than the single index model (17% versus 23% unexplained structural variation. First, there is nothing in the estimation than constrains $\hat{\gamma}$ to be lower in this case. Second, in practical terms, the fitted OLS equation may not well approximate the nesting condition (2.15); we likely don't have that $E[g(x)|x^T\hat{\beta}] \cong x^T\hat{\beta}$.

[11] LSTAT is the log of the percentage of adults without a high school education who are employed.

[12] The specifications used in Section 3 are discussed in Section 4.3 below.

[13] A kernel $\mathcal{K}$ is of order P if $\int \mathcal{K}(u)du = 1$, and "moments" $\int \Pi u_j^{\alpha_j} \mathcal{K}(u)du = 0$ when $\sum \alpha_j < P$; $\int \Pi u_j^{\alpha_j} \mathcal{K}(u)du \neq 0$ when $\sum \alpha_j = P$.

[14] The relevant bandwidth values were found via a grid search. They are as follows: (1) for the density estimate used for average derivatives and for trimming, and for the general kernel regression, $h = 1$, (2) for the univariate regression G_1 of the single index model, $h = .04$, (3) for the bivariate regression of PARTIAL1, $h = .07$, and (4) for the trivariate regression of PARTIAL2, $h = .10$.

References

Belsley, D.A., E. Kuh and R.E. Welsch (1980), *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, New York, Wiley.

Bierens, H.J. (1990), "A Consistent Conditional Moment Test of Functional Form," *Econometrica*, 58, 1443-1458.

Craven, P. and G. Wahba (1979), "Smoothing Noisy Data With Spline Functions: Estimating the Correct Degree of Smoothing by the Method of Generalized Cross-Validation," *Numerische Mathematik*, 31, 377-403.

Eubank, R. and C. Spiegelman (1990), "Testing the Goodness of Fit of a linear model via Nonparametric Regression Techniques," *Journal of the American Statistical Association*, 85, 387-392.

Ellison, G. and S. Fisher-Elison (1992), "A Simple Framework for Nonparametric Specification Testing," draft, December, NBER.

Härdle, W. (1991), *Applied Nonparametric Regression*, Cambridge, Cambridge University Press, Econometric Society Monographs.

Härdle, W. and Stoker, T.M.(1989), "Investigating Smooth Multiple Regression by the Method of Average Derivatives," *Journal of the American Statistical Association*, 84, 986-995.

Harrison, D. and D.L. Rubinfeld, (1978a), "Hedonic Housing Prices and the Demand for Clean Air," *Journal of Environmental Economics and Management*, 5, 81-102.

Harrison, D. and D.L. Rubinfeld, (1978b), "The Distribution of Benefits from Improvements in Urban Air Quality, *Journal of Environmental Economics and Management*, 5, 313-332.

Hong, Y. and H. White (1991), "Consistent Specification Testing via Nonparametric Regression," draft, Department of Economics, University of California at San Diego, December.

Hong, Y. and H. White (1993), "M-Testing Using Finite and Infinite Dimensional Parameter Estimators," draft, Department of Economics, University of California at San Diego, January.

Horowitz, J.L. and W. Härdle (1992), "Testing a Parametric Model Against a Semiparametric Alternative," Department of Economics Working Paper No. 92-06, University of Iowa.

Lewbel, A. (1992), "Consistent Tests of Nonparametric Regression and Density Restrictions," draft, Department of Economics, Brandeis University.

Newey, W.K. (1991), "The Asymptotic Variance of Semiparametric Estimators," Working Paper No. 583, Department of Economics, MIT, revised July.

Newey, W.K. (1992), "Kernel Estimation of Partial Means and a General Variance Estimator," draft, Department of Economics, MIT, January.

Rodriguez, D. and T.M. Stoker (1993), "Semiparametric Measurement of Environmental Effects," MIT CEEPR Working Paper, June.

Stoker, T.M. (1991), "Equivalence of Direct, Indirect and Slope Estimators of Average Derivatives," in *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, Barnett, W.A., J.L. Powell and G.Tauchen, eds., Cambridge University Press.

