AVOIDING MISSPECIFICATIONS AND IMPROVING EFFICIENCY IN HEDONIC AND CONSUMPTION MODELS: APPLICATIONS OF SEMIPARAMETRIC METHODS

KUO CHUEN LEE

A THESIS SUBMITTED FOR A DEGREE OF

DOCTOR OF PHILOSOPHY

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

JANUARY 1990

ABSTRACT

The objective of this thesis is to avoid misspecifications and to seek efficiency improvements in cross sectional and time series econometric applications using semiparametric methods. We restrict our attention to single equation models and the use of conditional moment restrictions as well as maximum likelihood methods. The first part of the thesis deals with cross sectional studies on the United Kingdom car market and the second part deals with time series studies of the United States consumption function. There are five main contributions of the thesis.

First of all, we have suggested minor extensions of existing semiparametric models; secondly, we have suggested the use of a dimensional reduction method prior to nonparametric estimation; thirdly, we have investigated the use of various rules of subjective and automatic bandwidth selection methods using real and simulated data; fourthly, we have suggested a new approach to overcome problems in the hedonic approach for cross sectional studies,; and finally, we have established a relationship between expected real interest rate and consumption using US time series data.

Dedicated to

my parents and my wife,

for their love, support and encouragement.

ACKNOWLEDGEMENTS

In preparing this thesis, I am indebted to many people. Foremost on the list is Prof. Peter M. Robinson. The project would not have been undertaken at all had it not been for the encouragement and support from him at the initial stages. The intellectual debts to him are the most important and hardest to measure, and his tactful criticisms, helpful advice, thoughtful guidance and constant encouragement have had the greatest effect on many parts of this research.

I would also like to thank Prof. James Durbin and Dr. Martin Knot for being my supervisors and taking care of my welfare. I would like to thank my colleagues Miguel Delgado, Javier Hildago and Masako Kurosawa for some stimulating discussions and assistance in computer programming. The comments and ideas of Adrian Pagan, Lonnie Magee, John Lane, Steve Pudney, and Orazio Attanasio on some chapters of the draft have led to the improvement in the present thesis. I am also indebted to Dr. Charles Bean, Robert Blackman of Emmerson Hill Associates, Rosemary Roberts of Haymarket Magazine Limited and Vivien Marles of the Association for Consumer Research, for their help in various ways during the collection of the data.

The financial support from the LSE and the University of London Postgraduate Studentship is gratefully acknowledged. Last, but not the least, the support given by my parents and the sacrifices made by my wife Gwinnie, have made my research less difficult and are dearly acknowledged.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACRONYMS

ACE	Alternating Conditional Expectations
ARCH	Autoregressive Conditional Heteroscedasticity
ARE	Asymptotic Relative Efficiency
ARMA	Autoregressive Moving Average
ASPE	Average Squared Prediction Error
BC	Box–Cox
CCAPM	Consumption Capital Asset Pricing Model
CRRA	Constant Relative Risk Aversion
CV	Cross–Validation
EVM	Errors–In–Variables Method
FIML	Full Information Maximum Likelihood
GARCH	Generalised ARCH
GCV	Generalised CV
GLS	Generalised Least Squares
GMM	Generalised MM
GN2SLS	Generalised Nonlinear 2SLS
HBC	Heteroscedastic Box–Cox
HHBC	Heteroscedastic Hyperbolic Box–Cox
HIHS	Heteroscedastic Inverse Hyperbolic Sine
i.i.d	independent identically distributed
ISE	Integrated Squared Error
IV	Instrumental Variable
LCPIH	Life Cycle–Permanent Income Hypothesis
LM	Lagrange Multiplier
LR	Likelihood Ratio
MA	Moving Average

MISE	Mean Integrated Squared Error
ML	Maximum Likelihood
MLS	Multiple Least Squares
MM	Method of Moments
MPG	Miles Per Gallon
MSE	Mean Square Error
NL2SLS	Nonlinear Two Stage Least Squares
NLIV	Nonlinear Instrumental Variable
NLLS	Nonlinear Least Squares
N–W	Nadaraya–Watson
OCE	Ordinal Certainty Equivalence
OLS	Ordinary Least Squares
QTM	Quadratic Transformation Model
RSS	Residual Sum of Squares
RT	Rules–of–the–Thumb
SARCH	Semiparametric ARCH
SML	Semiparametric Maximum Likelihood
TS	Two Stage
VAR	Variance

CHAPTER 1

INTRODUCTION

1.1 PRELIMINARIES

Econometrics are useful for policy analysis and counter factual purposes. For policy studies in general, economists are interested in finding the effects of the change of some controlled instruments on the dependent variables. Notable examples of policy studies are the measurement of environmental benefits of improving certain neighbourhood qualities and the impact upon car attributes brought about by the government restriction on fuel efficiency, in view of the hike in the price of petrol. As for counter factual purposes, diagnostic or specification tests are usually employed to cross examine empirical facts with theory. For example, economists are interested in finding whether anticipated and unanticipated inflation variables and stock prices matter in the consumption model.

One approach of econometric modeling is to formulate an economic model and end up with the relationship in a parametric form relating the variables of interest. In the process, econometricians almost always place some additional restrictions on the economic model under study for empirical tractability with little economic motivations. Generally, the restrictions are in the form of parametric assumptions regarding functional form or distribution of some variables or the disturbances. The stronger the assumptions, the simpler the structure of the econometric model. For easy interpretability, the resulting models are usually of a very simple form, e.g., a linear or log linear model. However, if one proceeds to conduct policy analysis based on the model, there are always some doubts on whether these additional assumptions embodied in the model are valid. These doubts will inevitably affect one's faith in the estimates.

If one's interest is in hypothesis testing and the null is rejected, one cannot know for sure whether the theory is at fault or that the additional parametric assumptions are being incorrectly imposed on the model. It is in fact a joint test of parametric assumptions and the theory.

Sometimes, these additional assumptions imposed by the econometrician can be tested using the data. Diagnostic and specification tests are usually employed to perform the task. However, in many instances, successive rejections of these tests leave the econometrician stranded with little idea of how to proceed.

Another approach known as measurement without or with little theory is to specify a very general statistical model and seek the best representation of the relationship and then reconcile the results with existing economic theory. The existing theory is not necessarily consistent with the best fitted models. Indeed, it may indicate that the best model is impossible within a certain theoretical framework. If the "best" model is inconsistent with the theory, then new explanations has to be offered and indeed a new theory. Unfortunately, the refined theory will often lead to empirical intractability unless some additional parametric assumptions are added. This leads us to the same problems as in the first approach.

Therefore, it would be desirable to proceed with policy studies or hypothesis testing without further assumptions besides those imposed or implied by the economic model. In this study, we use the economic framework as a bench mark for our studies; when economic information is lacking in assisting us in functional form selection, we resort to nonparametric technique which relieves us from making any further assumptions. When this is not possible, we will compromise by relaxing at least some restrictions so that we have a more flexible model. Since we combine the use of parametric and nonparametric components, the approach can be termed semiparametric.

In order to understand the work in this thesis, we will have to present the econometric models in this chapter. The motivations are discussed in details in the following chapters as we encounter them. This chapter is intended to serve only as an introduction to the empirical work in later chapters.

Let Y be a continuous random variable, X a kx1 vector of continuous variables and Z a px1 vector of predetermined continuous independent variable. The econometricians observed the values of Y, X and Z, where $\{y_i;i=1,\ldots,N\}$, $\{x_{ij};i=1,\ldots,N,j=1,\ldots,k\}$, $\{z_{ij};i=1,\ldots,N,j=1,\ldots,p\}$, $x_i = \{x_{ij};\ j=1,\ldots,k\}$, and $z_i = \{z_{ij};\ j=1,\ldots,p\}$. Let us assume for the moment that the y_i, and $w_i = (x_i,z_i)$ are related by the model:

$$T_y(y_i;\lambda) = T_w(w_i;\theta) + \epsilon_i, \qquad i=1,\ldots,N$$

where λ and θ are some parameters to be estimated, $E[\epsilon_i|w_i] = 0$, $E[\epsilon_i\epsilon_j|w_i] = 0$ and $E[\epsilon_i^2|w_i] = \sigma(w_i)^2$. T_y and T_w are transformations or functions to be defined.

1.2 AVOIDING MISSPECIFICATIONS

Misspecification of the model usually leads to loss of consistency or efficiency. In general, consistent estimates can usually be obtained at the expense of loss of efficiency. The existence of a true model is important in theoretical work. However, many believe that in empirical work, the true model is generally unknown and the trade off between consistency and efficiency does not really exist. Therefore, the preferred strategy for modelling should be to allow for possible misspecifications with minimal assumptions. This brings us to focus on some statistical concepts which we will now discuss.

These three aspects of statistical modelling are best summarized by Stone (1985):

Flexibility is the ability of the model to provide accurate fits in a wide variety of situations, inaccuracy here leading to bias in estimation.

Dimensionality can be thought of in terms of variance in estimation, the "curse of dimensionality" being that the amount of data required to avoid an unacceptably large variance increases rapidly with increasing dimensionality, or, as usually put, between bias and variance.

Interpretability lies in the potential for shielding light on the underlying *structure.*

While flexibility is positively correlated with dimensionality, there is always some trade off between these two aspects and interpretability in model building. Our foremost concern is to have "enough" flexibility in order to give us an accurate fit. Unfortunately, the more flexible a model is, the higher the dimension of parameter space that one has to deal with. Furthermore, it would generally be much harder for one to visualize or comprehend the relationship between the dependent and independent variables. Therefore, the three aspects are not entirely independent of each other and having a more flexible model will in general incur inflated variance and reduced interpretability. The particular class of semiparametric models that we employ in general, trade off inflated variance (efficiency) for flexibility (consistency). We are particularly concerned with in the following models in our applications:

(1) Method of Moments

The following models all have the common conditional moment restrictions of the form $E[\epsilon_i(w,y;\theta)|\ w_i] = 0$, where θ is the vector of parameters of interest. The most efficient of the method of moments (MM) estimator requires the estimation of $E[\partial\epsilon_i/\partial\theta|w_i] = h(w_i;\theta)$ which is generally unknown.

(a) Transformation model

A transformation model can be expressed as

$$T_y(y_i;\lambda_y) = \Sigma_{j=1}^{q}\beta_j T_j(w_{ij},\lambda_j) + \epsilon_i, \qquad i=1,\ldots,N \qquad (1)$$

where $E[\epsilon_i^2|w_i] = \sigma$, T_y and T_i are some known transformations, λ_y and λ_j's are some transformation parameters to be estimated, and β_j's are some unknown parameters to be estimated.

One of the reasons for using transformation is to induce independence of the independent variables. Although there are models which consider cross product terms between the w's, we restrict ourselves to the additive form. Transformation models have the advantage of inducing flexibility but as we have mentioned, this will increase the dimension of parameters of interest. In the case of a parametric model with a single transformation parameter for each T, the increase in the number of parameters to be estimated is two fold. Interpretability is complicated by the fact that the relationship is between functions of y and functions of the w's.

Various transformations will be introduced in Chapter 3. It should be mentioned that the model is semiparametric only in the sense that no distributional assumptions are imposed on ϵ_i. The contribution from this relaxation should not be overlooked, the reasons being that many transformation models, e.g., logarithmic transformation, exclude certain distributional assumptions and estimators obtained by imposing distributions are not robust. The use of MM will at least guarantee consistent estimates and correct inference.

Nonlinear two stage least squares, a special case of MM, is an appropriate estimation procedure for transformation models. It involves finding "optimal" instruments $E[\partial\epsilon_i/\partial\theta|w_i]$ which is usually assumed to be linear in w. However, the expectation is unknown in this case and we can apply the method of Robinson (1988e) in obtaining the optimal instruments rather than estimating the conditional expectations nonparametrically as in Newey (1987).

(2) Partial Linear Model

Now, consider the following model which is partly linear

$$y_i = \Sigma_j\beta_j^* x_{ij} + T_z(z_{i1}, ,..,z_{ip}) + \epsilon_i, \qquad i=1,...,N$$

Compare this model with (1), we have T_y and some T_j's all equal to 1. The transformations T_y and some T_j's, are known from economic theory or have good reasons to be linear in this case. But T_z is an unknown function which maps the remaining variables z onto the real line. This model is applied to cross sectional study in Chapter 4. We can then nonparametrically estimate T_z and then proceed to estimate β after eliminating T_z. If the interest lies in the shape of T_z, an estimate of

T_z can be recovered. In the event that we have a large number of z's in T_z and only a set of medium sized data, we may have to employ a more restrictive model of the form

$$y_i = \Sigma_j\beta_j^* x_{ij} + T(\Sigma_j a_j' z_{ij}) + \epsilon_i, \qquad i=1,\dots,N$$

where a_j is a vector of known coefficiencts. This device is to overcome some of the problems in nonparametric estimation. The issues of high dimensionality and inflated variance encountered in multivariate estimation are discussed in Chapter 2. This partly linear model is used in the time series study in Chapter 7. z's cannot be known with certainty in this case and rules out any dummy variables or constant. This is in consequence of relaxing the assumption that T_z is unknown. However, this consequence has a very useful purpose as we shall see in Chapter 4. In many instances, one is only interested in the policy changes of some or all of the z's on the y_i. In other words, we are concerned with distributional changes of y from, e.g., z to z^*. Let D be a vector of dummy variables and δ the coefficients. We may then be able to deal with dummy variables easily since we are only interested in $E[y]-E[y^*] = E[D\delta+T_z(z)] - E[D\delta+T_z(z^*)] = E[T_z(.)-T_z^*(.)]$. Thus, a nonparametric policy analysis can be conducted using the suggestions of Stock (1985a). Unlike the case of MM, where it is known that it is most efficient within a class, the partly linear model estimator is only known to be root–N–consistent and asymptotically normally distributed (Robinson (1988a)).

1.3 EFFICIENCY IMPROVEMENTS

Allowing for flexibility as in the above models may have the consequence of inflated variance. For example, if the error term in the transformation model is assumed to be of a parametric form, efficiency can be improved; and if T_z is assumed to be known in the partly linear model, efficiency can again be improved. However, in both cases, if these assumptions are not correct, then one has to pay a heavy price of having inconsistent estimates.

As mentioned above, the problem in applied work is that the true functional form is seldom known although economic theory may provide or impose some restrictions on the functional form. In many other instances, there is virtually no information to guide one in selection. Given the level of generality, the usual approach is to use an ad hoc functional form. In many instances, especially when the functional form itself is not of intrinsic interest, e.g., when the nuisance function is involved, linear functional form is employed. A more respectable approach is to use diagnostic and specification tests in aiding one to select the correct functional form. The problem arises when successive tests are rejected, one finds it difficult to suggest a suitable functional form.

Recent work has progressed in the direction of adaptive estimation. In simple terms, the adaptive semiparametric estimator has the same efficiency under unknown distribution or/and functional form as a parametric estimator under known distribution or/and functional form. Some of our models are of this class, but we limit ourselves to maximum likelihood (ML) estimation when the error term is known to be from a parametric family.

Our main concern under the heading of efficiency improvement is efficiency of the estimators and consistency is taken for granted. In other words, if we use the incorrect parametric functional form for the functions, we will still have consistent estimates. The nonparametric technique is only used to estimate the nuisance functions, i.e., the functions are itself not of intrinsic interest. Of course, in the case of misspecification of the nuisance function, the standard errors are usually inconsistent and therefore the semiparametric model has advantage.

The main motivation of efficiency improvement here is that statistical inference is the main objective. In the case of conditional maximum likelihood estimation of heteroscedastic models, e.g., autoregressive conditional heteroscedastic (ARCH) models, misspecification of the conditional variance may invalidate the variance–covariance matrix. Therefore, flexibility of the nuisance functions in ARCH and other conditional heteroscedastic models is important. The other estimators which belong to this class are the method of moments estimators which include the linear and nonlinear heteroscedastic, transformation and errors–in–variables models.

(1) Maximum Likelihood Method

Maximum likelihood estimators for the following models are not robust to slight misspecifications of the error distribution or nuisance functions. However, efficiency improvement can be attained if additional information, such as the information on the error distribution, is used for estimation.

(a) Transformation model

The error term of the transformation model can be known to have certain distributions, e.g., normal, Gamma or t–distribution. The normal distribution is favoured if it is permitted by the transformation used. In fact, one of the purposes of transformation is to reduce skewness. These distribution assumptions can be checked but if it is known with certainty, then making use of the information will usually lead to efficiency improvement.

Although one of the intentions of transformation is to induce homoscedasticity, a direct heteroscedasticity correction may still be desirable. This can be done by nonparametrically estimating the conditional variance and a model is introduced and applied to real data in Chapter 3.

(b) ARCH

In the ARCH model, the conditional variance is usually assumed to be a function of the squared lagged residuals. However, there is no reason to believe that this is known a priori. It is natural to consider nonparametric estimates for the conditional variance. We also argue that inclusion of the lagged dependent variable in the conditional variance is a good practice to avoid possible inconsistency of the so called Engle's ARCH model (Engle (1982)).

(2) Method of Moments

In the following models, $E[\epsilon_i|z_i] = 0$ and possibly $E[\epsilon_i\epsilon_j|z_i] = \sigma_i$. Since our MM estimators have the interpretation of IV (Instrumental Variable) estimators, efficiency improvement can be attained by using more efficient instruments.

(a) Transformation model

In cross sectional studies, the efficiency of most transformation models can be improved by taking into account heteroscedasticity. It is easy to see that the generalized non–linear two–stage least squares (GNL2SLS) estimator is more efficient than the non–linear two–stage least squares (NL2SLS) estimator. If the heteroscedasticity is of an unknown form but $\sigma_i = \sigma(z_i)$, then it is natural to employ nonparametric estimates for the conditional variance. This class of model is dealt with in Chapter 3. Of course, failure to take heteroscedasticity into account in this case will also invalidate the statistical inference.

(b) Errors–in–variables

By imposing suitable restrictions on the T's in (1), we can have the linear GLS model and this model is used in Chapter 3. However, if the explanatory variables contain conditional expectation terms, we have an errors–in–variable model. Consider the linear rational expectations models as in Chapters 6 and 7,

$$y_i = \beta_0 + \Sigma_j\beta_j E[x_{ij}|z_i] + \epsilon_i, \qquad i=1,\ldots,N$$

In some models where the conditional expectations are unknown functions, we may use the nonparametric estimates as instruments. Of course, it is well known that in IV estimation, consistency is taken for granted while efficiency can be improved by finding the instruments closely related to the explanatory variable.

1.4 THE SCOPE AND OUTLINE OF THE THESIS

The thesis is organized into nine chapters and we now provide a brief outline.

One of the most important nonparametric techniques that we employed is the method of kernels. In particular, kernel nonparametric "regression" will be the essence of the nonparametric technique in the applications. Of course, regression analysis in economics has usually been referred in relation to parametric linear or nonlinear regression ones. In fact, regression analysis extends beyond this class of parametric models and many other topics should be included in this broad heading, e.g., biased estimations (Stein, Ridge, principal components regressions and indeed the nonparametric regression that we are interested in). It is therefore important to understand the concept of nonparametric regression and the distinctions from parametric regression. We have presented and discussed the formulae, motivations and properties of the nonparametric estimators in Chapter 2. This will give us some understanding of the working of the nonparametric techniques and will serve as a reference in later chapters.

Two areas of economics will particularly benefit from the use of semiparametric models mentioned in Section 2 and 3. The first is the hedonic price function and the second is the consumption function.

The hedonic price and related functions were modelled by linear models in the early 70's and it is not until recently that transformation models have been favoured. One of the reasons is that most of the theories of the hedonic literature do not provide any information on the functional form of the price function. The usual way to approach it is to select a formula that will provide adequate description of the data. For application purposes and practicality, linear functions of the variable and the transformed variables (e.g., logarithmic) are preferred and indeed perform reasonably well in most studies. However, there are always some doubts and concern over the simple functional form. In the late 70's, the discussions of the choice of functional form were the focus of research in the hedonic literature. The results were mixed and it is still difficult to draw any firm conclusion as there is little uniformity among the different studies. This is in part due to the large number of independent variables in the studies and multicollinearity is often encountered. In the hedonic literature on vehicles, there is however a clear pattern evolving regarding the choice of variables. But this is still somewhat dependent on the functional form employed.

In Chapter 3, we have introduced a different methodology which makes use of the two–stage hypothesis of Ohta and Griliches (1976). The procedures overcome the the problem of multicollinearity and extract more information from the fuel efficiency and rental cost function then other approaches, such as the discrete choice model. In this study of petrol price elasticities of fuel efficiency as well as attributes, an intertemporal economic model is constructed. The fuel efficiency function and the rental cost function need to be estimated before we can construct the elasticity. However, there is virtually no information on the functional forms. Transformation models may be a desirable robust approach to overcome the problem because of the relatively simplicity of the model structure. In this case, efficiency improvement can be attained by correcting for heteroscedasticity which is of an unknown form, but known to be function of the parameters and the attributes. Further efficiency improvement can be achieved in some cases using maximum likelihood method. Of course, this efficiency improvement is at the expense of making the additional distribution assumption on the error term.

However, recent interest on the hedonic literature lies in the choice of instruments in Rosen (1974)'s two–stage estimation procedure discussed in Chapter 4. The absence of appropriate instruments in single car market data makes Rosen's approach undesirable. If only policy studies are concerned, the partly linear model will offer partial solution and minimize the impact of misspecifications. For example, if one is interested in the fuel efficiency standard, then the variable which one should change is the fuel efficiency variable. While historically, linear functional form is prevalent, the benefits estimates are not robust to misspecifications. The more flexible partial linear model provides one of the solutions to this problem. In our investigation into the change in capitalized value of new cars , we have avoided making too many assumptions on the hedonic price function.

In order to better understand and motivate the work in Chapters 6 and 7, Chapter 5 is devoted to the review of some of the surprise consumption literature. The applications in this second part of the thesis will use only time series data.

In Chapter 6, we have tested the consumption capital asset pricing model by assuming that the conditional variance of inflation follows an ARCH process. We have suggested and applied a Pseudo–Gaussian maximum likelihood criterion for bandwidth selection. We found no evidence of expected inflation being an explanatory variable in this consumption model. We have also attempted to establish a relationship between consumption and expected inflation in the presence of possibly non–linear rational expectations formation. The one quarter expected interest rate has no role in explaining consumption. We have also discovered some interesting properties of the cross–validation function.

In Chapter 7, a different data set and an extended model is used to analyze the surprise consumption function. Our intention is to examine if lagged variable and possibly non–linear expectations should matter in a linear rational expectations model. We have also suggested preliminary non–linear projection to lower dimensional space prior to nonparametric estimation. Bean (1986)'s model is used as a bench mark and following Blinder and Deaton (1985)'s rational expectations approach, we have relaxed the assumption of linear expectations formation in the surprise consumption model. The anticipated terms can be viewed as estimates from the errors–in–variables models and unanticipated terms from the partially linear model. In the formal case, we are only interested in efficiency improvement as instrumental estimates are consistent. In the latter case, we are interested in attaining consistency estimates as any slight misspecification of the expectations will generally lead to inconsistent estimates. The resulting semiparametric surprise consumption may provide an approach to rational expectations of unknown form in other linear rational expectations models. Two semiparametric test statistics, namely R test (Robinson (1988c)) and Hausman (1978)'s test, are used for testing zero–type restrictions and diagnostic checking. We have obtained the estimates of interest under less restrictive environment with results opposite to those obtained under the more restrictive regime. We have also established a relationship between expected real interest rate and consumption in this extended model.

In Chapter 8, we have evaluated the performance of the automatic bandwidth selection criterion and various subjective rule–of–the–thumb methods by means of Monte–Carlo simulations. The results favour the use of automatic bandwidth selection method that we have suggested, but only with reasonably large sample size based on the simulation results.

Finally, in Chapter 9, we have suggested further extensions and potential areas for research.

CHAPTER 2

THE METHOD OF KERNELS
AND TECHNICAL COMPLEMENTS

2.1 INTRODUCTION

Consider the regression function or curve

$$y_i = m(x_i) + \epsilon_i \qquad i=1,\dots,N \qquad (1)$$

where y_i is a scalar, $x_i = (x_{1i},\dots,x_{di})'$, $E[\epsilon_i | x_i] = 0$, $E[\epsilon_i \epsilon_j | x_i] = 0$ for $i \neq j$, $E[\epsilon_i^2] = \sigma^2$. m(x) is the regression function and can be estimated by parametric or nonparametric methods. While parametric models is familiar to most economists, the use of nonparametric models are a relative new idea in the field of economics.

The most common approach to modelling (1) is to specify a parametric form for m thus giving rise to a linear or nonlinear regression model. It is assumed that the form of the regression function is known and there are a finite number of parameters, e.g., slope coefficients, to be estimated. In this case, the practitioner selects one particular curve from a whole family. Whatever results and inference obtained from the data is heavily dependent on the choice of the functional form of m(x).

An alternative approach to estimating the regression function is to use a nonparametric regression model. In this case, we make no parametric assumptions on m(x) except that it belongs to some infinite dimensional collection of functions. However, some weak smoothness assumptions may still be imposed on m(x), e.g., m is rth times differentiable. But the nonparametric model is less restrictive than the parametric models and relies more on the data for information.

While a nonparametric model is less restrictive than a parametric model, it has little practical interest in economics for various reasons. The most important of these is that economic applications generally involve many independent variables. The nonparametric estimators have a very slow rate of convergence and this is usually worsened when the dimension of x increases. The second reason is that there is usually physical economic meaning associated with simpler models. These reasons, among others, have prompted economists to use parametric models even when there is inadequate information on the functional form of m(x).

From the theoretical point of view, it is believed that parametric models have desirable properties over the nonparametric models, e.g., gains in asymptotic efficiency. However, this belief may in fact be unwarranted if the parametric model is misspecified. Furthermore, any subsequent policy analysis and conclusions from hypothesis testing may be erroneous if the model is indeed misspecified. Therefore, it is unwise to adhere to parametric models when in fact there are undesirable side effects as opposed to possible efficiency improvements and ease of interpretability.

The methodology which underlies our study is dominated by the use of semiparametric models. In other words, the model under study has two components: a parametric and a nonparametric component. The main difference between its parametric component and nonparametric component is that the former has finitely many unknown parameters while the latter has infinite number of parameters to be estimated. The most interesting property of the semiparametric model is that not only does it allow for unknown functions to be estimated, it sometimes has the same rate of convergence, and indeed the same efficiency, as parametric estimators.

There are numerous methods of nonparametric estimation including Fourier inversion, histogram, nearest neighbour, orthogonal series, penalty functions, splines, delta sequences, kernels and others. These methods have been surveyed in Prakasa Rao (1983), and brief discussions of some methods are given in Silverman (1986). Eubank (1988) has also described various methods of estimating regression functions.

This chapter is devoted purely to the discussion of a particular method of nonparametric estimation: the method of kernels. Although kernel regression is only important as an input to our semiparametric estimation procedures, it is still essential to understand some basic concepts and the workings of nonparametric estimation and related procedures. It will also be useful to point out several shortcomings of purely nonparametric models and to motivate the use of semiparametric models in empirical work. Furthermore, the asymptotic properties of semiparametric estimators are much easier to understand after some discussion of the theorems on nonparametric estimators.

We begin by discussing the method of kernels in estimating the density function and regression function of the identical and independently distributed (i.i.d.) case. After presenting some asymptotic results and properties, extensions to other cases will also be discussed.

2.2 SOME PRELIMINARIES

Let us assume that $(x_i,y_i) = (x_{1i},\dots,x_{di},y_i)$, $i=1,\dots,N$ are identically and independently distributed as continuous multivariate random variable (X,Y). We are interested in the estimation of the density, f(x), as well as conditional expectations or moments, $E[y|x]$ and $E[g(y)|x]$. First, consider (1) again:

$$y_i = m(x_i) + \epsilon_i \qquad i=1,\dots,N$$

where m is the regression function or conditional expectation. Let us assume that the problem here is to construct consistent estimator of $m(x) = E[y|x]$. Extension to the conditional expectation of g(y) on x is straight forward. Assume that we have a joint density f(y,x). The associated marginal density for x is given by

$$f_X(x) = \int_{-\infty}^{\infty} f(y,x)dy \tag{2}$$

The conditional density of Y given X is

$$f_{Y|X}(y|x) = f_X(x)^{-1} f(y,x), \tag{3}$$

and appealing to the definition of conditional moment, we have

$$m(x) = \int_{-\infty}^{\infty} y\, f_{Y|X}(y|x)dy. \tag{4}$$

We can estimate the conditional multivariate density (3) and use (4) to obtain the conditional expectation. (4) is what we are really interested in for most of our applied work. Since we have to estimate the conditional multivariate density (2) first before we can obtain the conditional expectation (4), naturally, we begin our discussions by presenting the method of kernels for univariate and multivariate density estimation.

2.3 ESTIMATION OF UNIVARIATE AND MULTIVARIATE DENSITIES

The literature on the estimation of density functions is vast. The bibliography up to the late 70's is provided by Wertz and Schneider (1979) and up to the early 80's by Collomb (1981). The number of articles quoted in the two bibliographies will demonstrate the enormous interest shown in this field. We do not intend to survey the whole literature. We re–iterate that our main purpose is to present the methodology and sufficient asymptotic results in order to understand the properties of the estimators of interest to us. It will also serve as a useful reference chapter as the same techniques are repeatedly used in later chapters.

The work on multivariate density has its origin in the statistical literature of the 1950's. In particular, the first published paper on univariate kernel density estimation was by Rosenblatt (1956). Univariate kernel estimators are unlikely to be very useful in economic problems as multivariate applications are usually encountered in econometric studies. However, to give some initial insights into the working and the properties of the method of kernels, it is useful to discuss univariate estimators. The use of univariate kernels is also useful in simplifying the notations when we are discussing extensions and more complicated applications of the method of kernels. Let us give a very brief introduction of the working of a univariate kernel.

UNIVARIATE DENSITY

For this sub–section, let us consider the i.i.d. observations on the scalar random variables x_i, i=1,...,N drawn from the density function f(x). Our intention is to estimate f(x) and by definition

$$f(x) = d/dx\, F(x) = \lim_{a\to 0}(a)^{-1}[F(x+a/2)-F(x-a/2)]$$

where F(x) is the distribution function. A uniform kernel estimator can be derived as follows:

$$
\begin{aligned}
\hat{f}(x) &= [F(x+a/2)-F(x-a/2)]/a \\
&= a^{-1}\ [\text{average number of } x_j,\ j=1,..,N, \text{ in the interval a} \\
&\qquad\qquad \text{centred at } x] \\
&= a^{-1}[\text{number of } x_j \text{ in the interval } (x+a/2,\ x-a/2)]/N \\
&= (Na)^{-1}\Sigma_{j=1}^{N} I(|x_j-x| < a/2) \\
&= (Na)^{-1}\Sigma_{j=1}^{N} I(|x_j-x|/a<1/2\ ) \\
&= (Na)^{-1}\Sigma_{j=1}^{N} I(a^{-1}(x_j-x))
\end{aligned}
$$

where $a = a_N$ is a sequence of positive number which satisfies the condition $\lim_{N\to\infty}a_N = 0$, and is sometimes known as window width, bandwidth, or smoothing parameter. As these names suggest, a actually controls the number of x_j to be averaged. I(u) is the indicator function which takes the value of 1 if $|u| < 1/2$ and zero otherwise, and has the properties that $\int I(u)du = 1$ and $I(u) \geq 0$.

Rosenblatt (1956) has suggested replacing the indicator function I(u) by any function K(u) which possesses the properties (a) $\int K(u)du = 1$ and (b) $K(u) \geq 0$. In his case

$$\hat{f}(x) = (Na)^{-1}\Sigma_j K(a^{-1}(x_j-x))$$

This kernel estimator has the advantage that more weight can be given to the observations closer to x and less weight to those further away. For example, the normal kernel, $K(u) = (2\pi)^{-1/2}\exp(-u^2/2)$, has the property of giving most weight to the observation x itself and the weight decays exponentially. Rosenblatt (1956) established the consistency of this univariate kernel estimator.

However, alternative kernels may be desired including relaxing the assumption that $K(u) \geq 0$ or/and symmetric K(u) for superior performance. In fact, Parzen (1962) was the first to generalize the results to non–negative kernels. Under the following regularity conditions on the kernel

(A1) $\int K(u)du = 1,$

(A2) $\text{Sup}_u |K(u)| < \infty$,

(A3) $\int |K(u)| du < \infty$, and

(A4) $\lim_{|u| \to \infty} |uK(u)| = 0$,

together with some conditions on the rate of convergence of the bandwidth a,

(B1) $\lim_{N \to \infty} a = 0$ and

(B2) $\lim_{N \to \infty} Na = \infty$,

the kernel estimator is asymptotically unbiased and mean square consistent at every continuity point of f. (B2) requires that as the number of observations increases, the bandwidth has to converge at a slower rate. This is just to make sure that as we have more observations. we must have a smaller bandwidth in order to delete observations further from the point of interest x.

Under conditions A1–A4 and B1–B2, $(\hat{f}-E\hat{f})/\text{var}(\hat{f})^{1/2}$ is asymptotically distributed as standard normal. Notice that we are centering around $E\hat{f}$ rather than the true f (see Cacoullos (1964)'s result below which deals with the latter case).

In fact, these conditions are fairly standard in the kernel literature though the following conditions are usually imposed on the kernels:

(A1a) $\int K(u) du = 1$,

(A2a) $\int uK(u) du = 0$,

(A3a) $\int u^2 K(u) du = c_1 \neq 0$,

(A4a) $\int K(u)^2 du < \infty$,

(A1a) is equivalent to the weights summing to one, (A2a) is automatically satisfied if K is symmetric about zero. (A3a) and (A4a) will be easy to comprehend as we come to the asymptotic properties of the density estimates. Kernels which satisfy conditions (A1a) to (A4a) are usually called second order kernels or simply kernel. The discussions of kernels with higher order than the second will be presented.

MULTIVARIATE DENSITY

It was Cacoullos (1964) who extended the univariate nonparametric estimator to a multivariate framework. Consider the case where we have i.i.d. observations on the d–vector x_i, i=1,..,N drawn from the density function $f(x_1,...,x_d)$. Let us ignore the i subscripts. The estimates $\hat{f}(x)$ of the density function of f(x) is:

$$\hat{f}(x) = (Na_1a_2...a_d)^{-1}\Sigma_jK(a_1{}^{-1}(x_1-x_{1j}),...,a_d{}^{-1}(x_d-x_{dj}))$$

where K(u) is a multivariate kernel satisfying certain conditions. For reasons best suited to our purpose, we restrict our attention to product kernels and a common bandwidth $a_1 =...= a_d = a$, where a is a sequence of positive constant satisfying $\lim_{N\to\infty}a = 0$. A diagonal bandwidth matrix or a full bandwidth matrix may be desired in some circumstances. Indeed, empirical worker usually use a diagonal bandwidth with pth diagonal element a_p = s.d. (x_p) x constant x N^{α}, where α is a negative fraction.

The bandwidth parameter controls the degree of smoothness of our estimates. In this case, the bandwidth is in fact the size of the neighbourhood which controls how many observations of x around x_i should be used for local regressions or averages. If we have a common bandwidth, we end up with a simple form

$$\begin{aligned}\hat{f}(x) &= (Na^d)^{-1}\Sigma_jK(a^{-1}(x-x_j)) \\ &= (Na^d)^{-1}\Sigma_j\{\Pi_pk_p(a^{-1}(x_p-x_{pj}))\} \qquad (5)\end{aligned}$$

where Π_p refers to the product from 1 to d and k_p is usually but not necessary a probability density function. For example, the normal kernel can be expressed as

$$k(u) = (2\pi)^{-1/2}\exp(-u^2/2) \qquad (6)$$

In some cases, the k are not confined to be non–negative.

In our empirical work, we usually divide each x_p by the individual standard deviation. Constraining the covariance matrix of x to be an identity matrix is important in empirical work for the kernel estimator to be operational. It is also difficult to justify the use of the method of kernels if each x_p is of different units of measurement. In the case of nonparametric regression discussed below, the Nadaraya–Watson estimator with product kernels and unit covariance matrix can be interpreted as an estimator using diagonal bandwidth matrix with p–pth element as [a x standard error (x_p)].

2.4 NADARAYA–WATSON KERNEL ESTIMATOR FOR CONDITIONAL MOMENTS

We shall return to our original problem of estimating the conditional moments. While there are a number of choices for kernel estimator of conditional moments, what we are about to present is known as the Nadaraya–Watson (N–W) estimator which originated independently from Nadaraya (1964) and Watson (1964). This estimator will be the most commonly used nonparametric technique in our work.

Let Y be a random variable and X be a dx1 vector of random variables. We consider the estimation of the regression function E[Y|X]. Extension to the second and higher moments, E[g(Y)|X], is straight forward. The joint density can be estimated by

$$\hat{f}(y,x) = (Na^{d+1})^{-1}\Sigma_j K(a^{-1}(x-x_j))K(a^{-1}(y-y_j)).$$

Using (2), the estimator for the marginal density is

$$\begin{aligned}\hat{f}_X(x) &= \textstyle\int_{-\infty}^{\infty} f(y,x)dy \\ &= (Na^d)^{-1}\Sigma_j K(a^{-1}(x-x_j).\end{aligned}$$

Using (3), the conditional density is estimated as

$$\hat{f}_{Y|X}(y|x) = \hat{f}_X(x)^{-1}\hat{f}(y,x)$$

Using (4), the estimator for the conditional moment is therefore

$$\begin{aligned}\hat{m}(x) &= \textstyle\int_{-\infty}^{\infty} y\, f_{Y|X}(y|x)dy = \hat{f}_X(x)^{-1}\int_{-\infty}^{\infty} y\, f(y,x)dy \\ &= (a\Sigma_j K(a^{-1}(x-x_j))^{-1}\Sigma_j K(a^{-1}(x-x_j)) \textstyle\int_{-\infty}^{\infty} yK(a^{-1}(y-y_j))dy\end{aligned}$$

By a change of variable, $u = a^{-1}(y-y_j)$, i.e., $y = (y_j+au)$ and $dy = adu$, we have

$$\hat{m}(x) = (\Sigma_j K(a^{-1}(x-x_j))^{-1}\Sigma_j K(a^{-1}(x-x_j)\textstyle\int_{-\infty}^{\infty} (y_j+au)K(u)du$$

The classical assumption in these literature is that the kernel K(u) integrates to 1 (see A(1a)). If we make the assumption (A2a) that $\int uK(u)du = 0$, we shall arrive at the N–W estimator

$$\begin{aligned}\hat{m}(x) &= (\Sigma_{j=1}^{N} K(a^{-1}(x-x_j))^{-1} \Sigma_{j=1}^{N} y_j K(a^{-1}(x-x_j)) \\ &= \Sigma_{j=1}^{N} (\Sigma_{t=1}^{N} K_t(u))^{-1} K_j(u) y_j \\ &= \Sigma_j K_j^* y_j \qquad (7)\end{aligned}$$

or in matrix form,

$$\hat{M}(x) = Ky.$$

where $K_j^* = (\Sigma_t K_t(u))^{-1}\ K_j(u)$, K is known as the smoothing matrix with K_{ij} elements and y is vector with elements y_i. Since $\hat{m}(x)$ is a linear combination of the y_j, the N–W estimator belongs to the class of linear estimators given a. If K_j^* is nonnegative and sum to one, it can be taken as a weighted average of y_j. In this case, it is not too difficult to understand how the N–W estimator works. If m(x) is a smooth function, then it is plausible that the observations close to x_i contain useful information about m at x_i. One would want to place more weight on the observations close to x_i and less or none of the weight on the observations further away from x_i. Since the N–W estimator is constructed by taking local average of the data close to x_i, it can be taken to be a local average estimator. The bandwidth therefore controls how wide the interval should be and so how many observations should be used for averaging.

To implement the estimator, we have to make several choices. One of these is the choice of the kernel. The most popular in applied work have been the normal density and uniform density. It is known from Monte Carlo results that the choice of kernel is not very important as long as more weight is given to observations nearer to x_i and less weight to observations further away. Notice that in the extreme case by choosing a large enough bandwidth and $K(u) = I(|u| \leq 1/2)$, we have $K(u) = 1$. We are in effect taking averages over all the observed y_j and end up with the sample mean $\hat{m} = N^{-1}\Sigma_j y_j$.

In some cases, in order to overcome technical difficulties, the N–W estimator is augmented to prevent the denominator from going to zero. The trimmed N–W estimator is defined to be

$$\hat{m}(x) = [I(\hat{f}_i \geq b)(\Sigma_j K(a^{-1}(x-x_j))^{-1}]\Sigma_j y_j K(a^{-1}(x-x_j)) \qquad (8)$$

where I is the usual indicator function and b is a user chosen trimming constant.

In other cases, in order to minimize the influence of outliers or for the construction of cross–validation or other automatic bandwidth selection criteria, we use the leave–one–out estimator

$$\hat{m}_{-i}(x) = (\Sigma_{j \neq i} K(a^{-1}(x-x_j))^{-1}\Sigma_{j \neq i} y_j K(a^{-1}(x-x_j)) \qquad (9)$$

In order to gain more insight into the estimators, we have to discuss the asymptotic properties. In particular, it is interesting to know under what regularity conditions are the estimators asymptotically unbiased, consistent and have a limiting normal distribution.

2.5 ASYMPTOTIC PROPERTIES OF DENSITY ESTIMATOR

First of all, let us discuss some of the asymptotic properties of $\hat{f}(x)$.

Theorem 5.1. (Asymptotic unbiasedness of $\hat{f}(x)$). Suppose

(i) $\sup_u |K(u)| < \infty$,

(ii) $\int_{R^d} |K(u)| du < \infty$,

(iii) $\lim_{|u| \to \infty} |u|^d K(u) = 0$,

(iv) $\int_{R^d} K(u) du = 1$,

(v) $\lim_{N \to \infty} a = 0$.

Then at every continuity point of x of f,

$$\lim_{N \to \infty} E[\hat{f}(x)] = f(x)$$

Proof: Cacoullos (1964)'s Theorem 3.1.

The result can be straightforwardly extended to the estimation of functional of y_i, i.e., $g(y_i)$ replaces y_i in the kernel density estimates. In that case, the estimate is asymptotically unbiased if $h(x) = E[g(y)|x]$ is continuous at x and E $h(x) < \infty$.

Theorem 5.2. (Mean square consistency of $\hat{f}(x)$). If (i) to (v) hold and

(vi) $\lim_{N\to\infty} Na^d = 0$.

Then at every continuity point x of f,

$$\lim_{N\to\infty} E[\hat{f}(x)-f(x)]^2 = 0.$$

Proof: Cacoullos (1964)'s Theorem 3.2.

The results can again be straightforwardly extended to the case of g(y) with additional regularity conditions on the moments of g(y) and h(x). Besides being asymptotically unbiased and consistent, the estimator has a limiting normal distribution.

Theorem 5.3. (Asymptotic Normality of $\hat{f}(x)$). If (i) to (v) are satisfied

Then at every continuity point x of f,

$$(Na^d)^{1/2}(\hat{f}(x)-f(x)) \sim N(0,\sigma^2)$$

where

$$\sigma^2 = f(x)\int K^2(u)du.$$

Proof: Cacoullos (1964) and Prakasa Rao (1983).

We have now centred around the true f as opposed to Parzen (1972)'s $E\hat{f}$. Therefore, one can now construct confidence intervals for f(x) at each point of x with the estimator of σ^2 as

$$\hat{\sigma}^2 = \hat{f}(x)\int K^2(u)du. \tag{10}$$

In fact, the density estimates are also known to be strongly consistent and uniformly strongly consistent. The last property is one of the measures of the global performance of the estimator.

2.6 CONDITIONAL EXPECTATION

There are a lot of results on the N–W estimator. Under some regularity conditions, the estimator possesses consistent properties and is asymptotically normally distributed. Let us present some results in more detail. Consider the more general estimators of $m(x) = E[g(x)|x]$, at distinct x_i, i=1,...N.

Theorem 6.1. If the conditions of Theorem 2 and for function of derivative hold, $\hat{m}(x)$ is mean square consistent.

Proof: Since both the numerator and denominator can be shown to be mean square consistent, we apply Slutsky's theorem to obtain the results.

Consider the N–W estimator as a ratio

$$\hat{m}(x) = (\Sigma_j K(a^{-1}(x-x_j))^{-1}\Sigma_j y_j K(a^{-1}(x-x_j))$$
$$= \hat{\delta}(x)/\hat{f}(x)$$

where $\hat{f}(x) = (\Sigma_j K(a^{-1}(x-x_j))$ (11)

$$\hat{\delta}(x) = \Sigma_j y_j K(a^{-1}(x-x_j)) \qquad (12)$$

Since the N–W estimator is a ratio, some authors have looked at the approximation of the ratios and obtained the asymptotic results. Let $w(x) = \int yf(x,y)dy$ and $v(x) = \int y^2 f(x,y)dy$. The following theorem demonstrates that the N–W and some other nonparametric regression estimators are asymptotically normally distributed.

Theorem 6.2. Let $x_1,...x_k$ be distinct point. If

(i)(a) $K(u)$ and $|uK(u)|$ are bounded,

(b) $\int uK(u)du = 0$, and

(c) $\int u^2K(u)du < \infty$,

(d) $\lim_{N\to\infty} Na^3 = \infty$ and $\lim_{N\to\infty} Na^5 = 0$.

(ii) $f(x_i) > 0$, $1 \le i \le k$,

(iii) $E|y|^3 < \infty$,

(iv) the derivative of v exists and is bounded, w and f are twice differentiable and bounded,then

$$(Na)^{1/2}[\hat{m}(x_1)-m(x_1),...,\hat{m}(x_k)-m(x_k)] \sim N_k(0,\Omega),$$

where Ω is a diagonal matrix with the ith element with

$$\Omega_{ii} = f^{-1}\text{var}[y_i|x_i]\int K^2(u)du, \qquad 1 \le i \le k.$$

Proof: Schuster (1972).

This suggests that interval estimates can be easily constructed for nonparametric regression estimates. We also know that $\mathrm{var}[y|x] = v(x)/f(x) - (w(x)/f(x))^2 = E[y^2|x] - \{E[y|x]\}^2$. It is therefore easy to suggest how to construct the estimate for the conditional variance since we can express the variance as two regression functions. Indeed, nonparametric variance estimates are useful for heteroscedastic problems.

2.7 DERIVATIVES OF CONDITIONAL EXPECTATION

In some of our problems, especially in those cases where we have to maximize or minimize a function with respect to the parameters of interest, we will have to consider the derivative of the N—W estimator. In particular, we have in mind the problem of constructing two step or linearized ML estimators. Suppose that m admits r derivatives and we wish to estimate the r_1th ($r_1 < r$) derivative.

Theorem 7.1 If the following conditions are satisfied,

(i) $E[y^2] < \infty$,

(ii) $f(x) > 0$,

(iii) $m(x)$ is r times differentiable,

(iv) If the characteristic function of K is ψ, then $\int |u|^r \psi(u)du < \infty$, i.e.,
$\psi(u) = \int \exp(iux)K(u)du$, where $i^2=-1$.

(v) $\lim_{|u|\to\infty} |uK(u)| = 0$,

(vi) $\lim_{N\to\infty} h = 0$,

(vii) f is continuous in [a,b],

(viii) $\int_c^d |f(u)|du < \infty$, for $-\infty \leq c \leq a \leq b \leq d \leq \infty$,

(ix) $\int |uK(u)|du$ is finite,

(x) $K^{(j)}(u)$ be a continuous function of bounded variation for j=0,1,..,r.

(xi) f and its first r+1 derivatives are bounded,

(xii) $\lim_{N\to\infty} a/\epsilon = 0$,

(xiii) K,f,g are r+1 times differentiable, then there is a constant C such that for any positive

$$\lim_{N\to\infty} P[\sup_{a\leq x\leq b}|\hat{m}^{(r)}(x) - m^{(r)}(x)| > \epsilon\,] \leq C(Na^{2r+2}\epsilon^2)^{-1}.$$

Proof: Schuster and Yakowitz (1974).

From the theorem, we can see that the rate of convergence depends on r. The problems that we will be concerned with only requires first derivative.

2.8 HIGHER ORDER KERNELS AND OPTIMAL BANDWIDTH

As we have mentioned above, it is sometimes desirable to relax the non–negative kernel and take advantage of the smoothness of m. Improved asymptotic rates of convergence can be attained via the use of a higher–order kernels. Of course, what we are about to describe is not the only way to reduce bias, the method of Jackknife (see Schucany and Sommers (1977)) has also been used by other authors in semiparametric models. In fact, it would be useful to introduce the higher order kernel of Barlett (1963).

Definition 8.1: Let f be ℓ times differentiable in the neighbourhood of x, kernel of order ℓ is to satisfy

(i) $\int u^i K(u)du = \delta_{i0}$, $\quad 0 \leq i \leq \ell-1$;

(ii) $\int |u|^{\ell} K(u)du \neq 0$;

(iii) $\mathrm{Sup}_u\,(1+|u|^{\ell+1})|K(u)| < \infty$;

where δ_{ij} is the Kronecker's delta. (i) ensures that we have enough zero moments. Sometimes, it may be desirable to relax (ii) and (iii) as in other higher–order kernel literature (see, Prakasa Rao (1983)).

Higher–order kernels are unlikely to be very popular in purely nonparametric models and indeed few empirical work has used the higher–order kernels because they can take negative values. However, as we shall see in later chapters, it is important in establishing root N consistency in some semiparametric problems. Higher–order kernels can usually be taken as functions whose lower moments are zeros. The following theorem will demonstrate the advantage of a higher–order kernels.

Theorem 8.1: Let f be ℓ times differentiable in the neighbourhood of x and K is a higher–order kernel given in Definition 1. Then

$$E\,\hat{f}(x) = f(x) + O(a^{\ell})$$

Proof: Robinson (1989).

