A Nonparametric Residual-based Specification Test: Asymptotic, Finite-sample, and Computational Properties

by

Sara Fisher Ellison

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Economics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 1993

Author ....
Department of Economics
April 22, 1993

Certified by ....
Whitney Newey
Professor of Economics
Thesis Supervisor

Certified by ....
Jerry Hausman
Professor of Economics
Thesis Supervisor

Accepted by ....
Richard Eckaus
Chairman, Departmental Committee on Graduate Students

A Nonparametric Residual-based Specification Test: Asymptotic, Finite-sample, and Computational Properties

by

Sara Fisher Ellison

Submitted to the Department of Economics
on April 22, 1993, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Economics

Abstract

The main purpose of this chapter is to present a general framework for consistent testing of the specification of parametric conditional means. We discuss in detail two examples of tests in this framework, one motivated by the Davidson-MacKinnon test for nonnested hypotheses using kernel regression estimators, the other a much simpler version, computationally and conceptually. We prove asymptotic properties of the general tests, exhibit finite sample properties through simulations, and present an application of the test to some parametric models of gasoline demand.

In the second chapter we present a nonparametric test for the rank of a demand system. It is nonparametric in the sense that the test does not depend on the parametric specification of the Engel curves which comprise the demand system. This test is based on the nonparametric specification test of Ellison and Ellison (1992), a variant of the test presented in the first chapter, necessarily adapted here for the case of measurement error in the regressors. In addition to examining the test results, this paper also compares bootstrapped and simulated null distributions of the test statistics to check how well the theoretical predictions of Ellison and Ellison hold in this particular application.

Finally, the third chapter looks at the problem, common to all nonparametric tests, of window width choice. we critique various methods found in the literature and propose original ideas for choosing window width.

Thesis Supervisor: Whitney Newey
Title: Professor of Economics

Thesis Supervisor: Jerry Hausman
Title: Professor of Economics

Acknowledgments

There are many persons without whose help this thesis would not have been possible. I would like to thank my advisors, Whitney Newey and Jerry Hausman. They generously read and commented on work for this thesis at all stages. In addition, I received valuable experience (not to mention money) in the employ of Jerry Hausman for two summers. I would like to thank other faculty members at M.I.T. who provided invaluable guidance at early stages of the project: Danny Quah, Tom Stoker, and Jeff Wooldridge. Guido Imbens and Roger Klein were kind enough to read the first chapter carefully and make helpful comments. Greg Leonard helped me obtain data for the second chapter. The National Science Foundation provided financial support. Also, there were several students with whom I discussed my work: Chunrong Ai, Yacine Ait-sahalia, Joseph Beaulieu, Wallace Mullin, and Diego Rodriguez.

Lest one think the only sustenance I received from others while at M.I.T. was intellectual, I offer a heartfelt thanks to my friends who made my four years in graduate school more enjoyable. Among them are Joe Beaulieu, Linda Bui, Clara Chan, Judy Chevalier, Maura Doyle, Jan Eberly, Chad Jones, Katie MacFarlane, Chris Mayer, Bill Miracky, Wally Mullin, Gerald Oettinger, Danny Quah, Makoto Saito, Ted Sims, Chris Snyder, Jeff Zwiebel, the "Blue Lounge" lunch crowd, Bureau friends, the lacrosse team, pottery class members, euchre partners, the Earthen Vessels group, Churchill Scholars, and many others.

My parents and siblings deserve a lion's share of the credit, supportive of me from the first mumblings that I wanted to go to graduate school in economics. (My mother even typed my applications!) I thank them for their unflagging faith in me. My husband's parents, too, were a font of support and encouragement.

Finally, my largest debt of gratitude is owed to my husband Glenn. He has been an advisor, a mentor, a taskmaster, a proofreader, a supporter, a referee, a coauthor, and a friend all of my years in graduate school.

After partitioning credit and assigning it to so many deserving persons, I regret only that the whole was not greater so as to enlarge every person's portion.

Contents

Introduction

Economists have always had practical and immediate questions to answer: Will interest rates go up next year? Will a gasoline tax effect my state's tourism industry? Do doctors in different states have different prescribing practices? Statistical model building alone cannot answer such questions. Therefore, econometric theory, as distinct from statistical theory, has answered the call of economists and has always had as its particular focus hypothesis testing. Far from limiting attention paid to model building, hypothesis testing has enhanced it in the following way. As economists built ever more complex statistic models, questions of immediate practical significance, like those above, were supplemented with questions about the validity of the statistical model: Is assuming linearity in the relationship between rainfall and farm production valid? Are production and investment jointly determined? Hypothesis testing which answers such questions, specification testing, is an important complement to model building. This thesis is a small contribution to the large and growing econometric literature of specification testing.

The overarching objective of this thesis is to provide econometric practitioners with a test for checking some aspects of their models, which is intuitive and easy to implement, in addition to assuming as little structure as possible. By assuming little structure, the hope is that the test will provide researchers with the means to detect misspecifications in their models which traditional specification tests might be unable to detect. Again, great care is taken to design a test which does not place undue computational burdens on researchers. The first chapter proposes such a test for the functional form of the economic model. We discuss variations of this test appropriate for different circumstances, discuss its asymptotic behavior, and present results from simulations of the test statistic and from a simple empirical application of the test. The second chapter is an involved empirical application of the test of the first chapter to the problem of testing the rank of a demand system. The problem is discussed in some detail, as is the way in which the test is to be applied to the problem. Also it is necessary to adapt the test to handle the case of error in the explanatory variable. Finally, the testing procedure is carried out and the results are interpreted. The third chapter deals with a problem inherent in all nonparametric tests, this one included: choice of the smoothing parameter. Although the results in this chapter are quite preliminary, we believe that the ideas therein are important and of potentially wide applicability.

Chapter 1

The Testing Framework

1.1 Introduction

One of the most important functions of econometrics is to test the validity of economic theories. In particular, an economic theory might imply a relationship among economic variables, and econometrics is the tool by which these relationships are empirically examined. Most often, these empirical tests take the form of tests of exclusion restrictions. In other words, conditional on a general model being true, one usually tests whether a model nested in the general one explains the data nearly as well. Sometimes, however, one might want to test the *form* of the relationship among economic variables, the functional form of the conditional mean. (Economic theory might imply a specific functional form of the conditional mean, the rejection of which is a rejection of the economic theory.) The test in this thesis is designed mainly as a test of the second type. Using nonparametric techniques, a test is constructed to detect sufficiently large deviations from the null model in all directions. First we discuss a procedure which uses the kernel regression estimate of a conditional mean function to test the goodness of fit of a null parametric model. Second we propose a similar test which is easier to compute than the kernel regression test.

Hausman (1978) was one of the first to discuss the use of specification tests in the econometric literature. He described a so-called test of orthogonality using a model consistent under both the null and alternative to test the specification of a model efficient under the null. Later, Davidson and MacKinnon (1981) discussed a specification test using a parametric alternative which is not nested in the null (or vice versa) to test the null specification. Also, Newey (1985) and others have discussed conditional moment tests, of which specification tests are a special case. These procedures are parametric and not ususally consistent against all deviations from the null.

Recent work on consistent testing includes several papers by Bierens ((1982,1990) and others) on consistent moment tests. In addition, several recent papers have provided comprehensive discussions on nonparametric specification testing. Nonparametric specification testing, in general, is testing using a nonparametric model in place of the usual parametric alternative to test the specification of the null model. Typically, an advantage of nonparametric tests are their consistency properties. A variety of approaches have been taken to the problem of constructing these tests. First, different nonparametric estimators could be employed: spline regression, kernel regression, series regression. Eubank and Spiegelman (ES) (1990) use primarily a spline regression estimator to test null specification. Azzalini, Bowman, and Härdle (ABH) (1989) as well as part of this paper use a kernel regression estimator. In addition, different testing frameworks are possible, such as likelihood ratio-type testing, Lagrange multiplier-type testing, and so forth. ES use a Lagrange multiplier framework. ABH use a likelihood ratio framework. We use a Lagrange multiplier, specifically Davidson-MacKinnon, framework to motivate our statistics, but quickly move beyond that framework. These differences in approach may lead to asymptotically equivalent tests, but they will have different finite-sample properties and different computational characteristics. We focus on those questions in an attempt to make our statistic accessible and easily applicable. First, this chapter will provide a less computationally complex version of our statistic, a version convenient enough that we hope for broad application. Also, we perform simulations to study the finite sample properties of our statistics, such as distribution and power.

Section 1.2 is an informal discussion of the motivations behind our choice of the specific form of the test statistic. In Section 1.3 we write down the model, introduce the version of the test which is simpler to compute and prove the asymptotic normality and consistency of them. Section 1.4 presents simulation results, both on distribution and power. Section 1.5 is a discussion of a simple application of the methods. We conclude in Section 1.6.

1.2 Discussion

This section will discuss two motivations for our test statistics. As discussed in the introduction, most specification tests require a model, an alternative, to test the null specification. These tests detect misspecification only in the direction of the alternative. Tests employing parametric alternatives cannot be consistent against all deviations from the null because fixed parametric alternatives only span a finite dimensional subspace of all possible alternatives. The first motivation for our test statistics is, then, to construct a statistic, using a nonparametric instead of parametric model as an alternative, that will be consistent against *all* sufficiently large deviations from the null. We will use a framework similar to that used by Davidson and MacKinnon (DM). The second way in which we will motivate our test statistic is as a systematic way to look at spatial patterns in residuals from the null model. We will now informally discuss the motivations. In section 3 we will show how they are related.

Assume we have a null model $y = X\beta + u$, where u is *i.i.d.* with mean 0 and variance σ^2. [1] Estimates from this model are denoted with a tilde: $\tilde{y}, \tilde{u}, \tilde{\sigma}$. The DM test of the null model is based on the statistic

$$T = \frac{1}{c_N} \sum_{i=1}^{N} \lambda_i^T \tilde{u}_i,$$

where λ_i, the misspecification indicator, is $\hat{y}_i - \tilde{y}_i$, where $\hat{y}_i$ is a fitted value from an alternative, nonnested model, and c_N is a normalizing factor, $\sqrt{N}\tilde{\sigma}_N$. Also, $\cdot^T$ denotes the transpose of a matrix or vector.

[1] Relaxing the assumption of normality of the errors was actually made in work subsequent to Davidson and MacKinnon's original paper.

Obviously, this test statistic will tend to be large (or small) if there is positive (or negative) correlation between the misspecification indicator and the residuals. It is essentially a test of orthogonality between the residuals and the parametric alternative.

Note that the DM procedure requires the specification of a parametric alternative model with which to test the null model. Davidson and MacKinnon make the point that the alternative model need not be one in which the econometrician has any faith; however, a particular model will only be useful in detecting misspecification of the conditional mean in the direction of that particular model. Use of a more general alternative, such as a nonparametric alternative, will allow one to dectect any misspecification. Of course, standard asymptotic results for parametric alternatives will not apply when nonparametric alternatives are used. We will develop an asymptotic theory in that case.

The second motivation for our test statistic is a less traditional, more intuitive, view of a specification test. We motivate it as a means to look at spatial patterns in the residuals. Let $\tilde{u}$ be residuals from the null model. Under correct specification, we would not expect the value of $\tilde{u}_i$ to give us much information about the value of $\tilde{u}_j$. Even though we would expect a small amount of negative correlation because the residuals add to zero, we would not expect any spatial or serial correlation. In other words, knowing x_i and x_j should not give us any more information about $\tilde{u}_j$ than just knowing $\tilde{u}_i$. Under incorrect specification, however, we would expect residuals close together to have expectations of the same sign. The pictures in Figure 1.1 illustrate that point. In the first illustration, the null is the truth. The residuals do not seem to be spatially correlated. In the second illustration, the alternative is the truth, and there is clear positive spatial correlation in the residuals. If a systematic study of the residuals found that residuals whose x's were close together had expectations of the same sign, we might suspect misspecification. One possible statistic that would allow us to do that could be based on $\sum_i \sum_j w_{ij}\tilde{u}_j\tilde{u}_i$, where the w_{ij}'s are weights. We choose the weights depending upon which correlations in the residuals we think are informative. Usually, for $\tilde{u}_i$ and $\tilde{u}_j$ residuals from observations far apart, w_{ij} would be small because correlation between $\tilde{u}_i$ and $\tilde{u}_j$ would not be very helpful in detecting misspecification. We would, however, be very interested in positive correlations among groups of residuals close to each other.

It seems clear, then, that w_{ij} for observations i and j far apart should be much smaller than w_{ij} for i and j close. It is not clear, however, at what rate the w_{ij}'s should decrease as observations become further apart. In other words, for a fixed i, we would like to determine the shape the w_{ij}'s ought to be over an area near observation i. (Obviously, this is a problem without a simple general answer. We will discuss it further in the chapter on window width choice.) Note that kernel weights satisfy our intuition about what general shape the weights should take. It will be clear shortly why it is useful to consider the kernel weights in this context.

1.3 The Test Statistic

Assume we wish to test a null model $y = X\beta + u$, where u is *i.i.d.* with mean 0 and variance σ^2. Like before, estimates from this model are denoted with a tilde. Assume also that an alternate nonparametric model is specified. The estimates from the alternative model will be denoted with a hat. For the specific case of a kernel regression estimator, recall that $\hat{y}_i = \sum_j w_{ij} y_j$, where w_{ij} are the kernel weights ($w_{ij} = \frac{1}{h}K(\frac{x-x_i}{h})/\frac{1}{N}\sum_i \frac{1}{h}K(\frac{x-x_i}{h})$). The DM-type statistic would be

$$T = \frac{1}{c_N}\sum_i((\sum_j w_{ij}y_j) - \tilde{y}_i)\tilde{u}_i,$$

c_N is a normalizing factor. We assume the X's have bounded support with the density bounded away from zero on the support. To simplify, we substitute in $x_j\tilde{\beta} + \tilde{u}_j$ for y_j and $x_j\tilde{\beta}$ for $\tilde{y}_j$. Then,

$$T = \frac{1}{c_N}(\tilde{\beta}\sum_i \tilde{u}_i \sum_j w_{ij}x_j + \sum_i\sum_j w_{ij}\tilde{u}_j\tilde{u}_i - \tilde{\beta}\sum_i x_i\tilde{u}_i).$$

Note that the third term within the parentheses is identically zero. Note, also, that the second term, a quadratic form in the residuals, is precisely the type term that would allow us to look at spatial patterns in the residuals, as mentioned in the previous section. The weights are kernel regression weights. The first term goes away asymptotically under certain conditions; therefore, the DM-type statistic is asymptotically equivalent, under those conditions, to a statistic based on a quadratic form in the residuals. See Ellison and Ellison (1992) for a proof and discussion. Below we will consider such a statistic, based on a quadratic form in the residuals.

As we observed before, kernel weights meet the conditions on the matrix W to be useful in detecting departures from the null. Also by choosing a different kernel, we can give the weights a somewhat different shape, and by changing the bandwidth, we can make the weights more or less concentrated around each observation.

Kernel regression provides us with just the type of weights we need, but we should point out that weights satisfying our conditions for being useful in detecting departures from the null need not be kernel weights. In fact, there is a good reason not to use kernel regression weights. A problem with the kernel regression weights being used in the statistic is the resulting computational complexity of the statistic. The task of computing the weights is ususally the most computationally intensive part of performing the specification test. Computing the statistic is an $O(s(N)N)$ operation, where $s(N)$ is the minimum over the observations of the number of points in the support of the kernel. (Some common kernels, such as the normal kernel, have infinite support. The computation then becomes $O(N^2)$.) For this reason we would like to propose use of weights that satisfy the general characteristics we discussed earlier, but are much easier to compute. The simplest approach is to divide the data into bins of size $k(N)$ and set weights equal to each other (all equal $1/k(N)$) inside the bin and equal to zero outside the bin. The computation of these weights is no longer a rate-datermining step since all of the nonzero weights are equal. In addition, a simple trick allows us to compute the statistic in $O(N)$ calculations once we have the weights:

$$\sum_{i,j} w_{ij}\tilde{u}_i\tilde{u}_j = \frac{1}{k(N)^2}\sum_{k=1}^{B}\Big(\sum_{i\in \text{bin}\,k}\tilde{u}_i\Big)^2,$$

where B is the number of bins.

Just as the kernel regression test corresponds to the kernel regression estimator, the bin test corresponds to a bin regression estimator. Given the computational simplicity of the bin test, it is important to keep in mind that, again, we are not suggesting performing a DM-type test with $\hat{y}_i$ the fitted values from a bin regression. Rather, we suggest just computing $\frac{1}{c_N}\tilde{u}^T W \tilde{u}$, with W being the bin weight matrix.