Stoker, T.M. (1992), *Lectures on Semiparametric Econometrics*, CORE Foundation, Louvain-la-Neuve.

Stoker, T.M. (1993a), "Smoothing Bias in Density Derivative Estimation," forthcoming *Journal of the American Statistical Association*.

Stoker, T.M. (1993b), "Smoothing Bias in the Measurement of Marginal Effects," MIT Sloan School of Management Working Paper, revised January.

Wooldridge, J. (1992), "A Test for Functional Form Against Nonparametric Alternatives," *Econometric Theory*, 8, 452-475.

Yatchew, A.J. (1988), "Nonparametric Regression Test Based on Least Squares," *Econometric Theory*, 8, 452-475.

TABLE 3.1: VARIABLE SPECIFICATION IN THE BOSTON HOUSING DATA

$y = \ln p$	LMV	log of home value
x_1	NOXSQ	nitrogen oxide concentration
x_2	CRIM	crime rate
x_3	RMSQ	number of rooms squared
x_4	DIS	distance to employment centers
x_5	RAD	accessibility to radial highways
x_6	TAX	tax rate
x_7	PTRATIO	pupil teacher ratio
x_8	B	$(B_k - .63)^2$, where B_k is proportion of black residents in neighborhood
x_9	LSTAT	log of proportion of residents of lower status

TABLE 3.2: COEFFICIENT ESTIMATES FOR THE HOUSING PRICE EQUATION

$y = \ln p$	LMV	Average Derivatives $\hat{\delta}$	OLS $\hat{\beta}$
x_1	NOXSQ	-.0034	-.0060
		(.0035)	(.0011)
x_2	CRIM	-.0256	-.0120
		(.0056)	(.0012)
x_3	RMSQ	.0106	.0068
		(.0025)	(.0012)
x_4	DIS	-.0746	-.1995
		(.0504)	(.0265)
x_5	RAD	.0669	.0977
		(.0468)	(.0183)
x_6	TAX	-.0009	-.00045
		(.0003)	(.00011)
x_7	PTRATIO	-.0175	-.0320
		(.0152)	(.0047)
x_8	B	-.0526	.3770
		(7.514)	(.1033)
x_9	LSTAT	-.2583	-.3650
		(.0370)	(.0225)

(Standard Errors in Parentheses)

WALD TEST OF $\delta = \beta$: $W = 13.44$, $\text{Prob}(\chi^2(9) > 13.44) = .143$

TABLE 3.3: REGRESSION TESTS OF FUNCTIONAL FORM

TESTS AGAINST GENERAL REGRESSION

Restricted	Unrestricted	$\hat{\gamma}$	t value	Prob $[\chi^2(1) > t^2]$
LINEAR	GENERAL	.1712	3.41	.0006
INDEX	GENERAL	.2314	5.96	0.0
PARTIAL1	GENERAL	.0718	4.52	0.0
PARTIAL2	GENERAL	.0116	2.19	.0291

PARTIAL INDEX MODEL TESTS

Restricted	Unrestricted	$\hat{\gamma}$	t value	Prob $[\chi^2(1) > t^2]$
LINEAR	INDEX	.0276	.52	.602
LINEAR	PARTIAL2	.1862	4.51	0.0
INDEX	PARTIAL1	.1975	4.59	0.0
PARTIAL1	PARTIAL2	.0893	3.72	.0002

TABLE 3.4: ADJUSTED AND UNADJUSTED STANDARD ERROR ESTIMATES

TESTS AGAINST GENERAL REGRESSION

Restricted	Unrestricted	$\hat{\gamma}$	Standard OLS	Hetero. Consist. (White)	Corrected for NP Estimation
LINEAR	GENERAL	.1712	.0211	.0268	.0500
INDEX	GENERAL	.2314	.0224	.0311	.0388
PARTIAL1	GENERAL	.0718	.0131	.0149	.0157
PARTIAL2	GENERAL	.0116	.0053	.0047	.0053