This theorem shows that the bias decreases "sufficiently fast" with a. We have assumed that K(u) integrates to 1 and that we have a sufficiently smooth function with enough zero moments. Thus, it is easy to suggest a $K_\ell \in K$. With ℓ even and φ a even function, we can suggest each $k_p(u)$ as a product of $\Psi(u)$ and $\psi(u)$ where $\Psi(u)$ is an even polynomial in u, i.e.,

$$k(u) = \Sigma_{r=0}^{(\ell-2)/2} c_r u^{2r} \psi(u), \tag{13}$$

we can easily find c_r's which satisfy (i), e.g.,

(a) if $\psi(u) = 1/2\ I(|u|\leq 1)$, then the moments are

$$m_r = 0 \qquad \text{if r is odd;}$$

$$m_r = \int u^r \psi(u)du = [r+1]^{-1} \qquad \text{if r is even;}$$

(b) if $\psi(u) = (2\pi)^{-1/2}\exp(-u^2/2)$, then the moments are

$$m_r = 0 \qquad \text{if r is odd;}$$

$$m_r = \int u^r \psi(u)du = r![(r/2)!(2^{\ell/2})]^{-1} \quad \text{if r is even;}$$

substituting k(u) into (i), we will have a system of $(\ell-2)/2$ equations of the form

$$M\,c = d,$$

where

$$c = \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{(\ell-2)/2} \end{bmatrix} \qquad d = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$M = \begin{bmatrix} 1 & m_2 & m_4 & \cdots & m_{(\ell-2)} \\ m_2 & m_4 & \cdots & \cdots & m_\ell \\ \vdots & \vdots & & \ddots & \vdots \\ m_{(\ell-2)} & \vdots & \cdots & \vdots & m_{2(\ell-2)} \end{bmatrix}$$

Thus, there is little problem in finding c_r. Conditions (ii) and (iii) may be replaced by a stronger condition $K(u) = O((1+|u|^{\ell+1+\epsilon})^{-1})$, for some $\epsilon > 0$. When $r = 2$, we have the simple density estimates. Although the theorem holds for a variety of distributions for $K(u)$, it rules out Cauchy distribution because its moments do not exist.

To understand the bandwidth selection problem in higher–order kernels, let us look at the MSE. The MSE of a higher–order kernel can be expressed as :

$$\text{MSE}(\hat{f}) = \text{bias}(\hat{f})^2 + \text{var}(\hat{f})$$

where

$$\text{bias}(\hat{f}) = \int K(u)\{f(x-au)-f(x)\}du$$

$$\text{var}(\hat{f}) = N^{-1}[a^{-1}\int K(u)^2 f(x-au)du - (\int K(u)f(x-au)du)^2]$$

With the Taylor series expansion, given that au is small, we have

$$f(x-au) = f(x) - auf'(x) + 1/2\, a^2u^2f''(x) + \ldots.$$

But from the definitions of higher–order kernel, the first r–1 moments of the kernel is zero. This implies the moments of $\hat{f}$ are all zeros except the term associated with ar. Therefore, we are left with

$$\begin{aligned}\text{bias} &= \text{constant x } f^r(x) \int u^r K(u)du \text{ x } a^r \\ &= C_1 a^r\end{aligned}$$

Similarly,

$$\begin{aligned}\text{var}(\hat{f}) &= (Na)^{-1}\int\{f(x)-auf'(x) \\ &\qquad + 1/2\,(au)^2f''(x)-\ldots\}K^2(u)du + O(N^{-1}) \\ &\approx (Na)^{-1}f(x)\int K^2(u)du \\ &= C_2(Na)^{-1}\end{aligned}$$

We can now consider the "optimal bandwidth" $a_{opt} = \min_a$ MSE with

$$\text{MSE} = \{C_2(Na)^{-1} + (C_1a^r)^2\}.$$

Consider the use of simple calculus,

$$\partial\text{MSE}/\partial a = -C_2(Na^2)^{-1} + 2rC_1^2a^{2r-1} = 0$$

From this first order condition, we have

$$a_{opt} = C\,N^{-\alpha}$$

where $C = (C_2/(2rC_1^2))\alpha$, $\alpha = 1/(2r+1)$.

C is a function of f(x) since both C_1 and C_2 are functions of f(x). As mentioned above, the problem in obtaining the optimal a in practice is that f(x) is unknown! It appears that we may be able to work out the unknown constant C_1 and C_2 if we take the average a_{opt}'s corresponding to two different N's: N_1 and N_2. The results on optimal bandwidth can be extended to d–dimension multivariate density estimates straightforwardly. In the case of Cacoulous (1964)'s multivariate kernel, we have a_{opt} = constant x ($N^{-\alpha}$) where $\alpha = d + 2r$.

Since MSE_{opt} depends on $\int K^2(u)du$, Epanechnikov (1969) has gone further by finding the optimum kernel by minimizing $\int K^2(u)du$ subject to the constraints that (i) the kernel integrates to one, (ii) a symmetric kernel and (iii) $\int u^2K(u)du = 1$. This produces what is now known as the Epanechnikov kernel which is non–negative. Notice also that replacing (ii) by (ii)' $\int uK(u)du = 0$ allows for non–symmetric kernels. However, it has been known from Epanechnikov (1969) that there is very little difference in using different kernels and the choice of kernels should be based on computational and technical considerations. For these reasons, the uniform and normal kernels have been favoured in most applied studies and indeed our work in this thesis.

2.9 DEPENDENT OBSERVATIONS

In economics, we often deal with time series. It is therefore useful to understand some dependence conditions used in the study of the asymptotic properties of the estimators. We generally assume that the observed economic variables $\{W_t, -\infty<t<\infty\}$ is strictly stationary. Let M_a^b denotes the σ–algebra of events generated by $W_a,..W_b$, for $-\infty \le a \le b \le \infty$. We say W_t is

(i) Strong Mixing (SM) if

$$\alpha_j = \sup_{B \in M_{-\infty}^{0}, A \in M_j^{\infty}} |P(A \cap B) - P(A)P(B)| \to 0, \text{ as } j \to \infty.$$

(i) Absolute Regular (ARE) if

$$\beta_j = E\{\sup_{A \in M_j^{\infty}} |P(A|M_{-\infty}^{0}) - P(A)|\} \to 0, \text{ as } j \to \infty.$$

(iii) Uniform Mixing (UM) if

$$\psi_j = \sup_{B \in M_{-\infty}^{0}, A \in M_j^{\infty}} |P(A|B) - P(A)| \to 0, \text{ as } j \to \infty.$$

It is known that (iii) $\Rightarrow$ (ii) $\Rightarrow$ (i) since . $(\psi_j) \geq (\beta_j) \geq (\alpha_j)$.While there are asymptotic results under various stationary processes for the kernel estimators, some results for strong mixing have been provided in Robinson (1983). Under some regularity conditions, additional to the usual conditions imposed on the kernels, rate of convergence of a as well as the condition that W is SM, $\hat{f}(x)$ and $\hat{m}(x)$ have the same limiting distribution as in the case where x_i's are i.i.d.. However, he warned against placing too much faith in the estimates because the performance of the kernel estimates clearly depend crucially on the choice of bandwidth. It is perhaps not surprising that stronger conditions have to be imposed for the central limit theorem, what is surprising is that the same rate of convergence and covariance matrix are identical to the case of i.i.d.. The results thus justify the use of kernel estimation in the time series case. Recent work of Robinson (1986) discovered that the bandwidth chosen should be larger than that for i.i.d case.

2.10 OUTLIERS AND TAIL EFFECT

Unfortunately, the N–W estimator is sensitive to outliers in the data. Take the case of a normal kernel. If the point x_i lies far away from the rest of the observations with the exception of one or two points, the N–W estimate will be heavily influenced by these two points as the normal kernel places little weight on other observations further away. In the extreme case, where x_s is far away from the rest of the observations, $\Sigma_j K_j y_j / \Sigma_j K_j = y_s$. The same problem will occur if one uses too small a common bandwidth which controls the number of observations to be used for averaging. This is exactly the problem one faces in the case of density estimates when estimating the tails where there are only a few observations around.

There are solutions to these problems, such as using the leave–one–out estimator to exclude the ith observation when summing over the N observations in the density or N–W estimates. Another solution may be to use a variable bandwidth which may generate further problems in practice.

2.11 HIGH DIMENSIONALITY

Applying the kernel estimator to multivariate economic problems has its shortcomings. One serious problem with the kernel estimator is the "curse of dimensionality". This problem arises when we have a large number of explanatory variables as we often encounter in economic applications. The scarcity of data in a multivariate environment will generally make the kernel estimators undesirable. As we shall see from the theorem, the bias of the N–W estimator reduces with increased bandwidth. If one desires a small bias, then a smaller bandwidth should be used. Unfortunately, the variance of the estimate increases (variance is proportional to Na^d in the N–W estimator) as we increase the bandwidth. At the same time, the variance increases as we increase the dimension of our explanatory variables. Conceiving the kernel estimator as a local regression or average estimator, we can see that unless we have a large number of observations, in a very high dimension space, we are unlikely to find enough observations around x_i to give us a good enough estimate. This empty space phenomena associated with rapidly increasing variance with increasing dimension has been termed the "curse of dimensionality".

Another problem with multivariate application is interpretability. One will find it difficult to comprehend the relationship between dependent and independent variables. To visualize the relationship between the dependent variable and just two independent variables is not too difficult as a plot of three dimensional diagram will give some idea of the dependency. However, it becomes more of a problem when one has more than four dimensions. Failure to detect certain features may in fact be counter productive in using nonparametric regression.

From the theoretical point, there are always some assumptions to be made when one is dealing with multivariate nonparametric estimations. If the variables are highly correlated or are of different unit of measurements, it may be hard to justify the use of multivariate kernels. Furthermore, from the computational point of view, a multivariate kernel is very expensive to compute.

One of the solutions is to apply a preliminary dimension reduction technique so that we may additively approximate m(x). We may then apply the N–W estimator to the additive approximation, e.g., we may find

$$m(x) = \alpha + \Sigma_j T_j(x_j)$$

The discussion of the this type of additive approximation is presented in Stone (1985). One possibility is to find T_j to minimize

$$E[m(x) - \alpha - \Sigma_j T_j(x_j)^2]$$

subject to the zero mean constraint, i.e., $Em_j = 0$, j=1,..,d. One may let $\alpha = \bar{y} = N^{-1}\Sigma_j y_j$ and m_j be estimated by spline nonparametric estimators. Under some regularity conditions, Stone (1985) has shown that the estimator is consistent.

An alternative method of additive approximation is suggested by Friedman and Stuetzle (1981). Their backfitting algorithm allows for different smoothers for different m_j's, such as splines, kernel, nearest neighbour, etc., to be used. The procedure involves estimating the T_j holding all other T_j's constant. Assume that the current estimates are $\hat{T}_j$. We can update $\hat{T}_k$ by any nonparametric method smoothing e over x_j, where e is partial residual vector such that $e = y - \Sigma_{j\neq k}\hat{T}_j(x_j)$. Further improvement is attained by repeating the process.

Another method is the projection pursuit regression of Friedman and Stuetzle

(1981) which is a generalization of the methods mentioned above. The technique is to apply nonparametric regression on the linear combinations of the explanatory variables rather than on the individual variable. The main advantage over the previous methods being that multiplicative form, e.g., Cobb–Douglas function $Y = X_1X_2$, can be accommodated. More specifically, we have

$$m(x) = \alpha + \Sigma_j T_j(\beta_j' z_j)$$

which can easily handle interaction terms between the explanatory variables and thus, it is more flexible.

But most of the methods discussed above involve heavy computing and it will be difficult to justify their use as a preliminary dimension reduction technique for subsequent kernel estimation. A conceptually easy to understand and computational efficient method is outlined in a later chapter. In particular, we have

$$m(x) = \alpha + m(\Sigma_j \beta_j' z_j).$$

The model is discussed in Chapter 7. Various additive models have been proposed and applied in the statistical literature. For recent developments and further discussions, see Buja, Hastie and Tibshirani (1989).

CHAPTER 3

A SEMIPARAMETRIC HEDONIC APPROACH

3.1 INTRODUCTION

The use of hedonic price methods in transportation research has been less favoured in recent years for various reasons. The most important of these has been the focus of discrete choice modelling in transport economics. The advancement of computing facilities and econometric theory has allowed more elaborate models to be used (e.g. Boyd and Mellman (1980)). Discrete choice modelling makes good economic sense and is intuitively more appealing but some argue that it does not utilize as much market information as in the case of hedonic method.

Rosen (1974)'s seminal contribution has given the hedonic price function a firm theoretical foundation. His theory suggests that the price function reflects the points where marginal bids of buyers and marginal offers of sellers are equal. The hedonic price function thus represents the valuation of the attributes of the commodity jointly supplied by the sellers and purchased by the buyers in the market. His view has led to a lot of discussions of the "two–stage" estimation procedures in recent econometric literature which we will discuss in Chapter 4.

In the semiparametric literature, Stock (1985a) has suggested a semiparametric hedonic two–stage model but his method does not allow for discrete and dummy variables which are frequently encountered. In another paper, Stock (1985b)'s semiparametric method for valuation of benefits allows for the presence of dummy variables and has been applied in the next chapter.

In this chapter, we adopt Ohta and Griliches (1976)'s "two–stage" hypothesis in asserting that there are two kinds of car attributes: the first enters the utility function directly since the consumer derives utility from them; and the second enters only through the budget constraint. As noticed by Atkinson and Halvorsen (1984), the main difficulty, among others, in applying Rosen's two–stage procedure is the problem of multicollinearity among the attributes. This problem has led to estimates with the wrong signs, especially in estimating the implicit price of fuel efficiency, e.g., Goodman (1983). In order to avoid the multicollinearity problem, Atkinson and Halvorsen (1984) have constructed a static model and suggested the use of an augmented hedonic price equation which excludes fuel efficiency, and allows the efficiency to enter the consumer's optimization problem via the budget constraint.

In our case, a non–stochastic intertemporal model is constructed and the comparative static analysis of the rental market is conducted. The elasticities can be constructed from the system of nonhomogeneous equations. Identification of the parameters of the utility function is achieved via a log–linear augmented semiparametric hedonic price equation by imposing the homogeneity assumption.

One of the main contributions of this chapter is the use of MM in obtaining the estimates for transformation models. These techniques are used to estimate the rental cost function and the fuel efficient function which have little, if any, a priori economic information in formulation of the function forms.

Recently, Robinson (1988e) and Newey (1987) have proposed methods to construct efficient instruments for nonlinear system of equations of which transformation models is a special class. Robinson's method is parametric in the sense that one has to be able to obtain a closed form expression for the dependent variable and forms the instruments by taking sample averages. Newey's method is semiparametric in the sense that the efficient instruments and conditional variance (or variance–covariance matrix) are formed by nonparametric regression using nearest neighbours. In both cases, no specification of the distribution of disturbance term is required. We augment Robinson (1988e)'s method by forming the instrument with sample averages and adapting for heteroscedasticity using kernel regression. We require a smoothing number as a consequence of nonparametrically estimating the conditional variance which is a nuisance function. This bandwidth, which controls the smoothness of the nuisance function, is chosen automatically by Gaussian pseudo log–likelihood criterion. For completeness, we have also presented a semiparametric maximum likelihood (SML) method for the estimation of the heteroscedastic hyperbolic Box–Cox model.

The results for the elasticities are consistent with previous cross section and time series findings, and indicate that the petrol price elasticity of demand for fuel efficiency is close to unity and the own price elasticity of fuel is likely to be elastic.

The organization of the chapter is as follows: Section 2 introduces the economic model; Section 3 discusses the inverse hyperbolic sine and Box–Cox transformations and associated models; MM and ML method for these models are presented in Section 4; the results are presented and analyzed in Section 5; The problems associated with the identification of the taste parameters and construction of the elasticities are addressed in Section 6; finally, the conclusion is presented in Section 7.

3.2 A NON–STOCHASTIC MODEL OF UTILITY MAXIMIZATION

Our intention is to construct the petrol price elasticity of demand for attributes. Fuel efficiency with respect to the change in the price of petrol can be constructed easily from these price elasticities. Application of comparative static analysis to an intertemporal model will avoid the problems of multicollinearity and identification, which affect the traditional approach of hedonic studies. Let us consider a non–stochastic model of an individual making consumption plan over an infinite period. The individual consumer is assumed to have a well defined preference ordering over car attributes demand z and consumption C in each period. The consumer maximization problem can be written as:

$$\mathrm{Max}_{z,C}\ U(z_1, z_2, \ldots\ldots, C_1, C_2, \ldots\ldots)$$

subject to

$$\Sigma_{t=1}^{\infty} \frac{P_t C_t}{d_t} + \Sigma_{t=1}^{\infty} \frac{R(z) M_t}{d_t} + \Sigma_{t=1}^{\infty} \frac{P_{gt} M_t}{d_t E(z)} = Y_z + Y_c$$

where z_t = a mx1 vector of attributes in year t

C_t = consumption of other goods in year t

P_t = price of consumption good in year t

P_{gt} = price of petrol in year t

R = rental cost per mile/ running and maintenance cost per mile

M_t = miles travelled in year t

E = car efficiency

$d_t = (1+r)^t$ = discount factor in year t

r = real interest rate

Y_z = budget allocation to car ownership

Y_c = budget allocation to other consumption

The model allows for lending and borrowing at the market rate of interest r. The crux of the model is that the individual faces only a single budget constraint, which is a function of car efficiency, running and maintenance cost. These variables are in turn functions of car attributes. Here, we employ the hedonic hypothesis that car prices or costs are a function of their characteristics. The other basic maintained hypothesis, which has been employed in Atkinson and Halvorsen (1984) and Ohta and Griliches (1986), is that there are two kinds of car characteristics: the first kind of attributes, namely $z = (z_1, z_2, \ldots z_m)$, enters the utility function directly; the second kind, namely, E and R, enters the utility function indirectly. Those physical characteristics which enter the utility function directly are attributes such as specifications, dimensions and performance variables of the car desired by the consumer. The latter hypothesis has been termed "Two–Stage" hypothesis by Ohta and Griliches (1976). At this point, we make the following assumptions on U:

(i) U is weakly separable in car attributes (z) and all other consumption goods (C);

(ii) U is separable over time and $U = \Sigma_t U/s_t$, where $s_t = (1+\rho)^t$, and ρ is the nonzero rate of time preference.

The assumptions permit us to write the subutility maximization problem as:

$$\mathrm{Max}_z \, \Sigma_{t=1}^{\infty} \frac{U(z)}{s_t}$$

subject to

$$\Sigma_{t=1}^{\infty} \frac{R(z)M_t}{d_t} + \Sigma_{t=1}^{\infty} \frac{P_{gt}M_t}{d_t E(z)} = Y_z$$

The Lagrangian function is

$$L = S\, U(z) + \ell\,[Y_z - G_1\, R(z) - G_2\, E(z)^{-1}]$$

where $S = \Sigma_t(1+\rho)^{-t}$; $G_1 = \Sigma_t M_t/d_t$; $G_2 = \Sigma_t P_{gt}\, M_t/d_t$

which gives the following m+1 first order conditions:

$$S\, U_i + \ell\,[G_1 R_i - BE_i] = 0, \qquad i=1,2,\ldots, m \qquad (1)$$

$$[Y_z - G_1 R(z) - G_2 E(z)^{-1}] = 0$$

$$B = G_2 E^{-2}$$

The first m equations are just the usual marginal conditions which state that marginal utility is equal to marginal cost, where marginal cost is the sum of marginal rental cost and marginal petrol expenditure. The (m+1)th equation is just the budget constraint. Given these first order conditions, we can proceed with our analysis the effect of a change of the price of petrol on the attributes by using comparative static. Differentiate with respect to the base period price P_0, we have

$$\Sigma_j[S\, U_{ij} - \ell\{-G_1 R_{ij} + [BE_{ij} - 2BE^{-1}E_j E_i]\}]\ \partial z_j/\partial P_0$$

$$[G_1 R_i - BE_i]\ \partial\ell/\partial P_0 = -\,\ell B_p E_i \qquad i=1,2\ldots,m$$

$$\Sigma_j[G_1 R_j - BE_j]\ \partial z_j/\partial P_0 = -\,B_p E$$

and in partitioned matrix form

$$\begin{bmatrix} A & a \\ a & 0 \end{bmatrix} w = b$$

where

$$A_{(ij)} = [S\, U_{ij} - \ell\{-G_1 R_{ij} + [BE_{ij} - 2BE^{-1}E_j E_i]\}]$$

$$a(i) = [G_1 R_i - BE_i]$$
$$b = -(\ell B_p E_1, \ell B_p E_2, .., \ell B_p E_m, B_p E)'$$
$$w = (\partial z_1/\partial P_0, \partial z_2/\partial P_0, ..., \partial z_m/\partial P_0, \partial \ell/\partial P_0)'$$
$$B_p = -\partial G_2/\partial P_0 \, E^{-2}$$

The negative of the partial derivative of total present value of petrol expenditure with respect to fuel efficiency, $-B$, can be interpreted as the marginal benefits of increased fuel efficiency. The derivative of B with respect to P_0, $\partial B/\partial P_0$, is the effect of marginal benefit of a change in the base period price, P_0. The comparative static equations form a nonhomogeneous system of equations and therefore one can solve numerically for the m+1 unknowns. We need, however, to specify the cost as well as the fuel efficiency function to obtain the estimates before we can construct the petrol elasticities of attribute with the help of the elements in w. We can also construct the elasticity of demand for fuel efficiency with respect to the change in the price of petrol which is of interest, i.e.,

$$\epsilon_d = \frac{\partial \log E}{\partial \log P_0} = \Sigma_j \frac{P_0}{E} \frac{\partial E}{\partial z_j} \frac{\partial z_j}{\partial P_0}$$

There are two implicit simplifying assumptions. First of all, miles driven is exogenously given. It is equivalent to the alternative assumption that M_t is endogenous and dependent on z and E(z), and that the increase in miles driven caused by the increase in an attribute is entirely offset by the decrease in fuel efficiency brought about by the increase in the attribute, i.e.,

$$\frac{\partial M}{\partial z_j} = \frac{\partial M}{\partial E} \frac{\partial E}{\partial z_j}$$

Secondly, we rule out technological improvement in fuel efficiency. In other words, fuel efficiency improvements are brought about strictly by decreasing at least one of the attributes desired by the individual.

The following sections will discuss the appropriate procedures for estimating the rental cost function, R(z), and the fuel efficiency function E(z). In the absence of a priori information on the functional form, a new class of heteroscedastic transformation models is introduced. But some discussions on the Box–Cox transformation model are presented to motivate the use of the new procedures.

3.3 TRANSFORMATION MODELS

There is no economic information to assist one in choosing the appropriate functional form for the cost and fuel efficiency functions. Most parametric methods will be too restrictive and unlikely to be consistent in the presence of slight misspecifications. Purely nonparametric methods are unsuitable for multivariate applications unless one has a very large data set. The usual approach is to employ robust techniques where flexibility of functional form is allowed.

However, the presence of dummy variables in cross sectional studies does not allow the use of some of the interesting and relevant techniques which allow the parameter to be consistently estimated up to a scalar multiple, e.g., Stoker (1986) and Powell, Stock and Stoker (1986).

Transformation models are favoured by many applied economists for these models embody most flexible functional forms. Let us define some notations: the summation sign, Σ_j refers to the summing from attribute 1 to m and Σ_i refers to the summing from observation 1 to N throughout the chapter.

There are four classes of transformation models:

(1) Single transformation model

$$T(y_i;\mu) = c + \delta D_i + \Sigma_j \gamma_j z_{ji} + u_i \qquad i=1,2,\dots,N$$

with $E(u_i|z_i) = 0$ and $E(u_i^2|z_i) = \sigma^2$. T is a transformation to be defined and μ is the transformation parameter, δ is a scalar parameter, and γ_j's are parameters to be estimated. D is a dummy variable and z_j's are attributes.

Single transformation model refers to the transformation of only the dependent variable. This has been an area of interest for theoretical research for many years. It originated from the biostatistical literature (see, e.g., Curtiss (1943)) and recent attention has been on nonparametric estimation. The most commonly used transformation model in applied economics has been the Box–Cox model which is discussed below.

(ii) Double transformation model

$$T_1(y_i;\mu_1) = T_2([c + \delta D_i + \Sigma_j\gamma_j z_{ji}];\mu_2) + u_i \qquad i=1,2,\ldots,N$$

with $E(u_i|z_i) = 0$ and $E(u_i^2|z_i) = \sigma^2$. T_1 and T_2 are transformations to be defined and the μ_j's are the transformation parameters. The double transformation model has only appeared recently in empirical work (See Carroll and Ruppert (1988) and references thereafter). It refers to models which have both the dependent variable and the mean transformed. This model is of interest if only one can justify that $y = f(D_i,z_1,..,z_m)$ with f linear, and wish to correct for skewness (induce normality) and/or correct for heteroscedasticity.

(iii) Multiple transformation model

$$T_1(y_i;\mu) = c + \delta D_i + \Sigma_j\gamma_j T_2(z_{ji};\mu_j) + u_i \qquad i=1,2,\ldots N$$

with $E(u_i|z_i) = 0$ and $E(u_i^2|z_i) = \sigma^2$. T_1 and T_2 are transformations to be defined and the μ and the μ_j's are the transformation parameters. This model is also of great interest to economists for reasons discussed below. Here we have each z_j transformed either by (a) the same μ as y_i; (b) the same μ for all other z_i; or (c) different μ.

(iv) Quadratic transformation model

$$T_1(y_i;\mu) = c + \delta D_i + \Sigma_j\gamma_j T_2(z_{ji};\mu_j)$$
$$+ \Sigma_k\Sigma_{\ell\neq k}\,\gamma_{k\ell}T_2(z_{ki};\mu_k)T_2(z_{\ell i};\mu_\ell) + u_i \qquad i=1,2,\ldots N$$

The quadratic transformation model (QTM) is highly parameterized and involves the cross product terms. This approach is now favoured in many empirical studies. The Gaussian Box–Cox QTM is without doubt, the most used (mis–used) models in economics.

For our purpose, we define a class of heteroscedastic multiple transformation model which takes the form

$$T_1(y_i;\mu) = c + \delta D_i + \Sigma_j\gamma_j T_2(z_{ji};\mu_j) + u_i \qquad i=1,2,...N \ (2)$$

with $E(u_i|z_i) = 0$ and $E(u_i^2|z_i) = \sigma(z_i)^2$. T_1 and T_2 are transformations to be defined and μ and μ_j's are the transformation parameters. The rental cost and fuel efficiency models can be described by

$$T_1(R_i;\lambda) = c_1 + \delta_1 D_i + \Sigma_j\gamma_j T_2(z_{ji};\lambda_j) + u_{1i} \qquad i=1,2,...N \quad (1)$$

$$T_1(E_i;\psi) = c_2 + \delta_2 D_i + \Sigma_j\pi_j T_2(z_{ji};\psi_j) + u_{2i} \qquad i=1,2,...N \quad (2)$$

with $E(u_{1i}|z_i) = E(u_{2i}|z_i) = 0$ and we are not ruling out heteroscedasticity through–out the discussions.

Of course, as Curtiss (1943), Barlett (1947) and others have noted, transformation can stabilize the variance. For example, if $E[y_i] = W_i$, $var(y_i) = \sigma_i = \sigma f(W_i)$ and $var(T(y_i)) = \sigma_i^*$, then by approximation using Taylor series expansion

$$\begin{aligned}
\sigma_i^* &= E[T(y_i)-ET(y_i)]^2 \\
&\approx E[T(y_i)-T(W_i)]^2 \\
&\approx \{(\partial T/\partial W_i)E[y_i-W_i]\}^2 \\
&= \{(\partial T/\partial W_i)\sigma_i\}^2 \\
&= \{(\partial T/\partial W_i)\sigma f(W_i)\}^2 .
\end{aligned}$$

Therefore, if $\partial T/\partial W_i$ = constant/f(W_i), then σ_i^* is constant. However, this is not always possible. If heteroscedasticity remains, a heteroscedastic model is more appropriate.

The attributes we have included to capture the essential elements are engine size, spaciousness, power and acceleration. The dummy D_i (denoted by $D_{U,i}$) is included to give some idea of the effect of introducing cars which can use unleaded petrol without adjustment on its cost and fuel efficiency, as there have been some recent discussions and controversies surrounding the use of unleaded petrol. Our discussions below centre around various transformation models.

3.3.1 HETEROSCEDASTIC BOX–COX (HBC) TRANSFORMATION MODEL

The scaled heteroscedastic Box–Cox (HBC) transformation model is obtained by replacing T_1 and T_2 in (2) by T_{BC}:

$$T_{BC}(y_i;\mu) = c + \delta D_i + \Sigma_j \gamma_j T_{BC}(z_{ji};\mu_j) + u_i \qquad i=1,2,\ldots N$$

where

$$T_{BC}(Z;\lambda) = I(\lambda \neq 0)(Z^{\lambda} - c^{\lambda})/\lambda c^{\lambda}) + I(\lambda=0)\text{Log}(Z/c)$$

I is the usual indicator function, c is the scaling constant which depends on Z, and λ is a transformation (or scaling parameter) to be estimated. The scaling constant is usually employed in empirical work (see Dagenais, Gaudry and Tran (1987)) in order to avoid numerical problems.

Notice that if $\lambda = 0$, then the transformation is scale invariant while the usual approach to deal with the transformation when $\lambda \neq 0$ is to multiply the coefficients by the appropriate scale in order to reflect the units of measurement (except under the inappropriate MLE framework which will be discuss below). We can take c to be 10 to the power of $[\log_{10} N^{-1}\Sigma_i Z_i]$, where $[Z^*]$ is the largest integer less than or equal to Z^*. Some alternative definitions for scaled transformation families can be found in Schlesselman (1971).

For various reasons, the Box–Cox transformation has received special attention in economics. Despite the original intentions of Box and Cox (1964) for transformation to induce normality, constancy of the error variance, independence of the observations, and additivity of the independent variables, the Box–Cox model has been extended to accommodate all the violations of these assumptions for valid reasons. In particular, the quadratic Box–Cox model listed in (iv) above with $\delta = 0$, embodies most of the functional forms, e.g., translog ($\mu = \mu_j = \mu_{k\ell} = 0$), quadratic ($\mu = \mu_j = \mu_{k\ell} = 1$), log–linear ($\mu = \mu_j = \gamma_{k\ell} = 0$), linear ($\mu$ and all $\mu_j = 1$, $\gamma_{k\ell} = 0$), semi–log ($\mu = 0$, $\mu_j = 1$, $\mu_{k\ell} = 0$), inverse log ($\mu = 1$, $\mu_j = 0$, $\gamma_{k\ell} = 0$), generalized square root quadratic ($\mu = 1$, $\mu_j = \mu_{k\ell} = 2$), generalized Leontief ($\mu = 1/2$, $\mu_j = \mu_{k\ell} = 1$), generalized Leontief with homogeneity imposed ($\mu = 1/2$, $\mu_{k\ell} = 1$, $\gamma_j = 0$), etc.

This is accepted as the model to use in various areas of applied microeconomics (Halvorsen and Pollakowski (1981)) though it is not without criticisms (Cassel and Mendelsohn (1985)). Extensions to heteroscedasticity and serial correlation within the ML principle have also been discussed and applied (see, e.g., Gaudry and Dagenais (1979), Lahiri and Egy (1981), Savin and White (1978) and others).

A desirable estimation procedure for the Box–Cox model is MM. Although MM is the more appropriate technique, the most commonly used technique is the so called Gaussian "maximum likelihood" which suffers from various defects discussed below. Most applied work in urban economics failed to mention any of the following defects when using the likelihood maximand.

First of all, while it is true that ML estimation is first order efficient, it is known that the disturbance term cannot be normally distributed, as the dependent variable has to be positive. There are various modifications, such as Bickel and Docksum (1981)'s approach to induce normality, or the approach of defining a proper ML function (e.g., Gamma distribution as assumed in Amemiya and Powell (1981) or truncated normal as in Poirier (1978)). However, these have not been adopted in empirical work despite the Monte Carlo evidence of Amemiya and Powell (1981) that MM out–performs the ML method in many cases, even when the true Gamma error distribution is known.

While some (e.g. Dagenais et. al. (1987)) acknowledged that the so called "maximum likelihood" function is a maximand rather than a Gaussian likelihood function, none of the applied econometric work defined a proper likelihood or devoted any attention to the issue of consistency. One argument has been that all economic variables are some sort of limited dependent variable and it is rarely in economic applications that the dependent variable admits non–positive values. Draper and Cox (1969)'s result has also been quoted frequently in support of the fact that the if the disturbance distribution is nearly symmetric, the bias is negligible. Many would conclude that the impossibility of normality is only a theoretical curiosum and of no practical significance. However, further observations described below remind us that more care should be taken when Gaussianity is assumed.

The second point which one needs to bear in mind is that the Gaussian ML estimator for models which are nonlinear in the dependent variable is unlikely to be consistent when u is in fact not normally distributed (Amemiya (1977)). In this respect, one should therefore use robust methods which can at least retain consistency if not efficiency.

The third point is related to the estimation of the variance–covariance matrix and statistical inference. The true asymptotic variance–covariance matrix from the method mentioned above is not the expression derived from the usual ML method, for the fact that

$$\lim T^{-1}E(\partial^2 \mathrm{LogL}/\partial\theta\partial\theta') \neq \lim T^{-1}E(\partial \mathrm{LogL}/\partial\theta \; \partial \mathrm{LogL}/\partial\theta'),$$

where $T^{-1}E(\mathrm{LogL})$ is the maximand and θ is a vector of unknown parameters. It is in fact equal to

$$\lim T[E(\partial^2 \mathrm{LogL}/\partial\theta\partial\theta')]^{-1}E(\partial \mathrm{LogL}/\partial\theta \; \partial \mathrm{LogL}/\partial\theta')[E(\partial^2 \mathrm{LogL}/\partial\theta\partial\theta')]^{-1}$$

evaluated at the estimates $\tilde{\theta}$ obtained using the maximand (Amemiya (1985)). Thus, one would assume that the standard errors reported in the literature are incorrect. Many empirical studies also ignore the fact that Gaussianity is logically impossible and proceed to discuss the selection of functional form in the classical ML framework. The test statistic so widely used in the literature is not in fact the likelihood ratio test statistic.

In view of these observations, we argue that MM, which is asymptotically efficient within a class of models with conditional moment restrictions, should be used in Box–Cox estimation. However, there are other transformation models which may be useful to economic applications and two such models are described below.

3.3.2 HETEROSCEDASTIC INVERSE HYPERBOLIC SINE (HIHS) TRANSFORMATION MODEL

Replacing T_1 and T_2 in (2) by T_S gives the heteroscedastic inverse hyperbolic sine (HIHS) model

$$T_S(y_i;\mu) = c + \delta D_i + \Sigma_j \gamma_j T_S(z_{ji};\mu_j) + u_i \qquad i=1,2,\ldots N$$

where

$$T_S(Z;\lambda) = I(\lambda \neq 0)\lambda^{-1}\sinh^{-1}(\lambda Z/c) + I(\lambda=0)(Z/c)$$

or

$$T_S(Z;\lambda) = I(\lambda \neq 0)\mathrm{Log}(\lambda Z+\sqrt{a})\lambda^{-1} + I(\lambda=0)(Z/c)$$

where $a = (1+(\lambda Z)^2)$. The $\sinh^{-1}$ transformation first appeared in Beall (1942) on the square root transformation literature and later discussed by Curtiss (1943), Johnson (1949) and many others. Johnson (1949) was the first to mention T_S, i.e., the use of λZ rather than $\sqrt{(\lambda Z)}$ as in previous studies. Recent interest on ML estimation has been shown in Burbidge, Magee and Robb (1988), MacKinnon and Magee (1989). The advantage is that the transformation applies to all real values and it embodies both the level and logarithm as special cases. Gaussian ML function is not ruled out but MM is the preferred procedure because of its robustness. However, this model is more prone to multicollinearity problem when the independent variables are transformed. Notice that the transformation is applied to the product of the scaling parameter λ and Z, thus causing problem when λ is very small or large. The following proposed model which combines both Box–Cox and $\sinh^{-1}$ transformation has advantages over the HIHS model.

3.3.3 HETEROSCEDASTIC HYPERBOLIC BOX–COX (HHBC) TRANSFORMATION MODEL

A desirable transformation model for many economic applications is the following:

$$T_S(y_i;\mu) = c + \delta D_i + \Sigma_j \gamma_j T_{BC}(z_{ji};\mu_j) + u_i \qquad i=1,2,...N$$

which is a model with both hyperbolic and Box–Cox transformations. This can be termed heteroscedastic hyperbolic Box–Cox (HHBC) transformation model. There are various advantages to this model. First of all, like the HIHS model, the transformation can be applied to all real values of the dependent variable. Secondly, the distribution of the u can be Gaussian and the usual classical theory on Gaussian ML can be applied if one desires. Finally, we can have more efficient instruments by employing Robinson's method. Although Robinson's method can also be applied to the HBC model and may indeed produce more efficient estimates when it works, the method will most certainly fail if the γ's are mostly negative. We will return to this point in a later section.

Thus, we have proposed a model which retains the desired properties of IHS transformation and yet less prone to multicollinearity. Both MM and ML method can be used for estimation when the dependent variable is transformed by IHS. However, as we have mentioned before, ML is unlikely to be robust and it should be standard practice to produce MM estimates together with the ML estimates. Let us present MM and ML method.

3.4. METHOD OF MOMENTS AND MAXIMUM LIKELIHOOD METHOD

We proceed to discuss MM in a general set up. The conditional moment restriction of a transformation model is

$$E[u(x_i,\theta)\,|\,z_i] = 0$$

where u can be a vector with $E[u_i u_i'\,|\,z_i] = \Omega$, $x_i = (y_i,z_i')'$, $z_i = (c, D_i, T_2(z_{1i},\mu_1), .., T_2(z_{mi},\mu_m))'$ and $\theta = (c, \delta, \mu_1,..,\mu_m, \mu)'$. The MM estimator of θ can be obtained

as $\hat{\theta}^{*} = \operatorname{argmin}_{\theta \in \otimes} [\Sigma_i V(z_i)u(x_i,\theta)]' A_N [\Sigma_i V(z_i)u(x_i,\theta)]$ (4)
where $V(z_i) = \Omega^{-1}S_i$, $A_N = (\Sigma_i S_i \Omega^{-1} S_i)$, $S_i = E[D(x_i,\theta)|z_i]$, $D(x_i,\theta) = \partial u(x_i,\theta)/\partial\theta'$. The method which can achieve the asymptotic variance–covariance bound E(A) for the model is in fact the best nonlinear three–stage least squares estimator. However, this is not feasible because S_i is generally unknown unless one is willing to assume a particular distribution for u. In that case, MLE may be preferred although the MM is still more robust.

There have been suggestions for a feasible version of (4). Newey (1987) has suggested using nonparametric regression by nearest neighbours because S_i is in fact a regression function. However, this method is unlikely to work well in our case because of the dimension of z_i relative to the finite sample that we have. Another way of approximating $S(u_i,\theta)$ is to use the sample analogue, i.e., approximate S_i by N^{-1} $\Sigma_{j=1}^{N} D(x_i,u_j,\theta)$ from simulating u N times independently. This is practically cumbersome and u nonlinearity in the dependent variable poses problems in suggesting a justifiable distribution for u though it is not impossible.

Recently, Robinson (1988e) has suggested using the u_j obtained from first stage consistent estimates. Robinson's method may be more efficient in small sample than that of nonparametrically estimating the nuisance parameters $\tilde{S}_i$, because the method makes use of the entire sample.

The problem of heteroscedasticity and the functional form of the conditional variance, $\sigma_i^2 = E[u_i|z_i] = g(z)$, which is unlikely to be known a priori, can be overcome by using nonparametric methods. There are usually some hints on the variables which will affect σ, but it is generally difficult to suggest the choice of the functional form. In such instances, nonparametrically estimating g may be one way to avoid guessing at the functional form of heteroscedasticity. In particular, we resort to nonparametric kernel estimation. The discussion of kernel nonparametric regression is presented in Section 4.4.

It is natural to consider the case that σ is an unknown function of z, e.g., σ is dependent on z via the mean, W_i. Though it may be a bit restrictive, it is computationally valuable and should register some improvement in efficiency in finite samples because of the reduction in dimension. Of course, there are devices to reduce z to a manageable dimension (see Lee (1988) or Chapter 7) for nonparametric estimation but we avoid using them here.

Our strategy is to extend Robinson's approach to take into account heteroscedasticity. We obtain the consistent estimates in the first stage, and then use the consistent estimates to obtain efficient estimates via Robinson's method taking into account heteroscedasticity. To induce homoscedasticity, consider the following model:

$$T_1(y_i;\mu)/\sigma_i = c_1/\sigma_i + \delta_1 D_i/\sigma_i + \Sigma_j \gamma_j T_2(z_{ji};\mu_j)/\sigma_i i + \nu_i$$

with $\nu_i = u_i/\sigma_i(z)$,

$$\sigma_i^2 = \mathrm{var}(u_i\,|\,z_i) = g(W_i),$$

$$W_i = (c + \delta D_i + \Sigma_j \gamma_j T_2(z_{ji};\mu_j)).$$

We estimate the rental cost function and fuel efficiency function separately rather than as a system for various reasons. First of all, in view of the fact that automatic bandwidth selection is used, estimating the equations as a system is ambitious and not practical. Secondly, we wish to guard against incorrect choice of bandwidth and avoid estimating the covariance nonparametrically. Finally, the rental cost function and fuel efficiency function may be of interest individually. By not estimating the equations as a system, we may be sacrificing efficiency but we do have more robust estimates.

3.4.1 CONSISTENT INSTRUMENTAL VARIABLE ESTIMATES

The problem with the transformation model is in finding the most efficient instruments for the last element of θ. For consistency, there is not too much of a problem as the choices of instruments are vast. For the transformation parameters of the independent variables, the problem also does not exist and the instruments can be formed straightforwardly. Notice that for the Box–Cox model,

$$\partial u/\partial\mu = \mu^{-2}y^{\mu}\mathrm{Log}(y^{\mu}) - \mu^{-2}(y^{\mu}-1)$$

and for the hyperbolic model,

$$\partial u/\partial\mu = y\mu^{-1}(1+(\mu y)^2)^{-1/2} - \mu^{-2}\sinh^{-1}(\mu y).$$

As discussed above, the difficulty is in finding efficient instruments for μ as we have to deal with the conditional expectations of $\partial u_1/\partial\mu$ on z. The number of instruments can be greater than the dimension of θ, and it has been suggested that the squared or cross product of z_i be used. However, this is unlikely to work well in practice because multicollinearity is often encountered. Therefore, it may be important to limit the number of instruments and in our case, we have limited the number to 2m+3.

We may resort to linear projection of $\partial u/\partial\theta$ on instruments Z and form the instruments for μ, i.e.,

$$S_i = Z_i(\Sigma_{j=1}^{N} Z_j'Z_j)^{-1}(Z_i'\partial u/\partial\theta)$$
$$Z_i = (z_i', \partial u_i/\partial\mu_1,.., \partial u_i/\partial\mu_m, W_i^2)$$

One iterative scheme for consistent estimates is

$$\theta^{(k+1)} = \theta^{(k)} - \tau\tilde{\Phi}^{-1}\tilde{\Psi}$$

where

$$\tilde{\Phi} = (N^{-1}\Sigma_i\tilde{S}_i'\tilde{S}_i)$$
$$\tilde{\Psi} = (N^{-1}\Sigma_i\tilde{S}_i'\tilde{u}(x_i,\theta)))$$
$$\tau = \text{step length}$$

and the terms with ˜ are evaluated at $\theta^{(k)}$ which means that iterated instruments (i.e. the instruments are updated in each iteration) are used. The step length is obtained by using a combination of stepback and golden line search. As noticed by various authors (see Robinson (1988f)), theoretically, golden line search is preferred, but in practice, it is too time consuming. The strategy to use in empirical work is generally to use the stepback (inexact search) initially, i.e., by taking "backsteps" until an acceptable value is found. When it fails after several times to reduce the objective function, then golden search (exact search) is called upon to perform the task. The starting values are important in obtaining the estimates. Near collinearity is hard to detect and in general will still give a very low value of the objective function. Therefore, it is a good practice to check for near collinearity by examining the transformed data.

Fortunately, economic theory provides some a priori information on the sign of the parameters. For example, good starting points will be those values obtained with μ and all the μ_i's of BC model set equal to zeros, i.e., from the log–linear model. We then start the iteration with different values of μ and μ_i's between 0 and 1 to ensure that the converged values do give the minimum. We denote the NL2SLS consistent estimates that we obtained here by $\bar{\theta}$.

3.4.2 EFFICIENT INSTRUMENTAL VARIABLE ESTIMATES

Having obtained the consistent estimates, we seek to improve the efficiency of the estimates. While we are not assuming homoscedasticity, the following iterative scheme is employed to obtain the efficient estimates, $\hat{\theta}$,

$$\theta^{(k+1)} = \theta^{(k)} - \tau \bar{\Phi}^{-1}\bar{\Psi}$$

where

$$\bar{\Phi} = (N^{-1}\Sigma_i\bar{S}_i'\bar{S}_i)$$

$$\bar{\Psi} = (N^{-1}\Sigma_i\bar{S}_i'\bar{\nu}(x_i,\theta)))$$

$$\bar{S}_i = \bar{Z}_i^{*}\ (\Sigma_{j=1}^{N}\bar{Z}_j^{*}{}'\bar{Z}_j^{*})^{-1}\Sigma_i(\bar{Z}_i^{*}{}'\partial\bar{\nu}_i/\partial\theta)$$

$$\partial\bar{\nu}_i/\partial\theta = \bar{\sigma}_i^{-1}\partial\bar{u}_i/\partial\theta$$

$$\bar{Z}_i^{*} = (z_i\bar{\sigma}_i^{-1},\ \partial\bar{u}_i/\partial\mu_1\ \bar{\sigma}_i^{-1},\ldots,\partial\bar{u}_i/\partial\mu_m\ \bar{\sigma}_i^{-1},\ h_i)$$

$$\bar{W}_i^{*} = W_i\ \bar{\sigma}_i^{-1}$$

$$\tau = \text{step length}$$

$$\bar{\sigma}_i^{2} = \bar{E}(u_i^{2}\,|\,z) = \bar{g}(W_i)$$

For HHBC and HIHS, we have

$$\bar{h}_i = N^{-1}\Sigma_{j=1}^{N} \{\bar{\mu}^{-2}\tanh(\bar{\mu}(\bar{W}_i^* + \bar{\nu}_j) - \bar{\mu}^{-1}\bar{W}_i^*\}$$

For HBC, we have

$$\bar{h}_i = N^{-1}\Sigma_{j=1}^{N} \{1 + \bar{\mu}(\bar{W}_i^* + \bar{\nu}_j))\bar{\mu}^{-2}\log(1 + \bar{\mu}(\bar{W}_i^* + \bar{\nu}_j)) - \bar{\mu}^{-1}\bar{W}_i^*\}.$$

The instruments here are fixed instruments evaluated at the consistent estimates, $\bar{\theta}$, obtained from the first stage and does not need to be updated. This is in fact the GLS analogue for nonlinear least squares, i.e., nonlinear generalized two stage least squares (NLG2SLS).

If the form of heteroscedasticity is known, then we can substitute the parametric estimates for $\bar{\sigma}_i^2$. If the parametric specification of the variance is unknown, it is natural to estimate the conditional variance nonparametrically by kernel estimator.

Newey (1987) has suggested estimating the instruments nonparametrically. Under some regularity conditions in his case, $\sqrt{N}(\hat{\theta} - \theta) \sim N(0, \sigma^2\phi^{-1})$, where $\sigma^2 = E[\nu_i^2 | z_i]$. We cannot do better than the lower bound $\sigma^2\phi^{-1}$ unless we have further restrictions (see Chamberlain (1987)). We have not worked out the regularity conditions in our case but it is reasonable to conjecture that $\hat{\theta}$ is asymptotically efficient.

An estimator for the asymptotic variance–covariance matrix can be estimated as

$$\hat{V} = \hat{\sigma}^2\bar{\phi}^{-1}$$

where $\hat{\sigma}^2 = N^{-1}\Sigma_i\hat{\nu}_i^2$ and it is possible to have inferences which are valid and almost optimal with respect to the bound $\sigma^2\phi^{-1}$ with a large sample size N. If one wishes to protect oneself against a poor choice of "bandwidth", a heteroscedasticity–robust covariance matrix estimator for $\sqrt{N}(\hat{\theta} - \theta)$ can be taken to be $\hat{\sigma}^2\bar{\phi}^{-1}\bar{\phi}^*\bar{\phi}^{-1}$, where $\bar{\phi}^* = (N^{-1}\Sigma_i\bar{S}_i'\bar{u}_i^2\bar{S}_i)$. It remains to construct the nonparametric estimates for the unknown variance function σ^2

3.4.3 NONPARAMETRIC SMOOTHING AND BANDWIDTH SELECTION

If the heteroscedasticity is of an unknown form but it is a function of W_i, then it can be written either as $\sigma_i^2 = E[T(y_i)|W_i] - \{E[T(y_i)|W_i]\}^2$ or $\sigma_i^2 = E[u_i^2|W_i]$. Using the latter, we can estimate the conditional variance by a weighted average of the residuals from a consistent estimate of θ. Recall the discussion in Chapter 2, we would like to give largest weight to those observations closest to W_i. Therefore, we can estimate the conditional variance $g(w_i)$ using the Nadaraya–Watson kernel estimator by regressing $\bar{u}^2$ on w_i. The u_j's are obtained from the consistent estimates. The formula is given by equation (7) in Chapter 2.

The computation of these estimates is straight forward but heavy. For FORTRAN program, we refer the reader to Delgado (1988a). For the estimation of the derivatives of the nonparametric estimates, see Chapter 2.

For nonparametric estimation, the choice of bandwidth is crucial. We prefer the less subjective way of bandwidth selection rather than the more popular so called "rule–of–the–thumb" method that sets the bandwidth proportional to a number which minimizes the mean squared error. However, we avoid calling our method of bandwidth selection criterion objective, as the criterion function is still chosen by the practitioner.