1.3.1 Asymptotic Distribution

We will now present asymptotic results on the distribution of a test statistic based on $\tilde{u}^T W \tilde{u}$. For a test to be easily implemented, it is important for the statistic to have a standard (tabulated) asymptotic distribution. Then, critical values can be easily computed or looked up in a table. We will first argue that $\tilde{u}^T W \tilde{u}$ is asymptotically equivalent to $u^T W u$. Then we will look at the asymptotic distribution of $u^T W u$, suitably normalized. In our discussion of asymptotic results, we will add a subscript of N when it is necessary for clarity. We first note that

$$\tilde{u} = M_X u,$$

where $M_X = I - P_X$, and $P_X = X(X^T X)^{-1} X^T$ is the projection matrix onto X space. Then,

$$\tilde{u}^T W \tilde{u} = u^T M_X^T W M_X u.$$

The distribution, both finite sample and asymptotic, of this quadratic form in independent random variables will depend on the characteristics of the matrix W, as we shall see. In addition, we will now argue that the asymptotic distribution will depend solely on W (not $M_X^T W M_X$), *i.e.*, Proposition 1.1 will show that $\tilde{u}^T W \tilde{u}$ and $u^T W u$ have the same distribution asymptotically.

Proposition 1 *$\tilde{u}^T W_N \tilde{u}$ and $u^T W_N u$ have the same asymptotic distribution.*

<u>Proof</u>

First, note that

$$\begin{aligned}\tilde{u}^T W_N \tilde{u} - u^T W_N u &= u^T(M_X W_N M_X - W_N)u \\ &= u^T(P_X W_N P_X - P_X W_N - W_N P_X)u.\end{aligned}$$

We may assume without loss of generality that W_N is symmetric. Then $P_X W_N = P_X P_X W_N = P_X W_N P_X$, so we can collect terms to get

$$= -u^T(P_X W_N P_X)u.$$

$$P_X W_N P_X = X(X^T X)^{-1} X^T W_N X (X^T X)^{-1} X^T.$$

As before let $\hat{X}$ be the nonparametric estimate of X, $X^T W_N$. Then,

$$\begin{aligned}P_X W_N P_X &= X(X^T X)^{-1} \hat{X}^T X (X^T X)^{-1} X^T \\ &= X(X^T X)^{-1}((X^T X)^{-1} X^T \hat{X})^T X^T \\ &= X(X^T X)^{-1} \hat{I}^T X^T,\end{aligned}$$

where $\hat{I}$ is the matrix of coefficients from a regression of $\hat{X}$ on X. We can now write

$$\tilde{u}^T W_N \tilde{u} = u^T W_N u - u^T P_X u - u^T X (X^T X)^{-1} (X(\hat{I} - I_k))^T u,$$

where k is the dimension of X. The right term clearly can be ignored asymptotically as it has plim zero. The second term has an asymptotic distribution because P_X is idempotent so $u^T P_X u \sim \sigma^2 \chi^2_{col(X)}$. Then, to show asymptotic equivalence, we need only to show that the first term times N^{-D} has an asymptotic distribution for some $D > 0$. We will see in the next lemma that this is true.

QED.

We have shown that we need only consider $u^T W u$ asymptotically, a quadratic form in independent random variables. If we can find the asymptotic distribution of $u^T W u$ (or a nomalized $u^T W u$), we can base a test statistic on it. The following lemma gives us the asymptotic distribution of a quantity based on this quadratic form.

Note that an assumption of normality of u_N is added to prove the lemma. The necessity of the assumption is discussed later in this section.

Lemma 1 *Let W_N be a sequence of $N \times N$ symmetric positive semi-definite matrices with eigenvalues $\tau_{1N} \leq \ldots \leq \tau_{NN}$ and $u_N \sim \mathcal{N}(f_N, \sigma^2 I_N)$.*

Then

$$\frac{u_N^T W_N u_N - \sigma^2 \mathrm{tr} W_N - f_N^T W_N f_N}{\sigma^2 (2\mathrm{tr} W_N^2)^{1/2}} \xrightarrow{\mathcal{L}} \mathcal{Z} \sim \mathcal{N}(0,1)$$

if

$$\text{a)} \quad \frac{\max_h \tau_{hN}^2}{\sum_{l=1}^N \tau_{lN}^2} \longrightarrow 0$$

$$\text{b)} \quad \frac{f_N^T W_N^2 f_N}{\mathrm{tr}(W_N^2)} \longrightarrow 0.$$

Proof

From elementary linear algebra, we can diagonalize W in the following way:

$$u^T W u = u^T \Phi \Lambda \Psi u,$$

where Λ is diagonal, Φ is orthogonal, and $\Psi = \Phi^{-1}$. For W symmetric we may further simplify as

$$\begin{aligned} &= u^T \Phi \Lambda \Phi^T u \\ &= (\Phi^T u)^T \Lambda (\Phi^T u) \\ &= v^T \Lambda v, \end{aligned}$$

where $v \sim \mathcal{N}(0, \sigma^2 I)$.

The final step is true since Φ is orthogonal and $u \sim \mathcal{N}(0, \sigma^2 I)$. The assumption of normality of the u's is necessary to obtain independence of the v's. Otherwise, the v's would merely be uncorrelated. Now, note that $v^T \Lambda v = \sum \lambda_i v_i^2$ is a weighted sum of independent χ^2 random variables. We can then apply a central limit theorem to obtain the limiting normal distribution.

QED.

We will use this lemma to show asymptotic normality of both the kernel regression test and the bin test. Given the lemma, it is easy to see how we can construct a $\mathcal{N}(0,1)$ test statistic based on $u^T W_N u$.

Our normalizing factor c_N will be of the form $\sigma^2(2\mathrm{tr}W_N^2)^{\frac{1}{2}}$. In addition, to obtain a statistic that is standard normal under the null, we will have to subtract off the mean of the statistic, $\sigma^2\mathrm{tr}W_N$. Recall from Proposition 1.1 that $\tilde{u}^T W_N \tilde{u} = u^T W_N u - u^T P_X u + o_p(1)$, and that $u^T P_X u \sim \sigma^2 \chi^2_{col(X)}$. The chi-squared term will not matter asymptotically, but it will in finite samples. To ensure that the expectation under the null of our statistic is correct in finite samples, we will add $E(\sigma^2\chi^2_{col(X)})$ to the numerator. Thus, our statistic is

$$\mathcal{T} = \frac{\tilde{u}_N^T W_N \tilde{u}_N - \hat{\sigma}^2 \mathrm{tr} W_N + \hat{\sigma}^2 col(X)}{\hat{\sigma}^2 (2\mathrm{tr}W_N^2)^{1/2}}$$

for $\hat{\sigma}$ a consistent estimate of σ.

We will now verify conditions of the lemma. Note that under the null, $f_N \equiv 0$, so we need only to verify the condition a) to show asymptotic normality. This condition on the eigenvalues of the weight matrix essentially says that no one eigenvalue can dominate the others.

Proposition 2 *For $\{W_N\}$ a sequence of matrices of bin or kernel weights satisfying conditions to be described below, $\mathcal{T} \xrightarrow{\mathcal{L}} \mathcal{Z} \sim \mathcal{N}(0,1)$.*

Proof

For simplicity, we will assume in the proof that the true σ is known.

Verification of condition a) for bin test:

We note first that the weight matrix for the bin test is symmetric and positive semi-definite. To verify the first condition of the lemma, we note that W_N is a block diagonal matrix with $\frac{N}{k}$ blocks of k^2 elements (each being $\frac{1}{k}$) along the diagonal. Such a matrix will have $\frac{N}{k}$ eigenvalues of one and $N - \frac{N}{k}$ eigenvalues of zero. (The eigenvalues of one correspond to the following eigenvectors: ones in the first k places, zeroes elsewhere; zeroes in the first k places, ones in the next k places, and zeroes elsewhere; and so on. Since a matrix only has the same number of nonzero eigenvalues as its rank, the rest of the eigenvalues of the matrix are zero.) So,

$$\frac{\max_h \tau_{hN}^2}{\sum_l \tau_{lN}^2} = \frac{1}{N/k} = \frac{k}{N}.$$

Note that we can simply choose k fixed, *i.e.,* take a fixed number of observations in each bin. Alternatively, we may let the number in a bin grow, but if we restrict the number of observations in each bin to go to infinity slower than N, we satisfy the condition.

We, therefore, have that the bin test statistic is $\sqrt{k(N)/N}$ asymptotically normal.

Asymptotic nomality can also be shown without the assumption of nomality of u_N by noting that the weight matrix of the bin test is block diagonal. A quadratic form in u_N will, therefore, be the sum of *independent* random variables, and a central limit theorem can be applied. Such an argument cannot be applied in the case of the kernel test because the weight matrix is not block diagonal.

Verification of condition a) for kernel test:

The weight matrix for the kernel test will not be symmetric in general. We symmetrize the matrix by letting $W_N^s = (W_N + W_N^T)/2$. Then $u^T W_N^s u = u^T W_N u$ and W_N^s symmetric. Since the weights in each column of W_N add to one, it is a Markov transition matrix, and a well-known result is that its largest eigenvalue is one. W_N^T's largest eigenvalue is then also one because the transpose of a matrix has the same eigenvalues as the matrix. We then have that the largest eigenvalue of W_N^s is at most one.

We also know that for W_N^s symmetric, $\sum \tau_{lN}^{s2} = \mathrm{tr} W_N^{s2} = \sum_{i,j} w_{ijN}^{s2}$, which is minimized when all nonzero w_{ij}^s's are equal. Let b be the largest number of nonzero weights in any column of W_N. Then, the maximum number of nonzero elements in the entire W_N^* matrix is $2 \cdot N \cdot b$. So, $\sum \tau_{lN}^2 \geq N \cdot b \cdot \frac{1}{4b^2} = \frac{N}{2b}$. Therefore,

$$\frac{\max_h \tau_{hN}^2}{\sum_l \tau_{lN}^2} \leq \frac{2b}{N} \longrightarrow 0$$

as long as the support of the kernel goes to infinity at a rate slower than N. (This condition is not necessary. For kernels with infinite support, we just need that the bandwidth goes to zero at a rate slower than N for the condition on the eigenvalues to hold. We will, though, be primarily concerned with kernels of finite support.)

The kernel regression test statistic is $(\text{tr}W_N^2)^{\frac{1}{2}}$ asymptotically normal, a rate for which we only have a lower bound. It will, of course, depend on the shape of the kernel and the size of its support.

QED.

1.3.2 Consistency

Recall that one of the advantages of nonparametric specification testing over traditional parametric specification tests is their ability to detect departures from the null in *all* directions. We will now show the consistency of our test statistics against all alternatives that shrink at a sufficently slow rate. Specifically, we will show the consistency of the kernel statistic against local alternatives. The proof of consistency of the bin statistic is very similar.

Theorem 1 *Let $\{W_N\}$ be a sequence of weight matrices with window width or bin width $\longrightarrow 0$. Let ζ be such that*

$$\frac{N^{2\zeta+1}}{\text{tr}(W_N^2)^{1/2}} \longrightarrow \infty.$$

Then the test is consistent if the true regression function is $f_0 + f_a N^\zeta$ for some f_a, where f_0 is the null regression function.

Proof

Let f_o be the null function and let $f_o + g_a$ be the alternative (true) function. $g_a = f_a N^\zeta$ for some f_a. We will find a lower bound on ζ such that the test is consistent against $f_a N^\zeta$. (For simplicity, we have assumed that g_a is orthogonal to X. If not, we have

$$\tilde{u}_i = \dot{u}_i + M_X g_a.$$

The argument carries through if we note that

$$M_X g_a(x_i) \xrightarrow{p} g_a(x_i) - x_i \frac{\int x g_a(x) du(x)}{\int x^2 du(x)}.)$$

Let $\tilde{u}$ be the residuals from the (incorrectly specified) regression of y on X.

$$\begin{aligned} \tilde{u} = y - X\tilde{\beta} &= M_X(X\beta + f_a N^\zeta + u) \\ &= \dot{u} + M_X f_a N^\zeta \qquad\qquad \dot{u} = M_X u \end{aligned}$$

The test statistic is

$$\begin{aligned} \mathcal{T} &= \frac{1}{c_N}(\dot{u} + M_X f_a N^\zeta)^T W_N (\dot{u} + M_X f_a N^\zeta) \\ &= \frac{1}{c_N}\dot{u}^T W_N \dot{u} + \frac{2}{c_N}(M_X f_a N^\zeta)^T W_N \dot{u} + \frac{1}{c_N} N^{2\zeta}(f_a^T M_X W_N M_X f_a). \end{aligned}$$

From the null distribution we know that the first term is asymptotically $\mathcal{N}(0,1)$. In the second term we write $(M_X f_a)^T W_N \equiv \hat{h}_a^T$, a nonparametric estimate of

$$h_a = f_a - P_X f_a.$$

By the central limit theorem and the consistency of $\hat{h}_a$,

$$\frac{\hat{h}_a \dot{u}}{N^{1/2}} \rightarrow \mathcal{N}(0, \sigma^2 E(h_a^2)).$$

So, $c_N / N^{\zeta+1/2}$ times the second term goes to a $\mathcal{N}(0, \sigma^2 E(h_a^2))$.

For reasonable assumptions on the weights w_{ij} (such as assuming the bandwidth goes to zero), we will have

$$\frac{(M_X f_a)^T W_N M_X f_a}{N} \xrightarrow{\text{P}} E(h_a^2).$$

And hence, $c_N/N^{2\zeta+1}$ times the third term goes in probability to $E(h_a^2)$.

Now let ζ be any constant such that

$$\frac{N^{2\zeta+1}}{c_N} \to \infty.$$

Then, the first term is bounded in probability, the third term goes to infinity, and the difference in the absolute values of the second and third terms goes to infinity with probability one.

Therefore, $\mathcal{T} \to \infty$ and the test is consistent.

QED.

1.4 Simulation Results

We performed Monte Carlo studies to answer two questions about our test statistics: whether our asymptotic results from the previous section provide a reasonable approximation to finite sample behavior of the statistics and also how powerful the bin test and the kernel regression test are in finite samples. We will not only compare the performance of the bin test with that of the kernel regression test, but also both with the optimal Lagrange multiplier tests. The Epanechnikov kernel is used to construct the kernel weight matrix.

1.4.1 Distribution

First, we provide graphs of smoothed finite sample distributions, Figures 1.2 through 1.5. We include the asymptotic distributions, $\mathcal{N}(0,1)$, for reference. The simulated finite sample distributions were obtained in the following manner: First, the statistics were computed on 1000 simulated data sets using OLS residuals from a model with one explanatory variable distributed uniform over [-1,1], and normal errors. The null hypothesis of no misspecification was imposed. Then a density for the test statistic under the null hypothesis was estimated from the 1000 test statistics using kernel density estimation. (The kernel density estimation was used for expositional purposes only.)

The first set of graphs we present are the results for the bin test run on data sets with 30 observations with two different bin sizes. Note that even with as few as 30 observations, the bin test with five observations in a bin is beginning to look normal. The bin test with ten observations in a bin seems quite skewed and quite far from a standard normal, though. It is not surprising that the test with five observations in a bin, six bins, does a better job approximating a standard normal. Recall that the rate at which these test statistics converge to a normal is proportional to the square root of the number of bins.

The second set of graphs are the simulated finite sample distributions for the kernel tests with three different window widths, again run on data sets with 30 observations. Again, the finite sample distribution of the test with the largest window width is not well approximated by the standard normal. The small window width does a much better job. The very small window width was included to illustrate that the normal approximation does not continue to improve as the window width goes to zero. Obviously, at some point, each window will contain only one observation. The test statistic, then, will have plim zero. Here, the very small window width is small enough so that each window contains an average of about 1.5 observations.

The next set of graphs are finite sample distributions from three different bin tests with 300 observations. Note again that the normal approximation is better for the smaller bins—there are no very small bins this time.

Finally, we present three different window widths for the kernel test on 300 observations in the last set of graphs. The same observations as before apply to this set of graphs.

Note that almost all of the finite-sample distributions are approximated quite well by the standard normal density, especially in the right tails. The right tails are, of course, the important tails for this specification test—it will always be a one-sided test. This evidence is encouraging for the prospect of carrying out this test without simulating critical values.

1.4.2 Power

The next set of simulations was performed to demonstrate the power of the various test statistics against a variety of alternatives in finite samples. The simulations were performed for two different sample sizes, 30 observations and 300 observations. In both cases, the y's were generated by

$$y = x + cf(x) + u,$$

where x is a one-dimensional explanatory variable, distributed uniform on [-1,1], $f(x)$ is the misspecification, and u is chosen randomly from a standard normal. Then y is regressed on x and a constant, and the specification tests are performed on the residuals. In other words, $y = x + u$ is the null hypothesis, and $y = x + cf(x) + u$ is the truth. We will choose various different $f(\cdot)$ functions for which we will try to test. For 30 observations, $c = 2.0$, and for 300 observations, $c = 0.2$. These constants were merely chosen to obtain a wide range of performance of the tests at both 30 and 300 observations.