PARTIAL INDEX MODEL TESTS

Restricted	Unrestricted	$\hat{\gamma}$	Standard OLS	Hetero. Consist. (White)	Corrected for NP Estimation
LINEAR	INDEX	.0276	.0252	.0279	.0530
LINEAR	PARTIAL2	.1862	.0186	.0255	.0413
INDEX	PARTIAL1	.1975	.0232	.0301	.043
PARTIAL1	PARTIAL2	.0893	.0122	.0146	.0240

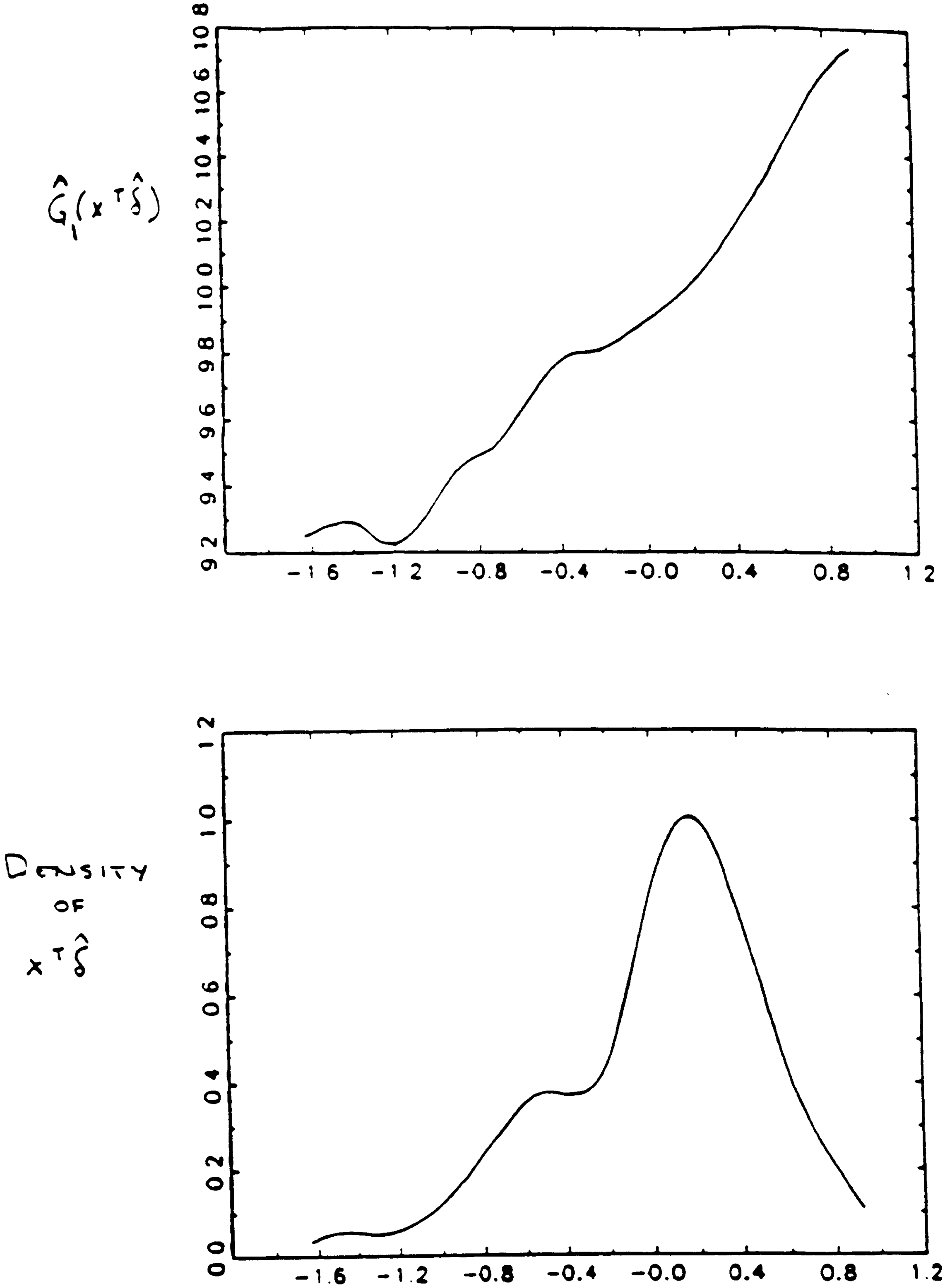

Figure 3.1
Single Index Function for Housing Data

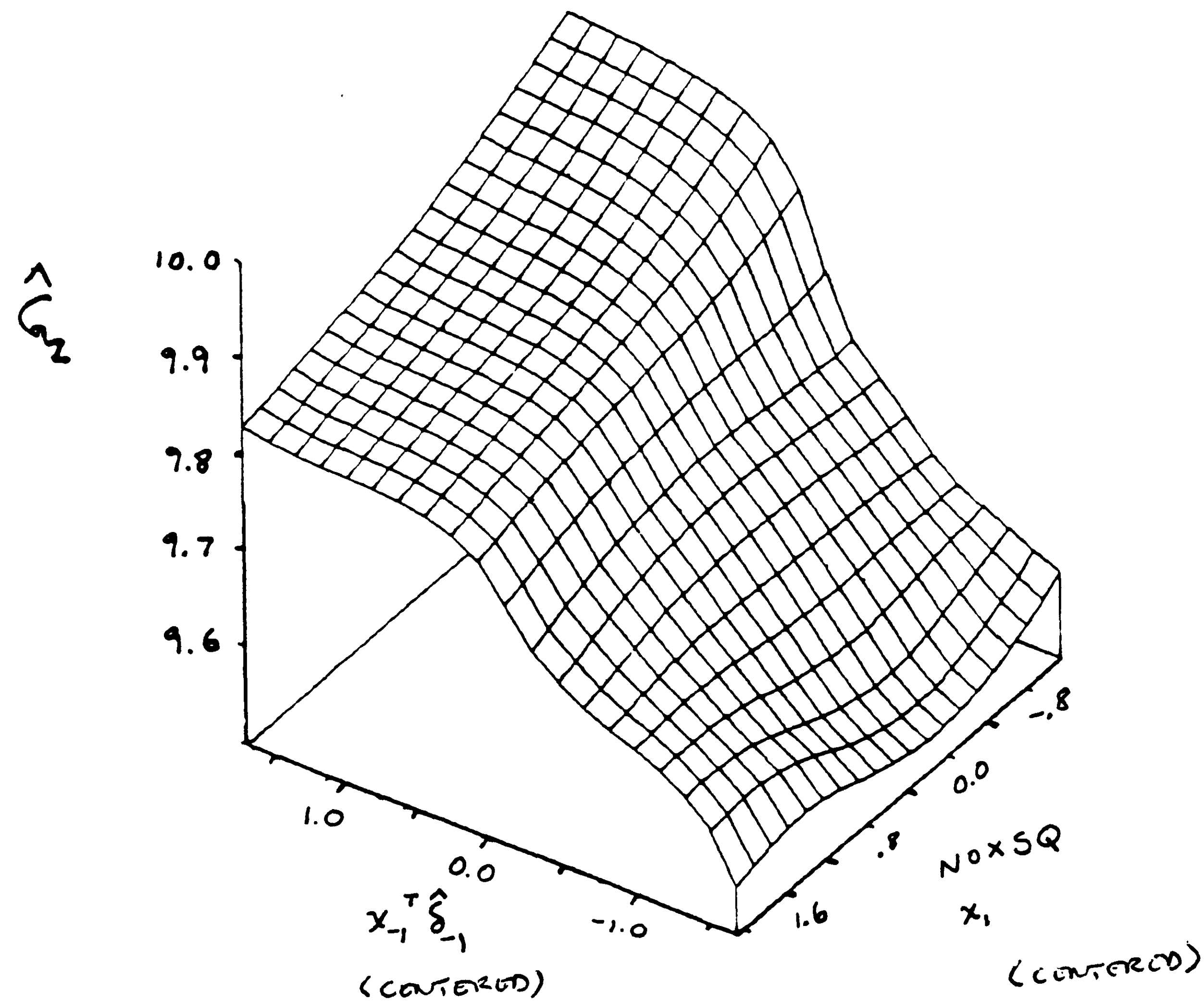