The "automatic" selected bandwidth is

$$a_{CV} = \operatorname{argmin}_a N^{-1}\Sigma_i(\log\bar{\sigma}^2_{(-i)} + \bar{\sigma}^{-2}_{(-i)}\, \bar{u}_i^2)$$

where $\bar{\sigma}^2_{(-i)}$ is the leave–one–out nonparametric kernel regression estimates for the conditional variance. The leave–one–out estimates used in the bandwidth selection criterion are just the same estimates with $\Sigma_{j=1}^N$ replaced by $\Sigma_{j=1,i\neq j}^N$ (see Chapter 2). The performance and discussions on this pseudo likelihood criterion will be studied in a much simpler setting in a later chapter. It is first suggested and is shown to be consistent by Robinson (1988d) in Hannan's GLS estimator in the frequency domain.

3.4.4 MAXIMUM LIKELIHOOD ESTIMATION

Transformation can also reduce skewness and in certain cases, to transform the original distribution to normal. Indeed, if the transformation T_y for the dependent variable is convex, left skewness is reduced. If T_y is concave, right skewness is reduced. But simple calculation or plotting of graphs, one can easily determine whether we have reduced right or left skewness. Of course, a single transformation parameter for the dependent variable may not simultaneously deal with both skewness and heteroscedasticity. Therefore, it is also wise to guard against heteroscedasticity in the unlikely event that the error distribution is known to be from a particular parametric family. In this instance, since the distribution is known, efficiency can be improved via ML estimation.

In the case of HIHS and HHBC, u can be normally distributed. If u is known to be conditionally normally distributed with zero mean and the variance $\sigma^2(z)$, then a SML estimation method can be employed. The log likelihood function is given by

$$\text{Log L} = \frac{1}{2N}\Sigma_i \{\text{Log } 2\pi + \text{Log } g(W_i) + u_i^2/g(W_i) + \text{Log } a_i\}$$

where $a_i = 1+(\mu y_i)^2$ is a term from the Jacobian $|J| = \Pi_i a_i^{-1/2}$. Consider the case where g is known and differentiable with $G = \frac{\partial g(W)}{\partial W}$, then we have

$$\frac{\partial \ln L}{\partial \alpha} = N^{-1}\Sigma_i \frac{1}{g_i}[u_i + \frac{1}{2}G_i(\frac{u_i^2}{g_i} - 1)];$$

$$\frac{\partial \ln L}{\partial \delta} = N^{-1}\Sigma_i \frac{D_i}{g_i}[u_i + \frac{1}{2}G_i(\frac{u_i^2}{g_i} - 1)];$$

$$\frac{\partial \ln L}{\partial \gamma_j} = N^{-1}\Sigma_i \frac{T_2(z_{ij},\mu_j)}{g_i}[u_i + \frac{1}{2}G_i(\frac{u_i^2}{g_i} - 1)];\ j=1,\ldots,m;$$

$$\frac{\partial \ln L}{\partial \mu_j} = N^{-1}\Sigma_i \frac{\gamma_{ij}\partial T_2(z_{ij},\mu_j)/\partial \mu_j}{g_i}[u_i + \frac{1}{2}G_{\mu_j}(\frac{u_i^2}{g_i} - 1)];\ j=1,\ldots,m;$$

$$\frac{\partial \ln L}{\partial \mu} = -N^{-1}\Sigma_i(\frac{u_i}{g_i}\frac{y_i}{\mu}a^{-1/2} + a^{-1}\mu y_i^2)$$

We can also suggest a Gauss–Newton iterative scheme for the model here, but a linearized ML or two step estimator, θ^{ML},is asymptotically as efficient

$$\theta^{ML} = \bar{\theta} - N^{-1}(DlogL(\bar{\theta})\ DlogL(\bar{\theta})')^{-1}DlogL(\bar{\theta})$$

Under regularity conditions, $\sqrt{N}\ \partial LogL/\partial\theta \sim N(0,\Lambda)$ at the true θ, where $\Lambda = E[DLogL\ DLogL'|z]$. The ML estimator can be shown to be asymptotically normal, i.e., $\sqrt{N}\ (\theta^{ML} - \theta) \sim N(0,\Lambda^{-1})$.

In the case where g is unknown, we use kernel methods for estimating g and its derivative G. Using kernel estimation, we can predict g by

$$\hat{g}(W) = \Sigma_{j=1}^{N} \bar{u}_j^{\,2} K_{ij}(\bar{\theta}) / \Sigma_{j=1}^{N} K_{ij}(\bar{\theta})$$

for given $\bar{\theta}$. K is the standard normal density. In order to estimate the derivatives of the regression curve, G, we can use the nonparametric estimation (see, e.g., Schuster and Yakowitz (1979)):

$$\hat{G}(W_i) = \frac{\partial \hat{g}\ (W_i)}{\partial W_i} = \frac{\Sigma_j \bar{u}_j^2 L_{ij}(\theta)}{\Sigma_j K_{ij}(\theta)} - \frac{\Sigma_j \bar{u}_j^2 K_{ij}(\theta) \Sigma_j L_{ij}(\theta)}{(\Sigma_j K_{ij}(\theta))^2}$$

where L_{ij} is the derivative of the standard normal density, i.e.,

$$\begin{aligned} L_{ij}(\theta) &= L(a^{-1}(W_i - W_j)) \\ &= \frac{d(K_{ij}(x))}{dx} \end{aligned}$$

where $x = a^{-1}(W_i - W_j)$ and a is the bandwidth here. This class of SML is not new and was pioneered by Robinson (1987). One of his suggestions is semiparametric ARCH model which has been modified and applied in Whistler (1988) and Chapter 6 (or Lee (1989)) with the later suggesting an automatic bandwidth selection method.

A semiparametric linearized two step estimator is

$$\theta^{SML} = \bar{\theta} - N^{-1}(DlogL(\bar{\theta})\ DlogL(\bar{\theta})')^{-1}DlogL(\bar{\theta})$$

Under regularity conditions, θ^{SML} is adaptive in the sense that it has the same asymptotic distribution as the case where g is known. The method also produces a consistent $\Lambda^{SML} = \{N\ DLogL(\bar{\theta})\ DLogL'(\bar{\theta})\}^{-1}$ as a by product.

3.5 RESULTS

We report the parametric estimates for constrained transformation models, i.e., linear, log–linear, and semi–log, as well as NL2SLS and ML estimates for selected unconstrained transformation models. The data set consists of 399 observations on different models taken from the What Car? (June, 1988) magazine. The data on market shares are obtained from Car Buyer's Guide 1988. The descriptions of these data together with the data are given in an appendix available from the author.

In Table 3.1 and 3.2, we have presented the parametric estimates for the cost function and fuel efficiency function respectively. These linear, semi–log and log–linear models are constrained transformation models because the Box–Cox transformation is equivalent to the log transformation when $\mu = 0$ and level if $\mu = 1$; similarly, IHS transformation is equivalent to the log transformation if μ is large and level if $\mu = 0$.

The diagnostic test statistics reported, where appropriate, are the Breusch and Pagan (1979) multiplicity heteroscedasticity (BPM) test, Breusch and Pagan heteroscedasticity (BP) test, F test for all the coefficients are zeros, and the Jarque and Bera (1980) normality test. The BPM and BP test statistic is LM = 0.5 $(r'Q\,(Q'Q)^{-1}\,Q'r)$, where $r_t=(\tilde{u}_t/\tilde{\sigma}^2-1)$ is constructed from ols residuals, with $Q_{BPM} = (1, (\log W_i)^2)$ and $Q_{BP} = (1, W_i^2)$. The test statistic is distributed as χ^2_1. Jarque and Bera (1980) normality test statistic is based on the formula JB = $N\{m_3/6m_2^3 + 1/24\,(m_4/m_2^2 - 3)^2\}$ where $m_i = \Sigma_{j=1}^{N} u_j^i/N$; i = 2, 3, 4 and JB ~ χ^2_2.

The estimates reported in Table 3.1 and 3.2 are statistically significant at the 5% level with the exception of the acceleration in the linear model of Table 3.1 and all models of Table 3.2, indicating that we may have constrained some of the variables unnecessarily. However, the signs of all the models are correct according to our a priori belief. In particular, the fit of the fuel efficiency function for all three different specifications is good and there is no sign of heteroscedasticity in the semi–log and log–linear specifications.

Let us now concentrate on the estimates for rental cost function reported in Table 3.3. We have allowed for more flexible functional form for the rental cost function. There are six different transformation models. The first column presents the consistent NL2SLS estimates for the hyperbolic Box–Cox (HPBC) model.

Besides the parameter μ_4, none of the estimates are significant at the 5% level. There are two reasons for this: first of all, it is reasonable to assume that by allowing for flexible functional form, one has to pay the price of inflated variance reflecting the uncertainty of the transformation parameters. There are debates on whether one should make use of the information that we possess regarding the transformation parameters and whether the results make any scientific sense. One view is that an experienced practitioner will have more information about the functional form than an inexperienced one, and it is reasonable to assume that the experienced practitioner will get a better fit. The question is whether one should make use of the information on the transformation parameters and obtain the variance–covariance matrix. This will give good t–ratios in our case but we do not follow this approach. We take into account the fact that the transformation parameters are unknown and estimated. Secondly, the heteroscedasticity is severe and the instruments used may be far from efficient. We believe that both reasons are valid in our case. We seek to improve efficiency of our estimates by concentrating on the use of efficient instruments and adjustment for heteroscedasticity using the method mentioned in the previous section.

We have assumed that the conditional variance depends on W_i and we have selected the bandwidth for estimating g via a Gaussian log–likelihood criterion. Although the variance may depend on the market share, it is not appropriate in the estimation of the cost function and fuel efficiency function because we are also interested in the effect of unleaded petrol.

The HHBC estimates in second column are obtained using the Robinson (1988e)'s method with adjustment to induce homoscedasticity. Thus we expect the choice of instruments and the adjustment for homoscedasticity will help to improve the accuracy of the estimates. This is indeed the case. These efficient estimates obtained using the "automatic" bandwidth have smaller standard errors and besides μ_2, all the t–ratios are statistically significant and positive. The t–ratios reported are the robust t–ratios to guard against incorrect choice of bandwidth and these t–ratios differ only slightly from those obtained
using $\bar{\phi}^{-1}$.

We have also produced the ML estimates for the HPBC and HHBC respectively in the third column and the fourth column. The HHBC estimates are obtained from one step update from the NL2SLS estimates reported in the first column and the same bandwidth is used. The log likelihood function has increased by a massive 71.75 and would give a likelihood ratio test statistic of 143.5. However, one should not place too much faith on the ML estimates as they are unlikely to be robust to very slight misspecifications.

The BC and HBC estimates are reported in the fifth column and the sixth column. They are both obtained using Wi as instruments for μ. The reasons for using this instrument are:

(i) Amemiya (1985) has recommended the use of the squares and the cross products of the independent variables but this does not work in our case;

(ii) Robinson's method, unfortunately, breaks down if there are negative $\psi(\bar{W}_i^* + u_j)$, and this is the case for the fuel efficiency function in our application because the estimates π are negative and ψ positive.

In view of this, W_i^2 will be a very good choice as it is a linear combination of the squares and cross products, and indeed Newey (1987) has argued that it is close to efficient. This perhaps is the case as the t–ratios of the fifth column are much better than those in the first column. However, when heteroscedasticity is corrected, the sign of the estimate for the acceleration variable is inconsistent with a priori belief as in the case of log–linear regression.

Finally, the $\sinh^{-1}$ (IHS) and the heteroscedastic $\sinh^{-1}$ (HIHS) models are reported in the last two columns. The t–ratios for the estimates for both models are not statistically significant. The function and rss values are extremely small. These are symptoms of multicollinearity, thus confirming our initial fear.

The picture for the fuel efficiency function is almost identical. The HHBC estimates seem to be the most plausible. The estimates for the independent variables are statistically significant and negative, with the exception of the acceleration variable which is inconsistent with our prediction.

We do not have any criterion for selecting among the different models. It is perhaps unwise to look at the objective function because different models and instruments are being used. However, the HHBC models have lower function values and rss, good t–ratios and the signs of the estimates are mostly consistent according to our a priori prediction. The HHBC estimates for both rental cost and fuel efficiency functions seem to be the most plausible. These estimates are therefore used to construct the price elasticities.

3.6 IDENTIFICATION AND CONSTRUCTION OF ELASTICITIES

There is a growing literature on identifying the parameters of the demand functions and thus the utility function in recent years. Since there are no observations on the individual household, the best that one can do is to estimate the parameters from the model level. This is a common problem in studies using model and make data.

There are two usual ways in the literature of obtaining the parameters. The first is the two–stage method of Rosen (1974) which has been applied by Goodman

(1983). However, this two stage (TS) method of Rosen (1974) has generated a lot of discussions recently because it suffers from identification problem. The problem is especially difficult to handle when dealing with the car market because there are virtually no appropriate instruments as we are dealing with a single market data here. This issue is discussed briefly in the next chapter. The second is the use of discrete choice modelling as in Boyd and Mellman (1980). This method requires sales data of all individual models and the cost of collection is extremely high. We therefore adopted the following method.

For identification and other purposes mentioned above, we have assumed that the subutility function is Cobb–Douglas, i.e., $U = b_0 + b_j \mathrm{Log} z_j$. The procedures that we have proposed does not specify the functional form of utility function except that it should be intertemporal separable. In other words, if Cobb–Douglas function is not suitable for some reasons, any other intertemporal separable utility function can be used. The reason why we have chosen this particular form of utility function is because it has been widely used in the literature and has certain well known properties. This is a maintained hypothesis and we merely use it to demonstrate the methodology.

To identify the parameters of the utility function, we have imposed the constant expenditure share restrictions and let $b_j = \alpha_j z_j / P$, where $\Sigma_j b_j = 1$ and α_j's are the slope estimates in the log–linear capital cost (P) function for new cars. In other words, in addition to assumptions (i) and (ii) in Section 2, we assume that

(iii) the utility function is of Cobb–Douglas form i.e.

$$\exp(U(z)) = \exp(b_0)\, \Pi_{j=1}^{m} z_j^{b_j} \qquad \text{where } \Sigma_j b_j = 1.$$

We obtain the approximation at the mean of the z_j's and C. Consider

$$T_S(P_i;\mu) = \alpha_0 + \delta' D_i + \Sigma_j T_{BC}(z_{ji};\mu_j) + u_i \qquad i = 1,2,\ldots,N$$

where $\delta = (\delta_1, \delta_2, \delta_3, \delta_4)'$ and $D = (D_U, D_B, D_L, D_S)'$. We have restricted μ to be fairly large, $\mu_{\mathrm{i}} = 0$ in the HPBC model which is the simple semiparametric linear GLS case. We have also included several 0–1 dummy variables to capture the consumer's perception of the makes and models which are not captured by the attributes that we are interested in. $\mathrm{D_U}$ is the dummy for car which can run on unleaded petrol without adjustment. The British made dummy ($\mathrm{D_B}$) is to capture the difference between locally and foreign made cars. The luxury ($\mathrm{D_L}$) and speciality ($\mathrm{D_S}$) dummies are used to capture the effect of the more distinguished and high performance cars. Observations on individuals would be helpful but we did not pursue along this line. In any case, at any one instant in time, the supply is assumed to be perfectly elastic and the consumer's preferences can be identified. We are also making the strong assumption that the preferences are the same across the rental and new car market.

The fit of the price function (see Table 3.5) is consistent with previous results. We allowed for heteroscedasticity and obtained the semiparametric GLS estimates using the automatic bandwidth selection method. The semiparametric GLS literature is now well developed (see e.g., Carroll (1982), Robinson (1987), Rose (1978) for using kernel estimates for the conditional variance which is relevant to our application). Since it is a special case of the heteroscedastic transformation model, we will not devote any more space for discussion.

The heteroscedastic estimates are obtained by using the bandwidth from minimizing the same criterion as before which has been suggested by Robinson (1988d) in a separate problem in the frequency domain. In this case, the conditional variance is assumed to be a function of the market shares. The elasticities are constructed based on the estimates in the last column in Table 3.5.

To obtain the estimates for the equation for the comparative static equations in Section 2, we have to consider the following approximations for the HPBC model:

$$\partial y/\partial z_i = \gamma_i z_i^{\mu_i - 1} a^{1/2}$$

$$\partial y/\partial z_i \partial z_j = (\gamma_i \mu z_i^{\mu_i - 1})^2 y + \gamma_i(\mu_i - 1) z_i^{\mu_i - 2} a^{1/2} \qquad i=j$$

$$= \gamma_i \gamma_j z_i^{\mu_i - 1} z_j^{\mu_j - 1} \mu^2 y$$

The advantage of using HPBC rather than BC is apparent because the BC model is essentially a limited dependent model. Huang and Kelingos (1979) and Huang and Grawe (1980) have discussed the various issues regarding the conditional mean and elasticity from Box–Cox model under the assumed truncated normal distribution of the disturbance term. In particular, one should realize that the conditional mean of the BC model does not not exist when μ falls in the the interval 0 and –1 under the assumed distribution. Notice that in the log–linear case, the second derivative is not equal to zero as claimed by Atkinson and Halvorsen (1984). In that case, the second derivative should have been $-\gamma_i z_i y^{-2} \partial y/\partial z_j$

It remains for us to work out the lagrange multiplier ℓ. It is obtained by using (1) summing over j from 1 to m.

$$U_i = -b_i/z_i \qquad\qquad i = 1,\ldots,m$$

$$U_{ij} = -b_i/z_i \qquad i=j \qquad i = 1,2,\ldots,m$$

$$= 0 \qquad i \neq j \qquad j = 1,2,\ldots,m$$

We are also interested in the effect of unleaded petrol on the dependent variables. One must be careful in interpreting the effect of dummy on the dependent variable in log–linear model and transformed models (see e.g. Blaylock and Smallwood (1983)). It will be incorrect to interpret the effect as δ since the dummy variable is clearly not continuous. The correct interpretation of the dummy D on a dependent variable y in our HPBC models is perhaps to look at the proportion change in y due to a unit change in D, i.e.,

$$\Delta = (y_1 - y_0)/y_0$$

where $y_1 = \mu^{-1}\sinh[\mu(\alpha + \delta + \Sigma_j \gamma_j T(z_j;\mu_j)]$

$$y_0 = \mu^{-1}\sinh[\mu(\alpha+\Sigma_j\gamma_j T(z_j;\mu_j)]$$

The log–linear case is just a special case of the formula. One could of course extend this formula to include 2 dummies and so on, but the analysis would be more cumbersome as the number of permutations increases. Dagenais, Gaudry and Tran (1987) have suggested other criteria. In this chapter, we will look at the effects and elasticities based on sample mean rather than conditional expectation as these other suggestions generally require one to assume a particular distribution which we have tried very hard to avoid in the first place. Of course, under conventional consumer theory, with Shepherd's Lemma and Roy's identity, various elasticities can be consistently estimated without knowing the form of the cost or indirect utility function, (see e.g., Elbadawi, Gallant and Souza (1983)) using Fourier series (Gallant (1981) to approximate the unknown function concerned).

Corresponding to the second column of Table 3.3 and Table 3.4, Δ are 0.003 and –0.026 respectively. The results indicate that there is a slight difference in cars which can use unleaded petrol without adjustment and suggest that rental cost will be 0.3% higher and fuel efficiency will be 2.6% lower with these cars.

In Table 3.6, we have presented the price elasticity of demand for attributes for the 5 most popular models, the two extreme cases of fuel efficiency with the highest and lowest fuel consumption in our data, and the overall mean value. The elasticities are constructed under various assumptions stated in the Table. We would just like to mention that S =.993 which corresponds to ρ = 2.3% estimated in Attanasio and Weber (1989) and r is set at 23% to reflect the high interest rate charged. A unitary elastic price expectation with respect to the base period price P_0 is assumed, i.e., $P_t = (1+f)^t P_0$. The justifications for the assumptions are presented together with the source of data in the appendix. We tried different assumed values and found that the elasticities are not sensitive to these assumptions but very sensitive to the estimates from the three equations. These observations are confirmed by evidence in Table 3.7 using different assumed parameters. In particular, we find that setting r at the treasury bill rate as well as taking the planning horizon to be the average length of the life of a car (T), makes little difference.

The program used in constructing these elasticities and all those written for this study can be obtained from the author. They are written in GAUSS programming language with the exception of the programs which involve considerable amount of element–by–element operations and take a fair amount of time in the matrix programming environment. These include programs for bandwidth selection, nonparametric regression and Robinson's method of forming instruments. They are written in FORTRAN for computational efficiency.

All the price elasticities of attributes are generally inelastic with the exception of the elasticity for spaciousness. The smaller cars seem to be generally more responsive than larger cars. The price elasticity of demand for size is generally slightly greater than –0.10 but less than –0.26, which is very inelastic. While the price elasticity for power is generally inelastic, the price elasticity for spaciousness is elastic and can be as high as –2.8 as in the Citron model. As for the price elasticity of acceleration, it is positive. This perhaps reflects that cars with faster acceleration is favoured. We have also reported the price elasticities of demand for fuel efficiency and they are generally close to unity.

As for Table 3.7, the results are very similar although we have used different r, T and b_i's. Furthermore, we find that the elasticity for fuel efficiency estimates from the linear, log–linear and semi–log are fairly similar. But the other attribute elasticities (not reported) are substantially different from each other.

Our approach is interesting in the sense that we are able to construct the elasticities, and by the fact that they are constructed and not estimated directly, we are not able to say much about their properties. In particular, we cannot obtain the standard errors. In the absence of the standard errors for these constructed elasticities, one perhaps should compare the results with those obtained elsewhere. As noted by Atkinson and Halvorsen (1984), Pindyck (1979) used pooled cross section data for 11 OECD countries and found that the long run fuel efficiency elasticity to be 1.43. Griffin (1979)'s results also indicate the long run elasticity of fuel efficiency has a lower bound of 0.79 and an upper bound of 1.43 using 18 OECD countries. Our elasticity lies in between.

We can also infer the own price elasticity from our estimates. Consider the total petrol consumption Q = M C/E, where C is the stock of cars. The relationship among the elasticities is $\eta_{\mathrm{p}} = \eta_{\mathrm{M}} + \eta_{\mathrm{S}} - \eta_{\mathrm{E}}$. Since the price elasticity of demand for miles travel (η_{M}) and stock (η_{S}) are both negative and η_{E} is close to one in magnitude, this implies that the own price elasticity is likely to be greater than unity.

3.7 CONCLUSION

We have presented an intertemporal model and apply comparative statics analysis to the consumer's choice of attributes. The main purpose of the approach is to avoid the problems of multicollinearity and identification which are common in previous hedonic studies. The method does not require variations in the price of petrol.

Recently, there is some interest surrounding the use of the IHS transformation model. Unlike the Box–Cox model, the IHS model can be normally distributed and can deal with observations with non–positive value. We argue in favour of the use of HPBC model because of the associated advantages.

We have presented the ML method and an alternative robust method of estimating the parameters from various heteroscedastic transformation models. The latter procedure does not make distributional assumptions and is asymptotically efficient within a class of models based on conditional moment restrictions. In the HPBC model, it is easy to find efficient instruments via Robinson (1988e)'s method.

In particular, the approach of Robinson (1988e) does not require nonparametric regression in forming instrument and makes fuller use of the sample of data. It can be combined with the use of nonparametric conditional variance estimates. We are not required to make any assumption of the form of heteroscedasticity when the assumption of homoscedasticity is violated. The bandwidths are chosen automatically using the pseudo Gaussian log–likelihood criterion suggested by Robinson (1988d). Although the asymptotic properties for data–dependent bandwidths are unknown in the transformation models, we can perhaps draw upon the frequency domain results of Robinson (1988d) to our time domain case in the linear GLS model by analogy. In any case, one believes that the rate of convergence will be very slow for the bandwidth estimate if it is consistent.

Our elasticities are constructed from these efficient estimates. This prevents us from giving the standard errors. Comparison with previous results is one way of indicating the accuracy of these elasticities. There was not a large discrepancy between our results and previous findings. The petrol price elasticity of demand for attributes is close to unity and the own price elasticity of petrol is very likely to be in excess of one according to our results.

Table 3.1

Parametric Estimates for Running Cost Function

	Linear	Semi–log	Log–linear
intercept	–0.2702	–3.4732	1.2113
	(–5.8773)	(35.2212)	(5.9849)
unleaded	0.0063	0.0101	0.0282
	(0.7714)	(0.5760)	(1.5050)
size	1.3797	2.2308	0.8022
	(16.3557)	(12.3310)	(14.647)
spaciousness	1.1247	7.7701	0.4912
	(3.5107)	(11.3094)	(3.7063)
power	0.2389	0.5876	0.2102
	(4.0648)	(4.6611)	(2.9459)
acceleration	0.0284	0.2951	0.1465
	(0.8357)	(4.0501)	(1.9846)
R squared	0.816	0.862	0.842
R bar squared	0.814	0.860	0.840
se	0.074	0.159	0.170
rss	2.152	9.896	11.331
tss	11.725	71.718	71.718
F(6 ,393)	349.670	491.012	418.869
BP χ_1^2	26.700	8.230	4.270
BPM χ_1^2	26.438		
JB χ_2^2	1255.122	178.525	200.579

Figures in parentheses are the estimated t–ratios.
F: Testing the null hypothesis that all the coefficients are zeros.
BPM: Breusch–Pagan Multiplicity Heteroscedasticity LM test based on the square of mean.
BP: Breusch–Pagan Heteroscedasticity LM test.
JB: Jarque–Bera Normality test.

Table 3.2

Parametric Estimates for Fuel Efficiency Function

	Linear	Semi–log	Log–linear
intercept	0.7415	–0.0691	–2.2437
	(35.1350)	(–1.1592)	(–19.7539)
unleaded	–0.0046	–0.0155	–0.0247
	(–1.2454)	(–1.4638)	(–2.3428)
size	–0.2970	–1.2639	–0.4611
	(–7.6709)	(–11.5450)	(–15.002)
spaciousness	–1.6362	–3.6397	–0.1776
	(–11.1278)	(–8.7540)	(–2.3878)
power	–0.1004	–0.2995	–0.0995
	(–3.7229)	(–3.9258)	(–2.4858)
acceleration	–0.0214	–0.0186	0.0681
	(–1.3759)	(–0.4238)	(1.6544)
R squared	0.755	0.782	0.786
R bar squared	0.752	0.779	0.783
se	0.034	0.096	0.095
rss	0.453	3.624	3.569
tss	1.851	16.641	16.641
F(6 ,393)	242.289	282.298	287.875
BP χ_1^2	6.202	0.642	0.091
BPM χ_1^2	14.642		
JB χ_2^2	33.269	89.166	106.984

Figures in parentheses are the estimated t–ratios.
F: Testing the null hypothesis that all the coefficients are zeros.
BPM: Breusch–Pagan Multiplicity Heteroscedasticity LM test based on the square of mean.
BP: Breusch–Pagan Heteroscedasticity LM test.
JB: Jarque–Bera Normality test.

Table 3.3

Box–Cox and Inverse Hyperbolic Sine Estimates for Running Cost Function

	NL2SLS HPBC	HHBC	MLE HPBC	HHBC	NL2SLS BC	HBC	NL2SLS IHS	HIH
^{a}CV		0.044		0.044		0.010		0.03
cons	1.359	1.365	1.452	1.364	1.053	1.031	–0.174	0.010
	(0.142)	(3.174)	(0.387)	(0.569)	(0.168)	(8.424)	(–0.93)	(0.084)
Dum.	0.004	0.001	0.005	0.004	0.001	0.000	0.004	0.001
	(0.533)	(2.272)	(0.720)	(0.959)	(0.385)	(16.94)	(0.833)	(0.486)
size	0.721	0.750	0.819	0.721	1.518	1.520	0.912	0.224
	(1.813)	(32.04)	(3.273)	(3.225)	(2.925)	(66.14)	(3.992)	(0.455)
spac	0.373	0.378	0.353	0.380	0.724	0.729	1.041	0.430
	(0.587)	(12.56)	(0.049)	(0.088)	(0.195)	(10.33)	(0.536)	(0.324)
pwr.	0.153	0.138	0.117	0.153	0.411	0.354	0.126	0.055
	(0.191)	(7.823)	(1.863)	(1.590)	(4.359)	(210.9)	(1.541)	(0.611)
acc.	0.143	0.090	0.058	0.143	0.123	–0.118	0.062	0.007
	(0.603)	(6.366)	(2.070)	(5.395)	(0.687)	(–38.2)	(1.358)	(0.441)
transformation parameters								
size	0.833	0.845	0.727	0.833	1.925	1.946	0.972	0.806
	(1.457)	(75.09)	(3.145)	(3.856)	(2.812)	(139.1)	(0.118)	(0.005)
spac	0.995	0.999	1.041	1.006	1.427	1.427	0.999	0.999
	(0.000)	(0.002)	(0.093)	(0.168)	(0.000)	(0.008)	(0.000)	(0.000)
pwr.	0.935	0.942	1.123	0.935	1.773	1.742	–0.551	–0.806
	(1.221)	(50.08)	(1.830)	(1.342)	(26.86)	(724.1)	(–1.27)	(–4.89)
acc.	1.005	0.977	1.006	1.006	0.980	1.005	0.028	0.024
	(3.206)	(45.10)	(0.580)	(1.626)	(7.579)	(441.7)	(0.020)	(0.006)
y	1.017	1.031	1.367	1.017	2.904	2.896	2.328	13.128
	(0.343)	(29.98)	(5.218)	(4.459)	(2.978)	(95.37)	(2.023)	(0.490)
ftn.	0.615	128.1	513.382	585.136	0.124	276.6	0.128	2.801
se	0.0049	0.0050			0.0011	0.0010	0.0022	0.0001

Note: The attributes are size, spaciousness (spac), power (pwr), accelearation (acc), and the dummy variable (Dum) is for unleaded petrol.
HPBC: Hyperbolic Box–Cox.
HHBC: Heteroscedastic Hyperbolic Box–Cox.
BC: Box–Cox.
HBC: Heteroscedastic Box–Cox.
IHS: Inverse Hyperbolic Sine.
HIHS: Heteroscedastic Inverse Hyperbolic Sine.
The NL2SLS estimates are converged estimates while the MLE estimates are one step update from the consistent NL2SLS estimates. Models with heteroscedasticity of unknown form are obtained using the "automatic" bandwidth and Robinson (1988d)'s method of forming efficient instruments.
Figures in parentheses are the estimated t–ratios for MLE robust t–ratios for NL2SLS.

Table 3.4

Box–Cox And Inverse Hyperbolic Sine Estimates For Fuel Efficiency Function

	NL2SLS HPBC	HHBC	MLE HPBC	HHBC	NL2SLS BC	HBC	NL2SLS IHS	HIH
^{a}CV		0.031		0.031		0.031		0.01
cons	–1.155	–1.156	–1.059	–1.061	–1.680	–1.676	0.524	0.646
	(–0.29)	(–12.0)	(–0.40)	(–0.53)	(–0.51)	(–14.8)	(3.103)	(150.3)
Dum.	–0.010	–0.009	–0.013	–0.010	–0.007	–0.006	–0.004	–0.004
	(–1.99)	(–59.6)	(–2.28)	(–2.91)	(–1.34)	(–31.0)	(–1.20)	(–43.7)
size	–1.308	–1.308	–1.040	–1.038	–1.090	–1.091	–0.618	–0.504
	(–0.49)	(–11.0)	(–0.21)	(–0.20)	(–0.71)	(–12.6)	(–3.25)	(–84.0)
spac	–1.106	–1.107	–1.046	–1.032	–0.689	–0.693	–0.096	–1.052
	(–1.09)	(–54.9)	(–0.25)	(–0.34)	(–1.23)	(–37.6)	(–0.07)	(–32.3)
pwr.	–0.496	–0.493	–0.444	–0.494	–0.339	–0.332	0.025	–0.083
	(–2.92)	(–70.4)	(–6.76)	(–8.05)	(–2.33)	(–49.7)	(0.334)	(–42.0)
acc.	0.105	0.103	0.074	0.105	–0.016	–0.004	–0.073	–0.009
	(0.609)	(15.97)	(4.968)	(6.414)	(–0.10)	(–0.55)	(–0.71)	(–2.61)
transformation parameters								
size	3.567	3.567	3.341	3.369	3.220	3.221	0.953	0.875
	(0.490)	(16.59)	(0.836)	(0.958)	(0.780)	(20.99)	(0.326)	(9.032)
spac	1.012	1.012	0.977	0.972	1.108	1.108	0.999	0.999
	(0.003)	(0.127)	(0.487)	(0.670)	(0.002)	(0.070)	(0.000)	(0.003)
pwr.	1.230	1.225	1.141	1.228	1.046	1.036	1.022	1.018
	(14.15)	(302.3)	(15.04)	(5.462)	(12.05)	(196.6)	(2.758)	(60.60)
acc.	0.860	0.865	0.911	0.851	1.076	1.060	1.075	1.174
	(12.70)	(353.0)	(1.155)	(1.512)	(8.990)	(174.0)	(0.499)	(152.0)
y	1.017	1.017	1.096	0.998	0.999	1.002	1.190	1.527
	(0.323)	(16.83)	(0.961)	(3.335)	(3.861)	(122.2)	(0.793)	(40.71)
ftn.	0.355	174.710	673.921	720.170	0.479	218.8	0.169	50.21
se	0.0018	0.0018			0.0022	0.0021	0.0012	0.0009

Note: The attributes are size, spaciousness (spac), power (pwr), accelearation (acc), and the dummy variable (Dum) is for unleaded petrol.
HPBC: Hyperbolic Box–Cox.
HHBC: Heteroscedastic Hyperbolic Box–Cox.
BC: Box–Cox.
HBC: Heteroscedastic Box–Cox.
IHS: Inverse Hyperbolic Sine.
HIHS: Heteroscedastic Inverse Hyperbolic Sine.
The NL2SLS estimates are converged estimates while the MLE estimates are one step update from the consistent NL2SLS estimates. Models with heteroscedasticity of unknown form are obtained using the "automatic" bandwidth and Robinson (1988d)'s method of forming efficient instruments.
Figures in parentheses are the estimated t–ratios for MLE robust t–ratios for NL2SLS.

Table 3.5

Estimates For Capital Cost Function

	Log–Linear	Heteroscedastic
bandwidth		5.0798
intercept	0.2760	0.1499
	(1.0799)	(0.2615)
unleaded	0.0339	0.0367
	(1.4285)	(1.5983)
British made	−0.0275	−0.0280
	(−1.1304)	(−1.1818)
luxury car	0.0988	0.1053
	(3.6440)	(4.1254)
specialty car	0.4942	0.4679
	(4.5449)	(4.5293)
size	0.9921	1.0315
	(14.6612)	(17.2117)
spaciousness	0.3683	0.1906
	(2.2554)	(2.3344)
power	0.2047	0.2304
	(2.3687)	(2.7843)
acceleration	0.3005	0.2357
	(3.3416)	(2.6682)
R squared	0.870	0.894
R bar squared	0.868	0.892
se	0.204	1.126
rss	16.185	494.114
tss	124.921	4674.219
F(6 ,390)	327.507	412.415
BP χ^2_1	9.653	
JB χ^2_2	72.782	

Figures in parentheses are the estimated t–ratios.
F: Testing the null hypothesis that all the coefficients are zeros.
BP: Breusch–Pagan Heteroscedasticity LM test based on market shares.
JB: Jarque–Bera Normality test.

Table 3.6

Constructed Price Elasticity Of Demand For Attributes

Vehicle Model	Ford Escort 1.6 L		Ford Sierra 1.6		Ford Fiesta 1.6		Austin /MG Metro 1.3	
Attrib.	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
size	1597.0	−0.143	1593.0	−0.128	1392.0	−0.139	1275.0	−0.159
space.	14.7	−1.688	16.2	−1.663	13.6	−1.822	12.4	−2.116
power	5.3	−0.421	3.7	−0.278	4.7	−0.432	4.3	−0.496
accel.	0.106	0.053	0.083	0.042	0.093	0.064	0.090	0.084
c.p.m	21.8		21.6		19.8		21.3	
eff.	41.0	0.981	36.6	0.984	39.7	0.983	42.0	0.981

Vehicle Model	Vauxhall Cavalier 1.6 L		Ferrari 412		Citron AX 10E		Mean value	
Attrib.	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
size	1598.0	0.981	4942.0	−0.032	954.0	−0.255	1914.1	−0.116
space.	16.0	−0.120	17.6	−0.350	12.6	−2.862	15.7	−1.433
power	4.5	−1.561	9.5	−0.151	3.6	−0.543	5.2	−0.342
accel.	0.084	0.040	0.149	0.024	0.056	0.033	0.098	0.049
c.p.m.	23.2		84.4		13.2		32.6	
eff.	36.0	0.985	14.7	0.996	55.9	0.968	35.1	0.985

c.p.m.: cost per mile
eff.: fuel efficiency
(1) refers to the observed values and (2) refers to the constructed price elasticities.
is the price elasticity of demand for fuel efficiency constructed from HHBC estimates. The values are constructed under the following assumptions:

Miles travelled,	$M = 8701.05$
Interest rate,	$r = 23.00\%$
Growth rate,	$f = 10\%$
Infinity life time set at	$T = 100$ years
Rate of time preference,	$\rho = 2.3\%$

and $b_1 = 0.34$; $b_2 = 0.05$; $b_3 = 0.20$; $b_4 = 0.39$.

Table 3.7

Constructed Price Elasticity Of Demand For Attributes

Vehicle Model	Ford Escort 1.6 L		Ford Sierra 1.6		Ford Fiesta 1.6		Austin /MG Metro 1.3	
Attrib.	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
size	1597.0	–0.134	1593.0	–0.115	1392.0	–0.130	1275.0	–0.149
space.	14.7	–1.231	16.2	–1.359	13.6	–1.348	12.4	–1.560
power	5.3	–0.686	3.7	–0.553	4.7	–0.713	4.3	–0.821
accel.	0.106	0.164	0.083	0.192	0.093	0.210	0.090	0.271
cpm	21.8		21.6		19.8		21.3	
eff.	41.0		36.6		39.7		42.0	
η_1		0.982		0.985		0.984		0.982
η_2		0.953		0.965		0.957		0.952
η_3		0.920		0.905		0.933		0.923
η_4		0.940		0.947		0.947		0.940

Vehicle Model	Vauxhall Cavalier 1.6 L		Ferrari 412			Citron AX 10E		Mean value
Attrib.	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
size	1598.0	–0.109	4942.0	–0.027	954.0	–0.236	1914.1	–0.106
space.	16.0	–1.222	17.6	–0.195	12.6	–2.334	15.7	–1.046
power	4.5	–0.567	9.5	–0.203	3.6	–0.994	5.2	–0.581
accel.	0.084	0.165	0.149	0.066	0.056	0.155	0.098	0.171
c.p.m.	23.2		84.4		13.2		32.6	
eff.	36.0		14.7		55.9		35.1	
η_1		0.985		0.997		0.966		0.986
η_2		0.965		0.992		0.921		0.965
η_3		0.777		0.916		0.948		0.946
η_4		0.945		0.919		0.950		0.925

c.p.m.: cost per mile
eff.: fuel efficiency

(1) refers to the observed values and (2) refers to the constructed price elasticities. η_1, η_2, η_3 and η_1 are price elasticity of demand for fuel efficiency constructed from HHBC, linear, semi–log and log–linear estimates respectively. The values are constructed under the following assumptions:

Miles travelled,	M = 8701.05
Interest rate,	r = 9.24%
Growth rate,	f = 10%
Life of vehicle,	T = 13 years
Rate of time preference,	ρ = 2.3%

and $b_1 = 0.61$; $b_2 = 0.11$; $b_3 = 0.13$; $b_4 = 0.13$.

CHAPTER 4

FURTHER SEMIPARAMETRIC POLICY ANALYSIS

4.1 INTRODUCTION

Fuel efficiency has been an important issue in the U.S. for more than two decades. The main interest on automobile efficiency was brought on by the two severe oil price shocks. This is because fuel efficiency, measured in miles per gallon (MPG), is an important component in energy conservation. It is also one of the most important attributes to potential buyers.

In the U.K., recent interest on fuel efficiency was brought about by the introduction of unleaded petrol. Better understanding of the damage of leaded petrol to the surrounding has led to the introduction of legislation that favours the use of unleaded petrol. Eventually, new laws will be imposed to bring about production of all new cars that will run on unleaded petrol.

In last chapter, there is indirect evidence that the use of unleaded petrol will lead to lower fuel efficiency, ceteris paribus. If the results are reliable, then one of the issues confronting the policy maker is the loss of efficiency in those cars that can use both leaded and unleaded petrol. The question then is whether one should introduce tax incentives to increase fuel efficiency.

There are two ways to improve fuel efficiency. First of all, one can reduce certain features of the car, e.g., one can reduce the weight of the vehicle to gain efficiency. Secondly, fuel efficiency improvement may be brought about by technical improvements independent of other attributes, e.g., the refinement of the petrol itself or the use of higher performance engineering parts. By introducing new regulations or providing tax incentives, both types of improvements can be achieved.

The benefits of the use of unleaded petrol to the environment is enormous. It is therefore not surprising that there is an asymmetric emphasis on the benefits and costs of using unleaded petrol. In the context of econometric policy analysis, it is of interest to find out the cost of such technical inefficiency associated with using unleaded petrol. If the benefits of increased fuel efficiency are high, it provides some justification for introducing legislations for improvement. This information may also be of interest to the producers as the average change in the capitalized values can be used as a guideline for pricing decision.

In this chapter, we evaluate the effect of an increase (a decrease) in the fuel efficiency on the hedonic price at the model level. It addresses the problem of estimating the average capitalized gains (or costs) brought about by increased (decreased) MPG. The procedure that we have adopted to estimate the "willingness–to–pay" for increased fuel efficiency has been adopted in the analysis of environmental benefits in the U.S. (see Stock (1985a)). There are advantages over the two–stage method of Rosen (1974) and most important of all, our semiparametric approach avoids misspecifications.

Econometric policy evaluation can be viewed as measuring the effects of a change of distribution of the exogenous variables on the mean value of an endogenous variable. The general approach that has been adopted in the profession is to parameterize a model, e.g., $P = f(x)$, and then evaluate the unconditional expectation of the change in the endogenous variable. However, in many instances, the parametric model itself is ad hoc, formulated with very little economic justification, e.g., assuming that f is approximately linear, and thus the consistency of the benefits or costs estimates is called into question.

Stock (1984, 1985a) suggested using a nonparametric approach to the problem of evaluating econometric policy which offers a solution when the conditional distribution of P on x^*, where x^* is the policy variable after the change, is unknown. By estimating the conditional expectation of y condition on x^* nonparametrically, one has the advantage of utilizing all the information concerning the structure of the policy and the effect of the policy change on the distribution of the exogenous variable.

There are two main contributions in this chapter. First of all, we have attempted to model the hedonic price function using a semiparametric approach. The resulting slope estimates for the dummy variables from the semiparametric hedonic model are interpreted as the characteristic or shadow prices. Although this semiparametric approach has the advantage in that the estimates are more robust than the traditional parametric approach, we have discovered that it is highly sensitive to the bandwidth used. Secondly, we have attempted to estimate the average benefits or costs associated with a change in fuel efficiency. We have discovered that these estimates are also very sensitive to the bandwidth used.

4.2 THE HEDONIC PRICE FUNCTION

The hedonic price function is originally proposed to reflect quality improvement in the price index. The first published paper on hedonic price function was by Court (1939). In the 60's, economists and national income statisticians alike, were worried that the price indices would be defective as they did not take fully into account quality changes. Much research, spearheaded by the efforts of Griliches and his colleagues (see Griliches (1971)), was then underway to construct price indices which would reflect quality improvements. The basic idea of an index incorporating quantity and quality change was to construct by dividing an index of money values by quality adjusted price index. The analysis is based on the price function

$$P = f(z_1,\ldots,z_m)$$

where P is the sales price, and the z's are the attributes of the car. The hedonic implicit or characteristic price of a component z is the partial derivative of P with respect to z. Typically, the function f is assumed to be linear, log–linear or semi–log in empirical work. The semi–log form was more popular for two reasons: the first was that the semi–log function was able to reflect varying marginal valuations; the second was that the statistical fit was thought to be much better based on R^2 in the earlier studies, and Gaussian Box–Cox likelihood criterion in recent studies (Goodman (1983)). However, there remain problems with the reasoning.

On the first point, semi–log form may not be consistent with the theory (Lucas (1975)). On the second point, we have known that the use of Gaussian Box–Cox transformation has various problems (see Chapter 3). The most troubled evidence from many of the results using Gaussian Box–Cox model has been the values of the transformation parameter of P lying inside the inadmissible range of –1 and zero. Goodman (1983) found that one of his two transformation parameters of the dependent variable was –0.6. The findings of negative transformation parameters is also common to the empirical studies in housing economics. Based on these reasons, the bias towards a semi–log formulation may be unwarranted.

In view of this, quadratic Box–Cox transformation has been used very often in the literature. However, this highly parameterized flexible functional form has the problem of giving imprecise estimates and is not suitable for prediction.

With both linear and semi–log forms rejected by either the statistical fit or economic theory, and the fact that the flexible functional from may not be desirable for prediction and yielding imprecise estimates, a large proportion of the literature on transport economics in recent years have reverted to the log–linear form. The log–linear transformation is favoured because of (i) easy interpretation; (ii) its ability to transform the error distribution to symmetry; and (iii) its ability to stabilize the variance as in most transformations. But, this functional from is still chosen subjectively and hardly satisfactorily. Many agree that the price function reflect underlying demand and supply in the market, but specifying the functional form remains very difficult. On the one hand, if one rules out certain arbitrage activities and abandons the divisibility assumption of the product (e.g. in Rosen (1974)'s model, he rules out the possibility of combining 1000cc car with a 2000cc car to form a new car), then one rules out the possibility of P(z) being linear. On the other hand, if one is interested in modelling a perfect rental market with no transaction or assembly costs, then the linear characteristic model (Lancaster (1971)) where the price of a typical good is formulated as a sum of the characteristic price multiplied by the quantity of the characteristic embodied in the good may be useful. This is of course consistent with earlier household production approach to quality variation which assumes that the consumers are also the producers. They purchase the goods not for consumption but as inputs into self production functions for characteristics (see, e.g., Lancaster (1971)).

In the latter case, Jones (1988) has shown that the hedonic linear decomposition of equilibrium prices need not hold, thus raising the possibility of nonlinearity. Indeed, some may argue that the former model is more realistic for the new car market in the sense that it rules out any repackaging attempts by the consumers. This implies that the hedonic function cannot be linear.

It was thought that if nonlinearity was the rule, then the econometricians were left with a difficult situation. There was no satisfactory framework to eliminate the element of subjectivity involved in functional form selection. As a consequence, there was usually a heavy bias in the choice of functional form.

Although the theory of Rosen (1974) presented the econometricians with a specification problem, it provided more insight into the hedonic price function than earlier studies. In his competitive market framework, consumers do not act as their own middle men. There is a clear distinction between buyers and sellers in the market for new goods simply for pure consumption. The consumers' basic problem is to maximize their utility $U(x,z)$ subject to a nonlinear budget constraint $x + P(z) = y$, where x is a composite good and y is the household income. The producers' basic problem is to maximize their profits $\pi = M\ P(z) - C(M,z)$, where M is the number of units and C is the cost function. In order to interpret and analyze the hedonic price function in this setting, it is useful to introduce the bid functions for consumers and the offer functions for producers. Extensions to allow for the consumers to purchase more than one unit of a model as well as different tastes across consumers are fairly straightforward in Rosen's model.

The general idea is that the consumers are bidding for goods and the bid functions depend on the attributes, given a level of utility, u, and income, i.e., BID(z;u,y). The bid function, BID(z;u,y), describes how much a consumer is willing to pay for different amounts of z given the level of utility and income. The hedonic function P(z) describes the corresponding minimum price the consumer has to pay to consume an amount z in the market. Consider a single dimension z_i, when the two curves are tangential to each other, utility is maximized. Notice that the derivative of BID wrt z_i, BID_{z_i}, is the marginal reservation demand price for attribute z_i, given u and y, and it is decreasing in z_i. In this case, $BID_{z_i} = P_{z_i}$, thus the derivative of P wrt z_i has the same interpretation in equilibrium.

On the other hand, we have the producers offering their models for sale with an offer function OFF(z; π, β). β is a shift parameter and depends on the prices of factor of production as well as the parameters of the production function. Given a level of M, and constant β , the offer function describes the the unit price of the car model corresponding to different designs of z the producer is willing to accept when M is optimally chosen. OFF_{z_i} is the marginal reservation supply price for attribute i at constant π and is increasing in z_i. Again, the profit is maximized when $OFF_{z_i} = P_{z_i}$, for all i, in equilibrium.

Rosen (1974)'s model suffers from the drawback that the consumer only purchases a single model but nevertheless, the implications are fairly interesting. First of all, the hedonic price function is the locus of points where the bidding and offer functions are tangential to each other. In other words, it reflects the equilibrium points where buyers and sellers are perfectly matched. This is an interpretation of hedonic price function in an equilibrium framework. Secondly, if the producers or firms are identical, i.e., β is the same across, the offer functions collapse to a single surface. This means that the hedonic function is identical with a unique offer function. On the other hand, if the consumers are identical, then the hedonic price function is identical with a value or bidding function.

The theory of Rosen (1974) predicts that the hedonic function is nonlinear but provides little information on the functional form. Since purely nonparametric methods are unlikely to work in the presence of dummy variables and because of the lack of a large sample size, a semiparametric model may be desirable. This is especially true when only a sub–vector of the attributes are of interest or one is interested only in the predicted price. Take the semiparametric model

$$P_i = \delta' D_i + g(z_{1i},\ldots,z_{(m-k)i}) + \epsilon_i \qquad i=1,2,\ldots,N$$

where D_i is a kx1 vector of continuous (possibly in log) or dummy variables where the characteristic price vector δ is of interest, and g is an unknown function of the remaining attributes. This is a more flexible model than the parametric ones and allows for nonlinear interactions among the remaining attributes. Robinson (1988a)'s method can be used to estimate the shadow price of attribute D_i. Besides having the desired property that the estimates are consistent and even efficient when f is indeed linear, the model can be used for prediction.