Along the top of Tables 1.3 and 1.4 are the various test statistics. The first statistic, NN, is a nearest neighbor statistic. Technically, the nearest neighbor is neither a bin or kernel test, but it is closely related to a bin test with a small bin. (The weight matrix of the nearest neighbor test has $\frac{1}{3}$ on the diagonal, the super diagonal, and the subdiagonal. On the other hand, the weight matrix of the bin test is block diagonal.) The second, third, and, in the case of 300 observations, fourth, tests are bin tests with increasingly larger bin sizes. The test labled Bin 5 has five observations in each bin, for example. The next three tests, KernS, KernM, and KernL, are kernel tests with small, medium, and large, window widths, respectively. Note that for 30 observations, the small window contains an average of three observations and the large window contains an average of ten observations. For 500 observations, the small window contains an average of six observations and the large window contains an average of 18 observations. Finally, the test labled Opt is the optimal Lagrange multiplier test for that particular alternative. The optimal test in each case was chosen with the knowledge of the particular form of misspecification; therefore, the nonparametric tests cannot be expected to perform as well. The optimal test will just serve as a benchmark for comparison of the nonparametric tests.

Along the left of both tables are the alternatives, or types of misspecification. The first six are Legendre polynomials two through seven. The Legendre polynomials are orthogonal on [-1,1], so an optimal test for one of the Legendre polynomials will only have power against that particular misspecification. Also included are the functions $\sin(2\pi x)$, and $\sin(10\pi x)$.

The numbers in the body of the tables are the proportions of simulated tests which reject the null at a 5 per cent level given the various forms of misspecification, or alternatives, on the left. In the case of 30 observations, 500 simulated tests were computed, but in the case of 300 observations, only 250 were computed. Note that in all cases the optimal Lagrange multiplier tests for the misspecification have the greatest power, as expected. For 30 observations, the bin tests seem to be significantly less powerful than the kernel tests. As expected, the performance of tests using different bin or window widths varies with the alternative—no bin or window width dominates over all alternatives. The Bin5 and KernM seem to do fairly well against a variety of alternatives. None of the tests does well against the particularly "wavy" alternative, $\sin(10\pi x)$, though. The window widths are not small enough, even NN, Bin5, and KernS. Such a result is not surprising. It would be difficult to find any specification test other than the optimal one that would reject in that case with such a small sample size.

Recall that for 300 observations, the size of the deviation from the null is much smaller than for 30 observations, which explains the lower rejection rates. These tests on 300 observations will obviously have much higher power than the test on 30 observations against the same size alternative. Note that in this table the bin tests perform very favorably compared with the kernel tests. In many cases the comparable bin test is only slightly less powerful than the kernel test. Again, all test have difficulty with $\sin(10\pi x)$.

The simulations seem to indicate that in small samples the kernel test outperforms the bin test. As the sample increases (and, incidently, the kernel test becomes much more difficult to compute), though, the bin test's power relative to the kernel test improves. Both tests fare reasonably well relative to the optimal tests.

1.5 Application: Gasoline Demand

As an illustration of how our test works, we present an application of it to survey data on household gasoline expenditures. We will look at gasoline demand as a function of price and income. The data sets we use are the U.S. Department of Energy's Residential Energy Consumption Survey from 1979-81 and its Residential Transportation Energy Consumption Survey from 1983, 1985, and 1988. We use them as a cross section with indicator variables for the years, yielding 18113 observations. The data are treated as in Hausman and Newey (1992), including the nonparametric "partialling out" of the demographic variables. We are thus left with a data set with price and income as the explanatory variables and gasoline demand as the dependent variable, all with the effect of the demographic variables taken out.

We would like to test parametric models of gasoline demand as explained by price and income using our nonparametric specification test. First, however, we will show two figures which lead us to believe that parametric specifications might not be sufficient for these data. Figure 1.6 is a kernel regression estimate of gasoline demand on price and income. The relationship does not appear to be regular or parametric in nature: the surface has humps and troughs that do not appear in all cross sections, either in the price direction or the income direction. Also, the cross sections are not an easily recognizable parametric function. For these reasons we suspect that a simple parametric specification might not be appropriate. Figure 1.7 illustrates the same point. It shows a cross section of gasoline demand as a function of price for the 50^{th} percentile of income. The middle line is a simple full log specification , the upper line is full log specification with quadratic terms and cross terms, and the lower line is the cross section from the nonparametric surface estimation.

Tables 1.5 and 1.6 show that our suspicions are confirmed by a nonparametric bin specification test. Using two different bin sizes we test three parametric specifications. Tables 1.5 and 1.6 report the test statistics and approximate p values for each test. The approximate p values were read from the theoretically predicted null distribution of the test statistic, the standard normal. This distribution should be a very good approximation due to the large sample size. First, the full log specification is rejected soundly by the larger bin test and very soundly by the smaller bin test. The large bin test would reject the quadratic specification at the 1% level and the cubic at the 2% level. The small bin test, however, rejects all three specifications with quite small p values. These results are not surprising. We expected to find strong evidence against the parametric specifications, which we did, and we would expect to find stronger evidence with the smaller bin test because it would be more able to pick up movements of the conditional mean function in small areas and not "smooth" them over. Note that with 17,000 observations, we are in no danger here of having too few bins (360 with the larger, 720 with the smaller) or too few observations per bin (causing too large a trace for the weight matrix) with either bin size. With the luxury of so many observations, we may then just choose bin size based on the type of misspecification for which we are looking.

For this particular application the bin test was really the only feasible test. The kernel test would have required larger computing facilities than were available to us. As noted earlier, though, the bin test performs very well for large sample sizes. And, of course, the computational savings over what would have been required to perform the kernel test were quite significant.

1.6 Conclusion

This chapter presents nonparametric specification tests based on the residuals from the null model, specifically, on a quadratic form in the residuals. The test statistics are asymptotically normal at a rate which depends on properties of the matrix in the quadratic form. The tests are consistent against all alternatives that shrink sufficiently slowly. Simulations performed on both the finite-sample distribution and power against various alternatives of the test statistics are encouraging. Finally, as exhibited by the application results, we are confident that these tests have the potential to be valuable tools in applied econometrics.

In this case rejection of a parametric null can have implications for estimation of, say, consumer surplus and deadweight loss, and eventually for policy questions such as the welfare loss under a gasoline tax increase. This point is made forcefully in Hausman and Newey, where estimates of these quantities based on both parametric and nonparametric models are computed. They are quite different. Our test points to the inadequacy of basing these estimates on the parametric models tested and suggests that a nonparametric modeling of the data might be more appropriate.

References

Azzalini, A., A. Bowman, W. Härdle (1989). On the Use of Nonparametric Regression for Model Checking. *Biometrika*, **76**, 1-11.

Bierens, H. (1982). Consistent Model Specification Tests. *Journal of Econometrics*, **20**, 105-134.

—- (1990). A Consistent Conditional Moment Test of Functional Form. *Econometrica*, **58**, 1443-1458.

Davidson, R., J. MacKinnon (1981). Several Tests for Model Specification in the Presence of Alternative Hypotheses. *Econometrica*, **49**, 781-793.

Eubank, R., C. Spiegelman (1990). Testing the Goodness of Fit of a Linear Model Via Nonparametric Regression Tecniques. *Journal of the American Statistical Association*, **85**, 387-392.

Ellison, G., S. F. Ellison (1992). A Simple Framework for Nonparametric Specification Testing. Draft.

Härdle, W. (1990). *Applied Nonparametric Regression.* Cambridge University Press.

Hausman, J. (1978). Specification Tests in Econometrics. *Econometrica*, **46**, 1251-1271.

Hausman, J., W. Newey (1992). "Nonparametric Estimation of Exact Consumers Surplus and Deadweight Loss." Draft, MIT Department of Economics.

Newey, W. (1985). Maximum Likelihood Specification Testing and Conditional Moment Tests. *Econometrica*, **53**, 1047-1070.

Serfling (1971). *Approximation Theorems of Mathematical Statistics.* John Wiley and Sons.

Wooldridge (1989). "Some Results on Specification Testing Against Nonparametric Alternatives." Draft, Michigan State University Department of Economics.

—- (1990). "A Test for Functional Form Against Nonparametric Alternatives." Draft, Michigan State University Department of Economics.

Figure 1.1

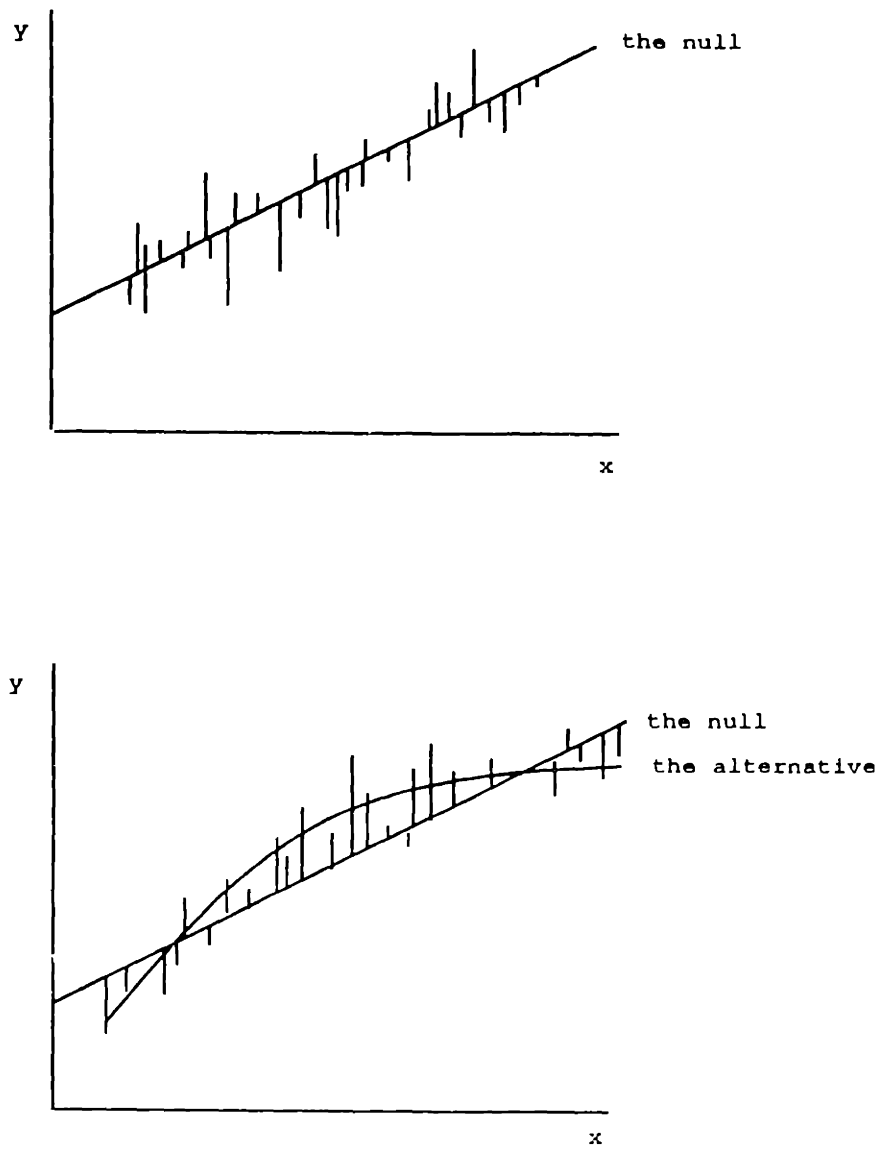
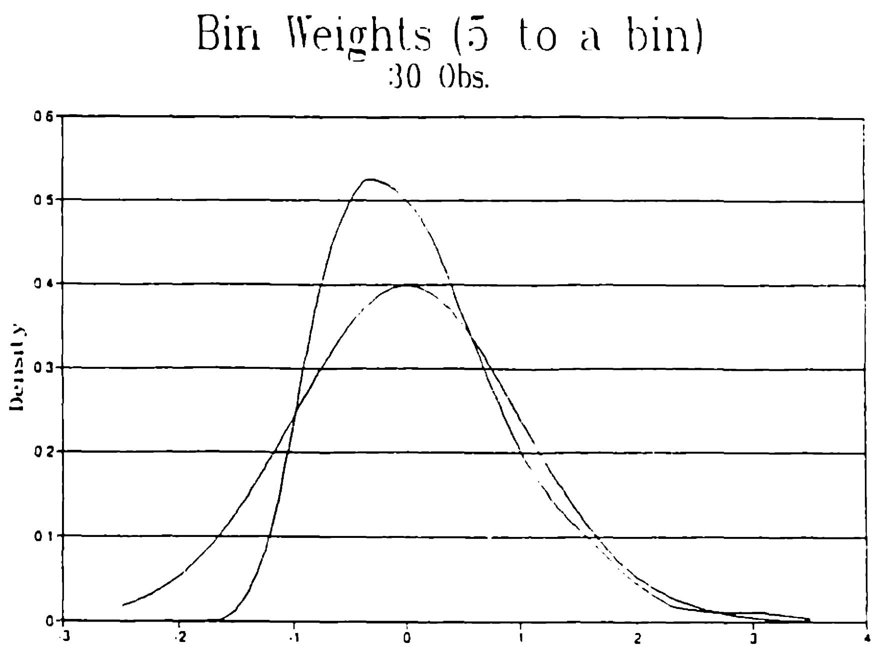


Figure 1.2


Bin Weights (5 to a bin)
30 Obs.
Density
0.6
0.5
0.4
0.3
0.2
0.1
0
-3
-2
-1
0
1
2
3
4

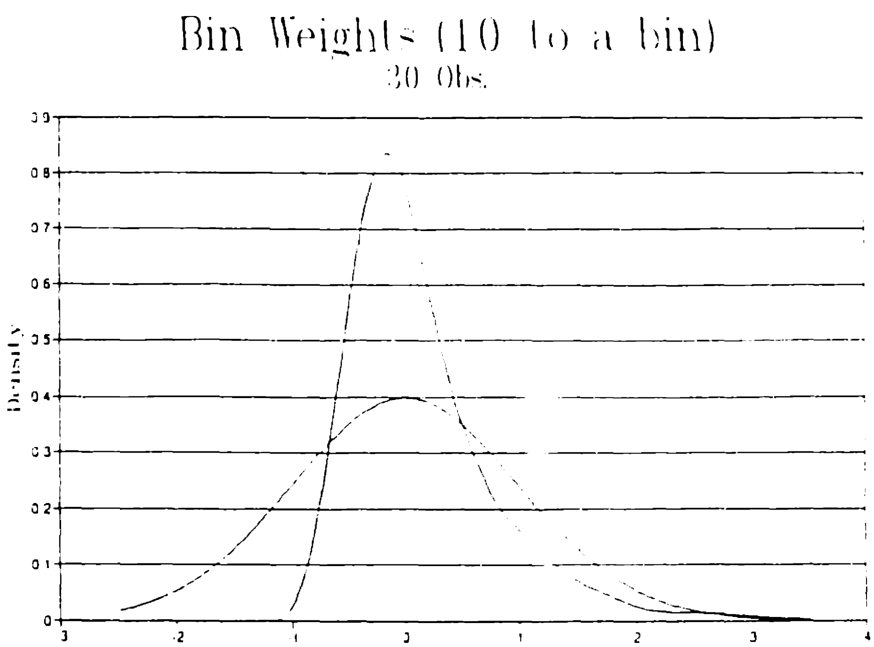
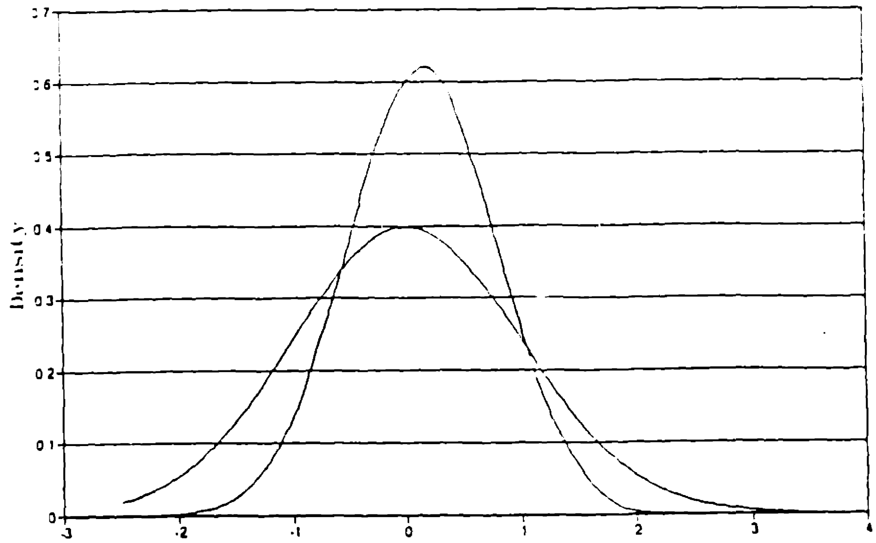
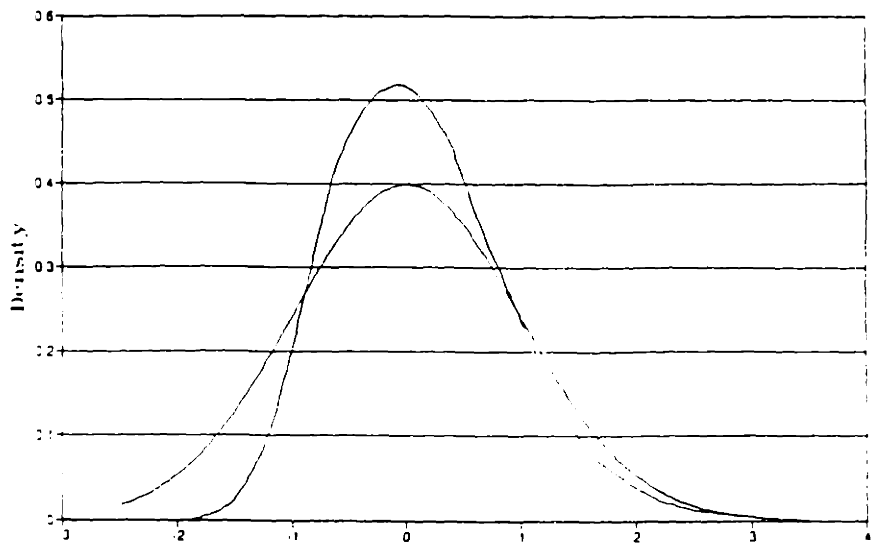

Bin Weights (10 to a bin)
30 Obs.
Density
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-3
-2
-1
0
1
2
3
4

Figure 1.3

Kernel Weights (Very Small)
30 Obs.