Figure 3.2
Effects of NOXSQ and Index Variable; Model PARTIAL1

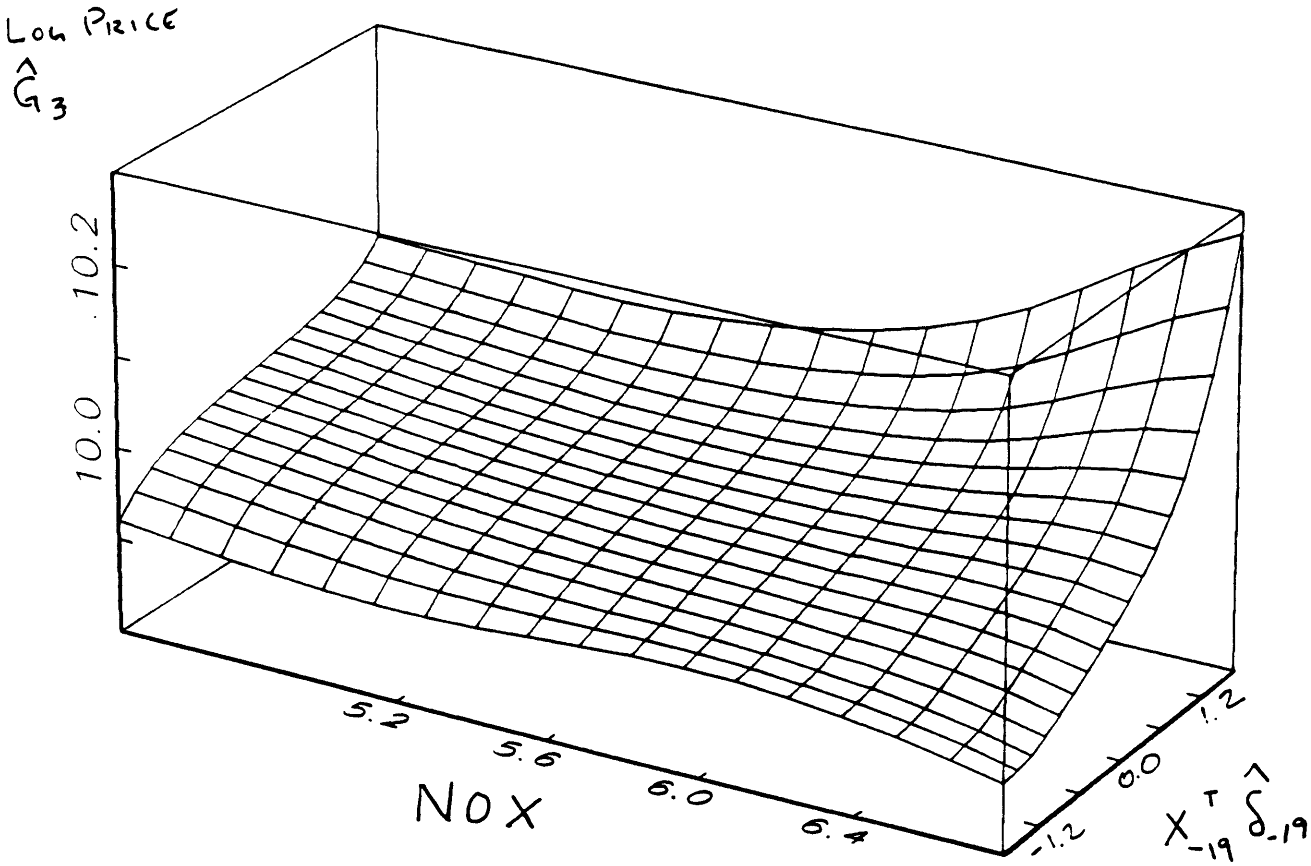

FIGURE 3.3a

EFFECTS OF NOX AND INDEX VARIABLE:

MODEL PARTIAL 3:

LSTAT CONSTANT.