4.3 PROBLEMS WITH THE TWO–STAGE PROCEDURES

Hedonic price function can also be used to estimate demand and supply functions and Rosen (1974) has suggested a two–stage procedure. In order to motivate the use of the new procedure, we have to understand Rosen's approach and its main problems. The two–stage procedure is as follows:

(i) the first stage involves estimating the marginal prices $\hat{P}_{z_i}$, usually from a parametric hedonic price model:

$$P = f(z) + \eta$$

(ii) the second stage of the technique is to identify the demand and supply functions and involves estimating the following simultaneous equations:

Demand: $P_{z_j}(z) = DD^j(z_1,\ldots,z_m,Y_d;\ \lambda_d) + \epsilon_1$ $\qquad j=1,\ldots,m$

Supply: $P_{z_j}(z) = SS^j(z_1,\ldots,z_m,Y_s;\ \lambda_s) + \epsilon_2$ $\qquad j=1,\ldots,m$

where Y_{dd} and Y_{ss} are the variables which characterize demand and supply respectively, λ_{dd} and λ_{ss} are the parameters of interest. The two equations in the second stage are derived from the first order conditions from both the consumers' and producers' optimization problem respectively. Basically, the equations describe that the marginal price is equal to the marginal rate of substitution and marginal rate of transformation respectively.

Two special cases are of interest. If we have identical consumers, but different producers, then the hedonic function identifies the value function. Y_{dd} drops out of DD^i's and single cross–sectional observations traces out the marginal rate of substitutions. Assuming constant utility of money, we have the inverse compensated demand functions. If consumers differ and producers are identical, we have Y_{ss} dropping out from SS^j's with P(z) identifying the offer functions. Therefore, we can identify the inverse compensated supply functions.

The more realistic and interesting question is whether we can interpret DD^j's as the inverse demand functions if consumers are not alike. In this case, we have to take into account the supply side. If we have perfectly elastic supply as in the new car market, then the implicit price P_{z_j} can be treated as exogenous. In this case, DD^j's still identify the demand functions. This suggests that new car prices be used rather than used car prices. In the event of perfectly inelastic supply, we cannot interpret the DD^j's as inverse demand functions because the DD_i^j for z_j and household i, are different for different households. Unless the consumers are identical, no easy solutions exist. In the event of sluggish adjustment of supply, we have to model the market using a system of simultaneous equations as above.

The problems of estimation of the simultaneous equations system are the focus in housing economics. There is a growing literature on the econometrics of structural hedonic price model, e.g., Witt, Sumka and Erekson (1979), Brown and Rosen (1982), Mendelsohn (1984, 1985), Bartik (1987a, 1987b), Epple (1988) and Kahn and Lang (1988).

We are only interested in one z_j (e.g., fuel efficiency variable) and proceed to discuss the estimation with a single z_j. The market for new car is assumed to consist of a large number of consumers who only purchase one model. We can write the typical consumer or household problem as maximizing the utility function subject to a budget constraint:

$$\operatorname{Max}_{x,t,t^*,z,z^*} U(x,t,t^*,z_1,..,z_m,z^*)$$

subject to $x + P(z,z^*) = y$, where

z = mxl vector of observable car attributes

z^* = car attributes not observed by the econometrician

t = kxl observable taste variables

t^* = unobservable taste variables

x = a composite good

y = income of the household or consumer

$P(z,z^*)$ = sale price of the car with attributes z and z^*

the first order conditions is

$$\partial P(z,z^*)/\partial z_i = U_{z_i}(.)/U_t(.) \qquad j=1,2,..,m$$

which describes that the marginal price is equal to the marginal rate of substitution. Assuming that we are only interested in estimating the MRS of z_1, and the marginal price can be written as $P^*_{z_1} = P_{z_1}(z,z^*)$. Then we can model the willingness to pay function as

$$P^*_{z_1} = g(z_1,t;\lambda) + \epsilon_1$$

with the possibly unknown nonlinear hedonic function

$$P(z,z^*) = f(z) + \eta(z^*)$$

where λ is the parameter of interest. Many studies including Goodman (1983), estimated the hedonic function at the model level using Box–Cox transformation model. In the second stage, ols is applied to obtain the estimates for λ. Unfortunately, ols estimates from the second stage are inconsistent. ϵ_1 will be correlated with z because of the simultaneous choice problem. This problem arises because the consumer can simultaneously choose the amount of z_1 to consume as well as the marginal price $\partial P(z,z^*)/\partial z_1$. So, z is usually a function of the attributes Y_d and ϵ_1, thus correlated with ϵ_1, which rules out ols estimation for consistency. Unobservable tastes, t^* (a source of error in ϵ_1), also rule out some of the Y_d and other z's as instruments as they are correlated with ϵ_1. Obviously, in a problem of this sort, IV has to be used in estimating λ. The two–stage procedures simplifies to the following:

(i) As Stock (1985b) has noticed, since we are interested in the derivative of P wrt to z_1 and not z in this problem, we can use nonparametric techniques to estimate the derivative. Denote these nonparametric estimates as $\hat{P}_{z_1}$.

(ii) Replace $P^*_{z_1}$ by $\hat{P}_{z_1}$ and we have

$$\hat{P}_{z_1} = g(z_1,t;\lambda) + \epsilon_1 + u$$

u arises because of the replacement, and $E\hat{P}_{z_1} - P^*_{z_1} \to 0$ almost surely (a.s). Then we can apply Amemiya (1974) NL2SLS estimator. Stock (1985b) has shown that

$$N^{1/2}(\hat{\lambda}-\lambda-b_N) \xrightarrow{d} N(0,\Omega).$$

where Ω is the variance–covariance matrix. The bias b_N is of $O(a_N{}^2)$ and may be a problem in finite sample. The use of higher–order kernels of Barlett (1963) to reduce bias may be desirable. However, there may be arguments against using kernels which admits negative values since some marginal prices are believed to be non–negative a priori in cases when the attributes are "goods" rather than "bads".

Rosen's approach suffers from two major problems. The first being that there may be problems with identification of the willingness–to–pay function. Even if one is able to solve the identification problem by imposing some restrictions on the nonlinearity of the hedonic price function, it is still an undesirable procedure if one has only data from a single market as correct instruments are still lacking. Therefore, in our case, the main difficulty is the availability of the appropriate instruments.

One solution is to use instruments from other markets. If however, multi–market data set is used, one has to make the implausible assumptions that the preferences across the markets are the same. This problem also relates to general equilibrium effects. If the changes in a particular attributes are large, then some owners will be tempted to move on to another model. If the changes in the attributes are across the models, then many owners will switch models and this may alter the hedonic price structure.

This suggests that an alternative approach has to be suggested for the estimation of the costs or benefits without resorting to Rosen's two–stage method. In particular, we refer to a procedure which involves no second stage and evaluate the willingness–to–pay for fuel efficiency directly using the hedonic price function.

4.4 THE MODEL

The use of the change in property values as a measure for benefits of improved environmental quality has been discussed by a number of authors (Freeman (1979), Chapter 6, Harrison and Stock (1984)). Our estimation strategy is that of Stock (1985a) and the main purpose is to produce estimates of the expected value and variance of the benefits of a change in the fuel efficiency. As in most transport problems, we have to work at the model level. Ideally, we should multiply the attributes of each model by its market share to reflect the importance of that particular model in the market. In the absence of reliable data on the sales of each model, we have not adopted this approach. So, all the models are assumed to be of equal weight in this study.

We will first estimate the semiparametric hedonic price function by Robinson (1988a)'s method using the data on individual models. Then, a change in the fuel efficiency is assumed and we evaluate the differences between the price of the model before and after the change. To obtain the average costs, we have to aggregate these differences and then divide the total difference by the number of models. We shall now outline the model formally.

Let P and x be scalar random variables, z be (m–k–1)x1 vector of identically and independently distributed random variables. A hedonic function can be represented as

$$P_i = D_i'\delta + g(z_i,x_i) + u_i \qquad i = 1,2,\dots,N$$

where P_i is the price of a new car, z_i is a (m–k–1)x1 vector of attributes which are continuous, x_i is the measure of fuel efficiency in MPG, D_i is kx1 vector of dummy specification or performance variables, u_i is assumed to be i.i.d. $(0,\sigma^2)$ independent of z_i and D_i, and g is continuous in z_i and x_i. Let F(z,x,D) be the distribution of (z_i,x_i,D_i), and H(z,x) be the unconditional distribution of the attributes (z_i,x_i).

Our policy can be stated formally as transforming x to x^*. After the shift of the policy, we have the distribution of (z_i,x_i,D_i) as $F^*(z,x,D)$ and that of (z_i,x_i^*) as $H^*(z,x)$. We are interested in evaluating the costs (benefits) of such a mapping as the unconditional expectation of the change in p, i.e.,

$$C = E_{F^*}(P) - E_F(P) = E_{F^*}(D_i'\delta + g(z_i,x_i)) - E_F(D_i'\delta + g(z_i,x_i))$$
$$= E_{F^*}(g(z_i,x_i)) - E_F(g(z_i,x_i)) \qquad (1)$$

The second inequality is achieved by exploiting that our model is partly linear and that the policy does not affect D_i. Let us pretend that g is linear in a scalar policy variable x_i for the moment with $g(x_i) = \gamma x_i$, and that $\delta = 0$. Then we have,

$$P_i = \gamma x_i + \eta_i \qquad i = 1,2,\dots,N$$

The cost estimator is

$$\hat{C} = N^{-1}\Sigma_i\{g(x_i^*)-g(x_i)\}$$

Denote $\hat{\gamma}$ as the ols estimate, the $\bar{x}$ as the sample mean. Replacing $g(x_i)$ by γx_i, we have

$$\hat{C} = N^{-1}\Sigma_i(\hat{\gamma}x_i^* - \hat{\gamma}x_i) = \hat{\gamma}(\bar{x}^* - \bar{x}) = \hat{\gamma}\Delta\bar{x} = \Sigma_j \Delta x(x_j/\Sigma_i x_i^2)\, P_j$$
$$= N^{-1}\Sigma_j q(x_j)\, P_j$$

with $\Delta\bar{x} = (\bar{x}^* - \bar{x})$ and $q(x_j) = N\Delta x(x_j/\Sigma_i x_i^2)$. This says that the estimate is a sum of the product of a function of x_j and P_j implying that $\hat{C}$ is a linear estimator. If the policy has the effect of altering the observation cell, than this has to be taken into account. For easy expositions, let us introduce a simplified model with only a scalar dummy variable d_i and policy variable x_i, i.e,

$$\text{Model 1:} \qquad P_i = \alpha d_i + \gamma x_i + \epsilon_i \qquad i=1,2,\dots,N$$

Assume that α is known for the time being. We may want to rewrite the model as

$$P_i - \alpha d_i = \gamma x_i + \epsilon_i \qquad i=1,2,\dots,N$$
$$p_i^+ = \gamma x_i + \epsilon_i$$

with $P_i^+ = P_i - \alpha d_i$ and the cost estimator which corrects for cell–specific effect is

$$\hat{C}^+ = N^{-1}\Sigma_i\{g^+(x_i^*) - g^+(x_i)\}.$$

But $g^+(x_i) = \hat{\gamma}^+ x_i$, while $\hat{\gamma}^+$ is the ols estimates from regressing $P_i^+ = P_i - \alpha d_i$ on x_i. From above, we have

$$\hat{C}^+ = \hat{\gamma}^+(\bar{x}^* - \bar{x})$$
$$= N^{-1}\{\Sigma_j q(x_j)(P_j - \alpha d_j)\} \qquad (2)$$

Now, assume that α is not known, then it can be estimated by ols,

$$\alpha^+ = \text{argmin}_\alpha\, \Sigma_i(P_i - \alpha d_i - g^+(x_i))^2 \qquad (3)$$
$$= \text{argmin}_\alpha\, \Sigma_i(p_i - \alpha d_i - \gamma^+ x_i)^2 .$$

Notice that in (2), we have the sum of the product of $(P_i - \alpha d_i)$ and q_i. Stock

(1985a) has outlined a similar strategy for the nonparametric cost estimates which is also linear in P. We now relax the assumption that g is linear and assume that it is fairly smooth by imposing some smoothness conditions on H and H^*.

It is also assumed that the conditional expectations of d_i on x and P_i on x_i are continuous in x_i. The unknown $E_{F^*}(g(x_i))$ and $E_F(g(x_i))$ can be estimated by the Nadaraya–Watson kernel estimator described by equation (7) in Chapter 2, i.e., :

$$\tilde{g}^+(x_i) = \Sigma_j w_j \tilde{P}_j^+ = \Sigma_j K_{ij} \tilde{P}_j^+ / \Sigma_j K_{ij} \qquad (4)$$

$$\tilde{g}^+(x_i^*) = \Sigma_j w_j^* \tilde{P}_j^+ = \Sigma_j K_{ij}^* \tilde{P}_j^+ / \Sigma_j K_{ij}^*$$

with the semiparametric cost estimates:

$$\tilde{C}^+ = N^{-1}\Sigma_i\{\tilde{g}^+(x_i^*) - \tilde{g}^+(x_i)\} = N^{-1}\Sigma_i\Sigma_j(w_j(x_i)^* - w_j(x_i))\tilde{P}_j^+$$

$$= N^{-1}\Sigma_j W_j \tilde{P}_j^+$$

where $\tilde{P}_j^+ = P_j - \tilde{\alpha}^+ d_j$ and the unknown α is estimated as

$$\tilde{\alpha}^+ = \operatorname{argmin}_\alpha \Sigma_i (P_i - \alpha d_i - \tilde{g}^+(x_i))^2$$

Using (4), we have

$$\alpha^+ = \operatorname{argmin}_\alpha \Sigma_i (P_i - \alpha d_i - \Sigma_j W_j (P_j - \alpha d_j))^2$$

In matrix from, with M = I–W and the corresponding D and P vectors for d_i and P_i, we have

$$\alpha^+ = \operatorname{argmin}_\alpha ((I-W)P-(I-W)D)'((I-W)P-(I-W)D)$$

$$= \operatorname{argmin}_\alpha (P-D)'M^2 (P-D)$$

$$= (D'M\,D)^{-1} D'M^2 P$$

It should be obvious now why we have to assume that the conditional expectations E[d|x] and E[P|x] have to be continuous in x. It is also important that the matrix $D'M^2D$ is positive definite. Returning to Model 1, taking conditional expectations on x, we have

$$E[P_i | x_i] = \alpha E[d_i | x_i] + \gamma x_i \qquad i=1,2,\ldots,N \quad (5)$$

Subtracting (5) from Model 1, we have

$$P_i - E[P_i|x_i] = \alpha(d_i - E[d_i|x_i]) + \epsilon_i \quad i=1,2,\ldots,N$$

$\tilde{\alpha}^+$ is then obtained by regressing $P_i - \tilde{E}[P_i|x_i]$ on $d_i - \tilde{E}[d_i|x_i]$, i.e., regressing My on (M D) if the conditional expectation is estimated by the method of kernel. The existence of the conditional expectations is essential for the non–singularity of (D'M^2 D). The positive definiteness of $E\{(d_i - E[d_i|x_i])(d_i - E[d_i|x_i])'\}$ is in fact given as the identifying condition for the Robinson (1988a)'s estimator that we will discuss below. Of course, if we have no cell specific effects, $\alpha = 0$, we are reduced to a nonparametric framework and the problem does not exists. As long as g is smooth, we have

$$\hat{C} = N^{-1}\Sigma_i\{g(x_i^*) - g(x_i)\} = N^{-1}\Sigma_i\Sigma_j(W_j(x_i)^* - W_j(x_i))P_j$$
$$= N^{-1}\Sigma_j W_j P_j$$

The parallel of the parametric and nonparametric estimators is apparent from the discussion above regarding the estimation of the expected value of dependent variable. Stock (1985a) has shown that, with non–negative kernel bounded above and below, and under the usual conditions on bandwidth, i.e., $a \to 0$, $Na^d \to \infty$ as $N \to \infty$, the fixed effect estimator, $\tilde{\alpha}^+$, and the semiparametric cost estimates, $\tilde{C}^+$, are consistent. Furthermore, he established that $N^{1/2}(\tilde{C}^+ - E[\tilde{C}|(x_i,d_i), \; i=1,2,\ldots,N])$ is asymptotically normally distributed with mean 0 and covariance matrix V where

$$V = P(1 + R'L^{-1}R)$$

with $R = N^{-1}W'D$, $L = N^{-1}(\tilde{D}'\tilde{D})$ and $\tilde{D} = MD$. The individual nonparametric estimate for conditional expectations, $(\tilde{m}_i^* - \tilde{m}_i)$, converges to $(m_i^* - m_i)$ at the rate of $(Na^d)^{1/2}$. The results of Stock (1984, 1985a) show that the average of these individual estimates converges to its expectation and achieves the usual rate of $\sqrt{N}$ – a fact which Stock (1984, 1985a) emphasized. This implies that although the individual estimates may converge at a slower rate, but taking the average of these estimates, we are able to exploit more information from the whole sample and thus get an estimate which converges at the usual rate.

The usual root–N convergence rate can also be achieved for the parameter of interests, α (for dummy variables d) or γ (for continuous variables x), in a partial linear model: linear in d and possibly nonlinear in z and x. Robinson (1988a), using higher–order kernels, has shown that this is possible. Speckman (1988), has also exploited the correlation of z and x with d, and showed that the estimator for α^+ achieves the usual rate of convergence.

It may also be useful to relax the assumption that w is a non–negative functions by using higher–order kernels. This suggests that if one is interested in the dummy variables, then as Robinson (1988a) has shown with higher–order kernels,

$$N^{1/2}(\tilde{\alpha}^+ - \alpha^+) \xrightarrow{d} N(0,\Omega)$$

where $\tilde{\alpha}^+ = (\tilde{D}'\tilde{D})^{-1}\tilde{D}'\tilde{P}$

$$\tilde{D} = M\ \mathrm{diag}(I_1,I_2,\ldots,I_N)\ D$$

$$\tilde{P} = M\ \mathrm{diag}(I_1,I_2,\ldots,I_N)\ P$$

$$I_i = I(\tilde{f}_i > b)$$

$$\Omega = (\tilde{D}'\tilde{D})^{-1}$$

The indicator function is a device to trim out small $\tilde{f}$, the density estimates, in order to avoid technical difficulties. Further discussions of the Robinson (1988a)'s estimator are presented in Chapter 7. At present, we confine our discussions to the estimation of cost in this chapter. Recall that the cost estimate is

$$\tilde{C}^+ = N^{-1}\Sigma_j W_j \tilde{P}_j^{\,+} = N^{-1}\Sigma_j W_j (P_j - \tilde{\alpha}^+ d_j)$$

In matrix terms, we have

$$\tilde{C}^+ = N^{-1}[W' - D(\tilde{D}'\tilde{D})^{-1}\tilde{D}'M\ \mathrm{diag}(I_1,\ldots,I_N)]\ y = N^{-1}\tilde{A}'\tilde{P}$$

where $\tilde{A} = [W' - D(\tilde{D}'\tilde{D})^{-1}\tilde{D}'M\ \mathrm{diag}(I_1,\ldots,I_N)]'$. A consistent covariance matrix estimator for $\tilde{C}^+$ is

$$\tilde{V} = \tilde{\sigma}^2 N^{-1}\tilde{A}\tilde{A}'$$

where $\tilde{\sigma}^2 = N^{-1}\Sigma_i(\tilde{P}_i - \tilde{\alpha}^+\tilde{d}_i)^2$. Stock (1985a) has also suggested and evaluated the performance of three other choices of consistent V, which have computational advantage. We have adopted Stock (1985a)'s third suggestion, i.e.,

$$\tilde{V} = \tilde{\sigma}^2 (1+R'L^{-1}R)(N^{-1}\Sigma_j W_j).$$

4.5 RESULTS

The same set of data and attributes as in Chapter 3 is used. The attributes are size (z_1), spaciousness (z_2), power (z_3), acceleration (z_4) and fuel efficiency (z_5). We have included several dummy variables to capture possible taste differences and to account for the two–stage hypothesis of Griliches. The individual dummy will take the value 1 if the car possess the following facilities or characteristics: air condition (d_1), sunroof (d_2), electric window (d_3), automatic transmission (d_4), 5 speed (d_5), power steering (d_6), unleaded (d_7) and British made (d_8). We have therefore allowed for the possibility that the change in fuel efficiency standard affects these characteristics. This assumption is crucial for the identification of the parameters. If the conditional expectations do not exist, the parameters are not identified and the estimates have no meaning. In that case, we can take advantage of the fact and a nonparametric policy analysis can be conducted instead. All the coefficients, with the exception of the unleaded and British made dummies, are expected to be positive as these are desired attributes.

Table 4.1 reports the results for the estimation of the hedonic price function. Following previous research, we present results for hedonic price functions of different parametric specifications. The first four columns of the results correspond to the hedonic function with level, log–log, semi–log and inverse–log specifications respectively. As expected, the semi–log hedonic model has the best fit based on R^2, F statistics and in terms of the number of coefficients which are consistent with a priori belief. Indeed, when the coefficients are negative in the semi–log specification, the t–ratios are not statistically significant at the 1% level. Unfortunately, the policy variable – fuel efficiency (z_5) has negative coefficients in three out of the four specifications. Two explanations are offered. First of all, multicollinearity may be the source of the problem. Secondly, the relationship may well be nonlinear and we have not allowed a flexible enough formulation. In particular, cross–product terms have not been allowed into the models. It is most likely that the problem arises from these two sources simultaneously giving rise to what is known as nonlinear singularity.

The last three columns are the results for the semiparametric hedonic model. The results reported are for second, fourth and sixth order kernel estimates corresponding to the bandwidth $a = N^{-(1/(2\ell+5))}$, where ℓ is the order of the kernel. This bandwidth is chosen because it is known that the MSE of the density estimates is minimized with a_{opt} = constant x $N^{-(1/(2\ell+5))}$. Of course, the bandwidth is only subjectively chosen to be proportional to a_{opt} and may not be the "optimum" bandwidth, because it may indeed be greater or smaller than the one that we have chosen.

The question is: how sensitive is the estimates to the bandwidth? We plotted the estimates for each dummy variable corresponding to H_2 and H_4 for the bandwidth interval [0.1, 2.0] in Figures 1 and 2 respectively.

First of all, let us look at the plots of the estimates in Figures 4.1(a) to (h). Although the estimates are very smooth over the range of bandwidth, these figures demonstrate that the choice of bandwidth is very crucial. The only consolation is that the estimates are fairly well behaved and they fluctuate less in the range of [0.5,0.7]. with the exception of five speed and automatic transmission dummy variables.

The picture from Figures 4.2(a) to (h) is very different. Since the higher–order kernels can take negative values, the plots are not as smooth as those in Figures 4.1. The plots are still fairly well behaved in the interval [0.5,0.7].

The most interesting results are those for H_2 and H_4. Higher–order kernels have the advantage of bias–reduction and perhaps as a consequence, the results in the last column are consistent with a priori belief. Comparing the estimates across the three kernels, it seems to be that the larger the standard error, the larger the change in the magnitude as we increase the order.

Some conclusions can be drawn from these results and plots. First of all, there is little doubt that most of the attributes are desired, with the exception of sunroof and automatic transmission. Secondly, it would be fair to say that the interpretation on the British dummy is unfavourable to the local manufacturers.

The interpretation on the unleaded dummy is not so clear cut although it appears that the evidence is against those cars which can run on both leaded and unleaded petrol.

Turning to the cost estimates in Table 4.2, we have reported results corresponding to both parametric and semiparametric specifications for general increases in the fuel efficiency. We consider a general increase of 10 and 20%. The level specification has the highest estimate with both the log–log and inverse–log failing to detect any changes thus giving zero estimates.

The semiparametric estimates are hard to interpret because of the sensitivity to the bandwidth choice. Figure 4.3(a) presents the plot of the cost estimates for H_2 (cost1) and H_4 (cost2). The upper bound for cost1 is 17.1740 and the plot is fairly smooth as compared to the more erratic cost2. We should remind the readers that Stock's theorem does not hold for non–negative kernels. Looking at Figure 4.3b for the plot of the absolute t–ratios, we can see that the t–ratios are fairly sensitive to the bandwidth used. The unpredictability of the t–ratios for cost2 is again caused by the use of trimming and negative weights.

Figure 4.4(a) and (b) are the plots of the cost estimates and t–ratios. We can see that the conclusion is fairly similar to the previous one: t–ratios of cost1 indicate that the estimates are statistically significant while that of cost2 are generally insignificant

We have also presented the estimates for the case where there is imposition of minimum efficiency standards in Figure 4.5. We can see that the cost estimates are also sensitive to the bandwidth. The message seems to be that for the data set on hand, the hedonic marginal price estimates and cost estimates obtained are actually not as ideal as it were unless some method of bandwidth selection criterion is defined. It is understood that the choice of bandwidth actually corresponds to choice of the functional form in finite sample. According to the asymptotic theory, the choice of bandwidth does not make any difference to the rate of convergence and the curse of dimensionality is not a problem in semiparametric problems. Empirically, these issues have to be dealt with before one can have any faith in the estimates. There are, however, exceptional circumstances, e.g., when one has a very large data set which itself poses a computational problem, or when the estimates are fairly insensitive to the bandwidth. Furthermore, we have to decide on which order of kernel to use which in turn depends on the smoothness of the unknown regression functions. The estimates here have really provided a very good example against a subjective rule–of–the–thumb method for bandwidth selection. Therefore, we use the cross–validation (CV) criterion function suggested by Robinson (1988a) for the selection of bandwidth:

$$\mathrm{a_{CV}} = \mathrm{argmin}_{\mathrm{a},\delta} \Sigma_{\mathrm{i}}[\tilde{\mathrm{P}}_{-\mathrm{i}} - \delta\,(\tilde{\mathrm{D}}_{-\mathrm{i}})]^2$$

The results provided in Table 4.4 are obtained using the bandwidth and the order of kernel which minimize the CV function. The signs of the semiparametric hedonic function estimates are all consistent to a priori belief. For policy analysis, we have investigated the case when the minimum fuel efficiency standard of 35 MPG is imposed. Since the semiparametric estimates can be downward biased and therefore, we have also reported the cross–validated nonparametric estimates. The semiparametric results suggest that the average willingness–to–pay is just 31.0 Sterling pounds per car. The statistical significant estimated value of 657.8 Sterling pounds per car seems too high a value.

4.6 CONCLUSION

In summary, our findings suggest that the semiparametric estimates are too sensitive to the bandwidth. We argue in favour of the less–subjective but computational more expensive cross–validated estimates. Unlike the semiparametric estimates, most of the ols estimates for the hedonic price function are not consistent with a priori belief. If this is an indication of misspecifications, then we should not place too much faith on the ols cost estimates.

The cross–validated bandwidths were used to calculate the hedonic price and cost estimates. The advantage of the semiparametric model over the quadratic transformation model is that we can use the semiparametric model for prediction. Since the semiparametric model is less prone to misspecifications, it is useful as a first stage estimation procedure for secondary hedonic applications in many areas of economics.

On the other hand, the semiparametric cost estimates could be downward biased due to the inclusion of unrelated dummy variables although we have chosen them carefully. However, the semiparametric estimates are all consistent with a priori belief and based on this, it is fair to suggest that the cost estimates are realistic and perhaps close to the true values. Finally, the nonparametric and semiparametric estimates suggest that there is at least 31 sterling pounds gain per car if a minimum standard of 35 MPG is imposed.