Kernel Weights (Small)
30 Obs.



Kernel Weights (Large)
30 Obs.

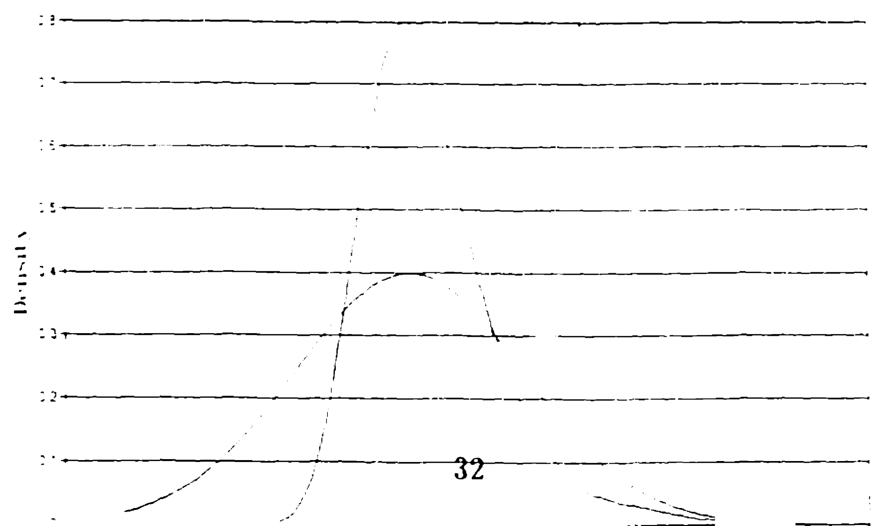
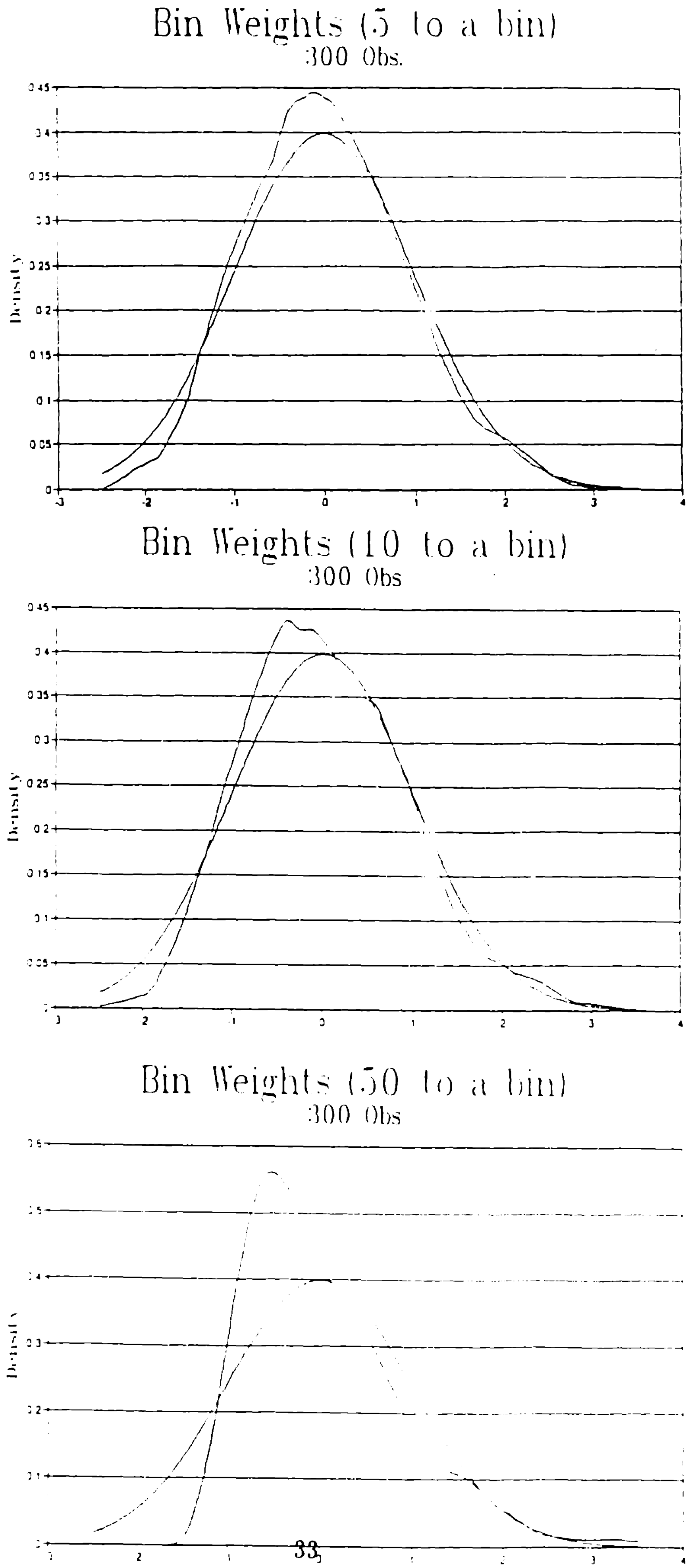
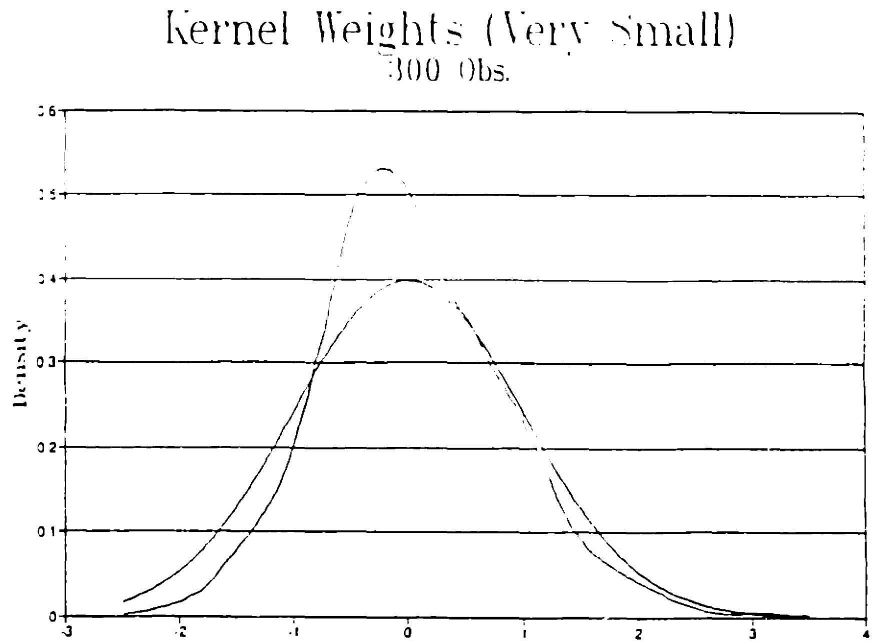


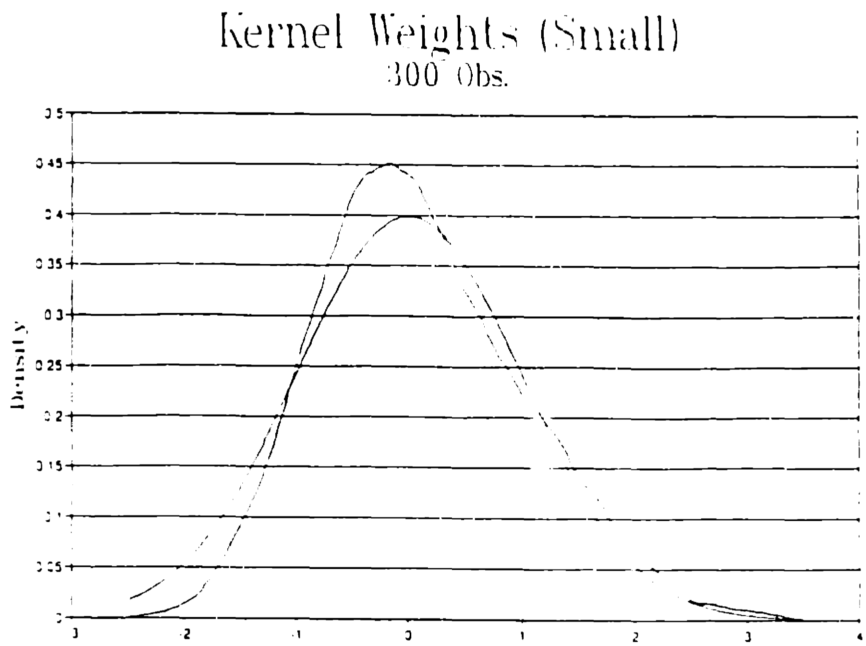
Figure 1.4



Bin Weights (5 to a bin)
300 Obs.
Density
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-3
-2
-1
0
1
2
3
4
Bin Weights (10 to a bin)
300 Obs
Density
Bin Weights (50 to a bin)
300 Obs
Density

Figure 1.5


Kernel Weights (Very Small)
300 Obs.
Density


Kernel Weights (Small)
300 Obs.
Density

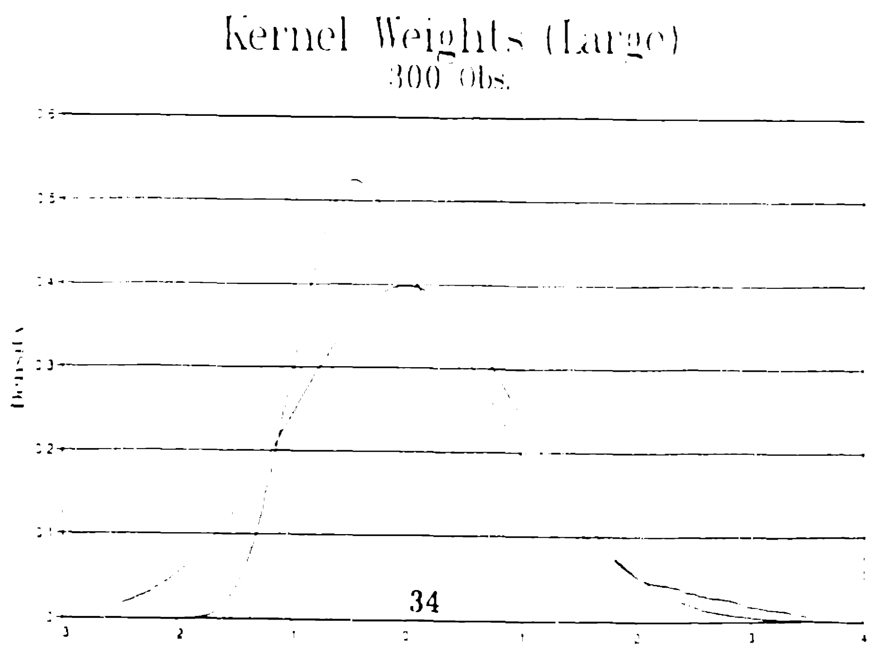

Kernel Weights (Large)
300 Obs.

Figure 1.6

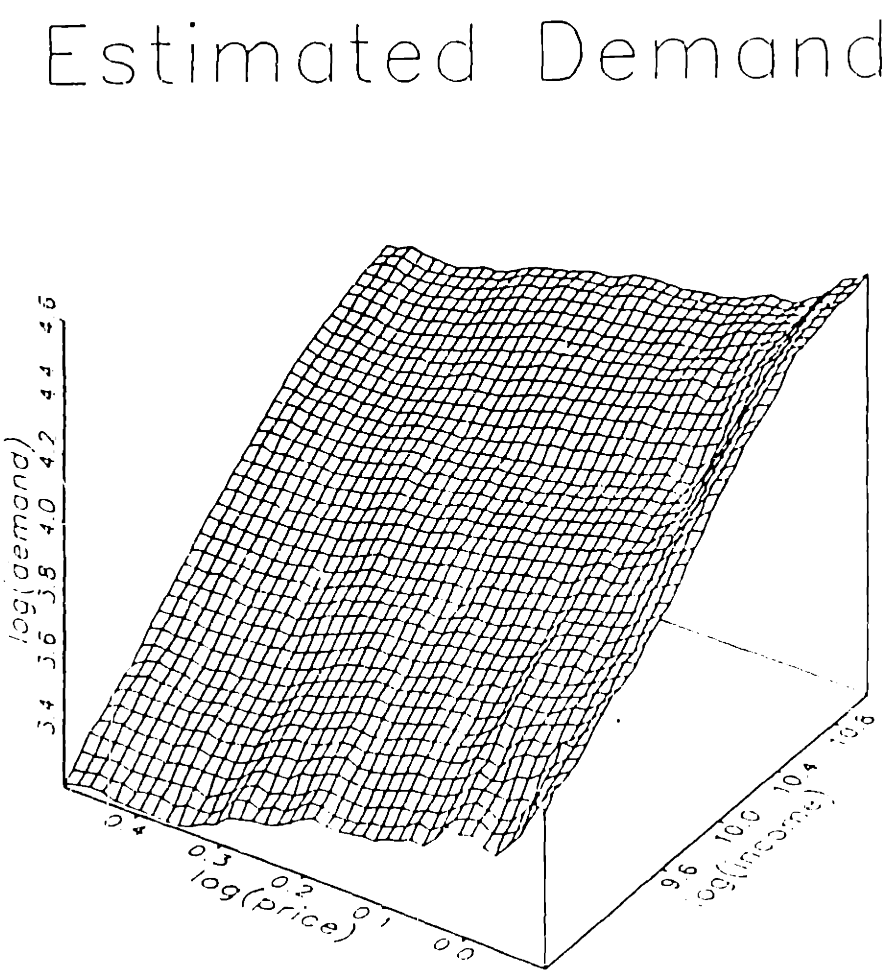

Estimated Demand
log(demand)
3.4
3.6
3.8
4.0
4.2
4.4
4.6
log(price)
0.4
0.3
0.2
0.1
0.0
log(income)
9.6
10.0
10.4
10.8

Figure 1.7

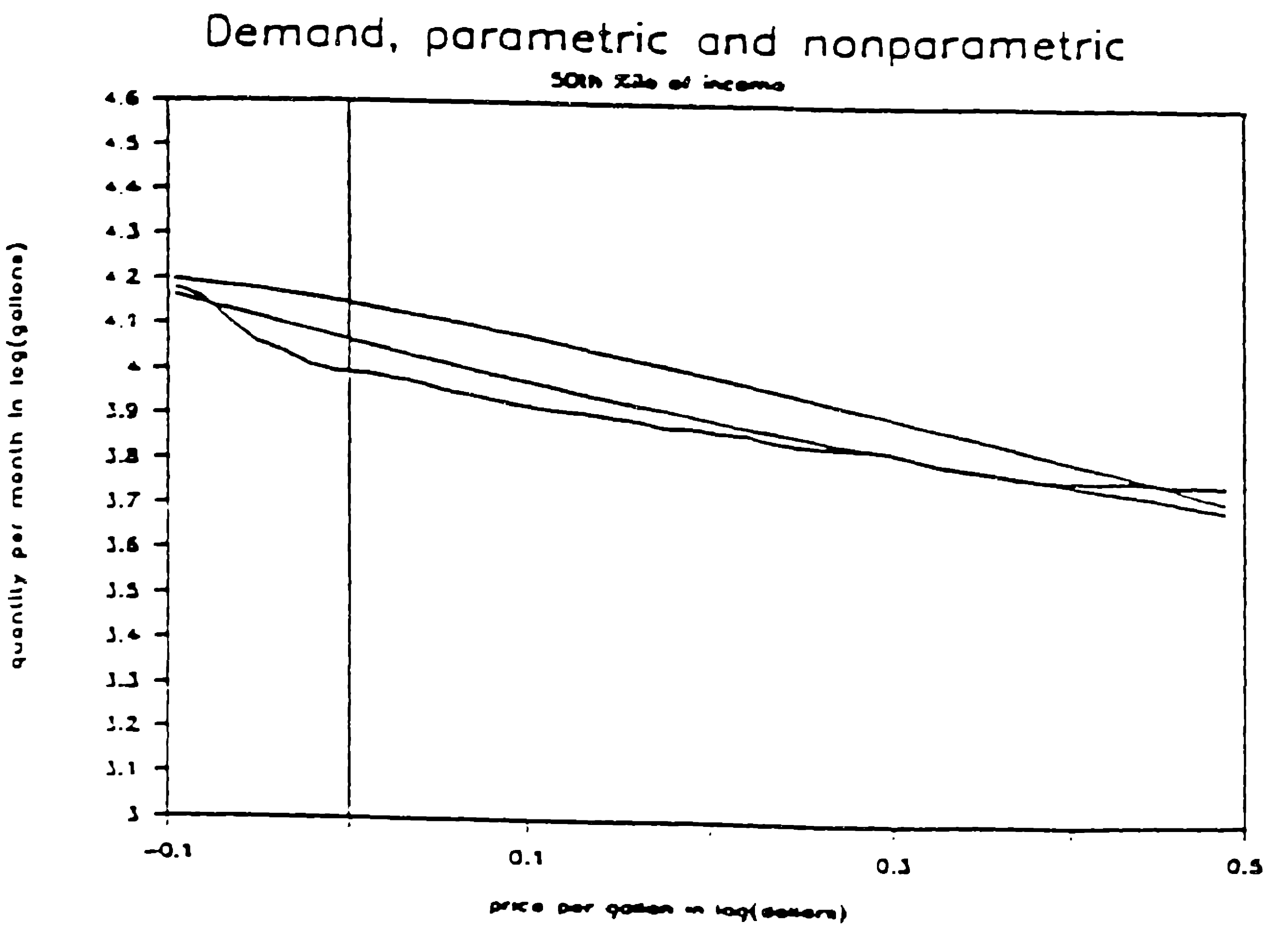

Demand, parametric and nonparametric
50th %ile of income
4.6
4.5
4.4
4.3
4.2
4.1
4
3.9
3.8
3.7
3.6
3.5
3.4
3.3
3.2
3.1
3
-0.1
0.1
0.3
0.5
price per gallon in log(dollars)
quantity per month in log(gallons)