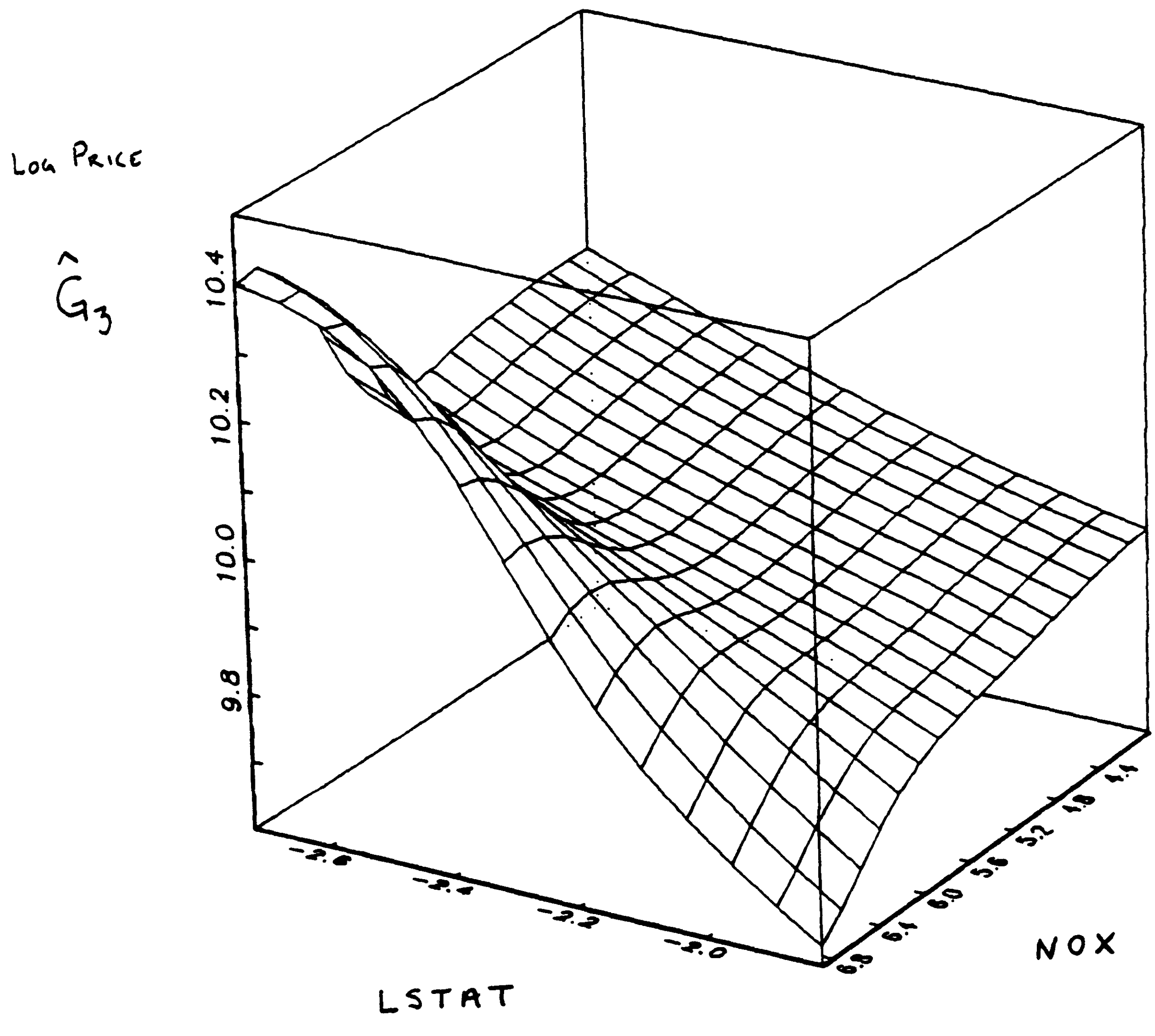

FIGURE 3.3b

EFFECTS OF NOX AND LSTAT:

MODEL PARTIAL2:

INDEX $x_{-19}^{T} \hat{\delta}_{-19}$ CONSTANT