FIGURE 4.1a PLOT OF SEMIPARAMETRIC H_2 ESTIMATES AGAINST BANDWIDTH: AIR CONDITIONED

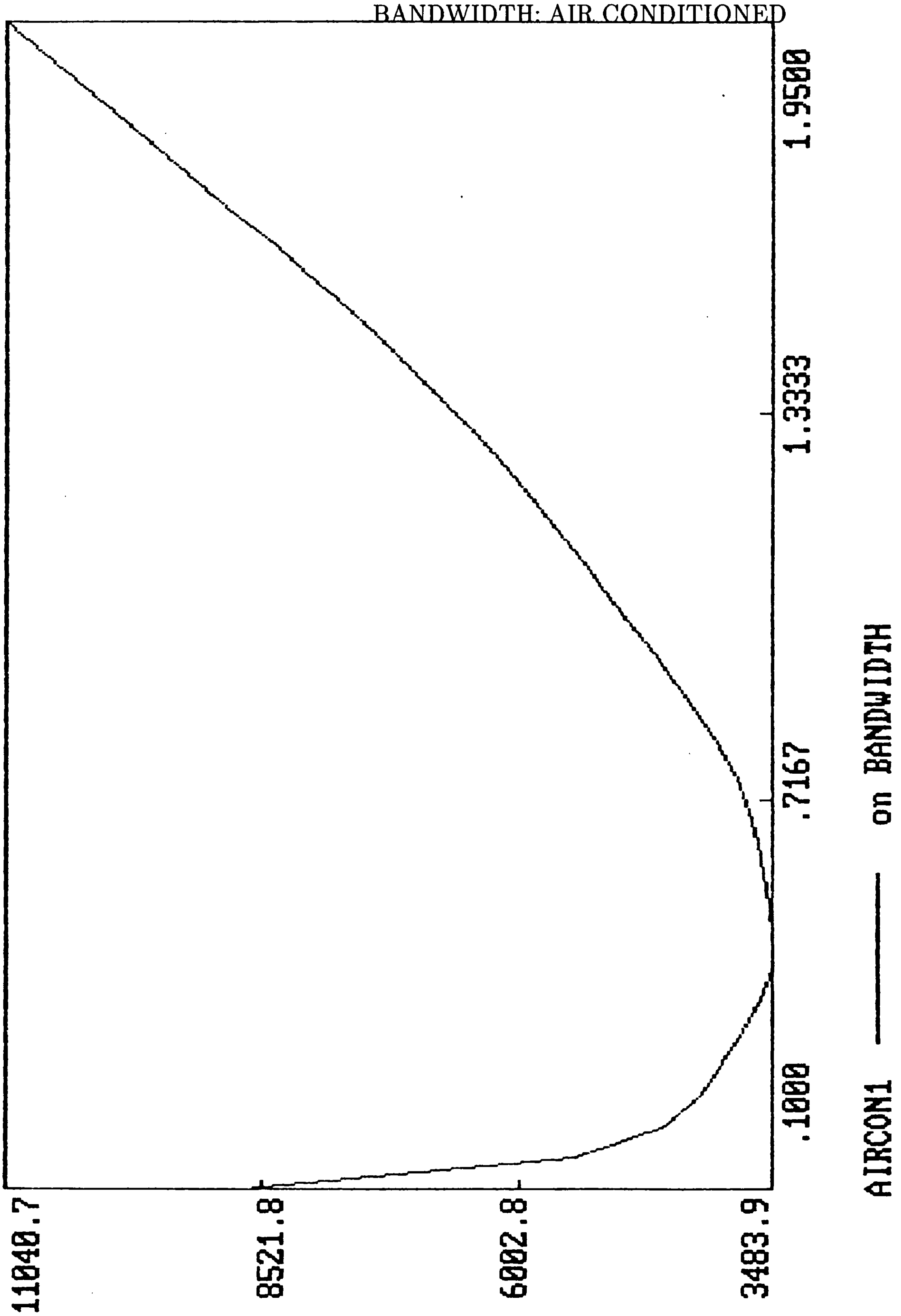


FIGURE 4.1b PLOT OF SEMIPARAMETRIC H_2 ESTIMATES AGAINST BANDWIDTH: SUNROOF

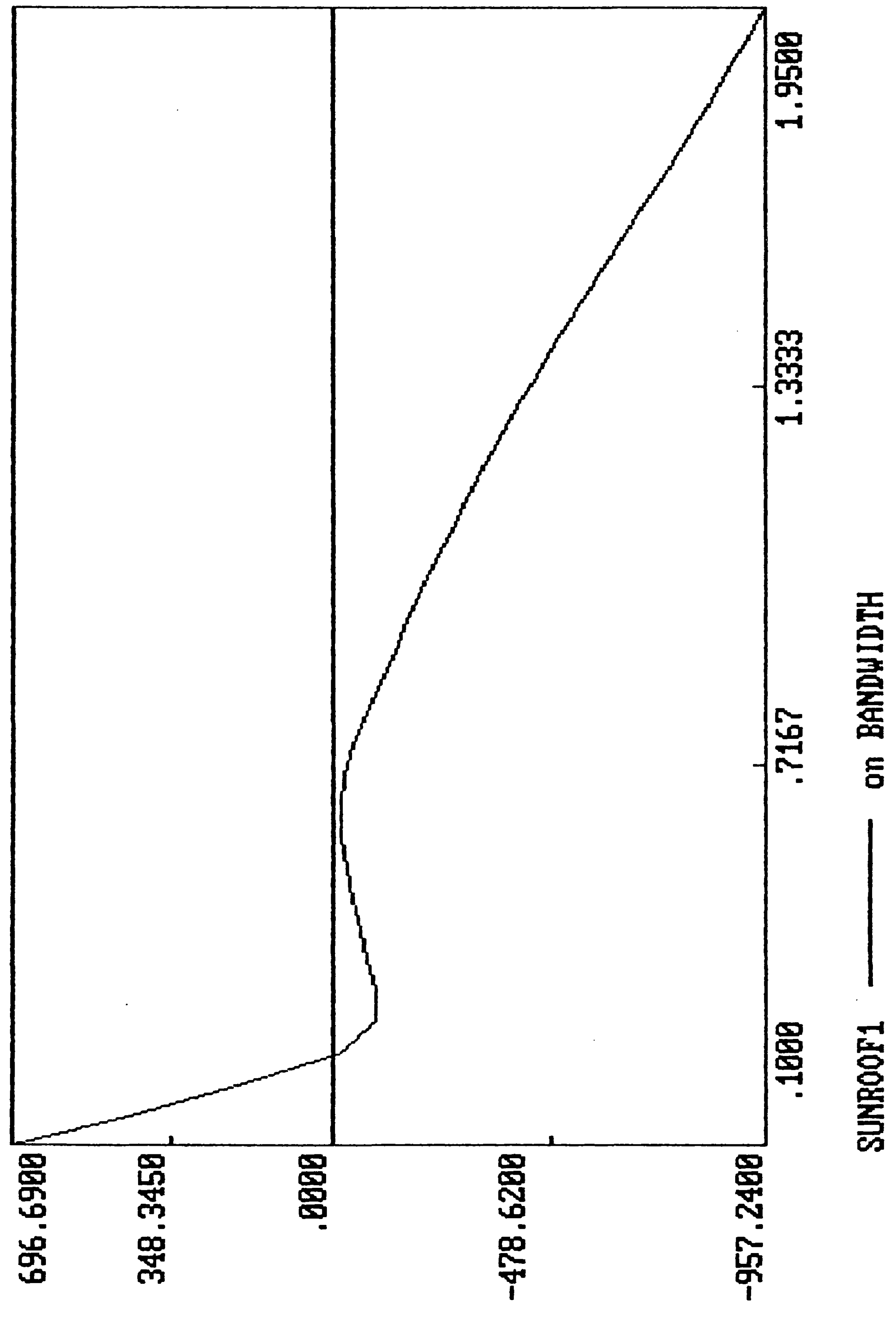


FIGURE 4.1c PLOT OF SEMIPARAMETRIC H_2 ESTIMATES AGAINST BANDWIDTH: ELECTRIC WINDOW

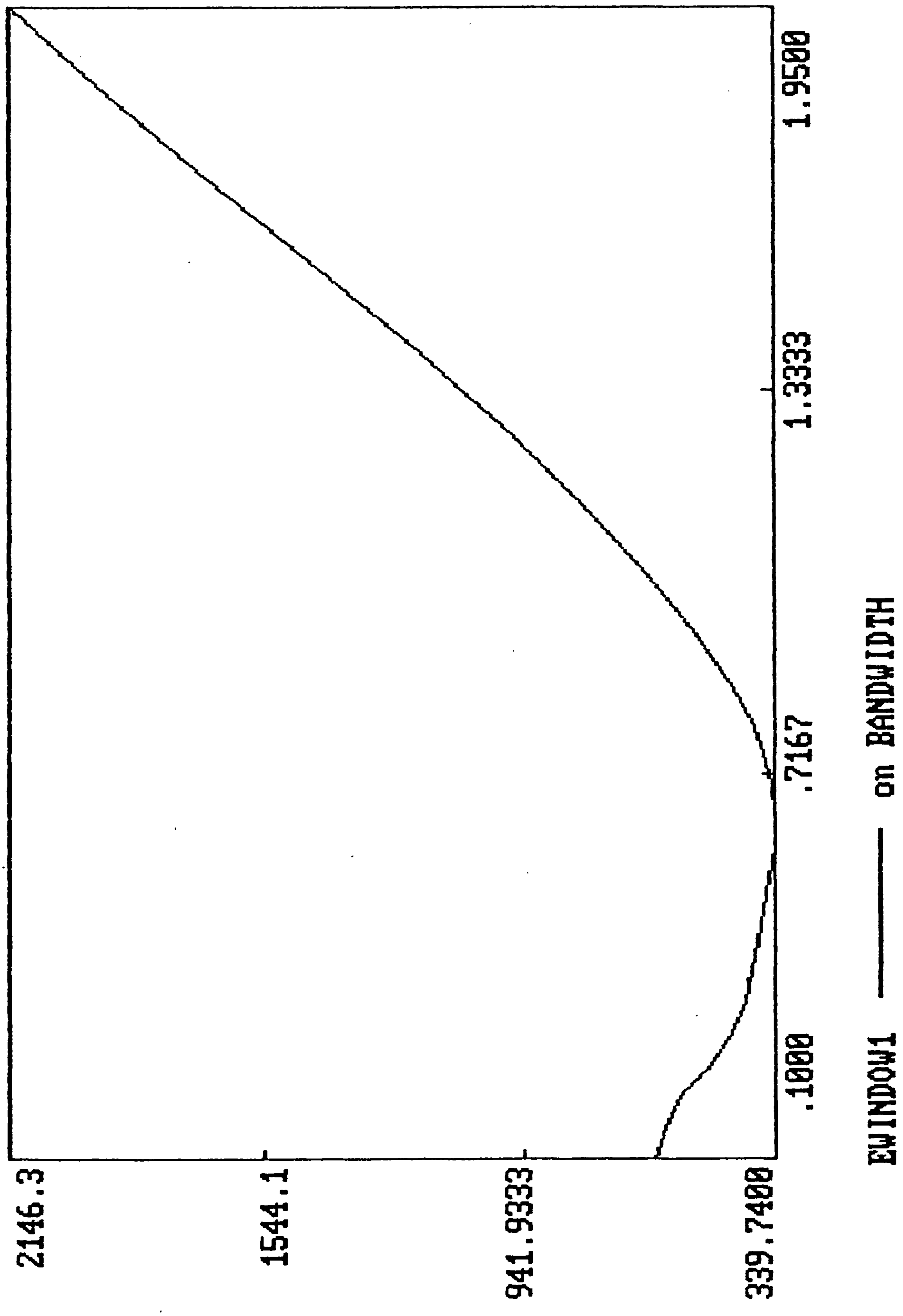


FIGURE 4.1d PLOT OF SEMIPARAMETRIC H_2 ESTIMATES AGAINST BANDWIDTH: AUTOMATIC TRANSMISSION

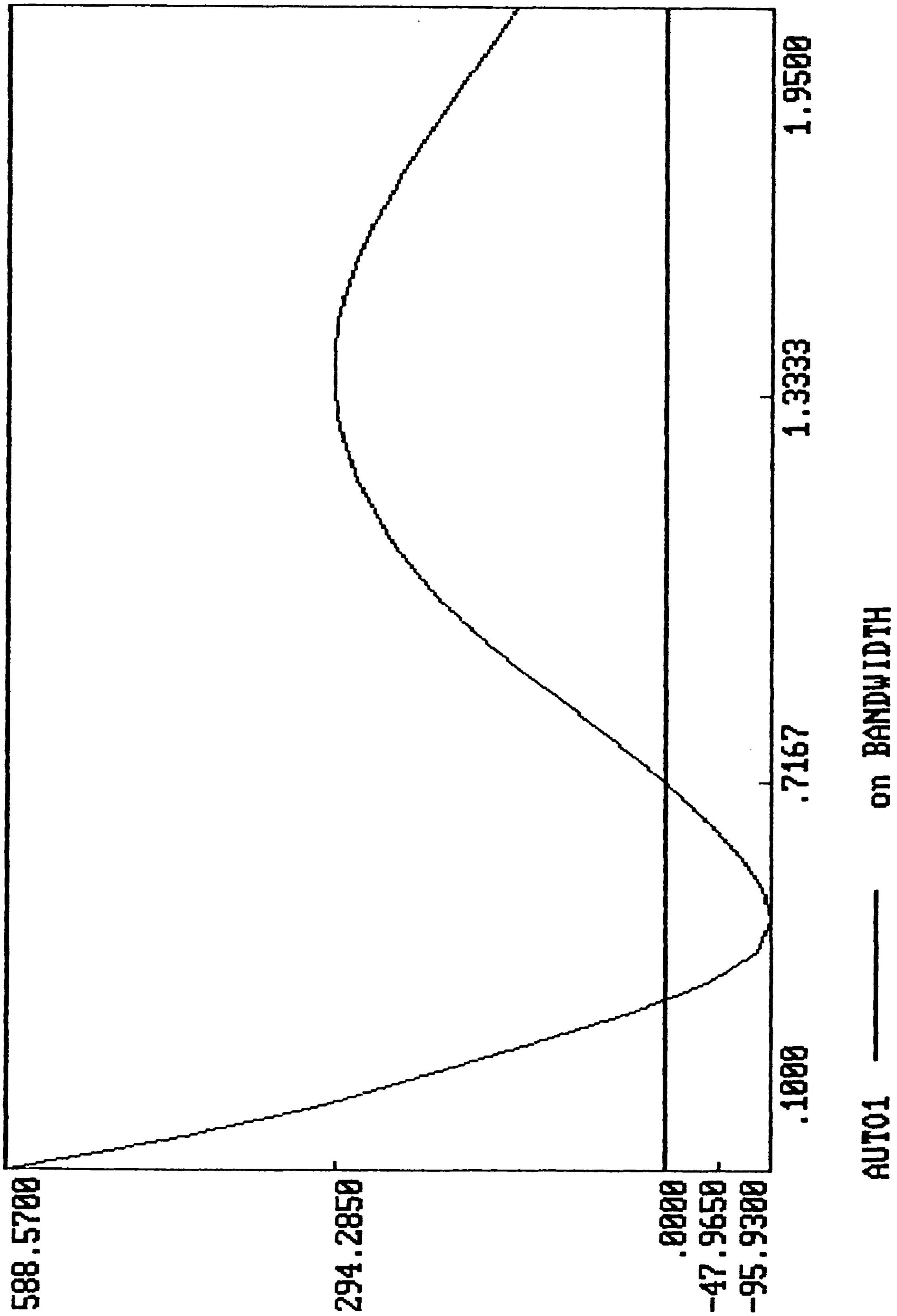


FIGURE 4.1e PLOT OF SEMIPARAMETRIC H_2 ESTIMATES AGAINST BANDWIDTH: 5 SPEED

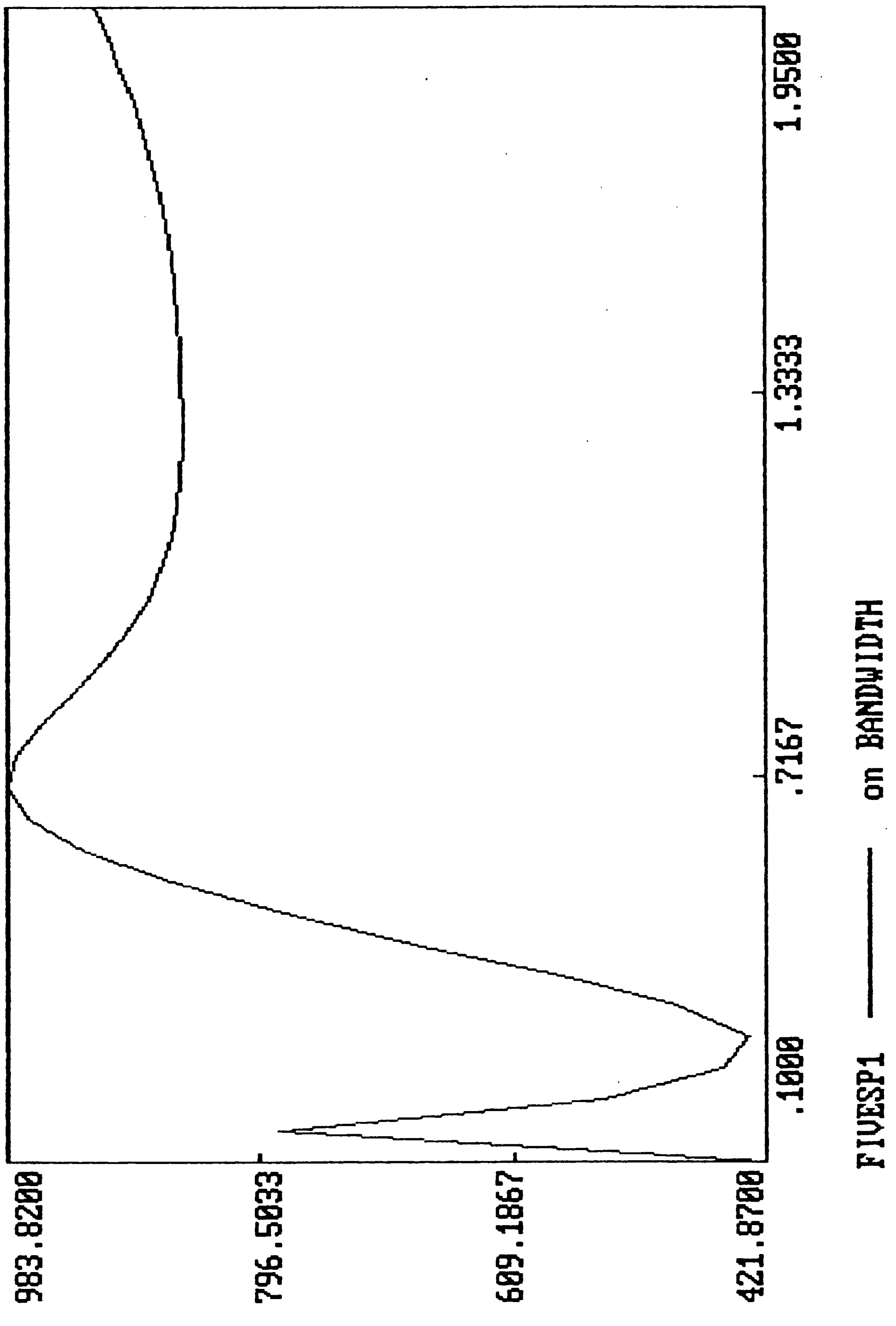


FIGURE 4.1f PLOT OF SEMIPARAMETRIC H_2 ESTIMATES AGAINST BANDWIDTH: POWER STEERING

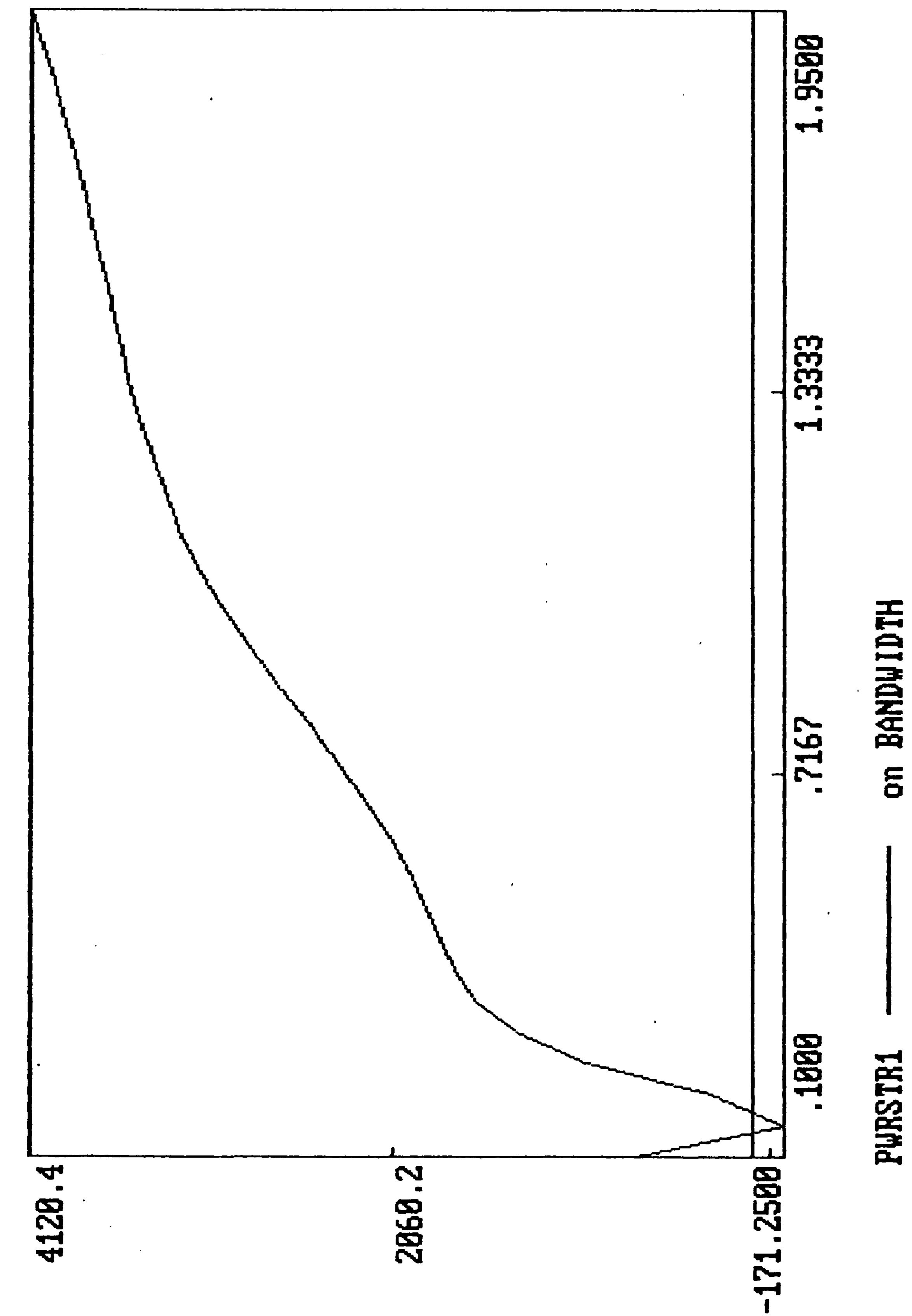


FIGURE 4.1g PLOT OF SEMIPARAMETRIC H_2 ESTIMATES AGAINST BANDWIDTH: UNLEADED

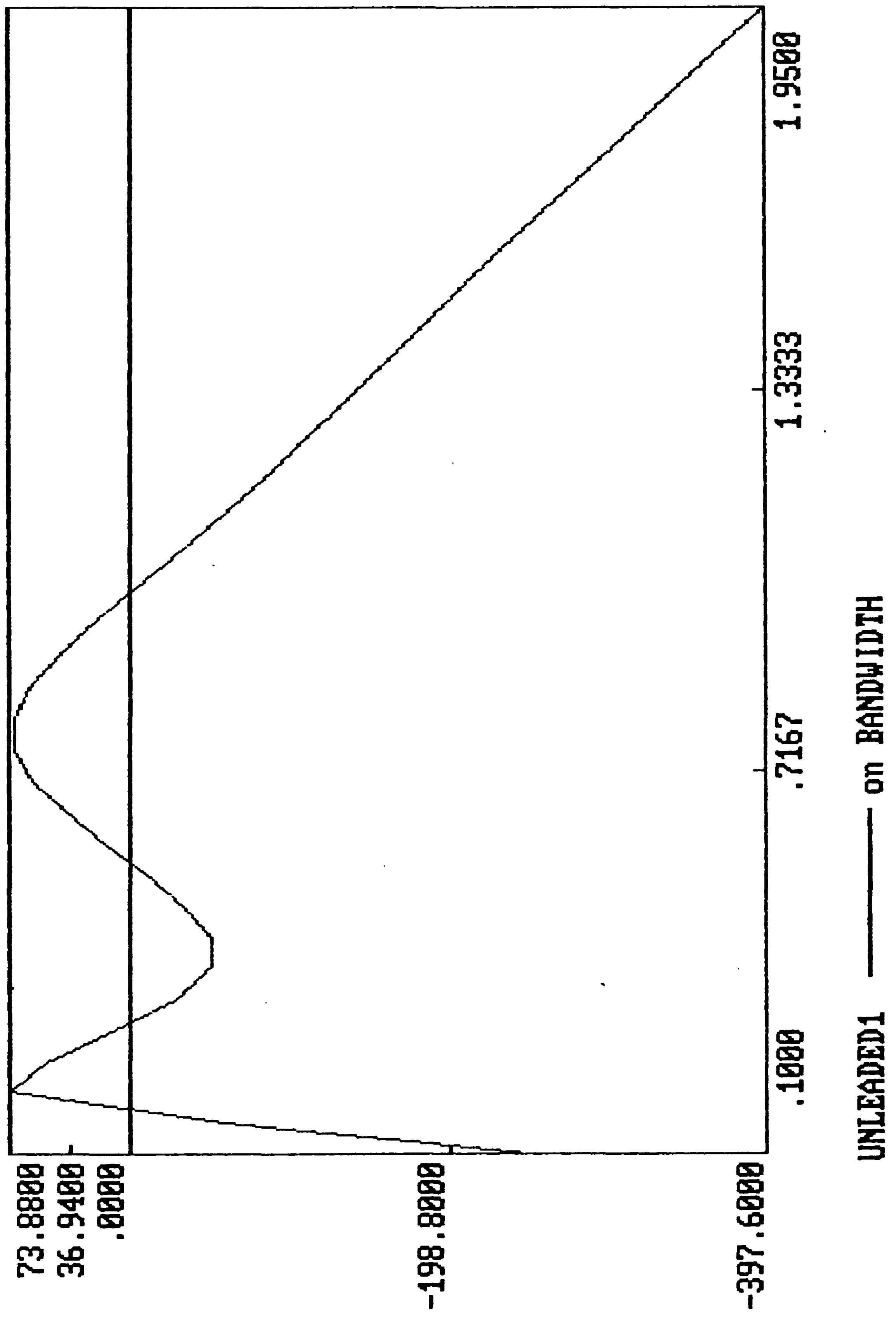


FIGURE 4.1h PLOT OF SEMIPARAMETRIC H_2 ESTIMATES AGAINST BANDWIDTH: BRITISH MADE

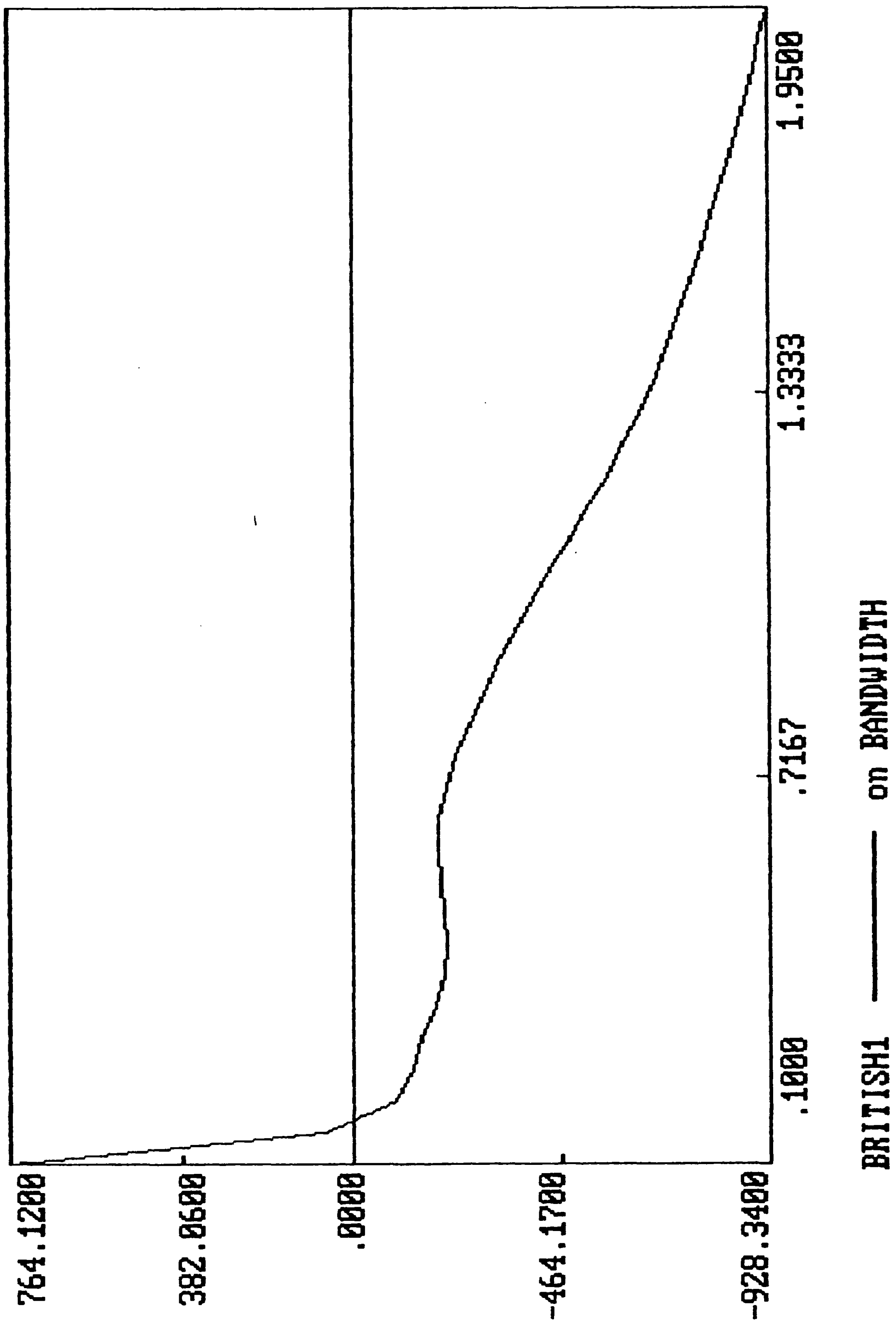


FIGURE 4.2a PLOT OF SEMIPARAMETRIC H_4 ESTIMATES AGAINST BANDWIDTH: AIR CONDITIONED

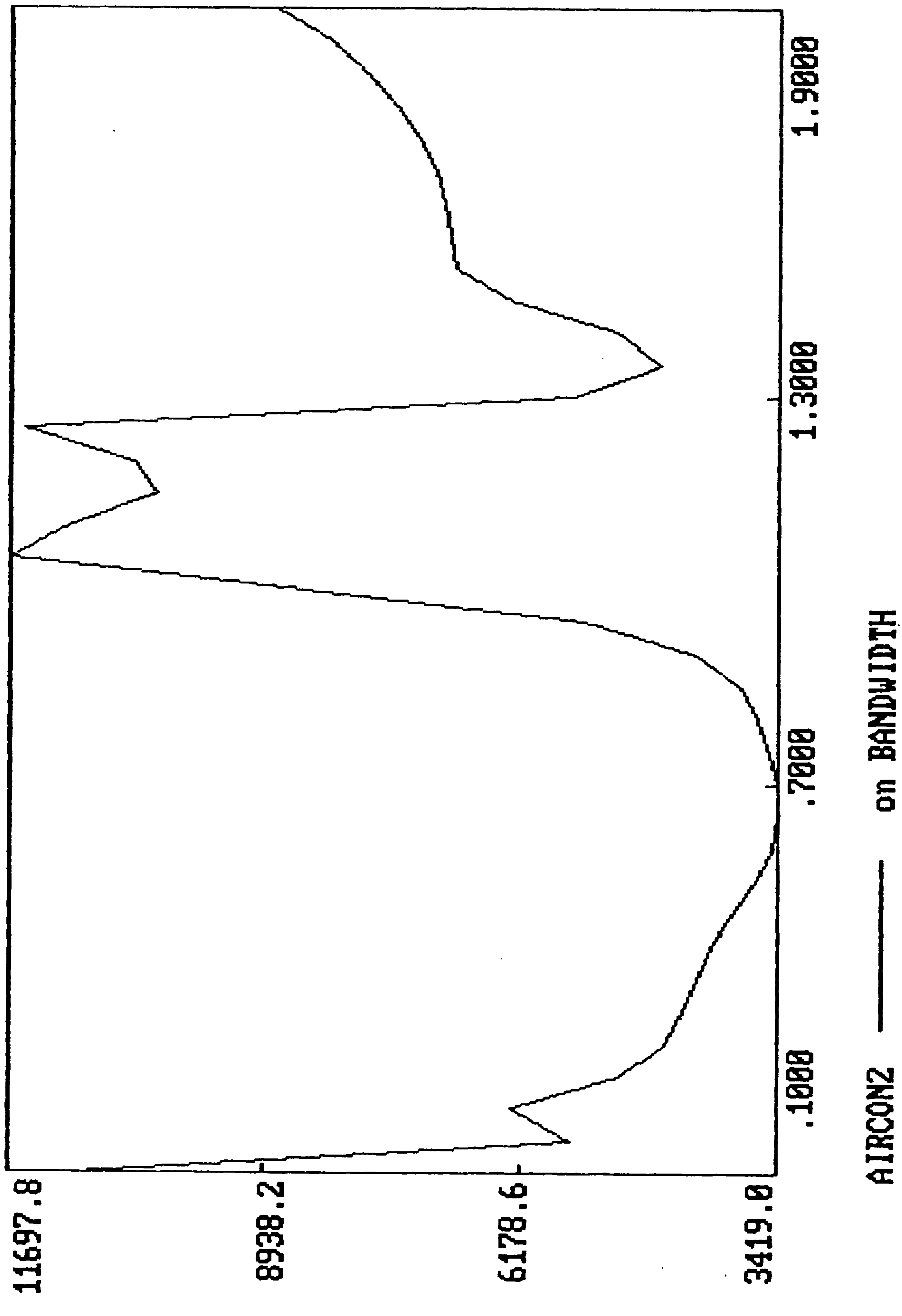


FIGURE 4.2b PLOT OF SEMIPARAMETRIC H_4 ESTIMATES AGAINST BANDWIDTH: SUNROOF

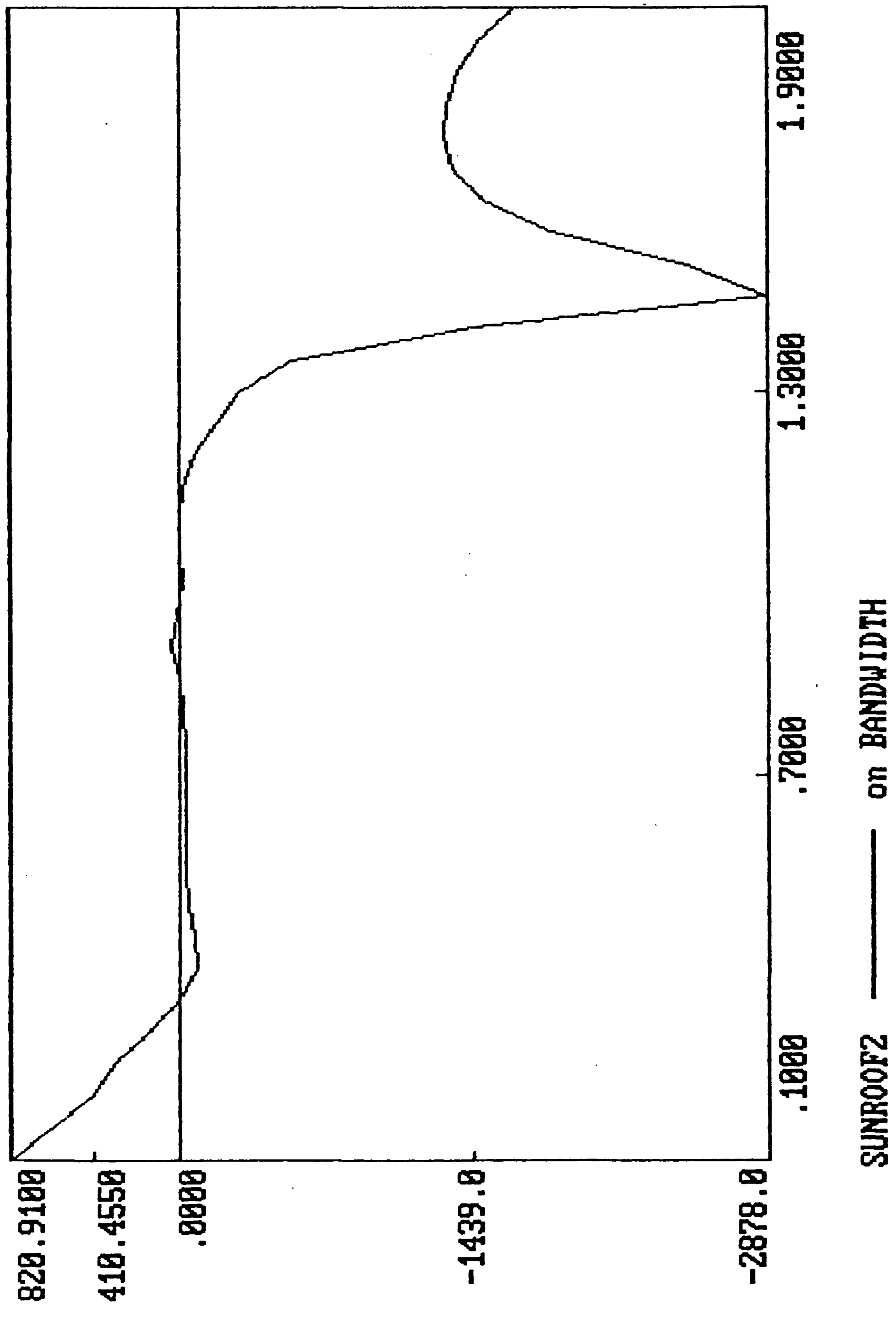


FIGURE 4.2c PLOT OF SEMIPARAMETRIC H_4 ESTIMATES AGAINST BANDWIDTH: ELECTRIC WINDOW

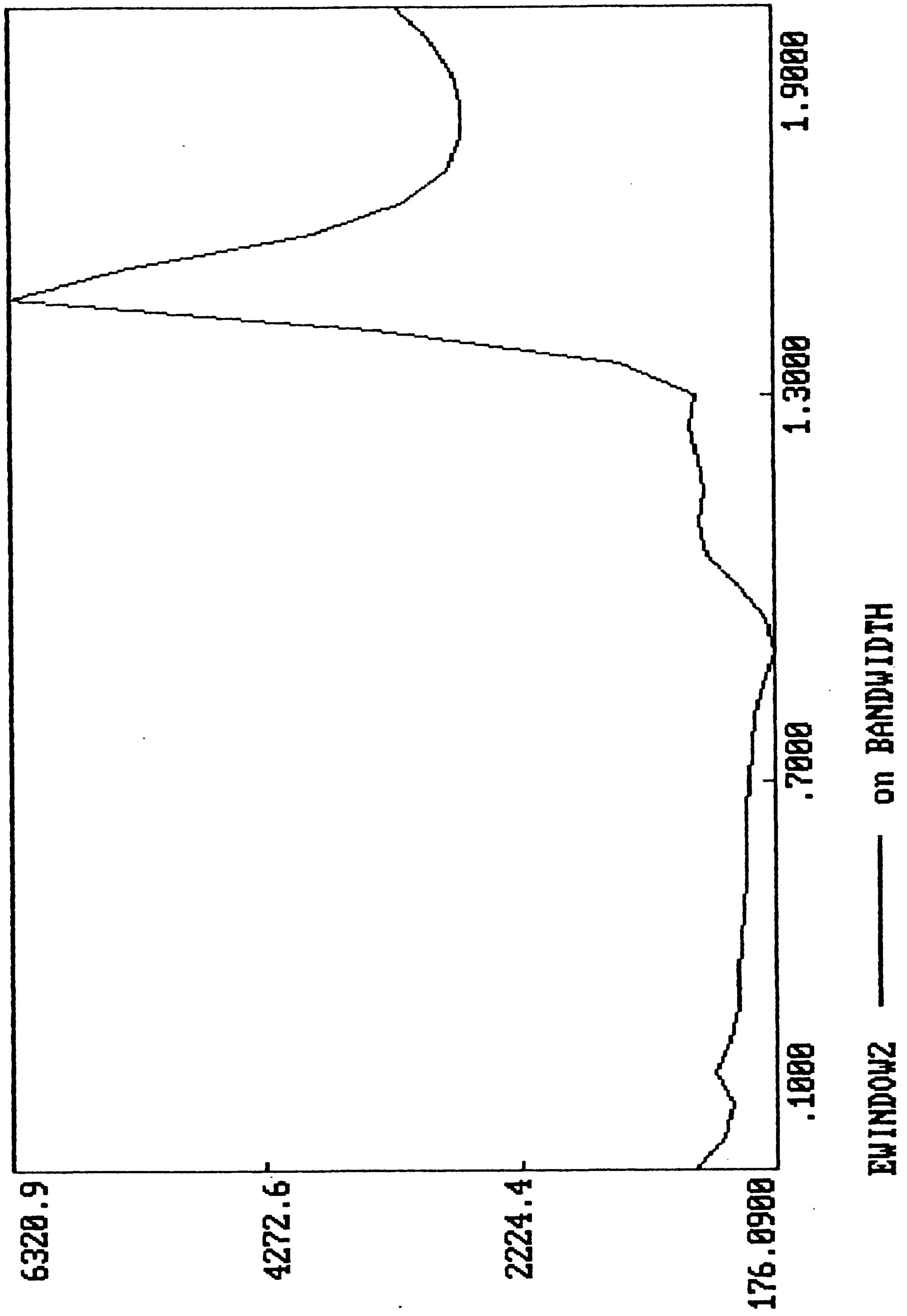


FIGURE 4.2d PLOT OF SEMIPARAMETRIC H_4 ESTIMATES AGAINST BANDWIDTH: AUTOMATIC TRANSMISSION

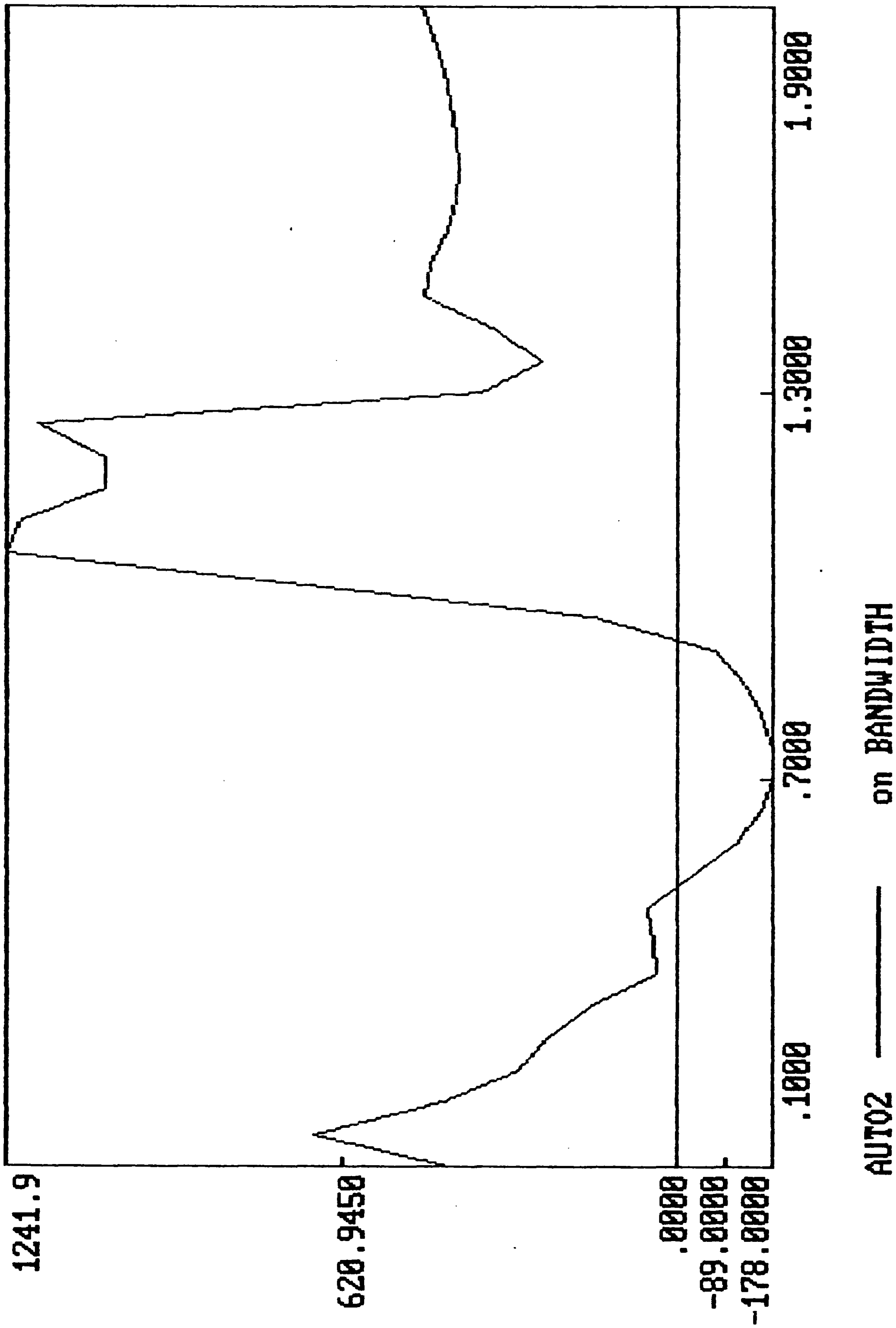


FIGURE 4.2e PLOT OF SEMIPARAMETRIC H_4 ESTIMATES AGAINST BANDWIDTH: 5 SPEED

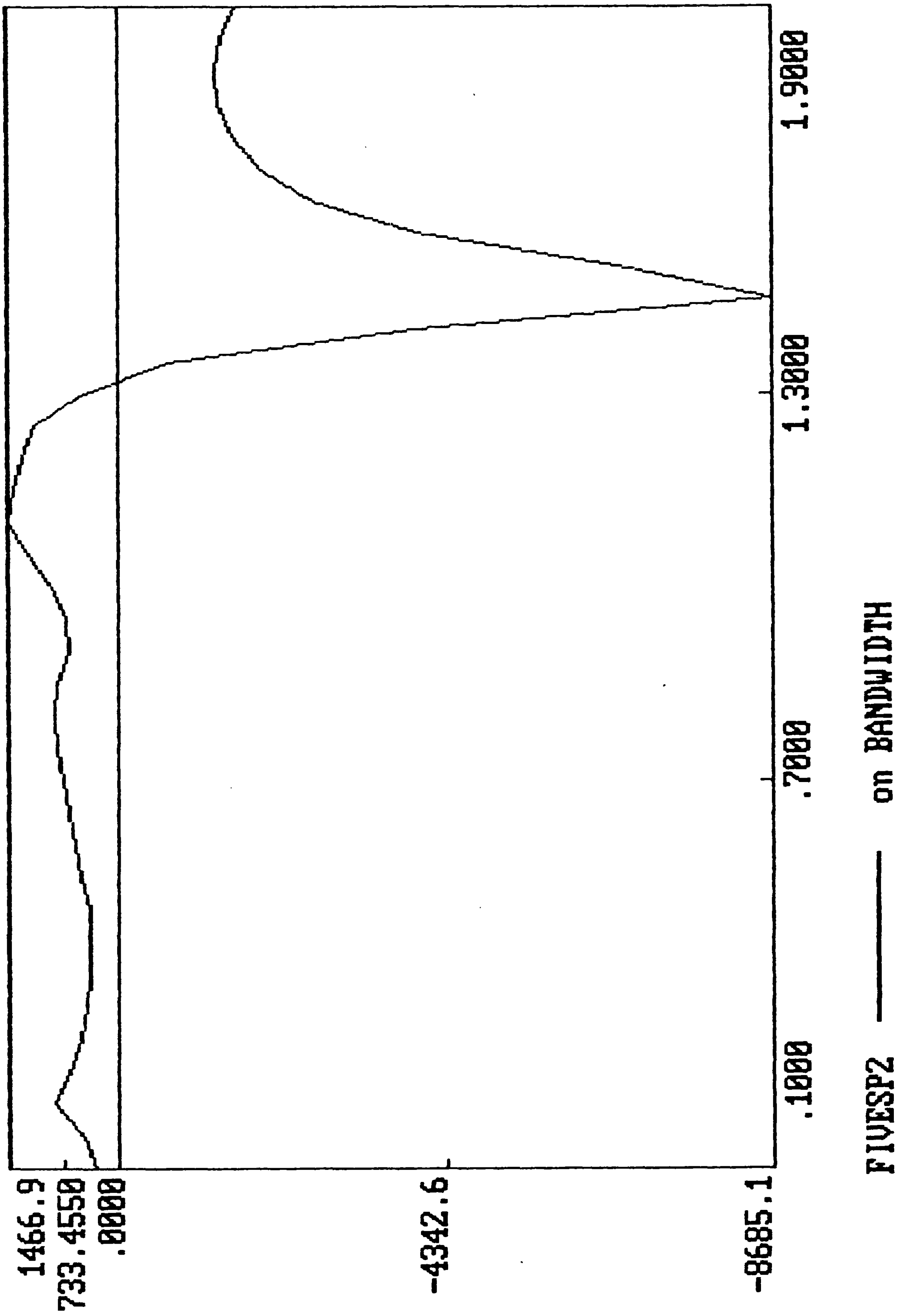


FIGURE 4.2f PLOT OF SEMIPARAMETRIC H_4 ESTIMATES AGAINST BANDWIDTH: POWER STEERING

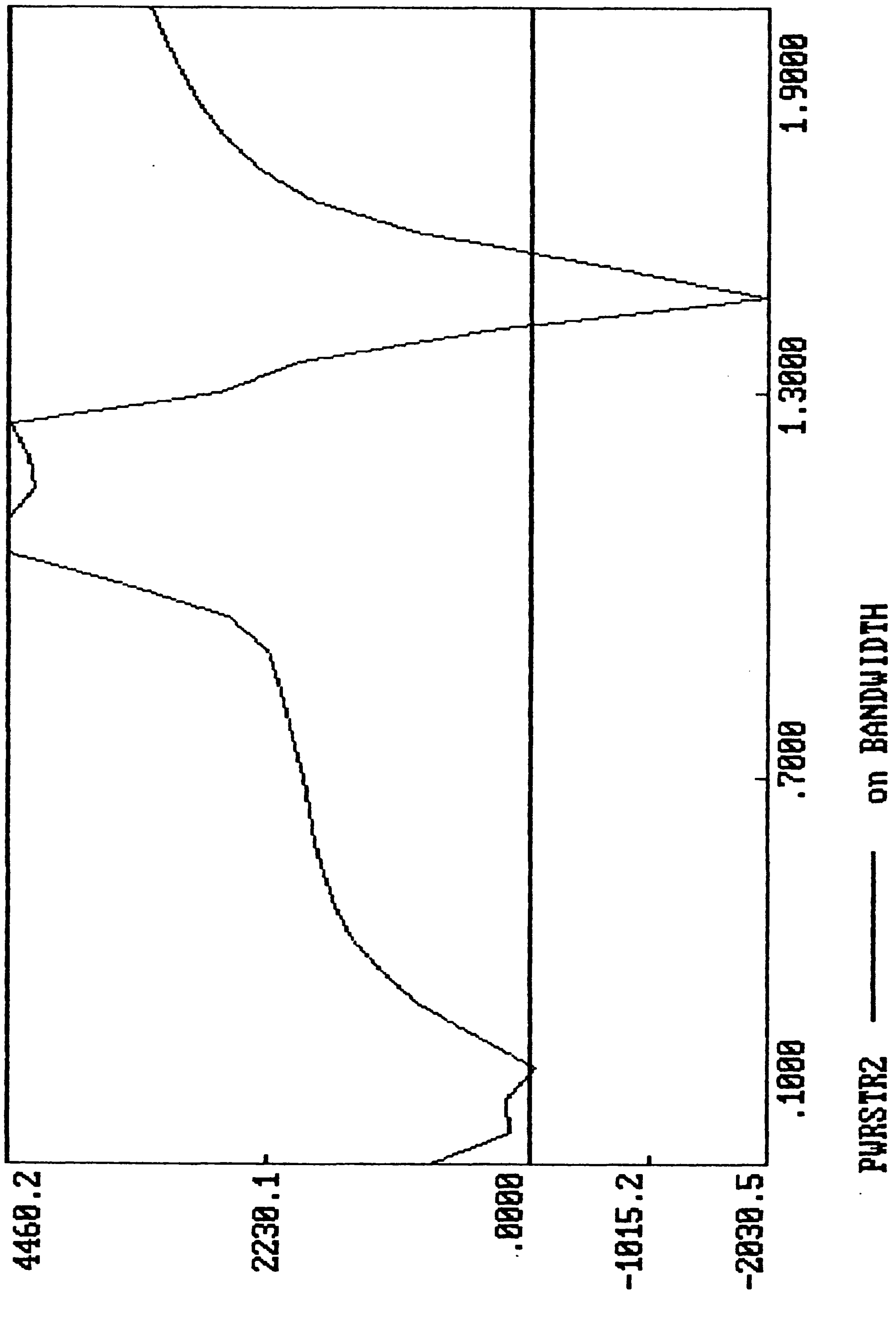


FIGURE 4.2g PLOT OF SEMIPARAMETRIC H_4 ESTIMATES AGAINST BANDWIDTH: UNLEADED

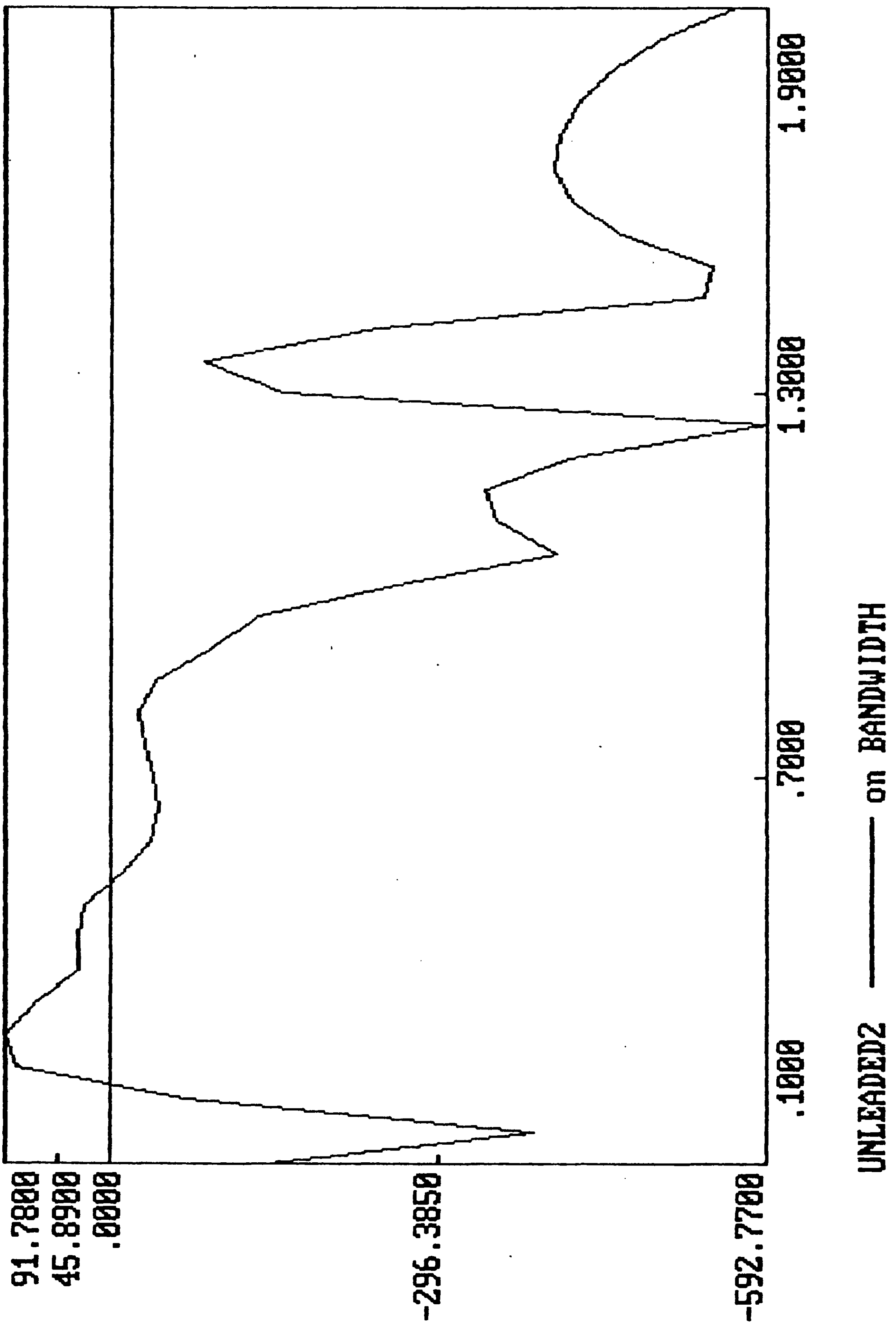


FIGURE 4.2h PLOT OF SEMIPARAMETRIC H_4 ESTIMATES AGAINST BANDWIDTH: BRITISH MADE

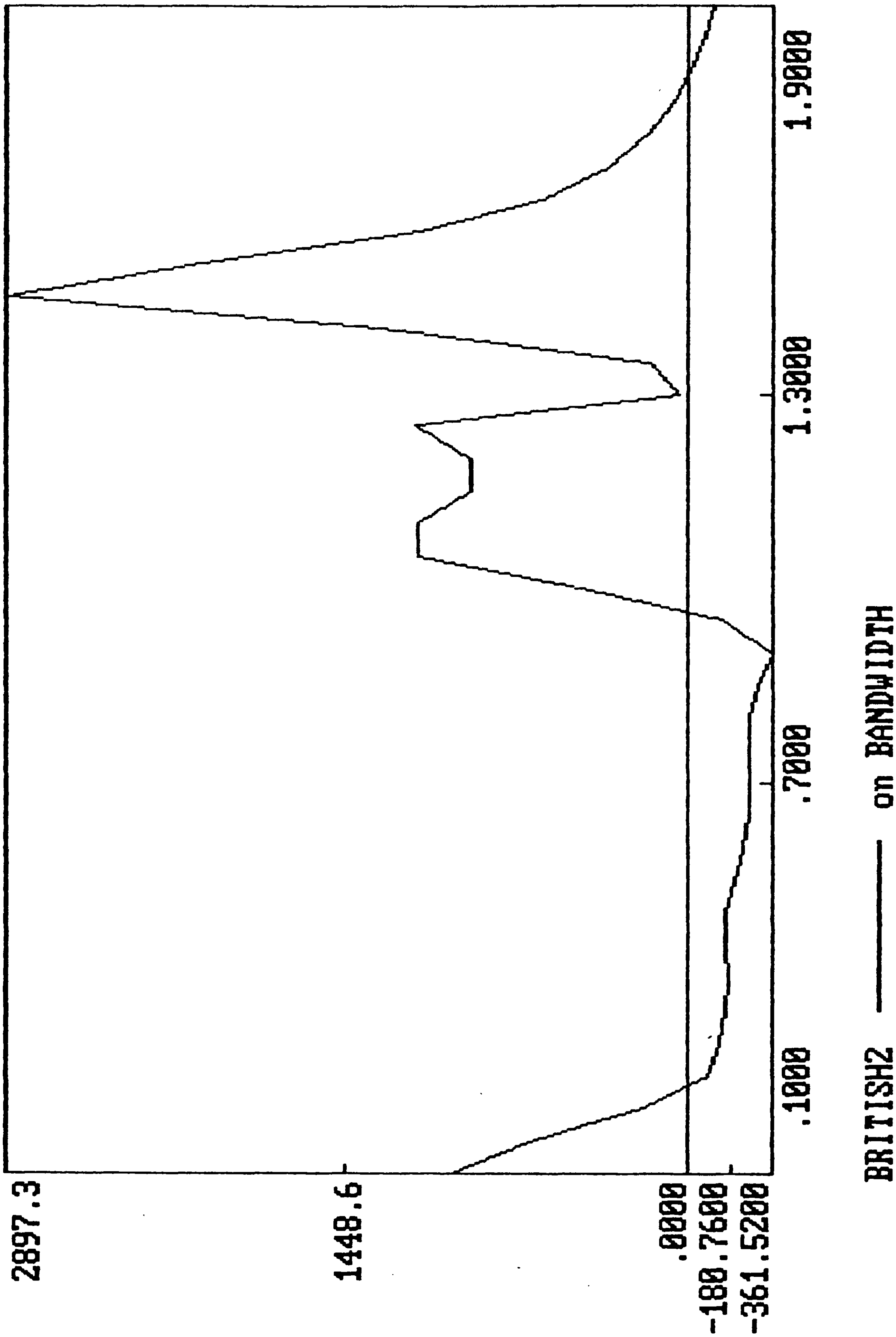


FIGURE 4.3a: PLOT OF COST ESTIMATES FOR H_2 AND H_4 AGAINST BANDWIDTH: 10% INCREASE IN FUEL EFFICIENCY

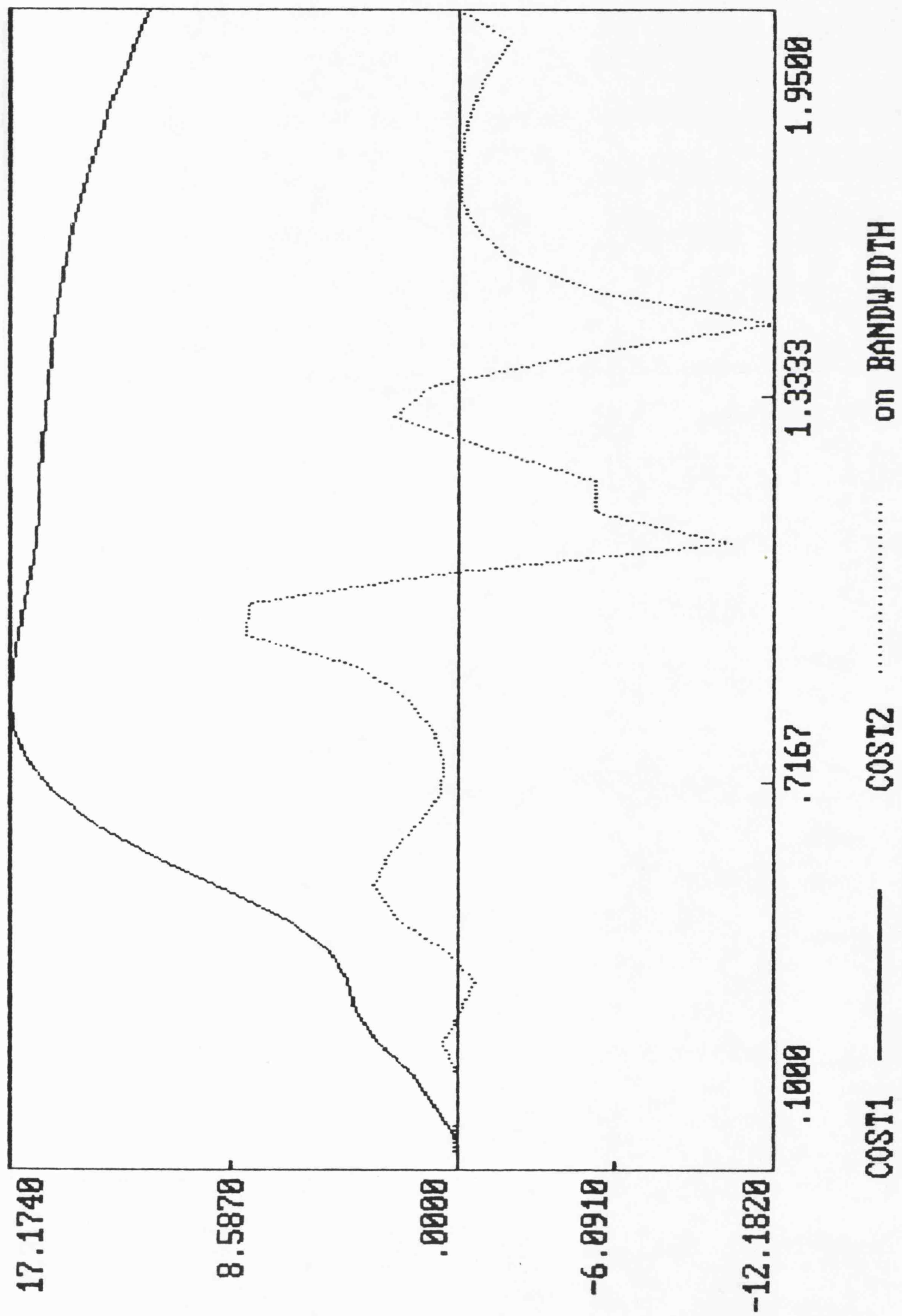


FIGURE 4.3b: PLOT OF t–RATIOS FOR H_2 AND H_4 AGAINST BANDWIDTH: 10% INCREASE IN FUEL EFFICIENCY

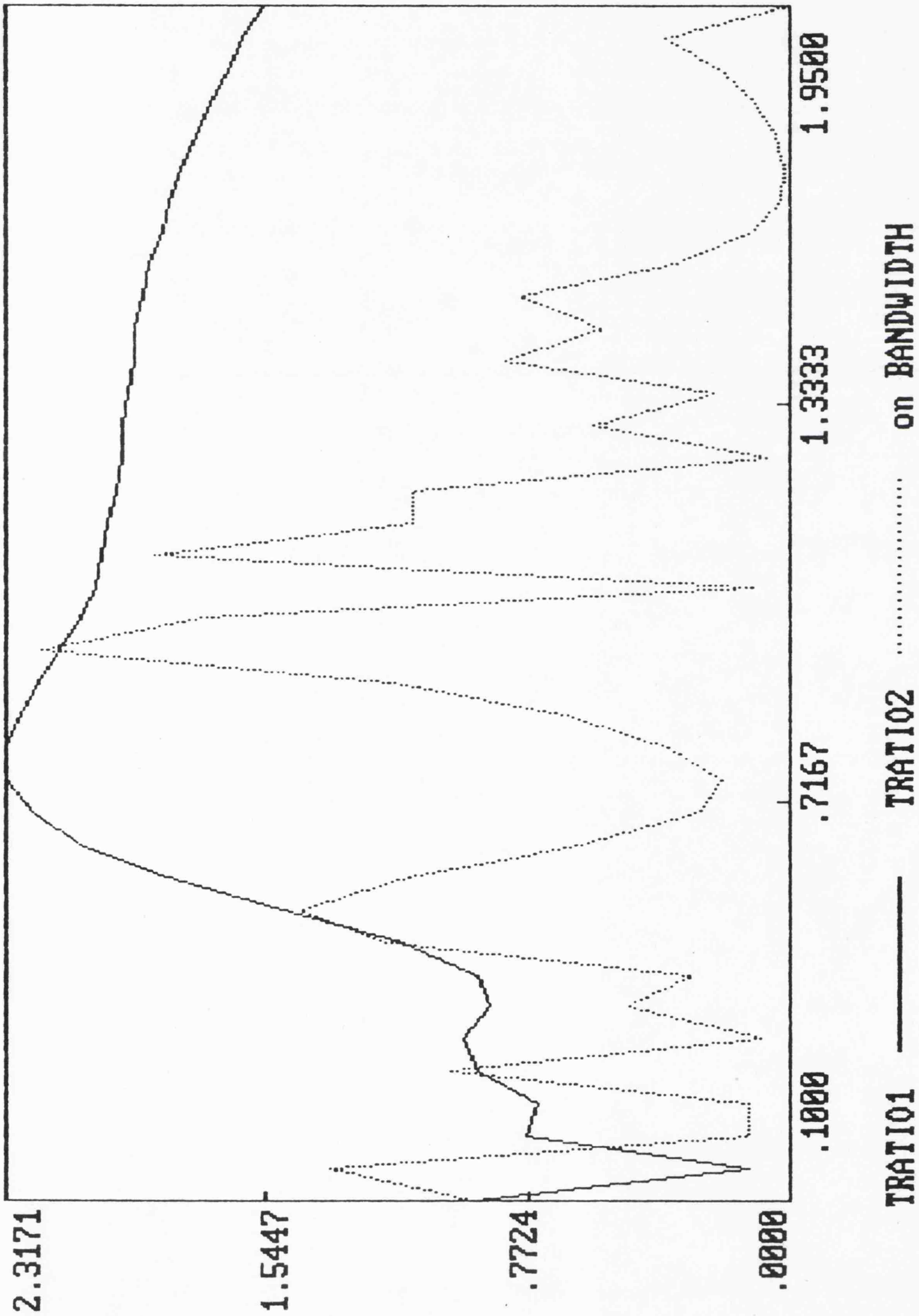


FIGURE 4.4a: PLOT OF COST ESTIMATES FOR H_2 AND H_4 AGAINST BANDWIDTH: 10% INCREASE IN FUEL EFFICIENCY

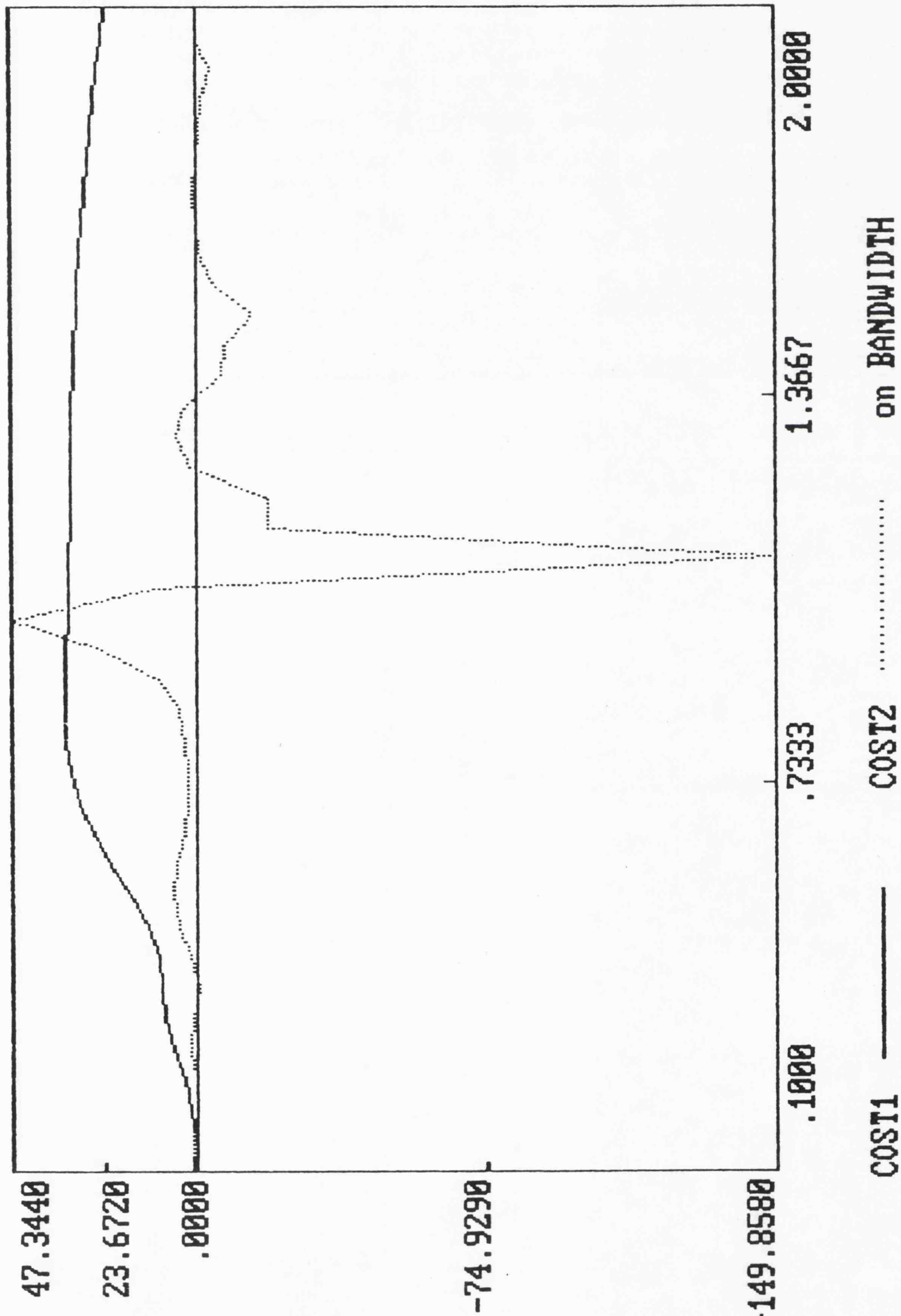


FIGURE 4.4b: PLOT OF t–RATIOS FOR H_2 AND H_4 AGAINST BANDWIDTH: 10% INCREASE IN FUEL EFFICIENCY

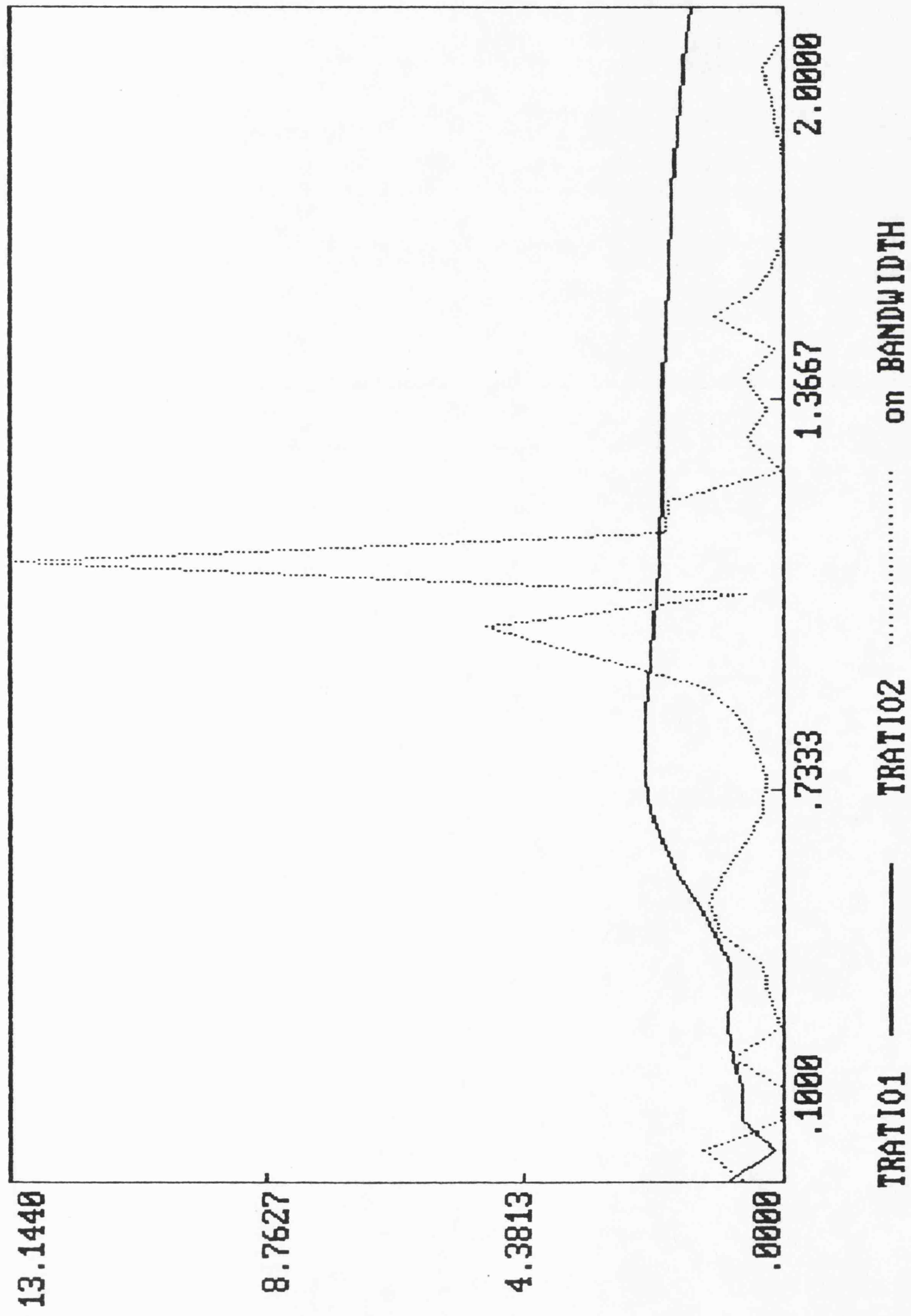


FIGURE 4.5: PLOT OF COST ESTIMATES FOR H_2 AGAINST BANDWIDTH: MINIMUM STANDARD OF 35 MPG

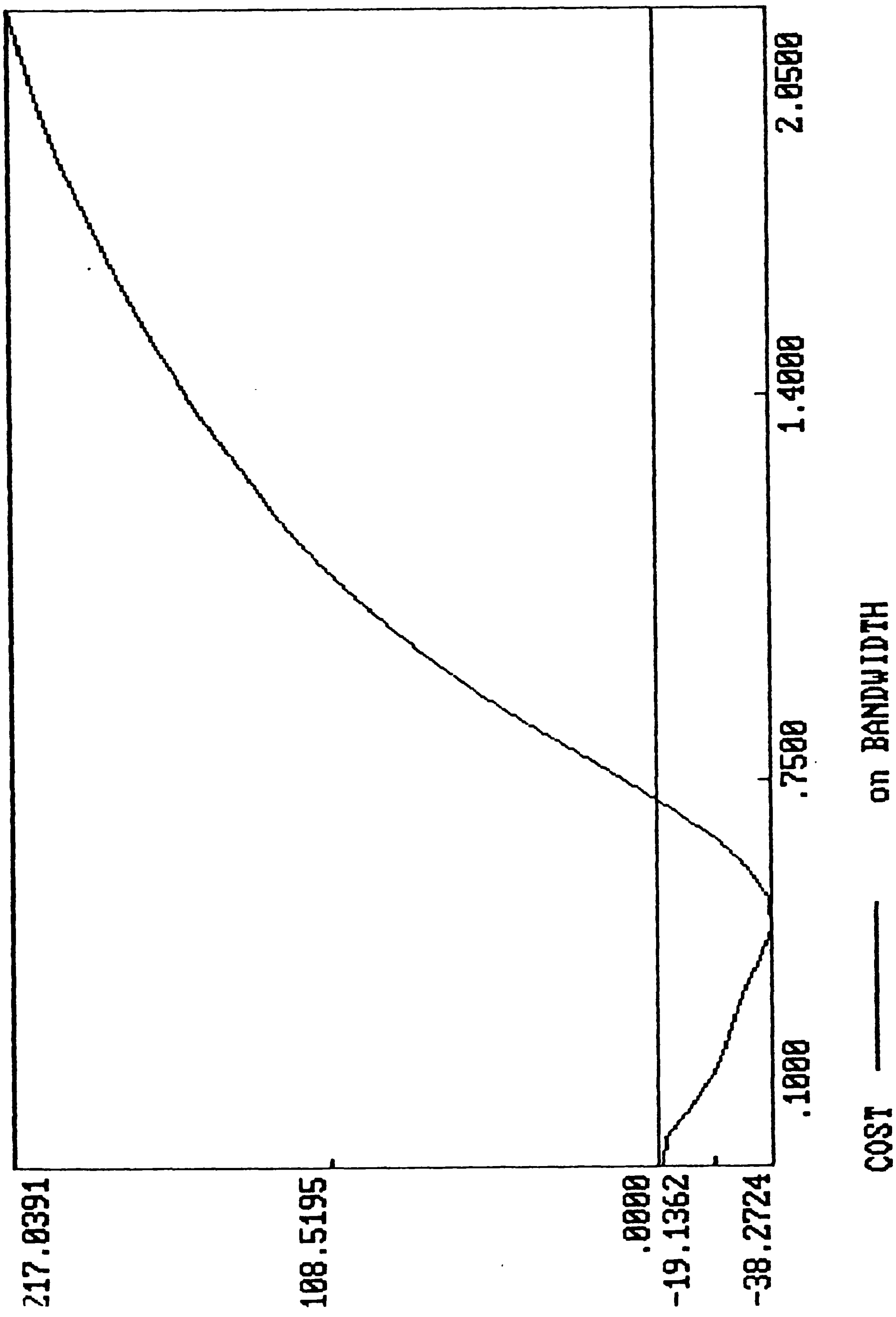


Table 4.1
Parametric Estimates for Hedonic Price Function

	level	log–log	semi–log	inverse–log
constant	–14729.008	3.743	7.048	–60488.303
	(–2.61)	(3.986)	(29.451)	(–2.556)
air con	8960.593	0.274	0.195	11016.140
	(8.58)	(5.963)	(4.423)	(9.495)
sunroof	–1163.959	–0.035	–0.025	–1347.444
	(–2.32)	(–1.571)	(–1.181)	(–2.391)
elec. wind.	239.931	0.054	0.051	386.434
	(0.40)	(2.094)	(2.079)	(0.587)
auto. trans	–826.067	0.000	0.024	–1415.464
	(–1.31)	(0.007)	(0.932)	(–1.982)
5 speed	–1784.539	0.044	0.073	–2545.595
	(–2.64)	(1.515)	(2.580)	(–3.411)
pwr. str.	1380.004	0.119	0.106	2003.631
	(1.80)	(3.527)	(3.269)	(2.349)
unleaded	615.293	0.014	–0.001	751.049
	(1.20)	(0.620)	(–0.068)	(1.304)
British	–1573.991	–0.017	–0.017	–1453.221
	(–2.88)	(–0.737)	(–0.771)	(–2.367)
size	6.682	0.720	0.000	16204.004
	(11.29)	(8.993)	(8.739)	(8.024)
spacious	337.941	0.383	0.067	–8538.774
	(1.41)	(2.314)	(6.681)	(–2.043)
accel.	1932.375	0.278	0.080	4651.727
	(5.480)	(3.392)	(5.384)	(2.250)
power	4263.804	0.260	3.616	368.801
	(0.195)	(2.789)	(3.900)	(0.156)
fuel eff.	3.336	–0.230	–0.003	–8347.505
	(0.050)	(–2.144)	(–1.129)	(–3.088)
R squared	0.834	0.887	0.897	0.791
RSS	7.1E+009	14.113	12.815	8.9E+009
TSS	4.2E+010	124.921	124.921	4.2E+010
SE	4299.377	0.191	0.182	4823.407
F(14,385)	149.125	232.520	259.086	112.396

Note:
t–ratios are in parentheses.

Table 4.2

Semiparametric Estimates For Hedonic Price Function

	H_2	H_4	H_6
Bandwidth	0.5	0.6	0.7
air condition	3483.858	3492.590	3581.135
	(4.921)	(4.973)	(5.121)
sunroof	–46.958	–32.487	2.797
	(–0.179)	(–0.130)	(0.011)
electric window	366.994	411.705	441.336
	(1.255)	(1.480)	(1.600)
automatic transmission	–95.929	–105.924	12.210
	(–0.264)	(–0.298)	(0.034)
5 speed	777.894	584.825	511.163
	(2.027)	(1.604)	(1.418)
power steering	1868.881	1810.794	1837.222
	(4.264)	(4.298)	(4.387)
unleaded	–35.216	–37.970	–3.110
	(–0.130)	(–0.146)	(–0.012)
British made	–208.592	–239.746	–166.583
	(–0.722)	(–0.868)	(–0.611)

Note:
t–ratios are in parentheses.

Table 4.3

Willingness–To–Pay Estimates

1. Parametric estimates

	level	log–log	semi–log	inverse–log
		10% increase in fuel efficiency		
Average	11.740	0.000	0.011	0.000
	(0.050)	(0.000)	(1.205)	(0.000)
		20% increase in fuel efficiency		
Average	23.481	0.000	0.00014	0.000
	(0.050)	(0.000)	(0.007)	(0.000)

2. Semiparametric estimates

	H_2	H_4	H_6
Bandwidth	0.5	0.6	0.7
		10% increase in fuel efficiency	
Average	6.589	2.713	5.045
	(0.057)	(1.129)	(2.239)
		20% increase in fuel efficiency	
Average	12.509	4.705	7.718
	(1.118)	(1.047)	(1.867)

Note:
t–ratios are in parentheses.

Table 4.4

Cross–Validated Hedonic Price and Willingness–to–Pay Estimates

	Semiparametric H_6	Nonparametric H_4
Bandwidth	0.34	0.36
air condition	4776.07	
	(7.74)	
sunroof	224.96967	
	(1.38)	
electric window	596.16173	
	(3.02)	
automatic transmission	237.64663	
	(0.92)	
5 speed	568.82756	
	(1.91)	
power steering	257.03949	
	(0.64)	
unleaded	105.83126	
	(0.56)	
British made	–141.62120	
	(–0.67)	
	10% increased in fuel efficiency standard	
Average	0.54	
	(1.29)	
	20% increased in fuel efficiency standard	
Average	1.00	
	(1.21)	
	Minimum Standard = 35 MPG	
Average	31.00	657.84
	(0.02)	(4.13)

Note:
t–ratios are in parentheses.

CHAPTER 5

A SURVEY OF CONSUMPTION MODELS WITH RATIONAL EXPECTATION

5.1 INTRODUCTION

In this chapter, we present the econometric methods used in previous studies of life cycle–permanent income hypothesis. There is a voluminous literature on econometrics studies and it will not be desirable to review all the models. The range of empirical studies covered is fairly broad, but we do not attempt to give a detailed description of the theory and methodology. Indeed, our intention is to provide an introduction for our empirical work in Chapter 6 and 7. We will focus on the interesting econometric testing procedures using time series and define the problems to be investigated.

The stochastic life cycle implications of the life cycle – permanent income hypothesis (LCPIH) were derived in Hall (1978)'s paper. Under rational expectations, the LCPIH model of consumption predicts that consumers' expenditure (C) should be a random walk with drift, i.e.,

$$\Delta C_t = \alpha + \epsilon_t \tag{1}$$

where Δ is the difference operator, α is a constant and ϵ_t is white noise. In general, Hall's empirical results conform to the random walk hypothesis. However, he did find a significant role for lagged stock prices thus violating the hypothesis.

The findings of Hall's paper motivated a whole new generation of empirical and theoretical work on the stochastic life cycle model and to a certain extent, the consumption capital asset pricing model (CCAPM). Two models are introduced below as our bench mark for discussions.

5.2 THE STOCHASTIC IMPLICATIONS OF THE LCPIH

We begin by looking at a model with stochastic interest rates and wage rate. There are two explanations for the rejection of the Hall model in response to the findings of significant coefficients for lagged stock prices. The first being that the interest rates were treated as constant and the second explanation is that the stock market has more information.

The model presented below for discussion can be easily modified and extended to a more general model. The LCPIH assumes that the consumer is forward looking. He is faced with the problem of dividing his consumption between now and the future. He can either save more for a possibly longer future life or enjoy the current consumption which is certain. In the model, there are other variables that are uncertain such as his future earnings, future rates of interest and changes in household consumption which have to be taken into consideration when planning. These variables are therefore stochastic.

The market structure is such that only spot markets exist in any period and the future markets do not exist. In this setting, current real wage will be known with certainty, whereas future wage rates and asset returns are uncertain and therefore stochastic. The consumer can invest in bonds, equity or banks and enter into any financial contracts. (There is a stochastic return $(1+\tilde{r}_{t+1})$ for some assets and there may also be a nonstochastic return $(1+\bar{r}_{t+1})$ for other nonstochastic assets which we call the risk free rate of return. Empirical evidence suggests that the US 3 months treasury bill rate is the best hedge against inflation and some researchers consider it to be the riskless rate of return. Hall's original model treated the return as nonstochastic. However, the r here is treated as stochastic.)

Consider an economic agent whose utility function within a time period is defined over consumption (C) and labour supply (L). This one period utility function is strictly concave. The optimization problem is to maximize the expected utility subject to the life time budget constraint, i.e.,

$$\text{Max } E_t \Sigma_{j=0}^{\infty} U(C_{t+j}, L_{t+j})/\tau_j$$

subject to the life time budget constraint

$$\Sigma_{j=0}^{\infty} \frac{C_{t+j}}{\delta_j} + \Sigma_{j=0}^{\infty} \frac{W_{t+j}\,(1-L_{t+j})}{\delta_j} = A_t + \Sigma_{j=0}^{\infty} \frac{W_{t+j}}{\delta_j} \qquad (2)$$

where

$E_t(.) =$ mathematical expectation condition on information set Ω_t.

$= E(.|\Omega_t)$

$\delta_j =$ the discount factor $= \Pi_{i=0}^{j}(1+r_{t+i})(1+r_t)^{-1}$

$\tau_j =$ the time preference factor $= \Pi_{i=0}^{j}(1+\rho_{t+i})(1+\rho_t)^{-1}$

$1+r_{t+i} =$ the rate of return between $t+i$ and $t+i+1$

$\rho_t =$ discount rate.

$C_{t+j} =$ consumption at time $t+j$

$L_{t+j} =$ labour supply at time $t+j$

$W_{t+j} =$ the real wage at time $t+j$

$A_t =$ assets at beginning of period t

Notice that in (2), the life time budget constraint states that the total expenditure on consumption and the total expenditure on leisure (the two terms on the left hand side) should exactly equal to the life time non–human and human wealth (the last two terms on the right hand side). There are no bequests in this model. The stochastic variables at period t are known at the beginning of period t (as we assume that they are in the information set Ω_t available to the agent), and all transactions take place in the beginning of the period.

The first order conditions for the consumer's optimal consumption plan consist of stochastic and nonstochastic Euler equations (necessary and sufficient conditions are provided in Lucas (1978), Breeden (1979) and others). But we are only interested in the following equations between period t and t+1:

$$\frac{W_{t+1}\, U_{C,t+1}}{U_{L,t+1}} - 1 = 0 \qquad (3)$$

(3) is the first order static condition and the following two equations are the stochastic Euler equations:

$$E_t\, U_{C,t+1} \frac{(1+r_{t+1})}{(1+\rho_{t+1})} = U_{C,t} \qquad (4)$$

$$E_t\, U_{L,t+1}\frac{(1+r_{t+1})W_t}{(1+\rho_{t+1})W_{t+1}} = U_{L,t} \tag{5}$$

These equations together describe an optimal plan for the individual under intertemporal optimization. U_C is the marginal utility of consumption. So, (4) states that the expected gain in utility from having an extra unit of consumption at period t+1 discounted to time period t (left hand side) should offset the lost in utility from sacrificing a unit of consumption at period t. This condition is derived from the assumption that the economic agent has free access to a capital market which he could either borrow or lend. (5) has the same interpretation with respect to leisure and require the assumptions that the individual is not constrained in both the labour and capital markets. The addition of variables to the utility function or the exclusion variables from the budget constraint can be easily accommodated as we shall see in the discussions below.

Macroeconomists are interested in the coefficient of elasticity of intertemporal substitution of consumption (σ) and the corresponding elasticity for leisure (λ) respectively. Therefore, it is useful to define the following before we proceed:

$$\sigma = \frac{d[\log(C_{t+1}/C_t)}{d[\log(1+r_t)]}$$

$$= \frac{\%\ \text{change in the ratio of consumption}}{\%\ \text{change in the real interest rates}}$$

$$\lambda = \frac{d[\log(L_{t+1}/L_t)}{d[\log(1+r_t)W_t/W_{t+1}]}$$

$$= \frac{\%\ \text{change in the ratio of leisure}}{\%\ \text{change in the real interest rates}}$$

5.3 ASSET PRICING AND CONSUMPTION

In financial economics, it is generally believed that asset prices react sensitively to economic news and that financial market agents are risk averse. These believes are brought about by the following reasons:

First of all, there is empirical evidence to suggest that macroeconomic variables and surprises in the macroeconomic terms affect asset premia, such as those studies of Chen, Roll and Ross (1986), and Tallman (1986).

Secondly, there are also empirical results to suggest that the joint hypothesis of efficient market hypothesis and risk neutral agents are rejected, such as those studies of Hansen and Hodrick (1980), Baillie, Lippens and McMahon (1983), and Fama (1984) for the foreign exchange market, Mankiw and Summers (1983), and Shiller (1979) for the bond market, Campbell and Shiller (1986) for the stock market.

Therefore, it is not surprising that many researchers are interested in the question of whether movement in asset premia is related to the development in the real economy and whether one should attribute any deviation of expected excess returns from zero to risk premia (there are many examples, see Wolfe (1986) for the foreign exchange market and the references thereafter). What we are interested here, however, is not the issue of which macroeconomic variables affect asset premia and which do not. We are interested in a special class of the asset pricing model which has been named consumption CAPM or Consumption–β model.

The CCAPM model describes a relationship between the returns of a financial asset and consumption from an intertemporal optimal portfolio composition setting. The assumptions of the model are identical to that of LCPIH and that individuals choose to invest in N financial assets so as to smooth their consumption. Although the motivations (see, e.g., Shiller (1982)) may be different in the LCPIH and the CCAPM, the time series econometric techniques used are common as most of them hinges on the Euler equation approach.

Assuming the utility function is separable in consumption and leisure, we may then derive the Euler equations with respect to consumption alone. The Euler equation between period t and t+M is:

$$P_{j,t}\, U_{C,t} = E_t[(1+\rho_t)^{-M}\, R_{j,t+M}\, U_{C,t+M}] \qquad (6)$$

where $R_{j,t+M} = (P_{j,t+M} + D_{j,t+M})$

= the real par value of the j^{th} asset at the maturity date t+M.

$P_{j,t}$ = the ex–dividend real price of asset j at time t.

$D_{j,t}$ = the dividend of j^{th} asset at time t.

N = Total number of assets held by the individual.

If we are interested in shares of stock of a firm or a one period index bond, then M=1. For example, the one period index bond where the rate of return $R^*_{j,t}$ is known at time t, (6) can be expressed as (e.g. see Grossman and Shiller (1981) and Shiller (1982)):

$$1 = E[R^*_{j,t}S_t \mid \Omega_t] \tag{7}$$

where $R^*_{j,t} = (P_{j,t+1} + D_{j,t+1})/P_{j,t})$

= the rate of return of the asset j at time t

$S_t = (1+\rho_t)^{-1}(U_{C,t+1}/U_{C,t})$

= the marginal rate of substitution between present and future consumption (MRS) between time t and t+1

The left hand side of (7) is a constant 1. We can take unconditional expectation of (7). This will yield the unconditional expectation of the return (on the R.H.S.), weighted by the MRS between now and future consumption, is always equal to 1 for all assets (on the L.H.S.), i.e.,

$$1 = E[R^*_{j,t}S_t]$$

Furthermore, a relation between expected return of an asset and the MRS can be easily derived using this relationship, and asset returns will have a predictive component. Consider

$$E[R^*_{j,t}] = E[S_t]^{-1}\ (1-\mathrm{cov}(R^*_{j,t},S_t)),$$

the covariance between R^* and S will be negative. The larger the magnitude of the covariance between payoff and MRS, the riskier the asset and the higher the rate of expected returns. Financial economists are more interested in the measurement of the coefficient of risk aversion. For example, if we make the assumption that the utility function is of the constant relative risk aversion type discussed below, then the coefficient of risk aversion is the inverse of σ.

5.4 SOME GENERAL PROBLEMS

Before we engage in any further discussion, it may be appropriate to give a general idea of the common problems concerning all early empirical studies in this area. While some recent studies have attempted to overcome these problems, there are generally four major problems inherent in the investigations and testing of these parametric models using either aggregate or household data.

(1) It is obvious that the changes in family size and composition will normally alter the optimal consumption path. But, it is difficult to identify the effects of household composition on consumption pattern in empirical studies. Most studies do not model for these effects explicitly. In effect, most of the models also assume that individual lives forever. Although some reasons have been offered, they may be hard to accept based on empirical results. These and other related problems have been discussed and investigated in Deaton (1986).

(2) The definition of consumption variable poses problems for econometricians as in many other works on demand analysis. The variable used is usually the value of consumption of non durables and services and inputed services flow from durable goods. Most studies exclude durables as it is believed that there is cost of adjustment. In order to justify a specification of consumption excluding durables, one has to make the assumption that the preferences are separable in consumption for non durables and durables. There are studies (e.g. Hayashi (1982) and Bean (1986)) which assume

that the individual derives services at a constant rate over time. This flow of services from durables can be modelled as depreciation at a specific rate. But the user defined rate is arbitrary and debatable.

(3) Studies using time series and panel data have to make assumption regarding the length of planning period of the consumers. If the data interval is different from the planning interval, then there will be aggregation bias arising from the serial correlation of the error term. So, if the consumer does his planning every month, then the use of quarterly and yearly data will not give consistent estimates. These issues have been discussed by Wickens and Molana (1984), Hall (1988) and Harvey (1988).

(4) It is assumed that the lifetime utility function is intertemporal separable in almost all the empirical studies so that a tractable form can be derived for estimation and test. Most of studies assume that the preferences are intratemporal separable between consumption and labour supply. This has been criticized by many researchers as too strong a restriction. There are studies which use intertemporal but not intratemporal separable utility function (Mankiw, Rotemberg and Summers (1985)). Though the assumption of intertemporal separable utility function can be accepted, the rejections using intratemporal separable preferences are not taken to be conclusive.

Most of these problems have been extensively and carefully discussed in surveys by King (1985) and Hayashi (1987). These problems are the focus of recent research and some studies discussed below have addressed and attempted to tackle these problems.

5.5 THE ECONOMETRIC METHODS

Recent empirical tests can be grouped into five main headings:

(i) constant interest rates

Some empirical studies have assumed that interest rates are constant as in the seminar paper of Hall (1978). We will describe Hall's approach below.

(ii) Sensitivity test

The applications here involves what is known as the excess sensitivity test. For example, we can include current labour income innovations in the regression to see if consumption is more sensitive to labour income than predicted by the LCPIH. Two studies, Flavin (1981) and Hayashi (1982) are of particular importance.

(iii) Euler equation approach with stochastic interest rates

We further break down the approach to three different categories:

(a) Approximation approach

We are referring to the certainty equivalent approach of Muellbauer (1983), and Wickens and Molana (1984) rather than the more recent approach of Hall (1988).

(b) Generalized method of moments

This group of applications embodies the CCAPM models as well as related extensions by including additional variables into the model. They usually involve the use of generalized method of moments, e.g., Hansen and Singleton (1982), Mankiw, Rotemberg and Summers (1985). One of the implications of the model is that any variable in the information set should be orthogonal to the consumption innovation. Therefore, in order to test this implication, an orthogonality test is usually performed.

(c) Log–normality approach

This group of applications makes use of the additional assumption that the macroeconomic variables are jointly log–normally distributed. It includes Attanasio and Weber (1989), Bean (1986), Hansen and Singleton (1983) and others. A test of the over identifying restrictions implied by the rational expectation hypothesis is usually conducted.

(iv) Consumption function approach

This different approach is more interested in finding a statistical relationship between consumption and other macroeconomic variables. Our attention is focused on the surprise consumption function of Blinder and Deaton (1985) and Deaton (1986).

As we are more interested in the econometric methods, our discussions are centred on the econometric models and techniques. The studies and techniques are by no means independent of each other as some of them employ more than one technique in the empirical work.

(i) Constant interest rates

In Hall's paper, he has assumed separability of consumption and leisure. In that case, the utility function can be taken to be

$$U(C_t,L_t) = U(C_t) + U(L_t)$$

Furthermore, it is assumed that the utility function is quadratic, and both r_t and ρ_t are constants. These assumptions give rise to a very simple regression model. Suppose that the utility function is

$$U(C_t) = -(C^* - C_t)^2/2$$

where C^* is the bliss level of consumption. Then the Euler equation can be simplified to (1) since r_t and ρ_t are constant and equal to r and ρ respectively.

$$C_t = C^* (r-\rho)/(1+r) + (1+\rho)/(1+r)C_{t-1}$$

The equation can be expressed as a regression model

$$C_t = \alpha_0 + \alpha_1 C_{t-1} + \epsilon_{1t}$$

where

$$\alpha_0 = (r-\rho)/(1+r)$$

$$\alpha_1 = (1+\rho)/(1+r)$$