Table 1.1: Characteristics on Null Distribution of Test Statistics, 30 observations

Test	Mean	Std Dev	5% Critical
Bin5	.1100	.7877	1.5758
Bin10	.1371	.6294	1.4006
KernVS	.1329	.5962	1.1328
KernS	.1213	.7401	1.4648
KernL	.1821	.5443	1.1681

Table 1.2: Characteristics on Null Distribution of Test Statistics, 300 observations

Test	Mean	Std Dev	5% Critical
Bin5	.0172	.8776	1.5731
Bin10	.0350	.8924	1.6174
Bin50	.0756	.8968	1.7199
KernVS	.0442	.8065	1.4621
KernS	.1068	.9053	1.7252
KernL	.0877	.8790	1.7664

Table 1.3: Power against Various Alternatives, 30 observations

Alternative	NN	Bin5	Bin10	Opt	KernS	KernM	KerL
2nd Legendre	.428	.354	.392	.972	.064	.692	.872
3rd Legendre	.358	.300	.132	.944	.044	.574	.348
4th Legendre	.374	.308	.106	.948	.030	.566	.274
5th Legendre	.232	.180	.130	.818	.026	.238	.028
6th Legendre	.258	.206	.006	.938	.036	.458	.032
7th Legendre	.222	.174	.124	.922	.036	.360	.008
$\sin(2\pi x)$	.260	.202	.374	.890	.030	.304	.080
$\sin(10\pi x)$	.048	.030	.076	.880	.018	.054	.068

Table 1.4: Power against Various Alternatives, 300 observations

Alternative	NN	Bin5	Bin10	Bin50	Opt	KernS	KernM	KerL
2nd Legendre	.120	.044	.172	.312	.596	.084	.224	.332
3rd Legendre	.076	.036	.204	.240	.696	.088	.212	.272
4th Legendre	.064	.032	.168	.148	.584	.088	.148	.208
5th Legendre	.132	.064	.152	.120	.612	.084	.152	.172
6th Legendre	.084	.052	.128	.132	.632	.100	.160	.192
7th Legendre	.136	.068	.136	.144	.644	.128	.168	.168
$\sin(2\pi x)$	.104	.064	.172	.244	.652	.092	.196	.240
$\sin(10\pi x)$	.128	.084	.148	.068	.712	.128	.152	.036

Table 1.5: Regression Diagnostic: Gasoline Demand (avg 50 obs/bin)

	log(price), log(income)	quadratic terms	cubic terms
statistic	3.243	2.234	2.069
app. pvalue	0.0005	0.01	0.02

Table 1.6: Regression Diagnostic: Gasoline Demand (avg 25 obs/bin)

	log(price), log(income)	quadratic terms	cubic terms
statistic	6.008	5.271	5.099
app. pvalue	very small	10^{-8}	10^{-8}

Chapter 2

An Application to Estimating Rank of a Demand System

2.1 Introduction

Economists have long been interested in studying individual consumption. Consumption is often quantified by means of an Engel curve, percent consumption of a particular commodity or group of commodities as a function of total expenditure. For example, an Engel curve would describe how the proportion of income spent on gasoline changed as income increased. We could then use the Engel curve to help predict the incidence of a gasoline tax increase. Obviously, the utility function, as a complete description of a consumers preferences for gasoline and other products, will determine the Engel curves for that consumer.

In addition to looking at Engel curves for a particular quantity in isolation, it is of interest to study demand systems, a system of Engel curves, one for each item or category of expenditure. Economists have sought to answer questions about the relationships among the equations in such a system. Certainly, patterns of consumption of some items might tell us about consumption of other items under some circumstances. One such relationship among Engel curves, rank, is defined to be the rank of the function space spanned by the Engel curves. For instance, a demand system with rank one would have Engel curves that were all multiples of one another. One application of the rank concept is in a demand system for risky assets. We might be interested in knowing how many different mutual funds would have to be available before all consumers could maximize utility by just purchasing shares in the funds. We could answer such a question, in principle, by observing which assets consumers do buy and seeing whether they buy them or groups of them in fixed ratios. (The case where consumers could maximize utility with two mutual funds available to them is known in the finance literature as two-fund separation and in the demand literature as a rank two system.) In addition to providing information about consumer behavior, such a finding would imply a concave utility function (and, incidentally, some things about the distribution of returns on the assets), an interesting economic result. In addition, macroeconomists in particular have been interested in the rank of demand systems for its implications about aggregability.

Gorman (1981) showed that the rank of a demand system will be at most three, assuming that the polymonial functions in expenditure do not contain price. Hausman, Newey, and Powell (1988) (HNP) and Lewbel (1991) have empirically studied the rank of demand systems to test Gorman's theory and implications of it. Lewbel constructs a direct test of the rank of a matrix which has the same rank as the demand system (a matrix of coefficients of the Engel curves) to examine Gorman's hypothesis. He performs a decomposition of the matrix and tests whether the pivots are equal to one. (Obviously, then, Lewbel's test, while applied to testing the rank of a demand system, is more general than that problem. It can be used as a test for the rank of any matrix.) He finds strong evidence for the rank of the demand system being at least three. HNP test for restrictions on the coefficients within Engel curves implied by the demand system having rank three. Again, strong evidence is found for Gorman's hypothesis.[1] Unlike those two papers, we test the rank of the demand system without having to make assumptions on the type of functions to appear in the Engel curves. In other words, we test the rank of the function space spanned by the Engel curves directly without first specifying and estimating the Engel curves and

[1] HNP correct also for the errors in variables problem caused by the fact that total expenditure is not observed and income is used instead. This issue should not present problems in our analysis because total expenditure will not actually be used as a regressor.

then testing the rank of the matrix of their coefficients. Although it is a test which exploits the specific structure of the problem and is therefore not of general use for testing the rank of matrices, it is a more general test than Lewbel in the sense that it will have power against linear independence of Engel curves no matter what their specification. This point will be made more clear in the next section.

2.2 Approach and Data

We now precisely define the problem. Suppose that Engel curves are specified as follows:

$$\begin{aligned} b_0 &= \beta_{00} + \beta_{01} f_1(\mathrm{i}) + \beta_{02} f_2(\mathrm{i}) + \ldots + \epsilon_0 \\ b_1 &= \beta_{10} + \beta_{11} f_1(\mathrm{i}) + \beta_{12} f_2(\mathrm{i}) + \ldots + \epsilon_1 \\ &\vdots \\ b_k &= \beta_{k0} + \beta_{k1} f_1(\mathrm{i}) + \beta_{k2} f_2(\mathrm{i}) + \ldots + \epsilon_k, \end{aligned}$$

where b_j are budget shares, and i is expenditure. Let B be the matrix of coefficients $\{\beta_{ij}\}$. Such a system of equations is known as a demand system. For example, let b_0 be the share of total expenditure on coffee. The first equation is then the Engel curve for coffee. The rank of this demand system, defined as the rank of the function space spanned by the Engel curves, will be exactly the rank of the matrix B.

In this section we will first discuss our approach to the problem of testing for the rank of such a demand system. We will then discuss in detail how the nonparametric specification test can be adapted for this problem. Finally, we discuss the data we will use to empirically examine this problem.

2.2.1 Description of the Testing Procedure

We will first discuss the general procedure in the context of testing for the rank of a demand system and then present a detailed example of the procedure with only two equations in order to help fix ideas.

The matrix B would have rank three if and only if there exist exactly three linearly independent vectors of Engel curve coefficients. We could, in principle, construct a sequential procedure involving testing for linear independence among groups of Engel curves to determine the rank of B, then. First, assuming rank three, say, we ought to be able to find two linearly independent Engel curves. Once found, we should then be able to find a third which is linearly independent of the group of two Engel curves. Finally, if the matrix actually has rank three, the rest of the Engel curves should be linear combinations of the group of three. Questions arise, however. For instance, we need a way to test whether a given Engel curve is linearly independent of a group of other Engel curves. We will use the fact that each budget share is equal to a fixed linear combination of coefficients of each Engel curve (ignoring for now the additive error). We can, therefore, test linear independence of the budget shares, not the Engel curves themselves. Such a test could be carried out by means of a specification test. A budget share b_{j_1} could be regressed on a group of budget shares, b_{j_2}, b_{j_3}, and b_{j_4}, and rejection of the linear specification would be tantamount to rejection of linear dependence between b_{j_1} and the group of three budget shares. We could, therefore, construct our sequential testing procedure by regressing budget shares on groups of other budget shares and testing the specification of the linear regressions.

The budget share equations are not deterministic, though. They do have additive errors. This situation creates an errors in variables problem because the dependent variables in the Engel curves, budget shares, will be used as regressors when testing for linear independence. Instead of each budget share being equal to the sum of their regressors times their coefficients, they are equal to the true Engel curve at that point plus a random disturbance. If they are to be used as regressors, then they are regressors "measured" with error. We can, of course, instrument for the regressors, using polynomials in expenditure or perhaps other variables as instruments. We will discuss the problem in more detail in the next subsection.

We will now step back for a moment from the application at hand, demand systems, and present a simple two-equation example. We hope to clarify how exactly the problem can be reduced to one of testing only budget shares and how one would perform this test.

We have two equations for which we would like to test linear indpendence of the parameter vectors, β_1 and β_2.

$$y_1 = f(x)^T\beta_1 + \epsilon_1$$

$$y_2 = f(x)^T\beta_2 + \epsilon_2$$

For each observation x_i is a scalar, but $f(x_i)$ is a k-vector-valued function, where k is the size of β_1 and β_2. Think of β_1 and β_2 as vectors, $f(x)$ as a set of coefficients, one set for each observation, and y_1 and y_2 as the value of the linear combinations of β_1 and β_2 with $f(x)$ as the coefficients, up to the additive errors ϵ_1 and ϵ_2. Now suppose that $f(x)$ is a $k \times n$ matrix with rank k. Then β_1 and β_2 being linearly independent k-vectors implies that y_1 and y_2 are linearly independent n-vectors. Therefore, there is no need to estimate β_1 and β_2 or even specify $f(x)$ to test linear independence of the vectors β_1 and β_2. We just have to know that $f(x)$ has rank k. For each observation, $f(x_i)$ has k elements. We need at least k observations for $f(x)$ to have rank k. (In other words $f(x)$ must be at least $k \times k$.) It is true that for k linearly independent functions $f(x)$, there exist k values x such that the matrix of values of $f(x)$ does not have full rank. The sets of these values x have measure zero, though. In any case, such matters should be of no concern because the number of sets of k coefficients, the number of observations, is usually much larger than the number of elements in the true β_1 and β_2, *i.e.*, n is usually much larger than k.

We have now reduced the problem of testing for linear independence of β_1 and β_2 to the problem of testing for linear independence of y_1 and y_2, if not for the additive errors. We have avoided having to specify the Engel curves and estimate the coefficients. Our biggest saving, though, is in the robustness of the test to the specification of the Engel curves. The power of this test is not limited by the specification of the Engel curves since the specification is never made explicit.

We now need a statistical procedure to test for linear independence of y_1 and y_2. One such procedure would be to regress one of y_1 and y_2 on the other and test whether that linear specification is correct. In other words, if we can reject the linear specification

$$y_2 = \alpha y_1 + \eta,$$

then we have rejected that y_2 is linearly dependent on y_1 and therefore that β_2 is linearly dependent on β_1.

2.2.2 Adapting the Test Statistic to the Case of Errors in Variables

As mentioned before, the additive errors in the equations cause an errors in variables problem when one dependent variable of a system of equations is regressed on another. This subsection will quickly review the test statistic and why it would be appropriate for such a problem, will discuss the problem of adapting the test to the case of errors in variables, and also will discuss the choice of instruments necessary to adapt this test to the case of errors in variables.

We will now be discussing the procedure in the context of the demand system. The procedure we will be using to test linear independence of sets of coefficient vectors is the nonparametric specification test of the first chapter. Recall that the statistic is based on the quadratic form $\tilde{u}^T W \tilde{u}$, where W is a weight matrix and $\tilde{u}$ is the vector of residuals from the null specification. In our case, the null specification is linear, of course. In other words, the null hypothesis is that b_1 is linearly dependent on b_0. If the statistic based on $\tilde{u}^T W \tilde{u}$ is larger than expected under the null, then we can reject linear dependence of b_0 and b_1 and, therefore, of β_0 and β_1.

From now on, we will use the notational convention of the previous chapter: X is the matrix of regressors and y is the dependent variable, for any regression whose specification we are testing. Bear in mind that some variables, the budget shares b_i, might be used as both dependent variables and regressors.

The theory of our specification tests was worked out in the case of OLS. Applying the tests discussed in this paper to the case of errors in variables is inelegant at best. The residuals from performing an instrumental variables regression in this case are not orthogonal to the regressors. As a consequence the nonparametric specification test statistics have extra terms that do not go away asymptotically. Recall that in the case of OLS,

$$\tilde{u}^T W \tilde{u} = u^T W u - u^T P_X W P_X u,$$

where u are the true errors, $\tilde{u}$ are the residuals, and P_X is the projection matrix onto X space. We see that in the case of errors in variables,

$$u = y - X\beta - \eta\beta,$$

where X^*, the true value for the regressor(s), is equal to $X + \eta$, the observed value plus an error. Then,

$$\begin{aligned} \tilde{u} &= (I - P_Z)((X + \eta)\beta + u) \\ &= \eta\beta + u - P_Z\eta\beta - P_Z u, \end{aligned}$$

where Z is the matrix of instruments and $P_Z = X(X^T X)^{-1} Z^T$. To obtain an expression in terms of the true errors of $\tilde{u}^T W \tilde{u}$, we then use the fact that the weight matrix W is symmetric (or can be symmetrized) and that P_Z is symmetric and idempotent. Then

$$\begin{aligned} \tilde{u}^T W \tilde{u} &= u^T W u - u^T P_A W P_A u + (\eta\beta)^T W (\eta\beta) \\ &\quad -(\eta\beta)^T P_A W P_A (\eta\beta) + 2u^T W (\eta\beta) - 2u^T P_A W P_A (\eta\beta), \end{aligned}$$

P_A behaves like a projection matrix, so the terms with $P_A W P_A$ in the middle will go away asymptotically, but might require a finite sample correction, like before. Since we will be simulating or bootstapping our critical values instead of relying on theoretical predictions in this empirical study, any such finite sample terms will not affect our results. There are now, however, two terms in addition to $u^T W u$ that do not go away asymptotically. We have two options for carrying out the test. First, note that $\tilde{u}^T W \tilde{u} = (u + \eta\beta)^T W (u + \eta\beta)$ asymptotically. One can then construct a test like before by merely replacing $\hat{\sigma}^2$, the estimated variance of u, with $\hat{\sigma}^2_{u+\eta\beta}$, the estimated variance of $u + \eta\beta$.

$$\mathcal{T}_1 = \frac{\tilde{u}^T W \tilde{u} - \hat{\sigma}^2_{u+\eta\beta} \mathrm{tr} W}{\hat{\sigma}^2_{u+\eta\beta} (2\mathrm{tr} W^2)^{1/2}}$$

This test, however, is now a *joint* test of the following hypotheses: correct specification of the model, zero covariance between u and η, and correct specification of the equation where the instruments explain the regressors. Usually one would not want such a joint test. One could, though, test the two "nuisance" hypotheses separately and reject correct specification of the model if the joint test rejected and the separate tests did not. Critical values for the joint test might be problematic.