$E_{t-1}(\epsilon_{1t}) = 0$ and ϵ_{1t+1} is the news in period t+1 or the surprise term which is orthogonal to the information set Ω_{t-1}. If $r \approx \rho$, then $\alpha_1 \approx 1$. This is known as the random walk with drift consumer expenditure model and it simply says that consumption in this period is approximately the same as the last period plus a constant. Besides consumption from the last period, only additional news from this period will affect current consumption.

Consider the following model

$$C_t = \alpha_0 + \alpha_1 C_{t-1} + a(L)X_t + \epsilon_{2t}$$

where $a(L) = \Sigma_{i=1}^{4} a_i L^i$

and L is the lag operator. Hall's procedure involved running regressions of C_t on C_{t-1} with a constant, plus one other set of variables (X) such as lagged consumptions, lagged incomes or lagged stock indices separately. The null hypothesis was that all the a_i's were zeros. The model can be interpreted as being rejected because the regression coefficients of the lagged stock indices were statistically significant. These results have therefore violated the prediction from the optimal consumption path that no variables other than lagged consumption, C_{t-1} should have explanatory power in predicting current consumption C_t.

(ii) Excess sensitivity test

(a) Indirect or Multivariate Least Squares

Since Hall's seminal contribution, other authors have extended the model and conducted similar tests by adding variables to the random walk plus drift model. In particular, Flavin (1981) reconciled the conflicting results obtained from Hall (1978) and Sargent (1978)'s model by modelling income as an ARMA (Autoregressive Moving–Average) process. Sargent's results were different from that of Hall, and Flavin argued that Sargent by omitting capital gains from his definition of permanent income, had in fact imposed incorrect restrictions across the parameters of the vector autoregression of (C_t, Y_t). Flavin's model is of the form

$$Y_t = \beta(L)Y_t + \xi_t$$

$$\Delta C_t = \alpha + k\Psi(Y_t - \beta + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_p Y_{t-p}) + \psi_0 \Delta Y_t + \ldots + \psi_{p-1} Y_{t-(p-1)} + \epsilon_{3t} \qquad (8)$$

where

$$\beta(L) = \Sigma_{i=0}^{p}\beta_i L^i$$
$$\Psi = (1 - \frac{\beta_1}{(1+r)} - \frac{\beta_2}{(1+r)^2} - \dots - \frac{\beta_2}{(1+r)^p})^{-1}$$

ξ_t is the income forecast error and ϵ_{3t} is the consumption disturbance term. The model suggests that the revisions in planned consumption are caused by the revisions in the expectation about future income (second term on the right of (8)). The ψ_i's are the measures of excess sensitivity of consumption to current income, and therefore should be zeros if the model is consistent with the LCPIH. The appropriate methodology for estimation and testing are the full information maximum likelihood (FIML) and likelihood ratio (LR) tests. But it turns out that the model is just identified. Therefore, the multivariate least squares (MLS) on the reduced form is equivalent to FIML on the structural form. All the coefficients can be recovered from MLS on the reduced form. Flavin also ran a Hall's reduced form test of the form

$$\Delta C_t = \pi(L)Y_t + \epsilon_{4t}$$

where $\pi(L) = \Sigma_{i=0}^{p}\pi_i L^i$

Flavin's evidence indicated that there was a strong excess response of consumption to current income thus against the joint hypothesis of rational expectations and life cycle. The procedure of Flavin has been questioned recently in the literature on unit roots. If income is in fact a random walk process, then there is a danger that the conventional test statistics give incorrect rejection probabilities. We shall return to the discussion of tests in the presence of unit roots in section 6.

(b) Non Linear Instrumental Variable Estimation

Hayashi (1982) dealt with constant relative risk aversion (CRRA) utility function $U(C)=U^{1-\gamma}/1-\gamma$ with $\gamma \geq 0$, where γ is the coefficient of relative risk aversion. He was the first to use a consumption series which included service flows from consumer durables. Previous studies had omitted durables altogether to avoid the problem with serial correlations. If future labour income is stochastic subject to a multiplicative shock e.g.

$$Y_t = Y^*(1 + e_t)$$

where e_t's are i.i.d. with mean zero. Hayashi asserted that the Euler equation would be the form (ignoring the time subscripts)

$$C(A,Y^*)^{-\gamma} = \frac{1+\rho}{1+r} E[C\{(1+\rho)[A+Y^*(1+e)-C(A,Y^*)] , Y^*\}]^{-\gamma}$$

The optimal consumption is a nonlinear function of initial nonhuman wealth, A, and Y^*. Even for this simple case, the closed–form solution to the Euler equation is tedious to derive. In general, it is even more complicated. So, the optimal consumption rule is assumed to take the form

$$C_t = \alpha(A_t+H_t) + u_t \qquad (9)$$

$$H_t = \Sigma_{k=0}^{\infty}(1+\mu)^{-k}E_t(Y_{t+k})$$

H_t, the expected future income, can be taken to be

$$H_t = (1+\mu)^{-k}(H_{t-1}-Y_{t-1}) + \xi_{2t} \qquad (10)$$

where μ = discount rate for expected future labour income.

u_t = transitory consumption which may be interpreted as a shock to the utility function or as measurement errors in Ct and At.

ξ_{2t} = the surprise term.

This discount rate, μ, is not equal to the risk free rate of interest, r. In fact, the discount rate is risk free rate of interest plus a risk premium. Since the expected future income, H_t, is not observable by econometricians, it has to be replaced by the observable variables, namely, consumption, and assets holdings using equation (9). It can be shown that by eliminating H_t from the equations (9) and (10), we get

$$C_t = (1+\mu)C_{t-1} + \alpha[A_t-(1+\mu)(A_{t-1}+Y_{t-1})] + \epsilon_{5t}$$

where

$$\epsilon_{5t} = u_t - (1+\mu)u_{t-1} + \alpha\xi_{2t}$$

not forgetting that ξ_{2t} is a surprise term which is a discounted stream of revisions of the expectations of Y_{t+k}. There are reasons to believe that the simple nonlinear least squares (NLLS) estimators will not be consistent. We have following points to consider:

(i) $\text{cov}(u,Y) \neq 0$ and hence $\text{cov}(\epsilon_5,Y) \neq 0$.

(ii) A_t may be correlated with Ω_{t-1} since it does not include A_t, which implies $\text{cov}(A_t,\epsilon_{5t}) \neq 0$.

(iii) Simultaneous bias associated with u, A, C and Y due to endogeneity and and may be $\text{cov}(\xi_2,u) \neq 0$.

Therefore, nonlinear instrumental variable estimation (NLIV) is the appropriate technique (see Amemiya (1974), Jorgenson and Laffont (1974)). Using a set of instruments from Ω_{t-1}, i.e. Z_{t-1}, where $E[Z_{t-1}u_t] = 0$ and $E[Z_{t-1}u_{t-1}]=0$ so that $E[Z_{t-1}\epsilon_{5t}]=0$, one can get consistent estimates (see Chapter 3 for the choice of efficient instruments). However, Hayashi estimated various models including one which took into account liquidity constrained households. Unfortunately, Hayashi's results suggest that a sizable proportion of the population are liquidity constrained and thus against the LCPIH.

(iii) variable interest rates

(a) Approximation or Errors–in–Variables Methods

Those studies which do not impose the log–normality assumption can be included in this group, e.g., Muellbauer (1983), and Wickens and Molana (1984). Muellbauer (1983) extended Hall's model to one which embodies time varying interest rates.

The first order condition can be approximated by

$$\Delta c_t = \alpha + \sigma E_{t-1} r_{t+1} + \epsilon_{6t}$$

where $\alpha = -\sigma \ln(1+\delta)$, σ is the elasticity of intertemporal substitution and c is the log of consumption. we shall assume that all lower–case letters represent natural logarithm.

Since there is a rational expectation term (i.e. $E_{t-1}r_t$), one can parameterize the expectations formulation for interest rates. One usually characterizes the income process as an autoregressive moving average (ARMA). The main argument offered being that the process is not to imply that individuals form rational expectations in this particular parametric setting, rather, the individuals exploit the serial correlation of the income variable. But, this argument may be difficult to justified with surprise models discussed below.

Empirically, this ARMA process can be approximated by a finite AR process thus giving an attractive approach from the practical point of view. The rational expectations hypothesis does not specify how individuals should form expectations but only assumes that individuals form their expectations conditional on all currently available information. Therefore, SM (substitution methods) can be employed to estimate the model. In other words, one simply needs to apply indirect least squares on the reduced form of the just–identified system and the estimates should be as efficient as FIML.

Muellbauer models can be written as

$$\Delta c_t = \alpha + \sigma r_{t+1} + \epsilon_{7t} \tag{11}$$

$$\epsilon_{7t} = \epsilon_{6t} - \sigma(r_{t+1} - E_t r_{t+1})$$

The estimation techniques and problems of this errors–in–variables model have been widely discussed in McCallum (1976a,1976b,1979), Nelson (1975), Wickens (1982), Pagan (1984, 1986) and Bean (1986). By choosing appropriate instruments for r_{t+1}, say r_t, IV estimation will yield efficient estimates for σ. This is known as the errors–in–variables methods (EVM).

The variables are in logarithm so that the model can also be taken to be from a CRRA utility function. There is a simple relationship between σ and γ, i.e., $\sigma = \gamma^{-1}$. But there is evidence of misspecifications in the model using data from the

United Kingdom as σ sometimes turned out to be statistically insignificant.

However, it may be interesting to note that consistent estimation of Muellbauer's model does not require the estimation of $E_t(r_{t+1})$. Indeed, simple IV (instrumental variable) will yield consistent estimates. The most commonly used efficient method is the two steps or two stage methods. One of the methods is to run an autoregression of r_t in the first step and then in the second step, regress Δc_t on the predicted values of r_t obtained from the first stage. However, the standard errors from the second stage are not the true standard errors. To obtain the correct standard errors, one needs to apply two stage least squares on (11) with the r values replaced by those obtained from the first step. Other procedures to improve efficiency are available, e.g. generalized methods of moments (GMM) of Hansen (1982) discussed below, semiparametric methods using conditional moment restrictions discussed in Chapter 6 and 7.

However, consistency of the surprise terms cannot generally be taken for granted in the surprise consumption function when there is misspecification in the parametric formulations of expectations. Therefore, other procedures such as semiparametric techniques are required if one were to safeguard against the possibility of nonlinear rational expectation formulations. These semiparametric models are characterized by the fact that the model has two components, parametric and nonparametric. One can model the unknown expectations by nonparametric regressions which involves infinite number of parameters and then run the usual ols technique in the second stage. This indeed is the method that we are to introduce in Chapter 7.

(b) The generalized methods of moments

Assuming that the subjective discount rate is constant as in the case of Muellbauer (1983), the first order condition (4) can be written as:

$$E_t[\frac{U_{C,t+1}}{U_{C,t}} \frac{1+r_{t+1}}{1+\rho} - 1] = 0$$

Depending on the formulation of the budget constraint in CCAPM model, we get slightly different forms. By letting the brackets equal to $h(X_{t+1},b_0)$ which is a vector function and where b is the vector of unknown coefficients and X_{t+1} as some forcing variables i.e. those observable in the information set of the agent and observable by the econometrician, then

$$E_t[h(X_{t+1},b)] = 0 \qquad (12)$$

$$E_t[u_{t+1}] = 0$$

Hansen and Singleton (1982) used the GMM to estimate and test the nonlinear rational expectations model directly from the stochastic Euler equations without making any strong a priori assumptions about the nature of the forcing variables X. The Euler equation can be expressed using the function h as in equation (12). The utility function was again assumed to be the CRRA type. Since marginal utility is equal to

$$U_{C,t} = C_t^{-\gamma}$$

Then (12) simplifies to

$$E_t[(X_{k,t})^{-\gamma}X_{j,t}/(1+\rho) - 1] = 0$$

where $X_{k,t+1} = C_{t+1}/C_t$

$$X_{j,t+1} = R^*_{j,t} \qquad j = 1, 2 ,3.$$

So, one can simply set

$$h(X_{t+1},b) = (X_{k,t})^{-\gamma}X_{j,t}/(1+\rho) - 1$$

$$b = (\gamma,\rho)'$$

and we have an equation which states that the unconditional expectation of u_{t+1} with any variable in the information set should be equal to zero. Hansen and Singleton (1982) suggested an estimator for b that minimized the weighted sum of squares of the product of instruments at t (say Z_t) and h. The optimum weight which minimizes the asymptotic covariance matrix is derived in Hansen (1982). There is also a J statistic for testing the overidentifying restrictions. These restrictions require that any extra instruments used in estimation should not increase the J statistics too much, because of the fact that the expectation of the product of h with the new instruments should be zero. The evidence presented is against the models as these overidentifying restrictions were rejected.

Mankiw, Rotemberg and Summers (1985) by making the additional assumptions that h was conditional homoscedasticity rather than heteroscedasticity as in Hansen and Singleton (1982), estimated both the three first order conditions (3), (4) and (5) as a system and separately. In fact, in this case the GMM simplified to NL3SLS (nonlinear three stage least squares). They were the first to use a utility function not separable in consumption and leisure

$$U(C_t,L_t) = \frac{1}{1-\gamma}\left[\frac{C_t^{1-1/\theta}-1}{1-1/\theta} + d\,\frac{L_t^{1-1/\lambda}-1}{1-1/\lambda}\right]^{1-\gamma}$$

It is interesting to note that when the coefficient of relative risk aversion is zero, we have an additive separable utility function in consumption and leisure. θ and λ represent the elasticity of intertemporal substitution in consumption and the corresponding elasticity for leisure respectively. They produced empirical evidence against the model. The estimated utility function appears to be convex and when it is concave, the overidentifying restrictions are often rejected.

Miron (1986) allowed for seasonal shocks to preferences and technology. It is sensible to assume that there is seasonality in consumer purchases and therefore one should allow the use of seasonally unadjusted data on real consumption purchases. Miron constructed the seasonally adjusted data. He divided the seasonally unadjusted data on nominal consumption purchases by the seasonally unadjusted components of the consumer price index. Since the model is nonlinear in parameters as well as variables, NLIV is again the appropriate technique for similar reasons.

The evidence he produced did not contradict the implications of the model as the J statistics were not rejected except for transportation at the 5% level but not at 1% (see below for the discussion on J statistics). The J statistics were however, rejected for most of the cases when seasonally adjusted consumption purchases were used (without seasonal dummies in the model). This suggests that seasonally adjustment may have accounted for most of the earlier rejections. This is not surprising as filtering may have caused the orthogonal conditions to be violated. So, lagged variables and the J test for overidentifying conditions will always be significant.

(c) The Log–normality Approach

Hansen and Singleton (1983) imposed restrictions on preferences and the joint distribution of consumption and returns to study the time–series behaviour of asset returns and aggregate consumption.

The important assumption one has to make in deriving a linear relationship is that consumption and asset returns are jointly log–normally distributed. In other words, Hall could have derived his random walk equation by imposing this assumption. Also, the maximum likelihood approach is actually Hansen's method of moments with some other conditions. One of them is the log–normality assumption, which states that $\{Z_t : Z_t = (x_t, r_{1,t}, \ldots, r_{N,t})\}$ is a stationary Gaussian process. With this strong assumption, the first order condition (4) can be simplified into a tractable form. We can get a simple linear regression model of the form:

$$E_t(r_{i,t}) = [\log(1+\rho) - (\sigma_i^2/2)] + \gamma E_{t-1}(x_t)$$

or $$r_{i,t} = a + \gamma E_{t-1}(x_t) + \zeta_t$$

with $$x_t = E_{t-1}(x_t) + \xi_t$$

where $r_{i,t+1} = \log R_{t+1} = \log$ of i^{th} asset return

$$x_{t+1} = \log X_{t+1} = \log (C_{t+1}/C_t) = \Delta c_{t+1}$$
$$a = \log (1+\rho) - (\sigma_i^2/2)$$
$$\sigma_i^2 = \text{var} [\log X_t^{-\gamma} R_{i,t+1} | \Omega_t]$$

and the error terms are normally distributed. Recently, Hall (1988) and Harvey (1988) have made consumption a dependent variable.

To estimate the model by maximum likelihood, $E_t(X_{t+1})$ is parameterized as a linear function of past values of y with a finite lagged length as in many previous empirical work that we have discussed. This linear parametric specification of the expectation term is a consequence of the log–normality assumption and rational expectations (e.g. see Bray (1981,1985) about the connection between linearity and normality assumptions). The maximum likelihood procedure assumes that the observable variables are log–normally distributed and that the lagged length specification of the vector autoregression is correct. So, if either of these assumptions is violated, the model can again be rejected. Of course, there are specification tests for these underlying assumptions.

Attanasio and Weber (1989) obtained unfavourable evidence under the ordinal certainty equivalence (OCE) framework of Selden (1978), using return on shares and return for building society respectively. They found implausibly low intertemporal substitution and high ρ for shares and a more plausible value for intertemporal substitution but a negative ρ for building societies. Furthermore, the coefficients for elasticity of intertemporal substitution were associated large standard errors. This suggests that their model is misspecified indicating that one or some of the underlying assumptions are inappropriate.

One of the important contributions of Attanasio and Weber (1989), among others, is the use of cohort data. In the model, individuals were assumed to live indefinitely so that aggregation problem could be ignored (see Deaton (1986)). However, as we have mentioned in point (1) above, the aggregate Euler equation would reflect demographic changes if there were real growth in the economy and if the households had finite time horizon. To alleviate this problem, they used the data from a single cohort of married couples between 25 and 40 in 1970. The results can be interpreted as unfavourable to the hypothesis.

Bean (1986) used similar approach adopted in Hansen and Singleton (1983). But, in his case, the utility function multiplicatively is not separable in consumption and leisure. Two extra variables, namely, leisure (L) and government expenditure (G), were added to the utility function taking into account the effects of wages and government spendings. It has been argued that government expenditure can act as substitute for private consumption. By assuming that $\{\Delta c_t, \Delta \ell_t, \Delta g_t, r_t\}$ is generated by a covariance stationary Gaussian process, the Euler equation can be simplified to a regression equation as in the case of Hansen and Singleton (1983).

$$\Delta c_t = a_0 + a_1 E_{t-1} r_t + a_2 E_{t-1} \Delta \ell_t + a_3 E_{t-1} \Delta g_t + \epsilon_{7t}$$

To test the LCPIH model, $E_t \Delta y_{t+1}$ was added to the equation as an extra regressor. The augmented equation was estimated jointly with a vector autoregressive system describing the evolution of the other variables of the system. A significant coefficient for $E_t \Delta y_{t+1}$ will indicate a rejection of the model because of sensitivity to expected income growth. Bean, like Miron, presented limited support for the hypothesis as the model was able to capture the salient features of the data but the overidentifying restrictions were marginally rejected.

(v) Consumption Function Approach

There are researchers interested in modelling the consumption function where the regressions used are taken as the test of the LCPIH. Deaton (1986) and Blinder and Deaton (1985) tested the LCPIH using a surprise consumption function. The method involves running the supplementary equation for expectations followed by the main regression with the expectations and surprises replaced by their calculated values from the first stage (the usual two stage method). Newey's (1984) method is used in calculating the covariance matrix because of the presence of the surprise terms. The standard errors are identical to the OLS standard errors for the surprised terms, and to the two stage least squares standard errors for the others. The procedure is asymptotically efficient as in the case of Pagan (1984), Bean (1986) and Baillie (1987).

Various nested models are proposed and the full model included both anticipated and unanticipated terms which are features of most rational expectations model. The full model of Deaton (1986) is:

$$\begin{aligned}
\Delta c_t = a_0 \quad & + a_1 \Delta c_{t-1} && + a_2 y_{t-1} \\
& + a_3\, E_{t-1} \Delta y_t && + a_4(\Delta y_t - E_{t-1} \Delta y_t) \\
& + a_5\, E_{t-1} \Delta wh_t && + a_6(\Delta wh_t - E_{t-1} \Delta wh_t) \quad (13) \\
& + a_7\, E_{t-1} R_t && + a_8(R_t - E_{t-1} R_t) \\
& + a_9\, t && \epsilon_{9t}
\end{aligned}$$

where wh is wealth. However, Deaton (1986) has argued that if the hypothesis is true, permanent income, measured income and consumption are of the same random walk. In this case, as Dickey and Fuller (1979, 1981) have pointed out, the t–distribution is a poor guide to the actual distribution of the t–statistics, even asymptotically. Deaton (1986) excluded the time trend from the regression and justified his action by a simple Monte Carlo experiment. The LCPIH is still rejected as the a_3 coefficient was still statistically significant indicating excess sensitivity of consumption to current income.

Blinder and Deaton (1985) have also a detailed discussion on the time series consumption function. Various formulations of the consumption function was put forward for discussions and a number of hypothesis was tested. Some of the equations are of the form described above and are tests of the LCPIH. The results are in general against the model.

Koskela and Viren (1987) have also rejected the model as they found significant roles for anticipated and unanticipated inflation, thus confirming the findings in Blinder and Deaton (1985) in various more general models with a lagged term added for each expected term. Further research needs to be done in this area as inflation should not matter within an intertemporal optimization framework.

Other recent contributions from this approach include Kugler and Bossard (1987) and Davidson, Hendry , Srba and Yeo (1978). Kugler and Bossard (1987) have obtained autocorrelation patterns for five different countries, and Davidson, Hendry, Srba and Yeo (1978) have conducted a study using the data on the United Kingdom. Notice that in (13), if the coefficients for the surprise terms are equal to the corresponding coefficients of the expected terms, then the model breaks down to an error correction model.

5.6 GENERAL OBSERVATIONS

The results from these tests have been disappointing. Many explanations have been given for the rejections of the LCPIH models and most of them are objections to the econometric specifications. In general, rejections of LCPIH using aggregate data includes studies by Hall (1978) for stock prices, Sargent (1978), Flavin (1981), Hansen and Singleton (1982,1983), Hayashi (1982), Mankiw, Rotemberg and Summers (1985), Deaton (1986) and others that we have mentioned above. However, more recent studies have produced favourable results. Bean (1986) and Miron (1986) both produced evidence to support the LCPIH. Our intention in the next two chapters is to examine various statistical issues, some of which are outlined below. The earlier rejections and empirical results lead to the following observations:

(1) It is difficult to determine the relationship between consumption and real interest rates using quarterly data. Many results have demonstrated that log–normality approach has given rise to the wrong sign, e.g. Koskela and Viren (1987). The rational expectation formation may not be linear and therefore it is

important to accommodate rational expectations of unknown form. An alternate source of problem is discussed in point (3) below. Our intention in the next two chapters is to investigate the proposition that there is no linear relationship between consumption and expected real interest rates and other macroeconomic variables using U.S. quarterly data. Our interests are statistical estimation and hypothesis testing, and we allow for possible nonlinear rational expectations formation.

(2) If indeed there is no relationship between consumption and expected interest rates, it may be important to include a separate parameter, independent of the risk parameter within the CCAPM, to measure the intertemporal elasticity of substitution of consumption. In this spirit, Hall's (1988) model derived a bivariate regression model:

$$c_t = \alpha + \sigma E_t r_{t-1} + \epsilon_{7t}$$
$$r_t = E_t r_{t-1} + \nu_t$$

where σ is interpreted as the intertemporal elasticity of consumption or the inverse of the risk aversion parameter $1/\gamma$ under the usual intertemporal framework. Consumption (C) is in log form. Both ϵ_7 and ν are assumed to be normally distributed. In Hansen and Singleton (1983)'s expected utility framework, σ is also the reciprocal of the coefficient of relative risk aversion. Under the OCE framework, the two equations do not give any information about risk aversion as it is independent of the coefficient of risk aversion.

Attanasio and Weber (1989) dealt with returns on two different assets and were able to identify γ. Using this two period model, one can compare the difference between the two estimates, i.e. $(\sigma - 1/\gamma)$ and see if indeed the difference is too large for the expected utility approach to be taken seriously. However, the resulting estimates from their work were not reliable enough to draw any firm conclusions. More research along the OCE approach has been fruitful and further discussions are presented in Chapter 6.

(3) If there is a component of transitory consumption denoted by C^T, then the random walk equation for quadratic utility will have a moving average error term i.e.

$$C_t = C_t^P + C_t^T$$

and $$C_{t+1}^P = a_0 + a_1 C_t^P + a_2 X_t + \eta_t,$$

then we have

$$C_{t+1} = a_0 + a_1 C_t + a_2 X_t + [\eta_t + (C_{t+1}^T - a_1 C_t^T)]$$

Therefore, the consumption model is a model with MA(1) error term. In aggregate studies, it has been assumed that the transitory consumptions are cancelled out. Most studies do not report the LM test for the serial correlation and normality test. It is not clear how well these models conform to the assumption and thus whether ols is consistent.

The time framework may also be an important issue in these studies as Wickens and Molana (1984) have discovered. If the planning is revised monthly and the econometrician employs quarterly data, then the error term will be correlated with the regressors. However, if one were to use the regressors which are lagged more than one quarter, one can avoid the problem (Hall (1988) and Harvey (1988)). In view of the possible presence of transitory components and incorrect timing, it may be useful to guard against the possibility of serial correlation in some of our estimations as in Chapter 6.

(4) Hansen and Singleton's (1982) method of moments allows for nonlinear expectations. All the other studies assume linear rational expectations implied by the log–normality assumption. Although every stationary series can be approximated by an ARMA series, rational expectations do not necessarily follow a linear stochastic specification as such. If one is interested in a statistical relationship, there is no reason why one should restrict rational expectations as a linear function. A more robust formulation seems appropriate and possibly one can employ some nonparametric techniques to infer rational expectations from the data itself.

In the case of the model consisting of surprise terms, the estimators will generally fail to be consistent if the rational expectations are not linear. Furthermore, most test statistics will give incorrect rejection probabilities. Therefore, it is vital to take a flexible approach to surprise consumption modelling.

The problem of nonlinearity has been ignored by econometricians as technically, it is difficult to tackle. But, in economics, nonlinearity is the rule rather than exception. One can test for the misspecification of the linear functional form by using a variety of the traditional techniques. In Chapter 7, we investigate the adequacy of the specification of the supplementary regression and the normality assumptions.

But once the null hypothesis of a particular parametric formation is rejected, there is no a priori reason to model nonlinearity in a particular form. Since the expectations are generally not of interest as far as estimation is concerned in some of the studies, it may be worthwhile to attempt to tackle the expectations in a non–traditional manner as in Chapter 6 and 7.

(5) There has been a lot of attention on cointegration and unit roots in recent years since the contributions of Nelson and Plosser (1982), who suggested that unit roots were common in macroeconomic time series. There are a lot of econometric papers on the subject of whether we reject the hypothesis too often. But it is of interest to mention the results of Mankiw and Shapiro (1985, 1986), Banerjee and Dolado (1987a, 1987b), Banerjee, Dolado and Galbraith (1987) and Nelson (1987). Let us consider a simple test of the LCPIH:

$$\Delta C_t = \pi_0 + \pi_1 Y_{t-1} + \epsilon_t$$

We wish to test the null hypothesis that there is no excess sensitive to income, i.e. H_0: $\pi_1 = 0$. The equation may in fact be the reduced form from Flavin's just–identified system of equations:

$$Y_t = \delta + \rho Y_{t-1} + \nu_t$$
$$\Delta C_t = \alpha + \beta(Y_{t-1} - E_t Y_t) + \epsilon_t^*$$

If Y_t is non–stationary, e.g. $\rho = 1$ or ν_t is MA(1), then as Mankiw and Shapiro (1985)'s Monte Carlo results have shown, the t–statistics are biased towards rejection of LCPIH (i.e. bias towards rejecting H). These results are not surprising to econometricians, as discovered earlier by other time series researchers such as Dickey and Fuller (1981). They obtained further results that if Y_t were to follow a strongly autoregressive or borderline stationary process (i.e. $|\rho| < 1$) , the bias would still be too large to be taken lightly. So, conventional test statistics are misleading in the presence of unit roots or strong autoregressive process because they do not give correct rejection probabilities. One can therefore find a significant time trend when in fact there is none, and worst of all, inappropriate detrending will give rise to spurious cycles when there are none.

Although Mankiw and Shapiro (1985)'s results are based on a small sample size of 100, it nevertheless is relevant to macroeconomic time series studies. Similar conclusion is obtained in a later paper by Mankiw and Shapiro (1986) for orthogonality tests. Barnerjee et.al. (1987a) provide an analytical Monte Carlo interpretation using Nagar expansions of the moments for the t–statistics.

However, Barnerjee et. al. (1987) present results which suggest that the critical values for the t and F tests are sensitive to the data generation process. The inclusion of extra regressors may in fact bring the critical values to the nominal ones. So, the initial fear of overrejections may be exaggerated. In a later paper, Banerjee et. al. (1987b)'s results confirmed their previous findings. The point here is that the magnitude of the bias of the test is sensitive to the inclusion and exclusion of certain regressors. One cannot place too much faith in the test results obtained in the presence of unit roots. Surely, such an issue will need to be settle in the future.

Although we do not address the issue of co–integration directly, we included lagged nominal consumption and income in our surprise consumption function in Chapter 7. It will be interesting to see whether the magnitude of these coefficients are close to one, thus suggesting that we may simply take the difference of the two variables as a regressor.

(6) It is easy to believe that a proportion of the consumer is liquidity constrained or simply myopic, i.e. they consume all of their disposable income. These consumers will explain why the models are rejected, as they cannot conform to the predictions of the model. However, empirical results have demonstrated that this is only a small proportion, perhaps much less than 20%. It is hard to explain for the rejections at the aggregate level.

(7) Most studies assume a very restrictive functional form for preferences e.g. CRRA, inter– and intra– temporal separability. Perhaps one other area of fruitful research is the use of semiparametric techniques which combine both parametric and nonparametric components. We, however, do not pursue along this line as one of our objectives is to compare the semiparametric estimates with previous parametric results.

(8) With the exception of Miron (1986) and Sargent (1978), the other studies use seasonally adjusted data. Sargent's definition of permanent income includes capital income and thus the results are not conclusive. Miron pointed out the rejections of the LCPIH may be due to the treatment of seasonal adjustments. He has presented evidence that is favourable to the LCPIH by using seasonally unadjusted data. This result is perhaps not too surprising. It is easy to argue that seasonal adjustments are predictable and thus must be taken into account in the model. Again, in order to compare our results with previous studies, we have not taken seasonality into account.

5.7 STUDIES USING HOUSEHOLD DATA

There are several studies which employed individual household data. Cross sectional studies allow the testing of the existence of liquidity constrained consumers. Panel data is now favoured by econometricians for investigation in consumption but suffers from problems which are only too well known.

The quality of data may not be as good as one would like it to be. Although panel data may have several advantages over the cross sectional data, there are problems such as measurement errors and household dropping out of the survey. The results would be sensitive to these defects. It may be necessary to study the method of collection so that one can adopt the appropriate techniques of estimation.

There may be difficulties in identifying the parameters of the model. As there are heterogeneous preferences and these preference parameters are correlated with observable variables, it may be hard to identify the model.

There are many studies which involved micro data, such as Hall and Mishkin (1982). Most of these works included a test of liquidity constraints. The survey by Hayashi (1987) gave some insights into the future directions of research in this area.

CHAPTER 6

SEMIPARAMETRIC ANALYSIS OF A CONSUMPTION MODEL

6.1 RELATIONSHIP BETWEEN CONSUMPTION AND REAL INTEREST RATES

As described in last chapter, there have been numerous empirical studies on the relationship between consumption growth and expected real interest rates. The linear rational expectations model which describes that expected consumption growth alone predicts real returns or vice versa has its foundation in the expected utility framework and is generally obtained by imposing distribution assumptions on observable variables in the Euler equation. While financial economists are more interested in the coefficient of relative risk aversion, macroeconomists are more concerned with the elasticity of intertemporal substitution between adjacent periods.

The earlier studies of the consumption capital asset pricing model assumed that the coefficient of relative risk aversion was inversely related to the coefficient of the elasticity of intertemporal substitution. Hansen and Singleton (1983) have applied the linear model, which is consistent with intertemporal separable isoelastic utility and the log–normality assumption of the data, to U.S. macroeconomic consumption series and asset returns and have found that the evidence were unfavourable to the model. In particular, the overidentifying restrictions were rejected, indicating misspecifications.

The earlier empirical studies using U.K. time series data on the life cycle–permanent income hypothesis have in fact arrived at the approximate linear relationship without resorting to the distribution assumption, e.g., Muellbauer (1983), Wickens and Molana (1984). These models, like many others, originated from the work of Hall (1978) which assumed that the interest rates was constant.

There are three contributions in this chapter. First of all, we have examined the role of inflation variables in the consumption model. Koskela and Viren (1987) have found roles for expected inflation and variability of inflation which contradicts the prediction of the model. In this chapter, we have modelled the inflation expectations using the general to specific methodology. We have allowed for the presence of an ARCH effect of unknown form. We call this a semiparametric ARCH (SARCH) model. The SARCH specification is more realistic for two reasons. First of all, inflation tends to persist and secondly, the true functional form of the ARCH process may not be known. If indeed the consumers form their inflation expectations in such a fashion, we find no evidence of the expected inflation or variability having any roles in the linear model.

Secondly, we have attempted to estimate the coefficient of intertemporal substitution using method of moments and allowing for rational expectations of unknown functional form. It is known that consistent estimate of the intertemporal elasticity of substitution can be obtained from a linear rational expectations model by two stage least squares. It is also known that the log–normality assumption implies linear formation of expectation and this has been the favourite approach. In view of recent contradicting results of Hall (1988) and previous studies, we assume that the approximate linear relationship holds and allows for possible nonlinear formation of expectation.

Finally, we have studied and compared the behaviour of two automatic bandwidth selection criteria with other rules–of–the–thumb (RT). Most of the semiparametric models are not truly "adaptive" because a bandwidth has to be selected subjectively. Our main contribution here is the use of automatic bandwidth selection which renders the two semiparametric models used in this study to be fully automatic and less subjective. We have compared these with other estimates obtained under various bandwidth selection rules which minimize the MSE in density estimation.

The plan of the chapter is as follows: in Section 2, we outline one derivation that will give rise to a linear regression model; in Section 3, we present the semiparametric ARCH model for the formation of inflation expectation; in Section 4, a pseudo Gaussian log–likelihood criterion is suggested for bandwidth selection; the results for ARCH modelling of inflation expectation are discussed in Section 5; Section 6 discusses the consistency of the two stage procedure and test; Section 7 presents the results of the test using different models for expectation formation; in Section 8, some semiparametric estimates of the intertemporal elasticity of substitution are presented for quarterly U.S. data; and finally, we give the conclusion and further comments in Section 9.

6.2 THE MODEL

The following is the simplest alternative derivation that will give rise to the linear regression model discussed in the last chapter. As in Hall (1988), we assume that the consumer maximizes the expected utility

$$E_t \Sigma_{j=0}^{\infty} \exp(-\rho_j)\ U_{t+j}$$

where E_t is the expectation conditional on the information set Ω_t, ρ is the discount factor and U_t is intertemporally separable and isoelastic, i.e.,

$$U_t = I(\alpha \neq 1) C_t^{1-\alpha}/(1-\alpha) + I(\alpha = 1) \mathrm{Log} C_t, \qquad \alpha > 0$$

where I is the usual indicator function, α is the inverse of the elasticity of substitution, σ. It is not necessary that α is the coefficient of risk aversion (γ) in this case. As Hall (1988) have argued, it is also not necessary to state the budget constraint explicitly as the model is consistent with a full contingent of commodity markets as well as the representative consumer holding a single risky asset. Assuming that the stochastic return on a single asset is $\exp(R_t)$ in year t, then the stochastic Euler equation between t+j and t is

$$E_t[(C_{t+j}/C_t)^{-\alpha} \exp(R_{j,t} + j\rho)] = 1$$

where $R_{j,t}$ is the real j–period return on the asset from time t to t+j. The equation states that the marginal rate of substitution is equal to the prices of present and future consumption. At this point, if one were willing to make the log–normality assumption on consumption growth and interest rate, one will end up with the following exact relationship between consumption growth and expected real interest rates for j=1:

$$\Delta c_t = \alpha_0(\rho) + \sigma E_t(R_t) + \epsilon_t \qquad t=1,\ldots,T \quad (1)$$

(1) may still hold approximately without the assumption of log–normality. The coefficient of intertemporal elasticity of substitution determines the relationship between consumption growth and expected interest rates with a constant. This describes that the rate of growth of consumption can be predicted by expected interest rates alone, and any past information should not have explanatory power. The intercept α_0 is a function of ρ which cannot be identified when a single asset is used. σ is the elasticity of intertemporal substitution. If σ is high, consumers will defer a great deal of their consumption to a later period when the real interest rates is expected to be high. While we are working with only j=1 here, in general, it may also be of interest to look at j of other values. Harvey (1988) has looked at the relationships between two, three and four quarter growth rates and expected real interest rate. First of all, we are interested in the role of inflation variables on consumption. In particular, we are interested in the expectation of inflation and variability of inflation.

6.3 SEMIPARAMETRIC ARCH (SARCH) MODEL AND BANDWIDTH SELECTION

There are numerous ways to obtain the expected real rates of interest. One can model the interest rates as an autoregressive process. Alternatively, the expected interest rates is the difference between nominal interest rates and expected inflation. We adopt the latter approach as in Harvey (1988) and Koskela and Viren (1987).

Inflation expectations can be taken from surveys as in Hall (1988) and others. Harvey (1988) has modelled inflation as an IMA(1,1) process while Koskela and Viren (1987) have used AR(3) and rolling AR(3) specifications. In order to allow for inflation to persist, we resort to modelling inflation using the the semiparametric

ARCH process. Consider the following model

$$\text{infl}_t = w_t'\delta + u_t \qquad t=1,\dots,T \quad (2)$$

where $E(u_t) = 0$, $E(u_t^2) = \sigma^2$, and w_t is a vector of weakly exogenous variables. Let Ω_t be the information set. In Engle (1982, 1983), u_t/Ω_t is normally distributed with mean zero and variance $h_t = h(u_{t-1},\dots,u_{t-p})$, where h is linear in the square of the lagged residuals. We shall refer to the model with linear ARCH process as Engle's model. Here, it is assumed that the u_t's are uncorrelated and w_t may include lagged pt. The appropriateness of conditional normal distribution can be checked by diagnostic test. In other financial models with nonlinearity, Engle and Bollerslev (1986) have demonstrated that a conditional t–distribution might give a better fit. The semiparametric ARCH (SARCH) model was first suggested by Robinson (1987) and an augmented version has been applied in Whistler (1988).

We seek to improve efficiency, and at the same time wish to guard against misspecification in Gaussian ARCH model. Parametric ARCH process is not only restrictive, it may also be inconsistent when the parametric form is incorrectly specified. When h is a function of δ, misspecification of the conditional variance will generally lead to inconsistent δ. Weiss (1984) has in fact found that, for many macroeconomic time series, terms such as infl_{t-1}^2 (a function of $w_{t-1}'\delta$) appear in the conditional variance.

Pagan and Sabau (1987) have shown that the Gaussian model of Engle is robust under various forms of misspecifications. Notably, if h is even function of $\{u_{t-j};j=1,\dots p\}$, u_t/Ω_t is symmetric around zero, δ remains consistent. This includes the log–linear formulation of Geweke (1986) and GARCH process of Bollerslev (1986). Under further restrictive assumptions, the presence of the strongly exogenous w_t in h_t, the Engle's specification may still retain consistency for δ. This is not true if w_t is weakly exogenous as in our case. OLS standard errors are also incorrect but a consistent variance–covariance matrix is available, e.g.,

$$(T^{-1}\Sigma_t w_t' w_t)^{-1}(T^{-1}\Sigma_t\, w_t' w_t \hat{h}_t)(T^{-1}\Sigma_t w_t' w_t)^{-1}$$

We assume h_t includes lagged u_{t-1} as well as lagged conditional mean, i.e., $h_t = h(u_{t-1},\ldots,u_{t-p},w_{t-1}'\delta,\ldots,w_{t-q}'\delta)$. The specification will be able to handle the case where lagged p's are present. However, extensions to GARCH and ARCH–M models which are much sought after in financial economics would be difficult if not impossible. Pagan and Ullah (1988) have instead suggested the use of IV semiparametric estimation for consistency in such cases, as the estimates from more complicated models are not likely to be robust.

The conditional log–likelihood function for SARCH(p,q) is

$$\text{Log } L(\delta) = -(2T)^{-1}\Sigma_{t=pq+1}^{T}\{\log 2\pi + \log h_t + \frac{(\text{infl}_t - w_t'\delta)^2}{h_t}\} \quad (3)$$

where pq = max(p,q). Our model also embodies the case where h_t depends on lagged infl_t (Engle's ARCH model is inconsistent in this case). It can be seen from the fact that if $u_t^2 = \alpha_0 + \alpha_1 \text{infl}_{t-1} = \alpha_0 + \alpha_1(w_{t-1}'\delta + u_{t-1})$, h_t is a function of lagged mean and residuals. This is the main advantage of making h_t a function of lagged u_t rather than lagged squared u_t. Consider the Gauss–Newton iteration scheme:

$$\tilde{\delta} = \hat{\delta} - \lambda(\hat{A}^{-1}\hat{d})$$

$$\hat{A} = T\, \text{DLogL}(\hat{\delta})\, \text{DLogL}(\hat{\delta})'$$

$$= T^{-1}\,\Sigma_{t=pq+1}^{T}\{\frac{w_t w_t'}{\hat{h}_t} + \frac{\hat{H}_t \hat{H}_t}{2\hat{h}_t^2}\} \quad (4)$$

$$\hat{d} = \text{DLogL}$$

$$= T^{-1}\,\Sigma_{t=pq+1}^{T}\{\frac{w_t u_t'}{\hat{h}_t} + \frac{\hat{H}_t}{2\hat{h}_t}(\frac{\hat{u}_t^2}{\hat{h}_t} - 1\} \quad (5)$$

H_t is the derivative of the conditional variance with respect to the unknown parameter δ. The second term of (4) and (5) on the right hand side reflects the presence of the ARCH effect. If ARCH is not present (e.g., independent of lagged u_t and w_t is strongly exogenous), H_t disappears and we are left with a GLS model. Denote $\omega_{t-k} = w_{t-k}'\delta$:

$$\hat{H}_t = \frac{\partial \hat{h}_t(u,\omega)}{\partial \delta} = -\Sigma_{\ell=1}^{p} w_{t-\ell} \frac{\partial \hat{h}_t(u,\omega)}{\partial u_\ell} + \Sigma_{k=1}^{q} \frac{\partial \hat{h}_t(u,\omega)}{\partial \omega_k}$$

Further iterations may be conducted and convergence can be achieved if λ is chosen appropriately (e.g. Berndt, Hall, Hall and Hausman (1974)). $\hat{\delta}$ is the OLS estimate and the one–step update with $\lambda = 1$ is asymptotically efficient. It can easily be seen that under some regularity conditions, $T^{1/2}$DLogL(δ) is asymptotically normally distributed with zero mean and covariance matrix A at the true δ. Then $\tilde{\delta}$ is adaptive as it has the same limiting distribution as parametric ARCH estimator obtained under the assumption that h_t is known with certainty.

The nonparametric elements come in when we try to estimate the conditional variance $E[u_t|\Omega_t]$ and its derivative. Although there are numerous methods for nonparametric estimation of the conditional variance, our preferred choice here is the method of kernel because of its ease in dealing with multivariate time series and its properties are well known as we have mentioned in Chapter 2 (see also Robinson (1983, 1986a, 1986b)). Using the OLS estimates, we can estimate h_t by nonparametric regression of $\hat{u}_t^2$ on $\hat{u}_{t-1},..,\hat{u}_{t-p}, w_{t-1}'\hat{\delta},..,w_{t-q}'\hat{\delta}$. In particular, we adopt the N–W estimator and its derivative described in Chapter 2. We estimate the conditional variance by regressing $\hat{u}^2$ on lagged u's and ω's. In terms of the notations here, we have:

$$\hat{h}_t = \hat{g}(u_{t-1},..,u_{t-p},\omega_{t-1},..,\omega_{t-q})/\hat{f}(u_{t-1},..,u_{t-p},\omega_{t-1},..,\omega_{t-q})$$

where

$$\hat{g}_t = (Ta^{p+q})^{-1}\Sigma_{\tau=pq+1}^{T} \hat{u}_\tau^2 K_{t\tau}(u,\omega)$$
$$\hat{f}_t \doteq (Ta^{p+q})^{-1}\Sigma_{\tau=pq+1}^{T} K_{t\tau}(u,\omega)$$

$$K_{t\tau}(u,\omega) = \Pi_{\ell=1}^{p} (2\pi)^{-1/2}\exp -1/2 \{(\hat{u}_{t-1-\ell} - \hat{u}_{\tau-1-\ell})/a\}^2 \times$$
$$\Pi_{k=1}^{q} (2\pi)^{-1/2}\exp -1/2 \{(\hat{\omega}_{t-1-k} - \hat{\omega}_{\tau-1-k})/a\}^2$$

To estimate the derivative, H_t, we need to consider the nonparametric estimation of derivative of regression curves. Consider

$$\frac{\partial\hat{h}_t}{\partial u_\ell} = \frac{\partial\hat{g}_t/\partial u_\ell}{\hat{f}_t} - \frac{\hat{g}_t \;\partial\hat{f}_t/\partial u_\ell}{\hat{f}_t^{\,2}}$$

$$\frac{\partial\hat{h}_t}{\partial \omega_\ell} = \frac{\partial\hat{g}_t/\partial \omega_\ell}{\hat{f}_t} - \frac{\hat{g}_t \;\partial\hat{f}_t/\partial \omega_\ell}{\hat{f}_t^{\,2}}$$

$$\partial\hat{g}_t/\partial u_\ell = -(Ta^{p+q})^{-1}\Sigma_{\tau=pq+1}^{T} u_{t\tau}\, \hat{u}_{\tau-\ell}^{2} K_{t\tau}(u)$$
$$\partial\hat{f}_t/\partial u_\ell = -(Ta^{p+q})^{-1}\Sigma_{\tau=pq+1}^{T} u_{t\tau}\, K_{t\tau}(u)$$
$$\partial\hat{g}_t/\partial \omega_\ell = -(Ta^{p+q})^{-1}\Sigma_{\tau=pq+1}^{T} \omega_{t\tau}\, \hat{\omega}_{\tau-\ell}^{2} K_{t\tau}(u)$$
$$\partial\hat{f}_t/\partial \omega_\ell = -(Ta^{p+q})^{-1}\Sigma_{\tau=pq+1}^{T} \omega_{t\tau}{}^{2} K_{t\tau}(u)$$

6.4 LIKELIHOOD BANDWIDTH SELECTION IN SARCH MODEL

The bandwidth in empirical work is generally subjectively chosen and set to a number proportion to n, e.g. Whistler (1988) which let a = std dev (u) x n x constant. There is no justification for choosing the standard deviation as the constant C except the consolation that it at least depends on the data. There are automatic methods of choosing bandwidth in density estimation and indeed for regression function. As for semiparametric models, very few results exist. Robinson (1988d) has shown that his bandwidth selection rule for Hannan (1963)'s GLS estimators gives a consistent bandwidth. It seems reasonable to motivate a similar criterion function following Robinson (1988d)'s results.

We therefore adopt the more objective way of selecting the bandwidth. Automatic bandwidth selection usually involves the use of the sample to help us in selecting a plausible bandwidth and then re–use the sample to estimate the parameters, i.e., we are using a subsample to estimate the bandwidth a, which in turn provides us with the nonparametric estimates.

The log–likelihood criterion of choosing a bandwidth is to find an a to minimize the function CV:

$$a_{opt} = \operatorname{argmin}_a \mathrm{CV}(a)$$
$$= \operatorname{argmin}_a T^{-1}\Sigma_{t=pq+1}^{T}\{\log h_t^*(a) + \frac{(\mathrm{infl}_t - w_t'\hat{\delta})^2}{h_t^*(a)}\}$$

The first derivative is given by

$$\frac{\partial \mathrm{CV}}{\partial a} = T^{-1}\Sigma_{t=pq+1}^{T}\frac{\partial h_t^*(a)}{\partial a} h_t^{*-1}[1 - \frac{(\mathrm{infl}_t - w_t'\hat{\delta})^2}{h_t^*(a)}]$$

$$\frac{\partial h_t^*(a)}{\partial a} = \frac{\Sigma_\tau \hat{u}_\tau^{\,2} L_{t\tau}}{\Sigma_\tau K_{t\tau}} - \frac{\Sigma_\tau \hat{u}_t^{\,2} K_{t\tau}\, \Sigma_\tau L_{t\tau}}{\Sigma_\tau K_{t\tau} \Sigma_t K_{t\tau}}$$

where Σ_t denotes summation from observation pq+1 to T for $\tau \neq t$ and

$$L_{t\tau} = \frac{u_{t\tau}}{a} K_{t\tau}$$

h_t^* is the leave–one–out nonparametric regression estimates, i.e.,

$$h_t^* = g_t^* / f_t^*$$
$$g_t^* = (Ta^{p+q})^{-1}\Sigma_{\substack{\tau=pq+1\\ \tau\neq t}} \hat{u}_\tau^{\,2} K_{t\tau}(u,\omega)$$
$$f_t^* = (Ta^{p+q})^{-1}\Sigma_{\substack{\tau=pq+1\\ \tau\neq t}} K_{t\tau}(u,\omega)$$

The leave–one–out estimates of the conditional variance h_t are the predictions from using the subsample of T–1 observations. In forcing the CV function to attain its minimum is to estimate δ well. Practical experience suggests that this function is fairly well behaved and not difficult to locate the global minimum in most cases. However, it is always wise to check for local minimum because sometimes they do occur. Of course, grid search can be used. In general, a simple quadratic interpolations optimization routine will be adequate in dealing with the function. The results and more discussions are presented in the next section.

6.5 INFLATION EXPECTATIONS

As in Koskela and Viren (1987), we have used seasonally adjusted U.S.

quarterly data on nondurable consumption (C) over the period 1951Q1 to 1986Q2 from the Business Conditions Digest. The consumption series at constant 1982 prices P_{1982}, the inflation rate is a standard measure, i.e.,

$$INFL_t = 400\ (1 + \log(P_t/P_{t-1})^4)^{-1}.$$

The interest rates variable (R^*) is the yield on 3–month Treasury bills. Real interest rates (R) variable is obtained by deflating the nominal rate by the inflation rate. To be consistent with Koskela and Viren (1987), we ignore any adjustment to take into account taxation.

We have obtained the OLS estimates using the general to specific methodology of Hendry (see Pagan (1987)). We included six lags each of log of real interest rates (r), consumption growth (Δc) and rate of change of inflation (Δinfl) with a constant in the information set. The parsimonious model is reported in Table 6.1 with the associated test statistics. The interest rates variable is not statistically significant at all but we will retain it for theoretical reasons. The t–ratios are presented in parentheses with heteroscedasticity–robust (Eicker (1963), White (1980)) t–ratios in the parentheses below.

While the model has captured the salient features of the data, the null hypothesis that the residuals follow Engle (1982)'s ARCH(1) process is rejected. There is no indication of any rejection of the hypothesis that the residuals are Gaussian. However, the ARCH test is unlikely to be powerful when there is misspecification of the conditional variance, e.g., when the conditional variance is not linear function. In view of this, we proceed to estimate the models with ARCH(1), SARCH(1) and SARCH(1,1) error processes and compare the results. Of course, one of the roles that semiparametric models can serve is diagnostic checking.

The estimates of the parametric ARCH model are obtained by the standard procedures described in Engle (1982) and the one step estimates are reported in Table 6.2. The ARCH(1) estimates are fairly similar to the OLS estimates except for lagged consumption growth variables.

The SARCH(1) estimates are also reported in Table 6.2. We shall denote the bandwidth from minimizing CV as a_{CV}, and denote a_{SD} = s.d.(u) x n and a_{UNIT} = n. The estimates obtained by a_{CV}, a_{SD} and a_{UNIT} are reported under the heading CV, SD and UNIT respectively. The optimized CV bandwidth is almost double that of a_{SD} and three and a half times that of a_{UNIT}. This suggests that the conditional variance from CV is much smoother than that from using the SD and UNIT rules.

It is worth looking at the CV function for SARCH(1) and SARCH(1,1) as plotted in Figure 6.1a and 6.1b respectively. Both functions are very well behaved convex functions and there is no problem in locating the minima.

Although it may be desired to scale the lagged u_t and ω_t to have unit variance, we did not employ this device here. There is, however, not too large a difference in the semiparametric estimates. The estimates are fairly similar except for the interest rates variable and consumption growth lagged five quarters. The standard errors of these estimates are larger than those of ARCH(1).

The SARCH(1,1) estimates are presented in Table 6.3. a_{CV} is again very large relative to a_{SD} and a_{UNIT} (we use $n = T^{-(1/5)}$ instead of $n = T^{-(1/6)}$ suggesting that ${u_t}^2$ is a fairly smooth function of u_{t-1} and the mean. This may be due to the fact that there is not a presence of strong ARCH effect. The CV estimates are almost identical to the CV SARCH(1) estimates although the standard errors are slightly larger for SARCH(1,1) estimates. The expectation of inflation from ARCH and SARCH are fairly similar since the estimates are similar.

Having obtained these estimates, it is relatively straightforward to form the interest rates expectation by subtracting the inflation expectation from the nominal rate of interest. The conditional variance can be used as a proxy for the variability or risk associated with inflation. We will use these estimates in Section 7 on examining the effect of inflation variables on consumption growth.

6.6 SURPRISES MODELS AND CONDITIONAL VARIANCE

Consider the following model:

$$y_t = \beta_1 E_t z_t + \beta_2(x_t - E_t x_t) + \beta_3 \sigma_{1t}^2 + \epsilon_t \tag{6}$$
$$x_t = E_t x_t + \eta_t = w_t'\delta + \eta_t$$
$$\sigma_t^2 = E_t(\eta_t^2)$$

The results below hold for the case where β_1 is a vector. If we replace $E_t x_t$, $(x_t - E_t x_t)$ and $\sigma_t 2$ in (6) by appropriate estimates, the estimates from regressing y_t on these generated regressors are consistent under some regularity conditions. Pagan and Ullah (1988) have recently studied the consistency of models containing risk terms which is similar to ours here. The difference is that the analysis is complicated by the presence of expectation and surprise terms. We are content to present the arguments that the OLS t–ratios will provide a consistent test for the model using fairly high level assumptions which exist under some regularity conditions. We can rewrite (6) as

$$y_t = \beta_1 \hat{x}_t + \beta_2(x_t - \hat{x}_t) + \beta_3 \psi_t^2 + [(\beta_1 - \beta_2)(\eta_t - \hat{\eta}_t)$$
$$+ \beta_3(\sigma_{1t}^2 - \psi_t^2) + e_t$$

The following proposition is a slight modification of Pagan (1984)'s Theorem 12.

Proposition: If we have $\psi_t^2 = w_t'\delta + u_t$ as in ARCH model and

(A1) x_t and w_t are jointly stationary and ergodic process.

(A2) $\text{plim}_{T\to\infty} T^{-1}\Sigma_t x_t' e_t = 0$, $\text{plim}_{T\to\infty} T^{-1}\Sigma_t w_t' e_t = 0$,

$\text{plim}_{T\to\infty}(\hat{\delta} - \delta) = 0$, $\text{plim}_{T\to\infty} T^{-1}\Sigma_t(\hat{\delta} - \delta)e_t = 0$

(A3) $T^{1/2}(\hat{\delta} - \delta) \sim N(0,V)$

Then

(i) $\text{plim}_{T\to\infty}\hat{\beta} = \beta$

(ii) $T^{-1/2}(\hat{\beta} - \beta) \sim N(0,BEB')$

where $B = (\text{plim}_{T\to\infty} T^{-1}\Sigma_t x_t x_t')^{-1}$

$$E = \sigma_e^2(\mathrm{plim}_{T\to\infty}T^{-1}\Sigma_t x_t x_t') + (\beta_2-\beta_1)^2\sigma_\eta^2(\mathrm{plim}_{T\to\infty}T^{-1}\Sigma_t x_t x_t')$$
$$+ \beta_3^2(\mathrm{plim}_{T\to\infty}T^{-1}\Sigma_t x_t w_t')\ V\ (\mathrm{plim}_{T\to\infty}T^{-1}\Sigma_t w_t x_t').$$

The proof is a straightforward extension of Pagan's and it is not worth repeating the process but a sketch is provided. Write

$$y_t = \beta_1\hat{x}_t + \beta_2\hat{\eta}_t + \beta_3\psi_{1t}^2 + Q_t$$
$$\hat{\beta} = \beta + (\Sigma_t x_t x_t')^{-1}\Sigma_t x_t Q_t$$
$$= \beta + (x_t x_t')^{-1}\begin{bmatrix}\Sigma_t\hat{x}_t Q_t \\ \Sigma_t\hat{\eta}_t Q_t \\ \Sigma_t\psi_{1t} Q_t\end{bmatrix} \qquad (7)$$

By construction, $\mathrm{plim}_{T\to\infty}T^{-1}\Sigma_t\hat{x}_t(\eta_t-\hat{\eta}_t)$, $\mathrm{plim}_{T\to\infty}T^{-1}\Sigma_t\hat{x}_t(\sigma_{1t}^2-\psi_{1t}^2)$, $\mathrm{plim}_{T\to\infty}T^{-1}\Sigma_t\hat{x}_t e_t$, $\mathrm{plim}_{T\to\infty}T^{-1}\Sigma_t\hat{x}_t(\hat{\eta}_t-\eta_t)$, $\mathrm{plim}_{T\to\infty}T^{-1}\Sigma_t\hat{\eta}_t(\eta_t-\hat{\eta}_t)$, $\mathrm{plim}_{T\to\infty}T^{-1}\Sigma_t\hat{\eta}_t(\sigma_{1t}^2-\psi_{1t}^2)$, $\mathrm{plim}_{T\to\infty}T^{-1}\Sigma_t\hat{\eta}_t e_t$, $\mathrm{plim}_{T\to\infty}T^{-1}\Sigma_t\psi_{1t}(\eta_t-\hat{\eta}_t)$, $\mathrm{plim}_{T\to\infty}T^{-1}\Sigma_t\psi_{1t}(\sigma_{1t}^2-\psi_{1t}^2)$, and $\mathrm{plim}_{T\to\infty}T^{-1}\Sigma_t\psi_{1t}e_t$ are all zeros. Thus $\hat{\beta}$ is consistent.

(ii) follows by looking at the distribution of $T^{1/2}(\hat{\beta}-\beta)$ which depends on $T^{1/2}\Sigma_t x_t Q_t$, and the latter is asymptotically normally distributed with zero mean and variance–covariance matrix E by assumption. Since $[\mathrm{plim}_{T\to\infty}T^{-1}\Sigma_t x_t x_t']^{-1} = B$, applying Cramer's Theorem completes the proof.

We are interested in testing the hypothesis that $(\beta_1,\beta_2,\beta_3)' = 0$. The standard errors from OLS are clearly understated and equal to the true standard errors under the null. Therefore, we will have a consistent test of the hypothesis that the individual coefficient is statistically insignificant from zero.

Similarly, if h is estimated nonparametrically by kernel estimation, then as Robinson (1983, 1986a, 1986b) has shown, under some regularity conditions, $\tilde{h}$ is $(Ta^{p+q})^{1/2}$ consistent and asymptotically normal. We can replace ψ_t^2 by $\tilde{h}_t$ in (6) and the consistency result holds. Of course, the rate of convergence would be expected to be slower in the nonparametric case. Nevertheless, a similar expression should exist for the case when $\tilde{h}_t$ is used except that the last term for E will be different. We have therefore established that the OLS standard errors are generally lower than the true standard errors and we have a consistent test for zero–type restrictions.

6.7 RESULTS

Returning to our model which is similar to that of Koskela and Viren (1987), i.e.

$$y_t = \alpha + \sigma E_t R_t + \beta_1 E_t \text{infl}_t + \beta_2(\text{infl}_t - E_t \text{infl}_t) + \beta_3 \sigma_{1t}^2 + \epsilon_t$$
$$E_t R_t = R_{t-1}^* - E_t \text{infl}_t$$

The results using conditional variance from ARCH(1) and SARCH(1) models are reported in Table 6.4 while that of SARCH(1,1) are reported in Table 6.5. The conclusion from using ARCH(1), SARCH(1) and SARCH(1,1) estimates is the same. If we use the conditional variance to measure variability of inflation, we find that the coefficient is not significant in both models. Semiparametric estimation reduces the conditional variance estimates by a great deal and this is reflected in the increase in value of its coefficient. Expected inflation variable is not statistically significant suggesting that it does not help to predict consumption growth. As for inflation surprises, there may be roles of unanticipated inflation, e.g., price confusion effect. There is therefore no significant evidence to suggest that expected inflation or inflation variability matters. The puzzle of negative estimates for expected interest rates in empirical studies is not a new one. The conclusion is unchanged if log of real rates rather than the level is used. This implies that the utility function is non–concave thus violating the assumption of the model. There are various explanations for the presence of negative estimates such as the violations of underlying assumptions of the model, e.g., separable isoelastic utility function and log–normality assumption. It may well be that the model does not allow for other behaviour, motives or the omission of some important variables in the utility function.

From the time series point of view, we make the following observations. First of all, the real rates that we used may be an inappropriate instrument. Secondly, we have used seasonally adjusted data which will not be able to capture seasonal change in taste. But recent results suggest that σ is insignificantly negative for one period growth rate using seasonally unadjusted data (Harvey (1988)). Thirdly, the incorrect choice of instruments due to incorrect assumption on revision periods will give rise to the time aggregation problem. Recently, Hall (1988) has argued that the time aggregation problem is severe and correction generally brings a lower non–negative estimate for σ. His results demonstrate that by using appropriate instruments, one should in fact find no evidence of strong intertemporal substitution. However, there is still a problem in establishing a relationship.

In the next section, we confine ourselves to the statistical relationship between consumption growth and expected real rates of interest. We seek to answer the question of whether the semiparametric estimates support the results of Hall (1988), that there is very little evidence to suggest the presence of strong intertemporal substitution. We therefore attempt to seek more efficient instruments in forming the expectations in the next section.

6.8 AUTOMATIC BANDWIDTH SELECTION AND ELASTICITY ESTIMATES

Consider the model

$$y_t = \beta_0 + \beta_1 E_t x_t + \epsilon_{1t}$$

$$x_t = E_t x_t + \eta_t$$

or

$$y_t = \beta_0 + \beta_1 x_t + (\epsilon_t + \beta_1(E_t x_t - x_t))$$

$$= \beta_0 + \beta_1 x_t + \epsilon_{2t}$$

The usual assumption in this model is that η_t is normally distributed and therefore $E_t x_t$ is linear in its mean. As we have mentioned above, the relationship holds approximately even without the log–normality assumption. We relax this assumption here and attempt to seek a more efficient instrument for $E_t x_t$ which may be nonlinear. In our problem, we are interested in finding appropriate instruments $\hat{x}_t$ such that the conditional moment restrictions $E[\hat{x}_t | \epsilon_{2t}] = 0$ is satisfied. The consistency and asymptotic normality of IV or method of moments estimators appropriate for time series has been studied in the econometric literature. For consistency, any appropriate instruments (parametric or nonparametric) would do as long as the instruments are not correlated with the error term ϵ_{2t}. In the case of independent observations, Newey (1987) has shown that we can achieve the same asymptotic efficiency if we replace $\hat{x}_t$ of known functional form with one which does not assume any functional form using nearest neighbours estimates for $\hat{z}$. Recently, Pagan and Ullah (1988) have discussed the use of instrumental variable estimation in the model with risk terms in the case of stationary time series. In brief, with the help of higher–order kernels (Barlett (1963)) and the device to trim out small density estimates, one should be able to show that under some regularity conditions

$$T^{-1/2}(\tilde{\beta}-\beta) \sim N(0,V)$$
$$\tilde{\beta} = \beta^* - (\Sigma_t \hat{x}_t x_t')^{-1} \Sigma_t \hat{x}_t \epsilon_{2t}$$
$$\hat{V} = \tilde{\sigma}^2 T(\Sigma_t \hat{x}_t \hat{x}_t')^{-1}$$

where $\hat{x}_t$ is the nonparametric regression of x_t on the information set, β^* is the initial consistent estimate and V achieves the semiparametric efficiency bound (Chamberlain (1987)). Consider the average squared prediction error (ASPE) cross–validation criterion, we have

$$a_{CV} = \operatorname{argmin}_a T^{-1} \Sigma_t (y_t - b^* x_t)^2 \tag{8}$$

b^* is the two–stage–least–squares estimates and $\hat{x}_t^*$ leave–one–out estimates as discussed above, i.e. for nonparametric regression

$$\hat{z}_t = \hat{g}(\Omega_t)/\hat{f}(\Omega_t)\ I(|\hat{f}|>\xi)$$

where

$$\hat{g}(\Omega_t) = (Ta^s)^{-1}\Sigma_{\substack{\tau=1 \\ \tau\neq t}}^{T} x_\tau K_{t\tau}(\Omega)$$

$$\hat{f}(\Omega_t) = (Ta^s)^{-1}\Sigma_{\substack{\tau=1 \\ \tau\neq t}}^{T} K_{t\tau}(\Omega)$$

I(.) is the usual indicator function, s is the number of instruments in Ω_t, and ξ is the user–defined trimming constant. This trimming device is needed to trim out very small values of density estimates which may cause problems both technically and in practice (e.g. Robinson (1988a)).

As we have mentioned, we also use the higher–order kernels of Barlett (1963) which exploit the smoothness of the function and take negative values. Here, $K_{t\tau}$ may not necessarily be a probability density function, as in some classes of K, it is not restricted to non–negative estimates. These higher–order kernels, H_ℓ, play the role of bias–reduction and ensure a certain rate of convergence.

The shape of the criterion function (8) will be of interest. The exact details of how these higher–order kernels are worked out are explained in an appendix. Incidentally, the ASPE CV criterion can also be used for models with risk terms.

Let us look at the parametric results presented in Table 6.6 before we discuss the semiparametric estimates. Four different columns of estimates corresponding to different instruments are reported. Column (1), (2) and (3) are results from using first, second or fifth lagged variables as instruments. Lagged 2 and 5 are used because of the time aggregation problems mentioned in Hall (1988) and Wickens and Molana (1984). According to Hall (1988), the lag 1 variables are not appropriate instruments and efficiency can be improved by noting that the error term is a MA(1) process. Wickens and Molana (1984) have also demonstrated that if the planning period is less than a quarter, serial correlation will again be present in the error term and variables from lag 1 to 4 are inappropriate instruments. Efficiency improvements can again be achieved by using further restrictions. We did not pursue efficiency improvement in this respect.

Notice that the normality assumption is rejected but there is no evidence of serial correlation or inappropriate use of instruments. The estimates of σ is around 0.2 for (1) and (3) but negative for (2). Clearly, the instruments set is too short to capture the features of the data and so we use the same instruments as in Hall (1988), i.e., real interest rates and consumption growth both lag 2 and 3. The results are in the last column. The fit for the model is the best of those reported and the elasticity of substitution is close to zero.

Next, we turn to the semiparametric estimates and see whether they are useful as an aid to diagnosis. The semiparametric results are presented in Table 6.7. We estimated the model using three different orders of kernel. The individual instrumental variable is prescaled by its standard deviation.

Let us look at some interesting observations on the cross–validation function. For model 1, the shape of the function H_2 is shown in Figure 6.2a. It illustrates the not so uncommon functional shape with multiple local minimum points one encounters in cross–validation. The function is reasonably smooth and the global minimum is located at 1.562. When the bandwidth is very small, we are actually evaluating at the tails of the normal distribution which is very close to zero. As the truncated normal kernel is generally used because of computation problems, some peaks reflect more of computation design rather than the underlying cross–validation function. What is typical of the cross–validation function of H_2 is that after attaining its global minimum, it reaches a peak and then starts to decrease very gently. The minimum occurs immediately after the disappearance of the very spiky curves. There is a lower bound at the far right but global minimum point is not difficult to locate.

We now turn to the functions of H_4 and H_6 which are not as well behaved. As one can see from Figure 6.2b and 6.2c that the functions are not as smooth as the case with H . There are also interesting spikes at various points especially when there is undersmoothing. When there is oversmoothing, the function is extremely flat thus creating problems for minimization routine if it is started at points too far right. There is also a tendency for the minimum to occur after the spiky regions but not in the examples that we have provided. The examples have revealed an interesting shape which occurs between 3.4 and 6.7. The curve starts to increase and then it reaches a peak. After that peak, its value drops and then increases again before the usual dying out effect takes place. Again, there is a lower bound at the far right of the curve. Not surprisingly, functions of this form, which may be unpredictable at times due to the fact that the kernel can take negative values, pose problems for minimization routines if used. This suggests that when higher–order kernels are involved, the ASPE criterion can give rise to problems in locating the global minimum, especially when one is minimizing with respect to a vector of parameters. In that case, the problem is more severe because line–minimization routine cannot be used. We now present the results.

The UNIT bandwidth, a_{UNIT}, is equal to $n = T^{-(1/(s+2\ell))}$ when s is the number of instruments and ℓ is the order of kernels. This bandwidth is proportional to the bandwidth which minimizes the MSE of the higher–order density estimates, not forgetting that the instruments are of unit variance.

The trimming constant ξ is set to 1.0e–16. All our semiparametric estimates suggest that the elasticity is not statistically significantly different from zero, regardless of which instrument set is used. In particular, our higher–order estimates are extremely small suggesting that the estimates for model (3) and (4) of 0.006 may be a lower bound for σ. The estimates are very much lower than those obtained by parametric estimates above and elsewhere.

The fact that there is a fixed relationship between risk aversion and intertemporal substitution makes it difficult to interpret the empirical results. The results of near zero elasticity will be very difficult to reconcile with theory if σ is the inverse of the coefficient of risk aversion. In particular, a very low elasticity of intertemporal substitution will imply a very large coefficient of risk aversion and vice versa.

However, recent theories suggest that the coefficient of relative risk aversion and the intertemporal elasticity of substitution need not be related. One can treat the two variables as independent and eliminate the fixed relationship between risk aversion and intertemporal substitution. Two interpretations have been offered in Attanasio and Weber (1989). The most plausible models have been the Selden (1978)'s OCE framework or that of Kreps and Porteus (1978).

Consider the problem of a consumer, with a choice of N assets and a single commodity, maximizing the following utility subject to a budget constraint

$$\Sigma_{j=0}^{T} \rho^j U\{V^{-1}[E_t V(C_{t+j})]\} \tag{9}$$

C_t is consumption in period t and ρ is the discount factor. The usual isoelastic functional form for V and U, i.e.

$$V(C) = I(\gamma \neq 0)C^{1-\gamma}/(1-\gamma) + I(\gamma=0)\log C$$

and $$U(C) = I(\alpha \neq 1)C^{1-\alpha}/(1-\alpha) + I(\alpha=1)\log C,$$

where $\alpha = \sigma^{-1}$ is assumed. Within the Selden's OCE framework, we are in effect converting future consumption to its certainty equivalence, C^*_{t+j}, which is the term in the curly brackets in (9). If U = V, the model is reduced to the familiar case where the coefficient of risk aversion, γ, is the reciprocal of the intertemporal elasticity of substitution, σ. The consumption of future C to its current certainty equivalence is dependent on the curvature of V which is determined by the coefficient of risk aversion. The ordering of the consumption choice is dependent on the curvature of U which depends on the elasticity of intertemporal substitution. Assuming the forcing variables, i.e., log of real return and consumption growth are jointly normally distributed, we can obtain a usual regression model from the Euler equation. All seems well except that the intertemporal optimization problem is not time consistent because it does not take into account that the consumer can revise his plan if news arrive.

An alternative interpretation involves the optimization of the following function subject to a budget constraint

$$U_t = [C_t^{1-\alpha} + \beta E_t(\tilde{U}_{t+1}^{1-\gamma})^{(1-\alpha)/(1-\gamma)}]^{1/(1-\alpha)}$$

where $\tilde{U}$ is stochastic and α, $\gamma \geq 0$. Attanasio and Weber (1989) have shown that the usual linear relationship holds. However, they did not explain how to interpret when the elasticity of substitution is unity. In that case, the Euler equations are not defined for $\alpha = 1$. It seems that one has to rule out the case of unit elasticity, but judging from our results, that does not seem to pose any potential problems.