The second option is to subtract off estimates of the two unwanted terms from the quadratic form. This method is the one we primarily used in order to avoid problems involved in performing the joint test of several hypotheses. Estimates of the two unwanted terms were constructed as follows. First, the regressors, X for now, in the regression of a budget share y on other budget shares were regressed on the instruments Z to obtain coefficient estimates $\hat{\alpha}$. Then the residuals were computed in the following way: $\hat{\eta} = X - Z\hat{\alpha}$. An estimate of β was calculated by regressing y on X in two stages using Z as instruments. Finally, the "clean" residuals were calculated as follows: $\tilde{u}_c = y - X\hat{\beta} + \hat{\eta}\hat{\beta}$. The quadratic form was then constructed with the $\tilde{u}_c$'s as follows:

$$\mathcal{T}_2 = \frac{\tilde{u}_c^T W \tilde{u}_c - \hat{\sigma}^2 \mathrm{tr} W}{\hat{\sigma}^2 (2\mathrm{tr} W^2)^{1/2}}$$

The choice of instruments was, in some sense, very natural. Given that we believe that each budget share can be expressed as some function of total expenditure—*ample* empirical evidence bears out the fact that total expenditure has explanatory power in that setting—polynomials in total expenditure seemed the natural choice. Obviously, we should only use those polynomials and just as many polynomials as have predictive power for budget share. Only polynomials which have predictive power for budget share will be correlated with the regressors. In this choice we were guided by results of our crossvalidated series regressions of budget shares on polynomials in expenditure which are discussed in Section 2.3.1. In addition, to be valid instruments, they must be uncorrelated with the "error" in the regressors and with the equation error. Recall, however, that since actual expenditure is not observed and income is used instead, the instruments are also measured with error. Our instruments, then, actually have two components, expenditure and the difference between income and expenditure, which we will call instrument error. We need, therefore, to be concerned with four possible correlations: both expenditure and the instrument error with both the equation and the regressor error. Expenditure should not be correlated with either the equation error or the regressor error, essentially by definition. The instrument error, however, is somewhat more problematic. In order for the instrument error to be uncorrelated with the regressor error, we would need to make the assumption that saving is not correlated with excess expenditure in any category (relative to total expenditure), *i.e.*, that savings is separable from other expenditure. We will make this assumption by necessity.

As mentioned before, the general testing procedure, in addition to the particular specification test used, is a nonparametric procedure because it does not require specification of the Engel curves. Of course, the particular test we are using is also nonparametric, in the sense that it does not look for departures from linear in any particular direction in the dependence of the budget shares. Therefore, there are, in fact, two ways in which other testing procedures impose stucture that this procedure does not.

2.2.3 Description of Data

The data that we will be using are from the 1982 Bureau of Labor Statistics Consumer Expenditure Survey. We have data on 1324 individuals and their consumption habits. In particular, we have their total dollar expenditures in each of five categories, food, clothing, recreation, health care, and transportation, and their income. Assuming that total expenditure is equal to income, as is customarily done, we can calculate expenditures in a sixth category: other. Table 2.1 contains the summary statistics, minima, maxima, means, and standard deviations, of the data set.

Lewbel uses the same source of data. He, however, treats the data differently. He estimates both a fixed effects model, assuming all consumers have the same preferences, and a random effects model, assuming taste parameters have some distribution. For the random effects model, it is necessary to exclude consumers who seem likely to have income correlated demographic variation. He, therefore, used the following selection criteria: married couples with the age of the head of household and the nomial annual income in certain ranges. Here we will assume fixed effects. We will, therefore, not need to use just a selected portion of the data.

2.3 Results and Comments

We present two types of results here. The first type of results give us information on the performance of the nonparametric specification test in the presence of errors in variables. They are results about the distribution of the test statistics under the null hypothesis that budget shares are linearly dependent. These distributions, both simulated and bootstrapped, can be compared with the theoretical null distribution of the test statistics, $\mathcal{N}(0,1)$. We will, of course, use these simulated and bootstrapped distributions to find critical values for rejection of the null.

The second type of results are, of course, the test statistics from the tests for the rank of the demand system. After examining those test results, we should be able to draw some conclusions about the rank of our demand system and consequently other interesting economic questions such as aggregability.

But first we will discuss in detail the simulation and bootstrapping techniques used to obtain the null distributions of the test statistics.

2.3.1 Simulating and Bootstrapping the Null Distribution

There are different possible methods for simulating and bootstrapping these distributions. To explain them, we first present the null model for testing for linear independence of one budget share from three others, say. Suppose that the fourth budget share b_3 is being tested for independence from the first three, b_0, b_1, and b_2. Then

$y = b_3$ and $X = [b_0\ b_1\ b_2]$. Also, our instrument matrix Z will be comprised of polynomials in total expenditure, $f(i)$. First we have

$$x_i = x_i^* + \eta_i, \qquad i = 1, 2, 3,$$

where x_i^* is the "true" budget share and x_i is the "observed" budget share, measured with error. Recall that each budget share is the dependent variable in its own Engel curve, so

$$x_i = \alpha_i Z + \eta_i, \qquad i = 1, 2, 3,$$

$$\text{or} \quad x_i^* = \alpha_i Z,$$

where Z are the regressors in the Engel curve (and also our instruments). The null model is then

$$y = \beta_0 + \beta_1 x_1^* + \beta_2 x_2^* + \beta_3 x_3^* + \epsilon$$

$$= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 - \beta_1 \eta_1 - \beta_2 \eta_2 - \beta_3 \eta_3 + \epsilon.$$

We now explain the method for simulating the null distribution that this paper employed. To simulate values of X and y, we first regress x_i on Z nonparametrically to obtain $\hat{\alpha}_i$.[2] Then we regress y on X using instrumental variables estimation to obtain $\hat{\beta}$. Next we draw values for η and ϵ from an assumed distribution. Finally to obtain simulated values of X and y we exploit the following formulae:

$$x_i^S = \hat{\alpha}_i Z + \eta_i^D \qquad i = 1, 2, 3$$

$$y^S = \hat{\beta}(\hat{\alpha} Z) + \epsilon^D,$$

where the D superscript denotes a draw from the assumed distribution of the corresponding random variable and the S superscript denotes the simulated variable.

Once we have y^S we can then perform the test as before—run regressions, calculate

[2] We use cross-validated series estimation. Using just a simple parametric procedure would not ensure that the residuals were orthogonal to the regressors because we do not claim correct specification of the Engel curves, thus invalidating our simulation and bootstrapping.

residuals, form the test statistic—as many times as we want in order to trace out the distribution of this test statistic under the null.

Finally, we will mention from where the actual simulations and bootstrap replications were drawn. For simulation we drew η_i^D and ϵ^D from a joint normal distribution with the means being zero and the variance-covariance matrix being the sample variance-covariance matrix of the residuals. For bootstrapping we drew η_i^D from the residuals from the regression of x_i on Z, $\hat{\eta}_i$. We drew ϵ^D from the pseudo-residuals $y - x^T\hat{\beta} + \hat{\eta}\hat{\beta}$. Analogous to transforming the simulated errors with the estimated variance-covariance matrix, we drew residuals from a whole observation for each bootstrap replication, not each residual separately. In that way we preserve the covariance structure.

2.3.2 Results on the Null Distribution

The results on null distribution of the test statistics, Table 2.2, were encouraging. Recall that applying the nonparametric specification test in this nonstandard setting, in the presence of errors in variables, produces several nuisance terms that will go away asymptotically but will effect the statistic in finite samples. No attempt was made to correct for these terms since the critical values were to be bootstrapped. These nuisance terms surely played some role in the way the bootstrapped null distributions deviated from the theoretically predicted null distribution. Still, in most cases the theoretical distribution was a fair approximation of the bootstrapped and simulated distributions. (Note that these tests were performed with over 1000 observations. Any conclusions drawn about the relatively small impact of the nuisance terms are drawn conditional on the sample size.) Some amount of nonsystematic deviation from the theoretical distribution can also be attributed to the small number of bootstrap replications and simulations performed, in all cases 500. The numbers were limited by computational considerations.

Note that in comparisons of null distributions between the bin and kernel tests, the kernel test distributions seem to be closer to the theoretical for the one dimensional test, but not for the two and three dimensional tests. (This pattern does not seem to be a coincidence. Although we only report a small number of the bootstrapped and simulated distributions here, we performed many others. The pattern seems to be borne out in them as well.) It could be that the kernel test does not perform as well in high dimensions due to problems of small amounts of data at the boundaries. The bin test might not suffer from the same problems because the bins are fixed in X space.

It should be pointed out that the critical values we have estimated here are estimated on the assumption that each of these specification tests will be performed alone. The true critical values for a sequential procedure such as ours would be lower than those for each test performed alone. There is also a problem of spurious rejections, though. For instance, if we performed a large number of one dimensional tests (a budget share regressed on one other) to find one that rejected, then the result might be spurious, and our critical values should be higher. The first problem should only make our test procedure more conservative, if it is not cancelled out by the second problem, so it should be of no concern. The second problem is potentially more troublesome, but certainly not for this application. First, very few combinations of budget shares were searched to obtain the results presented below. Second, as will be shown, the results were decisive enough to ease any concern.

2.3.3 Test Results

The results from the procedure for testing the rank of the demand system appear in Table 2.3. The test for the demand system being rank one rejected decisively. The values of the test statistics, 3.88 for the bin test and 13.48 for the kernel test, leave little doubt, despite the possible problem of spurious rejections mentioned earlier. (Assuming the theoretically predicted null distribution, the two tests would have p values of 0.00005 and much less than 10^{-7}, respectively.)

We found a third budget share to regress on the two from the first test to reject rank two of the demand system. In this case, the test statistics were 7.12 for the bin test and 9.77 for the kernel test, again a decisive result. (The p values for these two tests are both much less than 10^{-7}.)

Finally, a test was performed to see if rank three could also be rejected. Again it could, with the test statistics 6.34 and 4.38 for the bin and kernel tests. This final rejection is an interesting result. Both HNP and Lewbel found strong evidence in favor of rank three demand systems, although here we have found evidence for this demand system having rank at least four. [3]

One reason that our testing procedure might be rejecting rank three when other tests have been unable to do so is that our test will detect a lack of linear dependence in the Engel curve coefficient vectors regardless of Engel curve specification. In addition, our test should reject arbitrary deviations from linear dependence of the shares. There are certainly alternatives against which this test will have power and against which a given parametric test will have little or no power.

Our test, like all nonparametric tests, falls prey to the "curse of dimensionality," significantly lowering its power in higher dimensions. It is true, though, that as the space of alternatives becomes higher dimensional, conventional parametric tests "cover" less and less of that space, even though they retain high power against a fixed dimensional subspace of the alternative space; whereas, nonparametric tests will have greater coverage. Finally, we will mention an interesting property of this test arising from the fact that the Engel curves were not specified. It should not be viewed as unusual that this test might fail to reject linear dependence when b_{i_3} is regressed on b_{i_1} and when it is regressed on b_{i_2}, but then reject when b_{i_3} is regressed on both b_{i_1} and b_{i_2}. In fact this occurred in some of the tests we ran. The reason we might obtain such a result is that the relationship between b_{i_3} and b_{i_1} and the relationship between b_{i_3} and b_{i_2} might appear linear, but the relationship between b_{i_3} and b_{i_1}, b_{i_2} need not be planar.

[3] Aggregation of goods can cause the rank of the matrix of coefficients to be smaller than the actual rank of the demand system, as was pointed out in HNP and Lewbel. They, therefore, both state that they have found evidence for rank being greater than or equal to three. We have, however, used essentially the same data set and thus the same aggregation of goods as the previous two papers. Our finding of rank at least four, then, has nothing to do with a different aggregation of goods.

2.4 Conclusion

The results of this paper are interesting not only in their economic implications, but also in the fact that this nonparametric procedure of testing for the rank of the demand system yielded different results from previous empirical work, more parametric in nature. The fact that this nonparametric procedure has greater coverage of the space of possible alternatives may account for our rejection of rank three for the demand system.

References

Ellison, G., S. F. Ellison (1992). A Simple Framework for Nonparametric Specification Testing. Draft, Harvard Department of Economics.

Gorman, W. (1981). Some Engel Curves. *Essays in the Theory and Measurement of Consumer Behaviour in Honor of Sir Richard Stone.* Cambridge University Press.

Hall, P. (1986). On the Bootstrap and Confidence Intervals. *The Annals of Statistics*, **14**, 1431-1452.

Hausman, J., W. Newey, J. Powell (1988). Nonlinear Errors in Variables: Estimation of Some Engel Curves. MIT Working Paper.

Lewbel, A. (1991). The Rank of Demand Systems: Theory and Nonparametric Estimation. *Econometrica*, **59**, 711-730.

Stock, J. (1989). Nonparametric Policy Analysis. *Journal of the American Statistical Association*, **84**, 567-575.

Table 2.1: Descriptive Statistics of Budget Shares and Total Expenditures

Expenditure Category	minimum	maximum	mean	stand dev
food	0.00668	0.658	0.215	0.0920
clothing	0.000201	0.304	0.0554	0.0426
recreation	0.000293	0.842	0.206	0.139
health care	0.000283	0.673	0.0602	0.0730
transportation	0.000319	0.688	0.0539	0.0621
total expenditure	728	57064	5429	3526

Table 2.2: Comparison of Theoretical, Bootstrapped, and Simulated Null Distributions, bin/kernel tests

Budget Share Regression		Theoretical	Bootstrapped	Simulated
3 on 1	mean	0.00	-0.07/0.00	0.00/-0.01
	variance	1.00	0.85/0.98	0.89/0.96
	95% crit	1.64	1.31/1.64	1.48/1.57
2 on 1, 3	mean	0.00	0.02/-0.03	-0.01/-0.01
	variance	1.00	1.04/0.71	0.92/0.62
	95% crit	1.64	1.65/1.14	1.53/1.03
5 on 1, 2, 3	mean	0.00	-0.03/-0.04	0.02/-0.03
	variance	1.00	1.02/0.73	0.99/0.65
	95% crit	1.64	1.64/1.18	1.73/1.07

Table 2.3: Test Statistics and Bootstrapped Critical Values, bin/kernel tests

	statistic	critical value
test for rank 2 (6 on 1)	3.88/13.48	1.41/1.82
test for rank 3 (3 on 1, 6)	7.12/9.77	1.62/0.93
test for rank 4 (4 on 1, 3, 6)	6.34/4.38	1.79/1.28

Chapter 3

Window Width Choice

3.1 Introduction

The problem of providing theoretical guidance for window width selection in nonparametric testing has been largely swept under the rug in previous papers on the subject. In fact, most studies have relied solely on the established theory for window width selection in estimation despite its inappropriateness. In addition, while some studies have examined data-based methods for choosing window width, they have failed to present convincing theoretical or simulation results to justify their methods. This chapter is a preliminary attempt to investigate this issue. We will first present simulation results for various window widths to provide practical guidance for window width choice and also to provide intuition for the theoretical discussion to follow. Then we will argue that window width selection in testing and estimation contexts are different problems while examining methods of window width choice which other papers have employed. Finally, we will present a variety of preliminary ideas and conjectures along with simulation results on new methods for window width choice.

3.2 Simulations and Practical Guidance

To prevent confusion based on our perhaps nonstandard terminology, we now offer an explanation. "Smoothing parameter" refers to the most general class of roughly analogous parameters in nonparametric estimators. "Window width" describes the smoothing parameter in a narrower class, local smoothing estimators. This term will be used to describe the smoothing parameter in both the bin and the kernel tests. Finally, we might use even more specific terms, like bin width or bandwidth.

We first want to make clear the idea that different window widths work better with different alternatiaves. We will accomplish this goal by presenting two sets of simulations using the bin test statistic of Ellison and Ellison (1992), a slight variant of the statistic presented in the first chapter. We look at the power of the bin tests with various bin widths against different alternatives. In all cases, simulated critical values are used and 2000 simulated tests are performed. The explanatory variable is random, with the sample rechosen from $\mathcal{U}(-1,1)$ for each simulation.

The first set of power simulations simply employs a large variety of bin widths against a number of alternatives for a fixed sample size, 100 observations. See Figure 3.1. Obviously, the performance of the various bin widths is dependent on which alternative is used, the four alternatives being the second degree Legendre polynomial, the fourth degree Legendre polynomial, $\sin(10\pi x)$, and a line with one kink in it. The results are predictable. It is not at all surprising, for instance, that the small bin widths do much better than the large against a wavy alternative, $\sin(10\pi x)$, but lag behind the large for the fairly smooth alternatives. The tradeoff faced is that the larger bin widths "wash out" the wavy alternatives, whereas the small bin widths are unable to capture very much of the spatial correlation present in the fairly smooth alternatives. Figure 3.2 is a graph of the composite power of each different bin width. Here we define composite power as an average of percentage rejections against the four alternatives mentioned above, each at three different sizes. This is, of course, an arbitrary measure, but it might still be of some interest in assessing overall power.

For small sample sizes such as 100 observations, we saw that the tradeoff between power against smooth or wavy alternatives was fairly costly. Another question might be how that tradeoff changes with growing sample size. In other words, how do different rates of shrinkage of bin widths fare, as gauged by composite power. We performed a second set of simulations to examine this question. See Figure 3.3.

Three different rates of shrinkage were used: $k(N/100)^{-.2}$, $k(N/100)^{-.3}$, $k(N/100)^{-.4}$. Note that the bin widths are all equal at $N = 100$. Again, we used the measure of composite power to compare the different bin widths. As you can see, there is very little difference in the performance of the different shrinkage rates, at least until we have 1000 observations. At 1000 observations, the rate shrinking the fastest seems to gain advantage, certainly because it was the first to be able to detect wavy alternatives effectively. Essentially, though, this set of simulations offers us little guidance in window width choice.

3.3 Other Methods in the Literature

We will now present the main argument of this chapater: the context of testing provides us with a different objective function for choosing smoothing parameters than does estimation, thereby rendering much of the existing literature on window width selection in estimation inappropriate for the task of window width selection in testing.

We will first note that we would like a practical rule (or perhaps a few rules to choose among) for choosing window width. Before we can commence with the business at hand, though, we must first demonstrate what the optimal[1] window width would be for a given alternative. Since we do not know the alternative against which we are testing, the following rule is not a practical suggestion. We will need the result to exhibit inadequacies of other rules, though.

One possible criterion is to choose the window width h to maximize

$$\frac{\partial}{\partial c^2}\left\{\frac{E(\mathcal{T}|H_A, N)}{V(\mathcal{T}|H_0, N)}\right\}\Bigg|_{c=0},$$

where $\mathcal{T}$ is the test statistic, and c is the coefficient on the misspecification. Note that the derivative is taken with respect to c^2 because the derivative of that term is identically equal to zero at $c = 0$ due to its symmetry around $c = 0$.

[1]under a given criterion

Such a criterion will produce a window width that essentially maximizes the rate at which the expected value of the test statistic grows *relative* to the variance of the statistic under the null for infinitesimal departures from the null. Loosely speaking, the criterion ensures that the test statistic will grow quickly, thus making it easier for one to reject the null, at the point of infinitesimal departures from the null. We do not claim that this criterion is the only reasonable criterion. For instance, we could have chosen h to maximize the derivative of the probability that the test reject for infinitesimal c instead of maximizing its rate of growth.

All such criteria, of course, will depend on the particular form of the alternative, or departure from the null. For alternative $f(X)$,

$$y = X\beta + cf(X) + u,$$

and the proposed criterion is

$$\max_h \left\{ \frac{\sum_j f(x_j) \sum_i f(x_i) w_{ij}}{\sigma^2 \mathrm{tr} W^2} \right\}.$$

It is true that for a fixed alternative, the window width optimal under this criterion does not depend on N, the number of observations, once N is large enough.[2] This is because for a fixed alternative, once the window width is small enough to not wash out the alternative, it need not continue to become smaller to improve power.

To illustrate, we include an expository figure, Figure 3.4. Suppose that the line is our null model and the curve is the truth. The vertical lines represent residuals from the null. If we were to perform a bin test where the bins corresponded to the spaces between the double lines, it is clear that we would "wash out" much of the misspecification. Look, especially, at the second bin. Since the bin straddles the point where the sign of the expectation of the residual changes, the contribution of that bin to the statistic will be close to zero despite the misspecification. Suppose, instead,

[2]One can see this easily if we restrict h to the class such that $h \to 0$ and $Nh^d \to \infty$, really the only window widths we consider for the test statistic. Then, this criterion is bounded above by $c_1 Nh$, where c_1 is some constant. For h fixed, however, the criterion is asymptotically equivalent to $c_2 N$. Keeping h fixed, therefore, maximizes the criterion.

that we were to choose bins corresponding to the space between any lines, dotted or double. The contribution from almost all of the bins would be positive. If the bins were *too* small, however, much information on the spatial correlation of the residuals would be lost. Our intuition might lead us to the conclusion that the optimal window width for a given alternative would shrink as N increases only to the point where it will no longer wash out information. It would then stay constant as $N \to \infty$.

We will now discuss methods proposed in the literature. Rodriguez and Stoker (1992) choose the window width for their nonparametric regression test by generalized crossvalidation (GCV), although they offer the caveat that this method will not satisfy conditions necessary for their asymptotic results. By choosing the window width through GCV, a data-based method, they are essentially fitting the alternative model. We now present an example of why such a method does not produce the optimal window width for testing. Figure 3.5 presents two testing situations. GCV would provide quite different window widths for the two problems, a much smaller one in the second situation, trying to fit all of the "wiggles" of the true model. Since for testing we are not interested in the *shape* of the truth but rather just in whether it deviates from the null, we would optimally choose very similar window widths in the two situations. One can see this by referring back to the formula for optimal window width choice. It contains the alternative $f(\cdot)$, but no derivatives. Therefore, the wigglyness of one alternative matters very little as long as it does not stray far from the other alternative. We would want the window width to be some sizeable fraction of the periodicity with which the truth crosses the null. Any "wiggling around" either on top or below the null is irrelevant. One could also see that GCV will not choose the optimal window width for any alternative asyptotically because the window width will go to zero. As we saw before, the optimal window has postive width asymptotically. Eubank and Spiegelman (1990) also use GCV to choose their smoothing parameter.

Azzalini, Bowman, and Härdle (1989) have the most extensive discussion on window width selection. They propose two model-based selection techniques. The smoothing parameter is chosen to minimize the expectation under the null of, in one case, a pseudo-likelihood ratio, and, in the other, a normalized error sum of squares.

They note a shortcoming of such methods: a regression model with a slope near zero, for instance, leads to very large smoothing paramters, obscuring nonlinearities. This explanation is a roundabout way of making the point that choosing a smoothing parameter to optimally fit the null model in no way guarantees an appropriate smoothing paramter for testing. See Figure 3.6 for an illustration of the specific example mentioned by Azzalini, *et al.* Note that a procedure for window width choice based on the null might, depending on the procedure, yield different window widths in those two cases. For testing, though, we would want identical window widths in the two cases: the formula for optimal window width only contains the alternative, not the null. In addition, Azzalini, *et al.* also propose GCV for window width choice, obviously having the shortcomings mentioned before.

Other authors (*e.g.* Wooldrige, Bierens) have recognized that choice of smoothing parameter in the context of testing is a different problem than in the context of estimation. They commented on the possibility of tailoring their tests to detect certain alternatives by choosing a particular value for the smoothing parameter. Both noted, however, that such choices might affect asymptotic results.

We have now seen fairly compelling examples that lead us to reject methods of window width choice based on either fitting the null or fitting the alternative. The question now is what other solutions are available.

3.4 New Ideas

Earlier we derived an optimal window width choice rule for a given alternative. Our first proposal is to apply that criterion not to a particular alternative, but to a prior distribution that the researcher has over the alternative space. We believe that this Bayesian approach to window width selection would yield a test more powerful in finite samples against alternatives considered most likely, but still consistent against all alternatives with positive probabilities. It seems that this procedure would be similar to choosing among parametric tests, depending on which alternatives one believes are likely or possible, but it would be an explicit way of doing so based on stated prior beliefs.

Our second proposal for window width selection is to choose window width to maximize the test statistic. The motivation behind such a method is that finding *any* window width that results in a large statistic is evidence of misspecification. Also, the onus of choosing the window width, trying to ensure it is large enough to provide power but small enough to not wash out the misspecification, is taken off the researcher. Such an approach prompts several comments. First, some of the great computational saving afforded by the bin version of the test will be lost. Performing an actual maximization seems excessive, but even a grid search can be computationally taxing. Second, the question arises whether this test statistic exists. Even if the statistic exists, we would be much more pleased if we got an interior solution instead of, say, a solution at $h = 0$. A boundary solution would be strange because we think that a statistic resulting from $h = 0$ or h being a large as possible is fairly meaningless. Finally, the critical values of such a statistic seems problematic, to say nothing of the entire null distribution.[3]

We will now address the second comment, the one of existence.

Proposition 1 *Let h be the window width, $\mathcal{T}(h)$ be the test statistic of Ellison and Ellison (1992). For a fixed sample size N, $\mathcal{T}(h)$ is bounded. Therefore, a maximum of $\mathcal{T}(h)$ exists.*

Proof

$$\mathcal{T}(h) = \frac{u^T W(h) u}{\sqrt{2}\sigma^2 \mathrm{s}(W(h))},$$

where $\mathrm{s}(\cdot)$ is the square root of the sum of the squares of the elements of the matrix. Note that the numerator is bounded above by $u_{\max}^2 \sum_{i,j} w_{ij}(h)$. The denominator is bounded below by $\sqrt{2}\sigma^2(1/N^2)\sum_{i,j} w_{ij}(h)$. Therefore, there exists an upper bound on the statistic independent of h for a fixed N.

[3]One can imagine computing critical values directly if one knew the critical values of the tests over which the maximization was being performed. We will not attempt such a thing. We will look only at the null distribution and calculate critical values from that.

QED.

We are still not assured of an interior solution, but at least note that as the window width becomes too big, the weight matrix W will approach, essentially, a matrix of ones with zeroes on the diagonal. The statistic will, therefore, be negative, so we will not obtain a boundary solution there.

Once we know the statistic exists, it is crucial that we answer questions about its asymptotic behavior so that we can, among other things, compute critical values for it. We conjecture that such a statistic suitably nomalized has an asymptotic distribution and also that it is consistent against alternatives in all directions. Proofs are not included here,[4] but we do provide simulations of the null distribution.

To perform the simulations, a grid search was used instead of a maximization routine to save on computation. We performed a series of simulations of bin statistics under the null with increasing sample sizes. The goal was either to determine that the statistics were distributed identically or at least to detect a pattern of convergence to a particular distribution as the sample size increased. Table 3.1 contains the simulation results. As one can see, the first two estimated moments of the distribution, along with the estimated 5% critical values, are quite constant as the sample size increases. It is difficult to determine whether any convergence to particular values is taking place, but we are encouraged that the statistics seem to be properly normalized. These very preliminary results seem to indicate that future work in this area holds promise.

[4]The proof of asymptotic distribution could probably make use of results on the asymptotic distribution of order statistics, although those results would not be immediately applicable.

References

Azzalini, A., A. Bowman, W. Härdle (1989). On the Use of Nonparametric Regression for Model Checking. *Biometrika*, **76**, 1-11.

Bierens, H. (1982). Consistent Model Specification Tests. *Journal of Econometrics*, **20**, 105-134.

—- (1990). A Consistent Conditional Moment Test of Functional Form. *Econometrica*, **58**, 1443-1458.

Ellison, G., S. F. Ellison (1992). A Simple Framework for Nonparametric Specification Testing. Draft, Harvard Department of Economics.

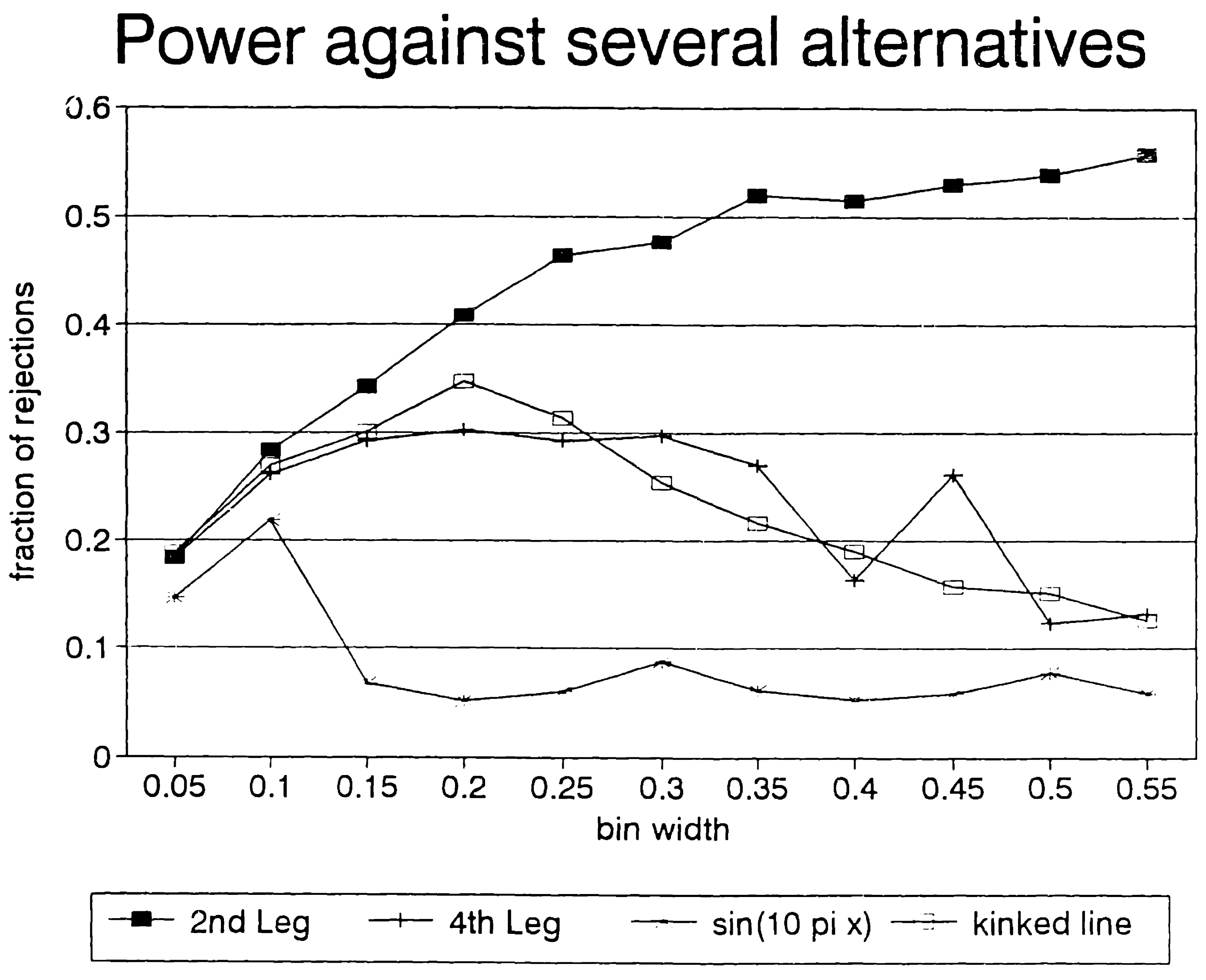
Eubank, R., C. Spiegelman (1990). Testing the Goodness of Fit of a Linear Model Via Nonparametric Regression Tecniques. *Journal of the American Statistical Association*, **85**, 387-392.

Rodriguez, D., T. Stoker (1992). "A Regression Test of Semiparametric Index Model Specification." Draft, MIT Sloan School.

Wooldridge (1989). "Some Results on Specification Testing Against Nonparametric Alternatives." Draft, Michigan State University Department of Economics.

—- (1990). "A Test for Functional Form Against Nonparametric Alternatives." Draft, Michigan State University Department of Economics.

Figure 3.1



Power against several alternatives
0.6
0.5
0.4
0.3
0.2
0.1
0
fraction of rejections
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
0.55
bin width
2nd Leg
4th Leg
sin(10 pi x)
kinked line

Figure 3.2

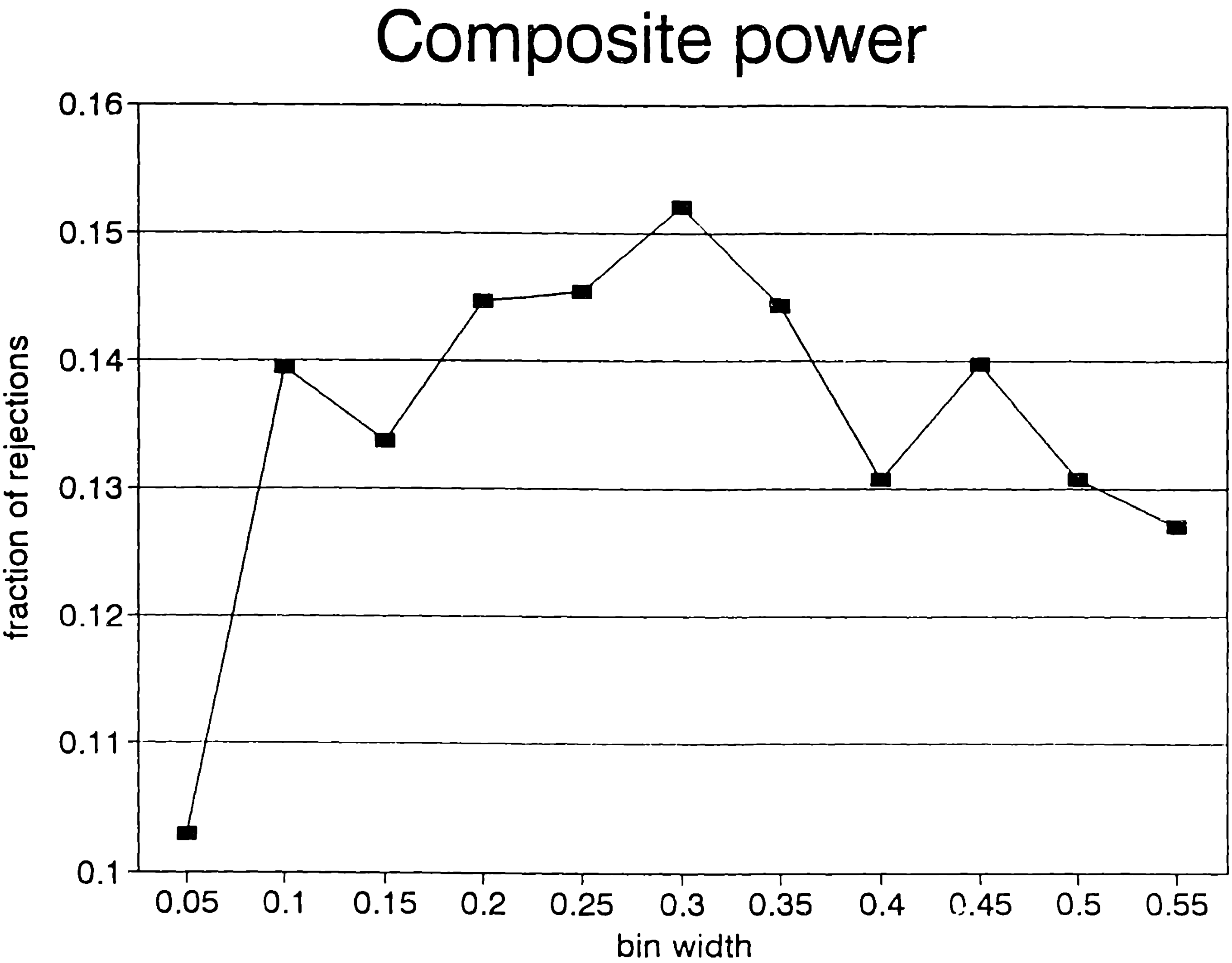

Composite power
0.16
0.15
0.14
0.13
0.12
0.11
0.1
fraction of rejections
0.05 0.1 0.15 0.2 0.25 0.3 0.35 0.4 0.45 0.5 0.55
bin width

Figure 3.3

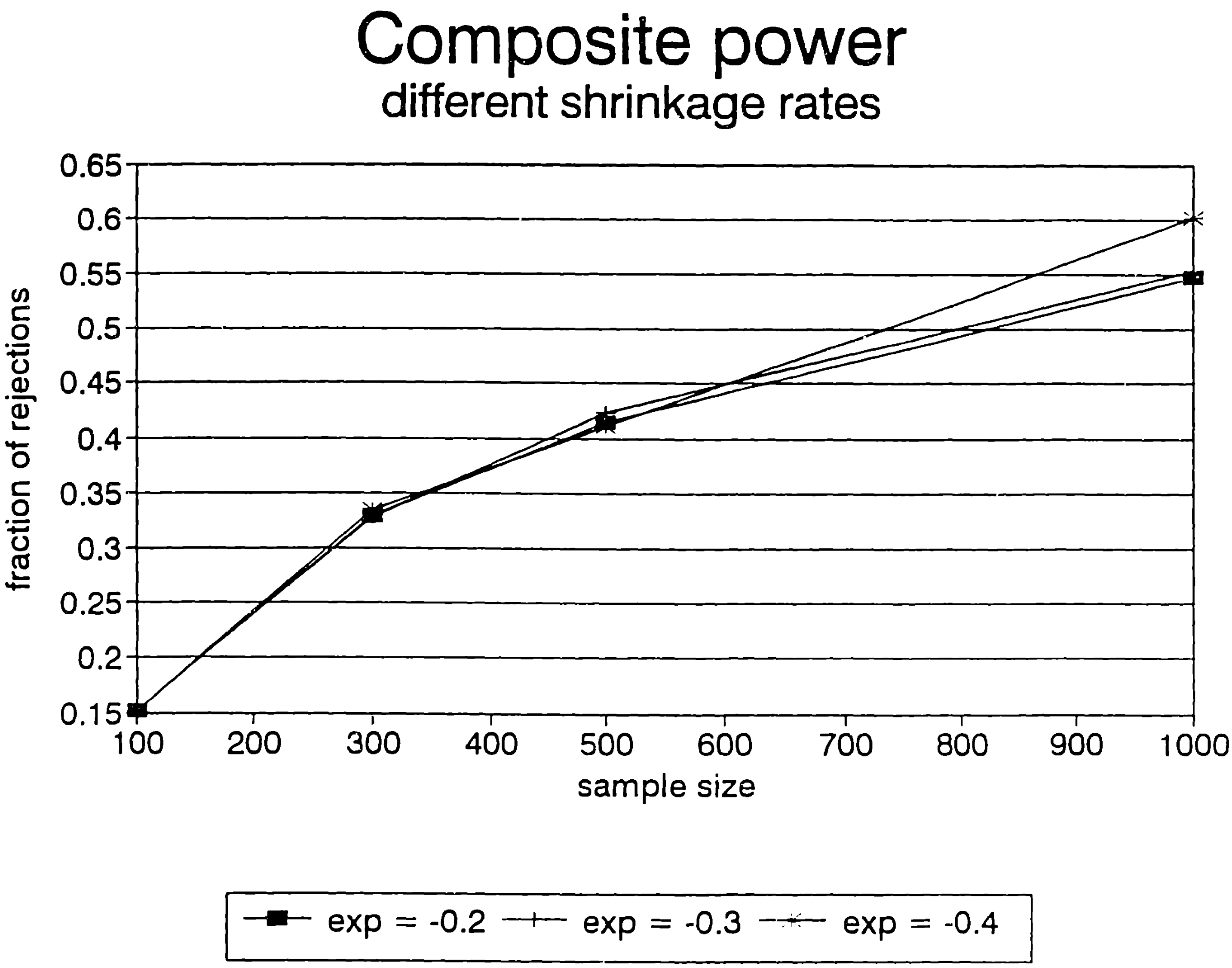

Composite power
different shrinkage rates
0.65
0.6
0.55
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
fraction of rejections
100
200
300
400
500
600
700
800
900
1000
sample size
exp = -0.2
exp = -0.3
exp = -0.4

Figure 3.4

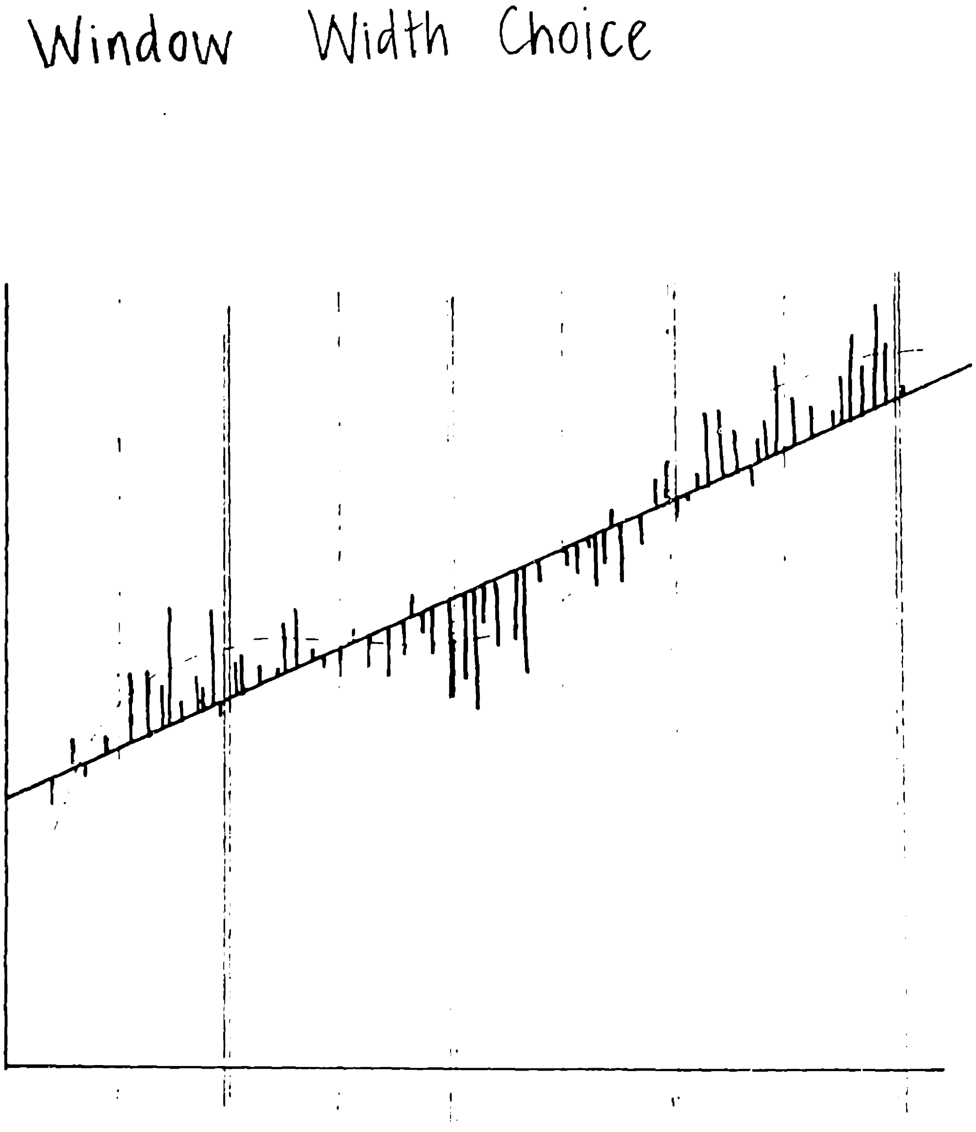


Figure 3.5

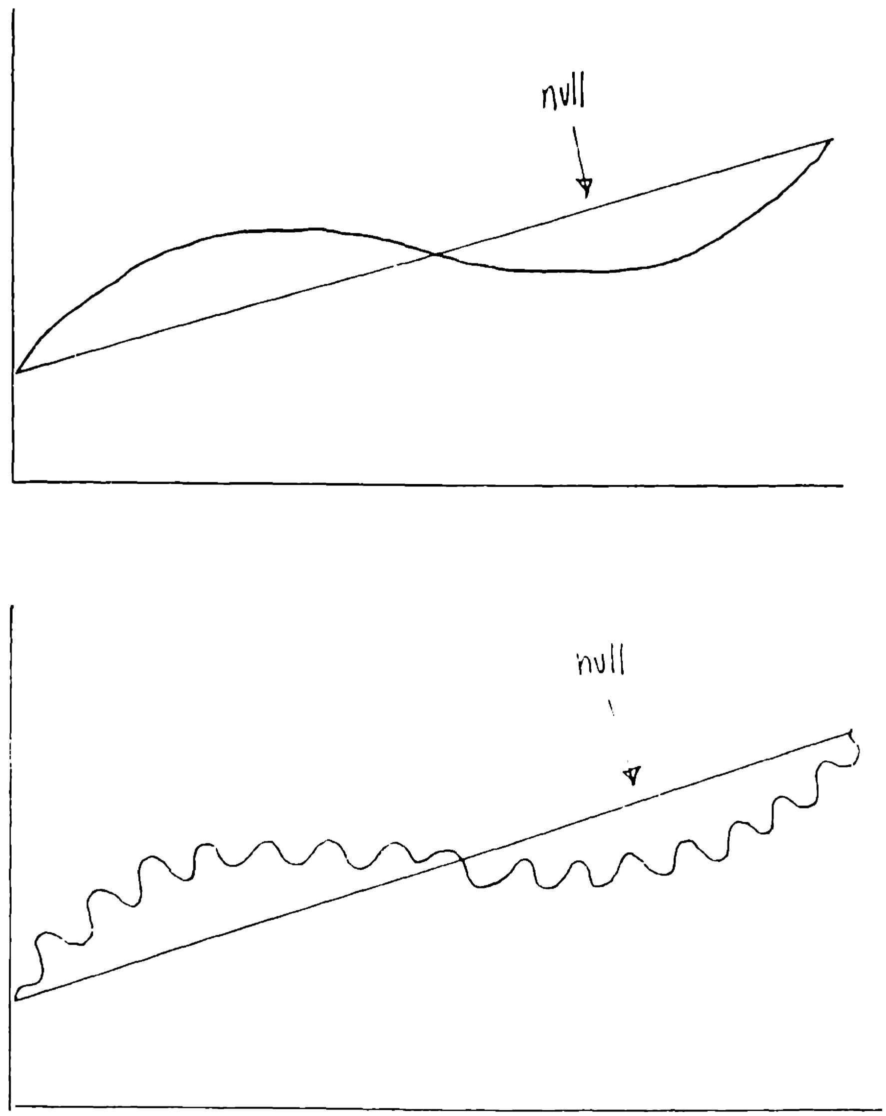


Figure 3.6

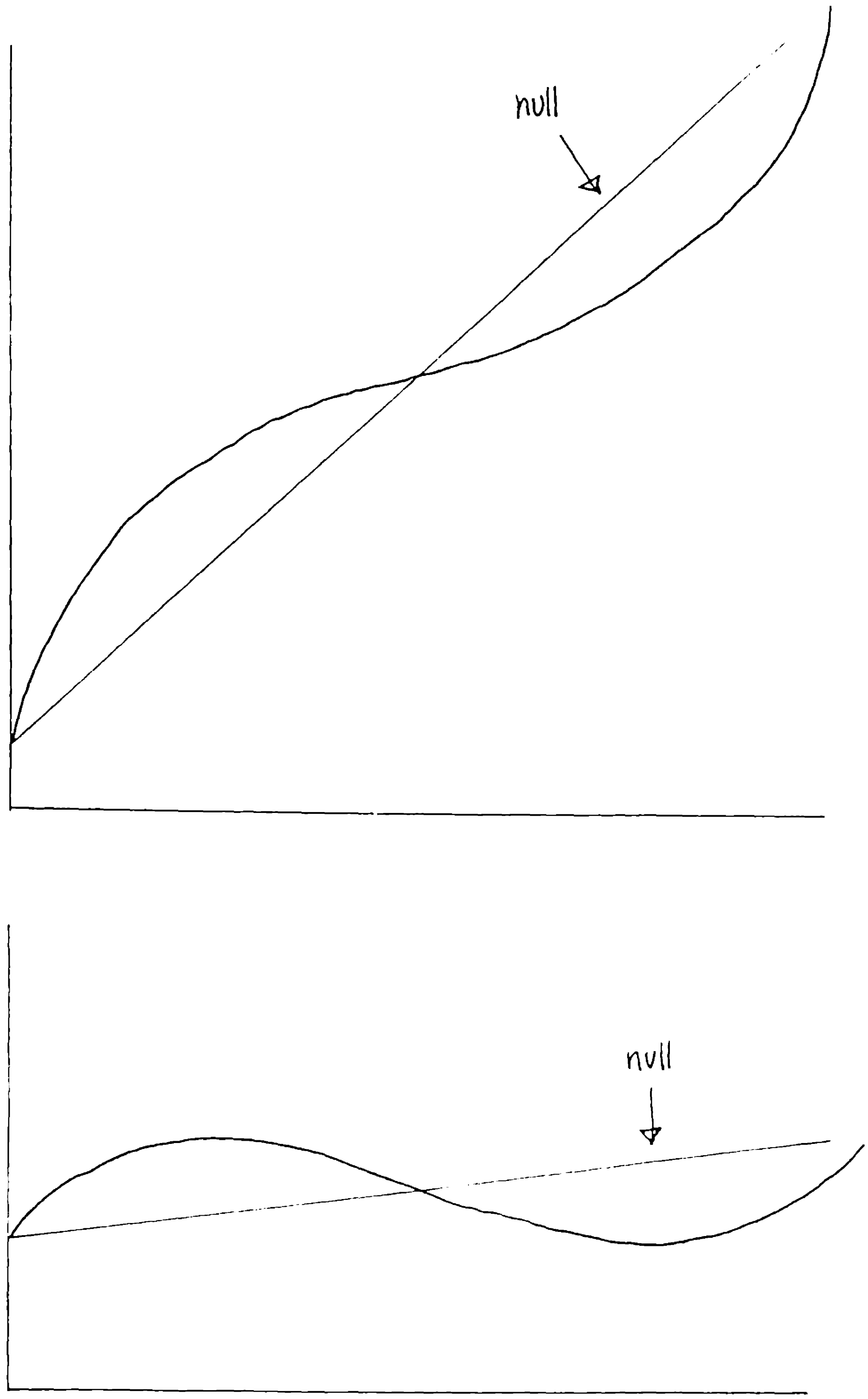


Table 3.1: Distribution of the Maximal Test Statistic

sample size	mean	std. dev.	5% critical value
50 obs.	0.98	0.54	2.13
100 obs.	1.00	0.59	2.23
200 obs.	0.92	0.50	1.83
300 obs.	0.96	0.56	2.16