6.9 CONCLUSION

We have conducted a test of the model by assuming that inflation follows an ARCH process. A new semiparametric technique is introduced to estimate the ARCH model with automatic bandwidth selection. These estimates are then used to form the inflation and interest rates expectations. When this slightly more flexible form of expectations is used in the test of the model, there is no convincing evidence of the presence of expectation variables besides the surprise variable in the model. The only evidence against the model comes from the negative significant estimates for elasticity of substitution. The model we have used is one which allows for a flexible modelling of the second conditional moment for inflation and yet retains the main parametric component implicit in the log–normality assumption.

While we have established some evidence of the importance of allowing more flexibility in obtaining generated regressors, we have also tried to model expectations by relaxing the assumption of normality and linearity. Although similar techniques have been used in Pagan and Ullah (1988), Pagan and Hong (1988) on model with risk terms, it is a first attempt in using cross–validation in determining the optimum bandwidth for the semiparametric models used. It appears that the earlier rejections cannot be attributed entirely to the rigid structure of the expectations formation although our technique here may have yielded more efficient estimates.

Our evidence demonstrates that the construction of instruments as well as the selection of instruments are important in parametric models as they both affect the efficiency of the estimates. While parametric techniques may be robust and give consistent estimates, the use of semiparametric techniques serves the role of diagnostic checking. In particular, our estimates from different information sets (some not reported) have led to the same conclusion. If the parametric and nonparametric techniques lead to identical conclusion, then one's faith in the results is increased. On the other hand, if the results are in contradiction to one another, then a more careful examination is required before a firm conclusion can be drawn. As the semiparametric estimates of the elasticity of intertemporal substitution are generally insignificant, we would like to conclude that the evidence here suggests that there is weak support for a large elasticity for intertemporal substitution. Furthermore, there is no strong evidence as in the previous parametric study that there is a significant relationship between consumption growth and real rates in a simple model.

We should mention that some assumptions of the model using expected utility framework can be relaxed as in Hansen and Singleton (1982, 1984) for distribution and Dunn and Singleton (1986) for separability. We have not directly address these issues here. There is another issue which we have not investigated here. This concerns the relationship between real interest rates and consumption when the model is extended to include other macroeconomic variables. In next chapter, using non–durable plus services from durable consumption data, we seek to answer this question.

Finally, we have presented an alternative method of bandwidth selection by minimizing either the log–likelihood or cross–validation functions. The method is less subjective but its properties need to be studied further in order to claim that we have truly adaptive estimates. This is done in Chapter 8. At present, we could at least be able to label the estimates as "automatic". The word "objective" is best avoided as the choice of criterion function, which minimizes either the log–likelihood function or average squared prediction error, is by no means objective. We have also found, unlike the likelihood function, the ASPE criterion function can be very well–behaved smooth functions in some instances but generally not. The use of higher–order kernels introduces spikes, but it is still not difficult to locate the minimum points. In particular, our optimization routine using quadratic interpolations performs reasonably well. All the programs used in this paper are written in FORTRAN with double precision and the results are all obtained using micro–computer within a reasonably short time.

Table 6.1

Ordinary Least Squares Estimates:
Dependent Variable is $infl_t$

140 observations used for estimation from 51Q3 to 86Q2

OLS	
Constant	–1.000
	(–2.840)
$infl_{t-1}$	–0.290
	(–4.028)
$infl_{t-2}$	–0.209
	(–2.89)
r_{t-1}	0.057
	(0.291)
Δc_{t-1}	0.961
	(4.080)
Δc_{t-2}	0.867
	(4.059)
R–Squared	0.310
R–Bar–Squared	0.285
rss/(T–k)	2.175
Diagnostic test	
F–statistic F(4,134)	12.080
Serial Correlation[1] χ^2_4	4.687
Functional Form[2] χ^2_1	2.401
Normality[3] χ^2_2	2.197
ARCH[4] χ^2_1	4.49
χ^2_2	5.22
χ^2_3	5.55
χ^2_4	6.70

1 LM test of residual serial correlation.
2 Based on the regression of squared residuals on squared fitted values.
3 Jarque and Bera's normality test.
4 Engle's Test of autoregressive conditional heteroscedasticity.

The F–statistic is for the test of the null hypothesis that the coefficients are zero. The t–ratios are in parentheses. Adjusted heteroscedasticity–robust t–ratios are in the bracket below.

Table 6.2

Autoregressive Conditional Heteroscedastic Estimates:
Dependent Variable is $infl_t$

140 observations used for estimation from 51Q3 to 86Q2

	ARCH(1)		SARCH(1)	
		CV	SD	UNIT
bandwidth		1.280	0.540	0.372
Constant	–1.007	–0.969	–0.963	–0.951
	(–3.72)	(–2.89)	(–2.94)	(–2.93)
$infl_{t-1}$	–0.290	–0.288	–0.289	–0.292
	(–4.79)	(–3.96)	(–4.07)	(–4.32)
$infl_{t-2}$	–0.211	–0.208	–0.208	–0.212
	(–3.76)	(–3.03)	(–3.14)	(–3.63)
r_{t-1}	0.054	0.063	0.064	0.063
	(0.36)	(0.34)	(0.35)	(0.35)
Δc_{t-1}	0.957	0.966	0.966	0.964
	(5.02)	(4.16)	(4.25)	(4.30)
Δc_{t-5}	0.899	0.775	0.758	0.741
	(5.89)	(3.86)	(3.88)	(3.89)
rss/(T–k)	2.174	2.184	2.186	2.187

CV refers to the automatic log–likelihood bandwidth. The SD bandwidth is equal to the product of the standard deviation of the data and $T^{-(1/5)}$. The UNIT bandwidth is equal to $T^{-(1/5)}$. The t–ratios are in parentheses.

Table 6.3

Autoregressive Conditional Heteroscedastic Estimates:
Dependent Variable is $infl_t$

140 observations used for estimation from 51Q3 to 86Q2

		SARCH(1,1)	
bandwidth	1.647	0.462	0.372
Constant	–0.968	–0.968	–0.961
	(–2.85)	(–3.09)	(–3.10)
$infl_{t-1}$	–0.288	–0.291	–0.293
	(–3.93)	(–4.52)	(–4.71)
$infl_{t-2}$	–0.208	–0.210	–0.213
	(–3.04)	(–3.68)	(–4.35)
r_{t-1}	0.063	0.061	0.059
	(0.34)	(0.35)	(0.34)
Δc_{t-1}	0.966	0.960	0.960
	(4.14)	(4.46)	(4.50)
Δc_{t-5}	0.772	0.788	0.779
	(3.81)	(4.30)	(4.39)
rss/(T–k)	2.184	2.181	2.182

CV refers to the automatic log–likelihood bandwidth. The SD bandwidth is equal to the product of the standard deviation of the data and $T^{-(1/5)}$. The UNIT bandwidth is equal to $T^{-(1/5)}$. The t–ratios are in parentheses.

Table 6.4

Ordinary Least Squares Estimates:
Dependent Variable is Δc_t

139 observations used for estimation from 51Q4 to 86Q2

	ARCH(1)		SARCH(1)	
		CV	SD	UNIT
Constant	.722	2.508	−5.006	−0.803
	(4.874)	(1.336)	(−1.919)	(−0.967)
	(5.19)	(2.48)	(−1.63)	(−1.29)
$E_{t-1}R_t$	−0.040	−0.042	−0.050	−0.044
	(−2.99)	(−3.12)	(−3.62)	(−3.32)
	(−2.05)	(−3.23)	(−3.86)	(−3.48)
$E_{t-1}infl_t$	−0.048	−0.022	−0.073	−0.016
	(−1.02)	(−0.45)	(−1.44)	(−0.34)
	(−0.89)	(−0.42)	(−1.72)	(−0.33)
$infl_{t-1}-E_{t-1}infl_t$	−0.099	−0.102	−0.102	−0.101
	(−3.33)	(−3.46)	(−3.50)	(−3.47)
	(−2.81)	(−3.05)	(−3.05)	(−2.96)
$Vart_{t-1}(infl_t)$	−0.016	−0.939	3.104	0.832
	(−.27)	(−.97)	(2.18)	(1.80)
	(−.34)	(−1.82)	(1.86)	(2.42)
R–Squared	.127	.137	.162	.153
R–Bar–Squared	.101	.112	.137	.128
rss/(T−k)	.257	.254	.247	.250
Diagnostic test				
F–statistic				
F(4,134)	4.883	5.358	6.493	6.065
Serial Correlation χ^2_4	7.297	5.949	8.285	6.121
Functional Form χ^2_1	5.726	5.024	5.370	3.110
Normality χ^2_2	2.901	2.504	2.273	2.727

CV refers to the automatic log–likelihood bandwidth. The SD bandwidth is equal to the product of the standard deviation of the data and $T^{-(1/5)}$. The UNIT bandwidth is equal to $T^{-(1/5)}$. The t–ratios are in parentheses and heteroscedasticity–consistent t–ratios are given below. See also the footnote of Table 6.1.

Table 6.5

Ordinary Least Squares Estimates:
Dependent Variable is Δc_t

139 observations used for estimation from 51Q4 to 86Q2

		SARCH(1,1)	
Regressor	CV	SD	UNIT
Constant	3.866	0.207	0.269
	(2.06)	(0.26)	(0.45)
	(1.54)	(0.28)	(0.65)
$E_{t-1}R_t$	−0.047	−0.043	−0.043
	(−3.40)	(−3.04)	(−3.08)
	(−3.66)	(−3.20)	(−3.21)
$E_{t-1}\text{infl}_t$	−0.032	−0.031	−0.029
	(−0.69)	(−0.65)	(−0.60)
	(−0.73)	(−0.61)	(−0.56)
$\text{infl}_{t-1}-E_{t-1}\text{infl}_t$	−0.105	−0.100	−0.100
	(−3.57)	(−3.36)	(−3.37)
	(−3.11)	(−2.87)	(−2.87)
$\text{Var}_{t-1}(\text{infl}_t)$	-1.611	0.287	0.258
	(−1.69)	(0.61)	(0.70)
	(−1.25)	(0.67)	(1.03)
R–Squared	.150	.133	.134
R–Bar–Squared	.124	.107	.108
rss/(T–k)	.251	.256	.255
Diagnostic test			
F–statistic F(4,134)	5.919	5.160	5.196
Serial Correlation χ_4^2	5.739	6.473	6.437
Functional Form χ_1^2	1.955	5.821	6.162
Normality χ_2^2	1.562	2.649	2.801

CV refers to the automatic log–likelihood bandwidth. The SD bandwidth is equal to the product of the standard deviation of the data and $T^{-(1/5)}$. The UNIT bandwidth is equal to $T^{-(1/5)}$. The t–ratios are in parentheses and heteroscedasticity–consistent t–ratios are given below. See also the footnote of Table 6.1.

Table 6.6

Instrumental Variable Estimates:
Dependent Variable is Δc_t

140 observations used for estimation from 51Q3 to 86Q2

Regressor	(1)	(2)	(3)	(4)
Constant	0.148	1.092	0.129	0.369
	(0.51)	(1.59)	(0.51)	(2.09)
$E_{t-1}R_t$	0.211	-0.400	0.223	0.068
	(1.16)	(-0.90)	(1.39)	(0.61)
rss/(T–k)	0.414	0.428	0.978	0.290
Sargan's test χ^2_{s-2}	3.725	1.364	0.322	6.750
Serial Correlation χ^2_4	3.112	2.841	3.095	12.064
Normality χ^2_2	81.75	2068.	116.9	1.990

List of instruments: For column (1), we use constant, R_{t-1}, Δc_{t-1}, $infl_{t-1}$ For column (2) we use constant, R_{t-2}, Δc_{t-2}, $infl_{t-2}$. For column (3), we use constant, R_{t-3}, Δc_{t-3}, $infl_{t-3}$. The Sargan (1964)'s test is the test of the hypothesis that the set of s instruments are valid instruments.

Table 6.7

Instrumental Variable Estimates:
Dependent Variable is Δc_t

140 observations used for estimation from 51Q3 to 86Q2

(1)						
	H_2	H_4	H_6	H_2	H_4	H_6
	CV	CV	CV	UNIT	UNIT	UNIT
bandwidth	1.562	0.401	1.786	0.493	0.638	0.719
Constant	0.443	0.436	0.503	0.443	0.438	0.435
	(2.14)	(6.82)	(4.94)	(6.70)	(7.05)	(7.21)
$E_{t-1}R_t$	0.011	0.015	0.029	0.010	0.013	0.016
	(0.08)	(0.56)	(0.48)	(0.36)	(0.53)	(0.66)
rss/(T–k)	0.312	0.313	0.317	0.312	0.312	0.313
(2)						
	CV	CV	CV	UNIT	UNIT	UNIT
Bandwidth	1.745	0.435	0.919	0.493	0.638	0.719
Constant	0.245	0.425	0.409	0.418	0.409	0.409
	(0.61)	(4.47)	(3.21)	(3.74)	(3.39)	(3.546)
$E_{t-1}R_t$	0.137	0.224	0.033	0.026	0.033	0.032
	(0.54)	(0.42)	(0.42)	(0.41)	(0.46)	(0.48)
rss/(T–k)	0.377	0.313	0.315	0.314	0.315	0.315
(3)						
	CV	CV	CV	UNIT	UNIT	UNIT
Bandwidth	0.158	0.299	0.367	0.493	0.638	0.719
Constant	0.449	0.449	0.450	0.458	0.471	0.484
	(7.31)	(6.96)	(6.90)	(5.48)	(5.17)	(·5.25)
$E_{t-1}R_t$	0.006	0.007	0.006	0.001	–0.006	–0.015
	(0.27)	(0.25)	(0.23)	(0.23)	(–0.13)	(–0.29)
rss/(T–k)	0.312	0.312	0.312	0.312	0.313	0.314
(4)						
	CV	CV	CV	UNIT	UNIT	UNIT
Bandwidth	0.129	0.223	0.139	0.539	0.662	0.734
Constant	0.451	0.450	0.451	0.438	0.432	0.431
	(7.43)	(7.32)	(7.44)	(5.51)	(4.71)	(4.67)
$E_{t-1}R_t$	0.006	0.006	0.006	0.014	0.017	0.019
	(0.24)	(0.24)	(0.24)	(0.34)	(0.34)	(0.37)
rss/(T–k)	0.312	0.312	0.312	0.312	0.313	0.313

List of instruments: The constant is excluded in the nonparametric estimation of conditional expectation. For (1), we use R_{t-1},Δc_{t-1}, $infl_{t-1}$. For (2), we use R_{t-2}, Δc_{t-2} $infl_{t-2}$. For (3), we use R_{t-3}, Δc_{t-3}, $infl_{t-3}$. H_ℓ corresponds to the order of higher kernel used in the estimation. CV refers to the cross–validated bandwidth. The UNIT bandwidth is equal to $T^{-(1/(s+2\ell))}$. The t–ratios are in parentheses.

Figure 6.1a Plot of Criterion Function Values Against Bandwidth:

for SARCH(1,1)

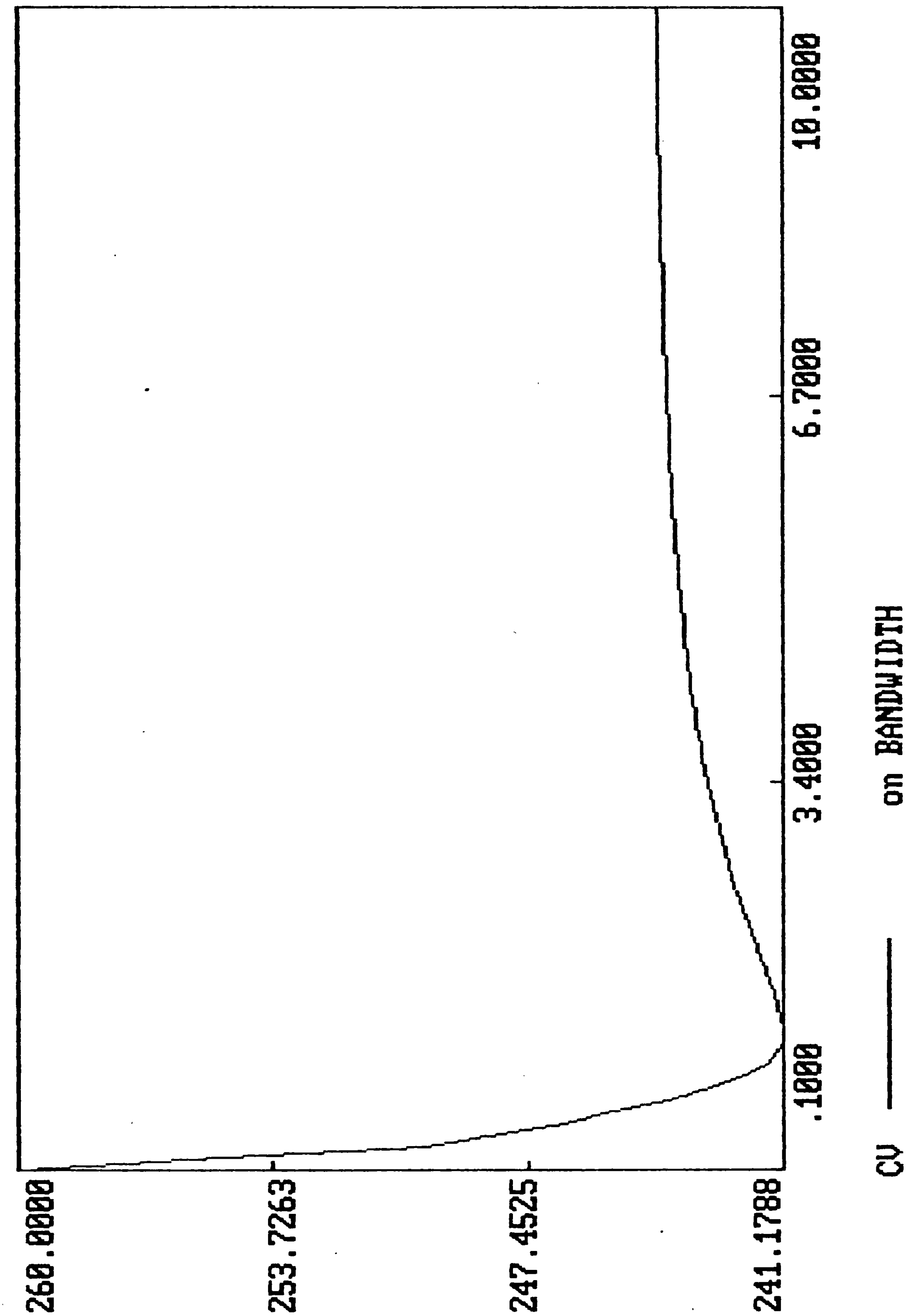


Figure 6.1b Plot of Criterion Function Values Against Bandwidth:

for SARCH(1,1)

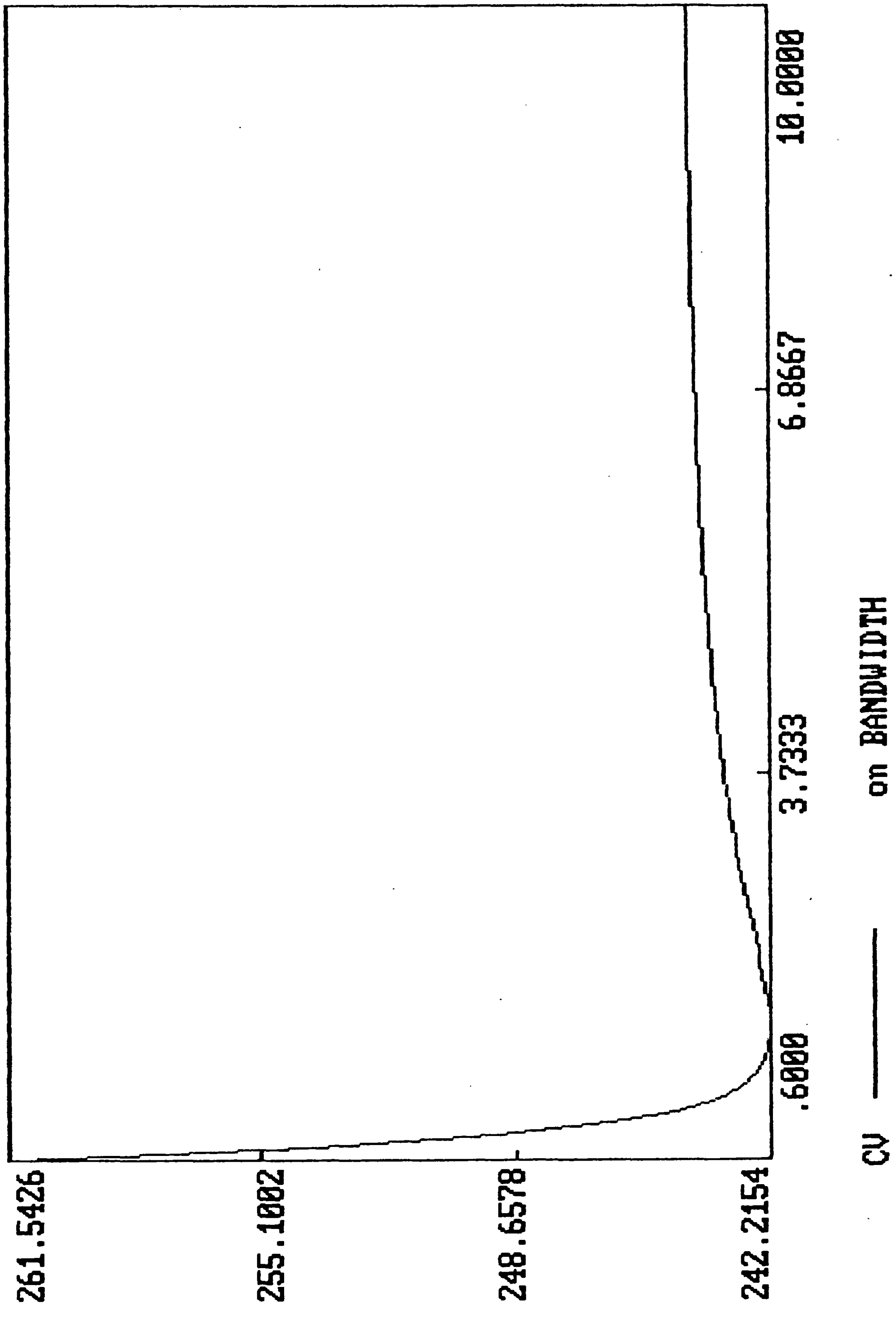


Figure 6.2a Plot of H_2 Criterion Function Values Against Bandwidth: for Model 1

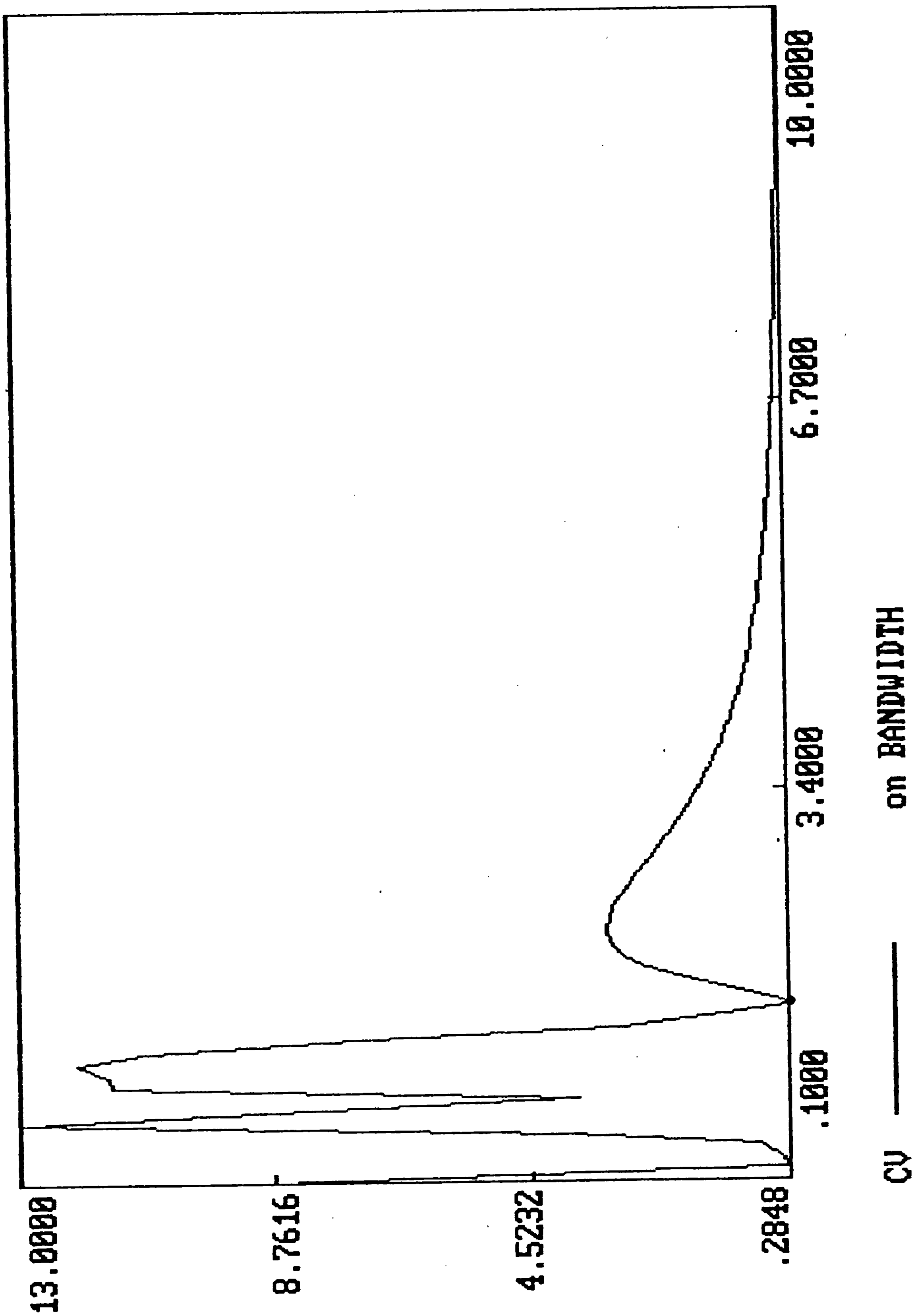


Figure 6.2b Plot of H_4 Criterion Function Values Against Bandwidth: for Model 1

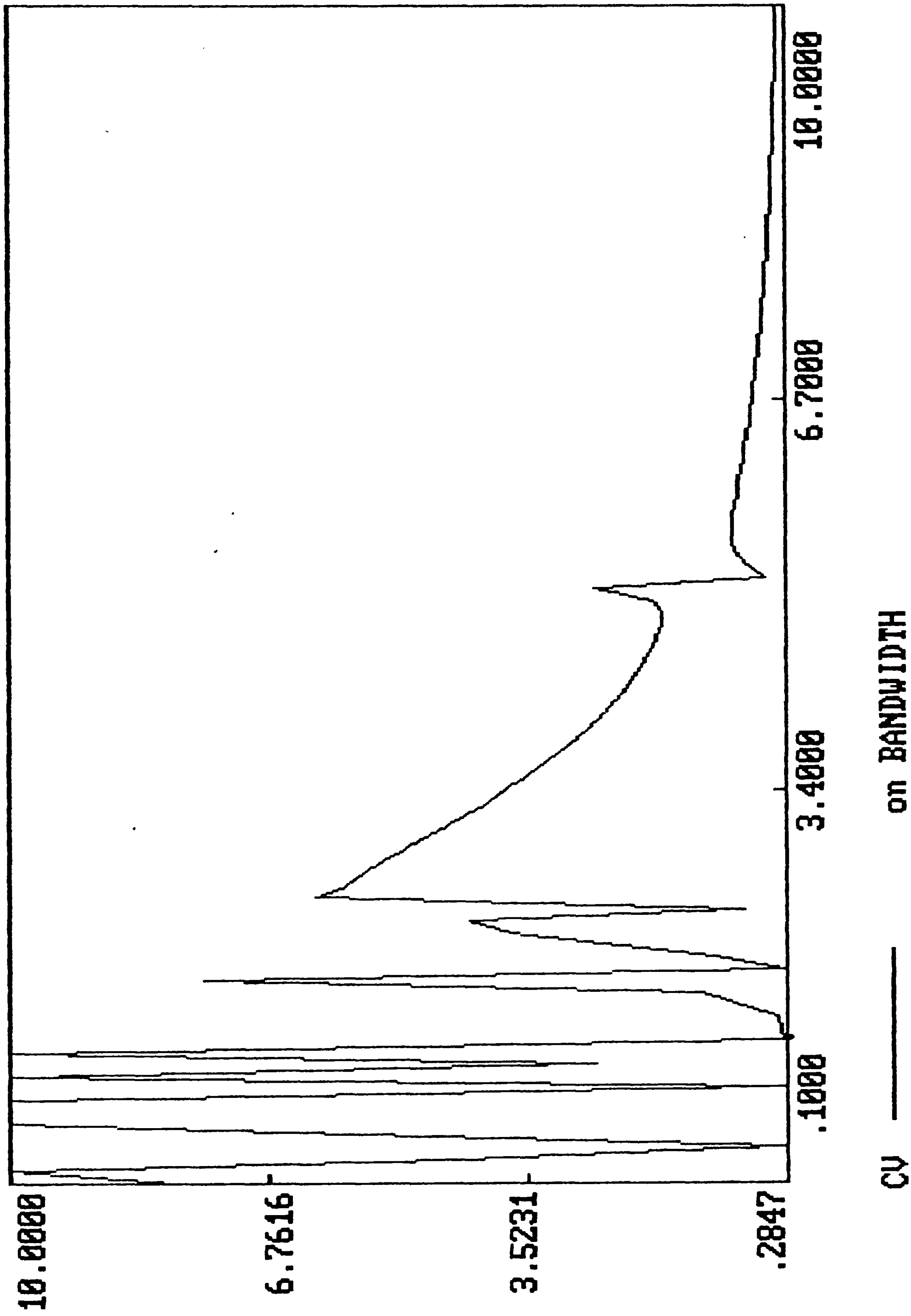
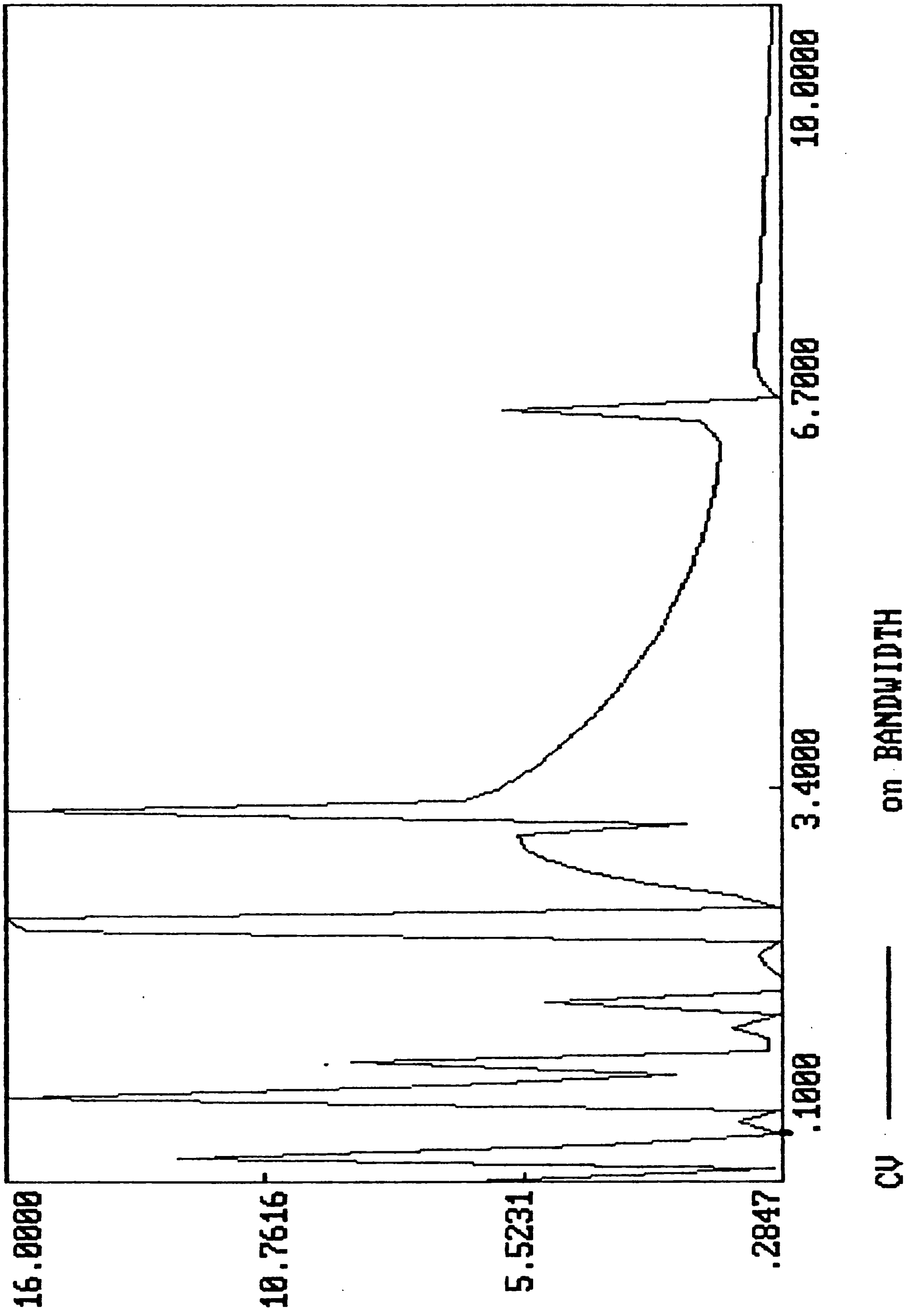


Figure 6.2c Plot of H_6 Criterion Function Values Against Bandwidth:
for Model 1



CHAPTER 7

FURTHER SEMIPARAMETRIC ANALYSIS OF "SURPRISE" CONSUMPTION FUNCTION

7.1 INTRODUCTION

The reliability of the results obtained from purely parametric models depends on the plausibility of the assumptions and this in turn depends on the quality and extent of information one can extract from economic theory. If the amount of economic information is not sufficient to assist us in forming certain functional forms, then we cannot treat the parametric model itself seriously.

In some rational expectations models, e.g., the surprise consumption function of Deaton (1986) and Blinder and Deaton (1985), the parametric specification of the main regression function is motivated from economic theory. In these models, we would place more faith on some specifications than others. In particular, the formation of rational expectations is one issue that one can debate on.

On the other hand, the models of Bean (1986) and Hansen and Singleton (1983) are derived within an intertemporal optimizing framework of an individual agent incorporating rational expectations and the log–normality assumption. In this case, the formation of expectation for some variables is well defined. However, many of these maintained hypothesis such as the linearity assumption in vector autoregression can be checked, and indeed there are many useful specification tests. But, if the hypothesis is rejected indicating that the functional form is misspecified, then it may well be a nonlinear autoregression. When nonlinearity is thought to exist, the question of which nonlinear function to employ for estimation is not generally easy to answer. To safeguard the vast possibilities in the specification of rational expectations, a semiparametric treatment of the models seems desirable in applied work.

In this chapter, we outline a procedure within the class of "generated regressor" models. The method requires nonparametric estimation of the unknown conditional expectations. These nonparametric generated regressors are used for estimating the coefficients for the anticipated and unanticipated terms in the surprise model respectively. The main contribution of this paper is the application of these semiparametric techniques to the surprise consumption function. However, the large number of lagged variables in the information set makes direct application of the nonparametric in finite sample undesirable. We have therefore resorted to applying nonlinear principal components analysis to the lagged variables, some of which are nonstationary, to reduce the dimension of the variables used for nonparametric estimations. We have also discovered that the resulting principal components are in fact stationary.

Two tests, namely, the Hausman (1978)'s type specification test and the Robinson (1988c)'s semiparametric test are of interest. In fact, many of the ideas developed here are motivated from the paper by Robinson (1988b).

7.2 BEAN'S MODEL

The economic model of Bean (1986) is presented below as a bench mark for discussions. As we have shown in Chapter 5, under rational expectations, the life cycle–permanent income model of consumption should be random–walk plus drift. Bean (1986)'s model is one of the many extensions with stochastic interest rates and allows for the possibility of government expenditure being substitutes for private expenditure. As in Chapter 5, we assume that the individual is forward looking. The individual will have to decide how much to consume now, which is certain, or save more for a possibly longer future life. The utility function is Cobb–Douglas. It is strictly concave and not (multiplicative) separable in consumption, leisure and government expenditure. His problem is to maximize the expected utility subject to his budget constraint, i.e.

$$\mathrm{Max}_{C_\tau} \; [\, \Sigma_{\tau=t}^{\infty} \, \delta^{\tau-t} U_\tau^{1-\gamma}/\gamma], \qquad \gamma \geq 0,$$

with

$$U_t = C_t^{\,1-\lambda-\theta} L_t^{\,\lambda} G_t^{\,\theta}, \qquad 0 \leq \lambda,\ \theta \leq 1,$$

subjected to the budget constraint $W_{t+i} = W_{t+i-1}R_{t+i}+Y_{t+i}-C_{t+i}$, where

U_{t+i} = Utility function in period t+i

C_t = Consumption in period t

L_t = leisure in period t

G_t = Government expenditure in period t

W_t = Assets held at end of period t

R_t = Rate of return between periods t–1 and t

δ = the discount factor

γ = the coefficient of relative risk aversion

λ, θ are parameters

$E_t(.)$ = the mathematical expectation conditioned on information set Ω_t

Ω_t = information set available to agent at time t

The first order conditions consist of stochastic as well as non–stochastic Euler equations. Since we are mainly concerned with consumption, we are only interested in the following equation between period t and t+1 (see Bean (1986) regarding the Euler equation for leisure):

$$E_t\,[(C_{t+1}/C_t)^{(1-\gamma)(1-\lambda-\theta)-1}(L_{t+1}/L_t)^{(1-\gamma)\lambda}(G_{t+1}/G_t)^{(1-\gamma)\theta}(1+r_{t+1})] = \delta^{-1}$$

If we let M_{t+1} denote the expression in the squared brackets, then we have

$$E_t[M_{t+1}] = \delta^{-1}$$

As mentioned in the Chapter 5, one can test the hypothesis by working with this Euler equation directly. We may also work with a system of Euler equations as in Mankiw, Rotemberg and Summers (1985). There has been considerable interest in the estimation by generalized method of moments and testing of the overidentifying restrictions using Hansen (1982)'s J test in the literature. For a more recent semi–nonparametric approach, see Gallant and Tauchen (1989).

But, the log–normality assumption will simplify the conditional expectation to a linear expectation model. This is a strong assumption but it has been argued that empirically the macroeconomic variables behave as if they are generated by a covariance stationary Gaussian process. Introduce the notation that $r_t = \log R_t$, $x_t = \ln X_t - \ln X_{t-1}$ and assuming that $Y_t = \{\Delta c_t, \Delta \ell_t, \Delta g_t, r_t\}$ is jointly normally distributed. Then

$$\ln M_{t+1} \sim N\left(E_t[\ln M_{t+1}], \sigma^2\right).$$

With the help of this normality assumption, and expressing leisure in terms of observable average per capita hours (H), we have a model of the form

$$\Delta c_t = a_0 + a_1 E_t r_t + a_2 E_t \Delta H_t + a_3 E_t \Delta g_t + \epsilon_t \qquad (1)$$

$$a_0 \lesseqgtr 0;$$

$$a_1 > 0;$$

$$a_2 \lesseqgtr 0 \text{ if } \gamma \lesseqgtr 0;$$

$$a_3 \lesseqgtr 0 \text{ if } \gamma \gtreqless 0.$$

where $\Delta H_t \approx -\Delta \ell_t / \overline{L}$, $\overline{L}$ is the total leisure endowment. Since we are only concerned with the change of consumption along the optimal path, equation (1) alone is of interest. Although (1) has been known as the "surprise" consumption function, it is not a consumption function. It only describes the optimal consumption plan. The name "surprise" consumption function arises because the error term represents "surprises" or "news" arriving between period t–1 and t. "Surprise" is defined as the discrepancy between the expected and the realized values of the particular variable or the innovations in the variable.

To proceed with estimation, one normally assumes that the true expectation formation is linear, e.g., Bean assumed

$$x_t = k_0 + a(L)'x_{t-1} + k_1 t + k_2 t^2 + \xi_t,$$

where $a(L)$ is a vector of finite order polynomials in the lag operator and k_i's are the parameters. The variables t and t^2 are included here because some of the variables in x are not stationary. So either the substitution method (SM) or errors–in–variables method (EVM) should be used. But since each equation is just–identified, two stage least squares (2SLS) estimates for (1) are asymptotically efficient and there is no need to estimate all the equations jointly.

Our interest lies in finding what macroeconomic variables have explanatory power in a surprise consumption function. Following Blinder and Deaton (1985) in the statistical study of the consumption function, we include expectations of macroeconomic variables to see if they have any explanatory power. Following that, $y_t = \log Y_{t-1}$ and $c_t = \log C_{t-1}$ can be included to test for sensitivity of consumption to income as well as habit persistence (see Deaton (1986)).

U.S. quarterly data are used and the descriptions of variables are clearly spelt out in Bean (1986). A brief description is given in the appendix. Throughout the text, the Σ sign shall refer to the summation from observation 1 to N unless otherwise stated.

Semiparametric results on three surprise models will be presented. These models are under the class of linear rational expectations models of the form:

$$y_t = b_0 + b_1'z_t^* + \beta' E_t x_t + \pi'(x_t - E_t x_t) + \epsilon_t \qquad (2)$$

$$t=1,2,..N$$

$$x_t = E_t x_t + \xi_t$$
$$E_t x_t = E[x_t | z_t]$$
$$E[\epsilon_t | x_t, z_t] = 0$$

where
- y_t is a scalar dependent variable
- x_t is px1 vector of exogenous variables
- z_t is qx1 vector of stationary variables
- z_t^* is sx1 vector and z_t^* is a subset of Ω_t
- b_0 is a scalar unknown parameter

b_1 is a sx1 unknown parameters

β, π are px1 unknown parameters

Notice that our z_t is not necessary the information set Ω_t here, but a linear or nonlinear combination of the weakly exogenous variables in Ω_t. The reason why we have taken z_t to be a linear or nonlinear combination is explained in the next section. z_t^* is a subset of the Ω_t and thus z_t^* may, for example, be lagged consumption or lagged income. The following notations are used in the following sections: $\theta = (b_0, b_1', \beta', \pi')'$; $X_t = (1, z_t^{*\prime}, x_t')'$.

7.3 RESULTS: USING STANDARD TWO STAGE METHODS

Three basic surprise models are of interest. Model 1 has two anticipated variables: anticipated real interest rates and income. Model 2 replaces anticipated income by anticipated inflation. Model 3 is the general model with six anticipated variables. When lagged consumption and inçome are included, each model is subscripted with an a, i.e., 1a, 2a and 3a respectively.

All the results that we have presented in the first three tables and discussed this section employ the usual parametric 2 stage method. In the first stage, the expectations and surprises are obtained by regression on a constant, three lags of consumption expenditure (c), nominal interest rates (r^*), average hours of work per week per capita (H), government expenditure (g), income (y), inflation (I), stock prices (S), time trend (t) and time squared (t^2). In the second stage, we have applied the ordinary least squares to the main equation containing both the anticipated and the unanticipated terms treating them as given. However, the standard errors are different for the anticipated and the unanticipated terms. The standard errors for the unanticipated terms are just the standard errors from ols and the standard errors for others are corrected using Newey (1984)'s method. In other words, they are the standard errors from applying 2SLS on the remaining variables with the expected terms replaced by those obtained from the first stage. Since each of these equations is exactly identified, our procedure is as efficient as 2SLS and 3SLS. The results reported in this section can be easily reproduced on the menu–driven computer package DATAFIT.

The diagnostic test statistics are reported at the bottom each table and all of them are Lagrange Multiplier (LM) tests. Some of them have been used in previous chapters and we have only very briefly described them. All these test statistics are standard output from the DATAFIT package. The serial correlation test is the standard LM test of serial correlation up to order 4 (see Harvey (1981), Godfrey (1978a, 1978b) and is asymptotically distributed as χ^2_4 under the null hypothesis. The RESET LM test of functional form (with quadratic terms) and is that of Ramsey (1969)'s, the statistic is asymptotically distributed as χ^2_1. The Jarque–Bera's (Jarque and Bera (1980)) LM test of the normality of regression residuals is asymptotically distributed as χ^2_2. The LM test of heteroscedasticity is based on the auxiliary regression of residuals on the squared of the forecast with an intercept and the statistic is asymptotically χ^2_1 under the null hypothesis. We have also reported the Chow test, which is a test of the stability of the regression coefficients where appropriate and the LM version is distributed as χ^2_k, where k is the number of periods used in the predictive test. The Sargan misspecification test is proposed by Sargan (1964) to test the null hypothesis that the equation with only anticipated terms is correctly specified and that the p_1 instrumental variables z are valid instruments. This general misspecification test is distributed as $\chi^2_{(p_1-p)}$. All the formulae used for these tests are described in detail in Appendix B of the manual of DATAFIT and we have decided not to give any further details.

Our findings are similar to those in the previous studies. In Table 7.1, we have reported the results for Model 1 and 2 in column 1 and 3 respectively. In column 2 and 4, we have also included additional surprise terms. Based on the estimates, there are two observations which are inconsistent with the theory. First of all, the sign of the expected interest rates variable is negative in both models and statistically insignificant. Secondly, consumption is too sensitive to expected income, as well as inflation. The diagnostic statistics in column 1 indicate misspecification when income is included, a high value for the normality test is recorded. The results confirm Flavin's findings and that of Koskela and Viren (1987).

We have reported the results for Model 3 and 3a in Table 7.2a and 7.2b. Model 3a is familiar as there have been various studies of consumption function using similar form, e.g. Blinder and Deaton (1985) and Deaton (1986). The models can be taken as a surprise consumption function though the wealth variable is not included. The wealth variable can be constructed as in Blinder and Deaton (1986). The inclusion of the lagged variables in Model 3a is a limited attempt to test for over sensitivity to income and habit persistence respectively.

In Table 7.2a, column 1 and 3 are results for the period 1949Q1 to 1982Q4 while column 2 and 3 are for 1949Q4 to 1979Q4. The sign of the interest rates expectation is now consistent with the theory and statistically significant. Unfortunately, all the additional variables to the models are highly significant. There are clear indications of over–sensitivity to current income. The conclusion is unchanged if we use the same period for estimation as Bean (1949Q1 to 1979Q4).

Let us turn to the diagnostic test statistics reported in Table 7.2b. One can see that the test statistics have not detected any misspecifications of the consumption function, with the exception of the Sargan's test. We have conducted two other hypothesis tests as in Blinder and Deaton (1985), i.e., the "anticipation only" and the "surprise only" hypothesis. The first one is a purely statistical test corresponding to testing the significance of all the surprise variables while the second one corresponds to testing the economic hypothesis that anticipated variables has no explanatory power. From the results reported in Table 7.2b, the conclusion is that we can neither omit the anticipated nor the unanticipated terms judging from the test statistics.

Clearly, the empirical evidence from the parametric models are not consistent with the theory. While many explanations have been offered in the previous chapter, we are interested in modelling a surprise consumption function allowing for expectation of an unknown form. In particular, we are interested in finding the magnitude of the surprise terms.

We argue that the use of the vector autoregression on inflation and stock prices may be inappropriate. In fact, it is generally difficult to infer the expectation of inflation using vector autoregression using 2 or 3 lags as shown by the results in Table 7.3. In fact, some believe that long lags are usually needed to "soak up" some model misspecifications in linear time series.

The results in Table 7.3 demonstrate that this may indeed be the case. We have allowed for 3 lags of each variable in (I) and 2 Lags in (II) to see how well each of the individual variable captures the features of the data. ΔY, ΔI and ΔS have enormous values for normality test in both cases indicating that the residuals may have departed from the normality assumption. From the results in Table 7.1, 7.2 and 7.3, it seems fair to conclude that the expectations formation is not consistent with the data.

In view of our observations, we repeat the similar exercise using the semiparametric techniques. The only difference between the parametric and semiparametric models is intended to be the formation of rational expectations. But there are minor differences. While we have included time trend and time squared in the information set for the parametric models, we have excluded these from the semiparametric models. Furthermore, we have taken logarithmic transformation for H and that we have included quadratic functions for S and I before applying PCA. It remains to see if our semiparametric estimates are very different from these parametric estimates.

7.4 NONPARAMETRIC REGRESSION AND REDUCTION OF DIMENSIONALITY

We have to estimate the conditional expectation as well as density in our application. As in Chapter 6, N–W estimator with trimming and higher–order kernel is used. The asymptotic distribution of many semiparametric estimators does not depend on dimensionality of z. This is indeed the case for the estimators that we are about to introduce and has been shown to be so in the i.i.d. case. However, the application in this chapter is complicated by the fact that we have a limited number of observations and a large number of variables in the information set. It is reasonable to believe that in finite sample, high dimensionality may have deleterious effect. We may encounter a problem in density estimation caused by the high dimension of z since we only have a finite sample. This problem has been discussed in Chapter 2 and known as the "curse" of high dimensionality in the nonparametric literature. The reason being that some regions of the high density may not be filled with observations thus giving rise to what is known as the 'empty space phenomena'.

Silverman (1986) has provided some hints on the number of observations required for density estimation in order to ensure that the relative mean squared error at zero is less than 0.1 using a normal kernel when the true density is a standard multivariate normal density. His results suggest that for a sample size of around 130 observations (say quarterly data for around 32 years), the maximum desired dimension for q is between three and four for density estimation. One may argue that since density estimation may be used only for the nuisance functions in some of the economic applications and we are not interested in the density estimates itself, the problem of 'empty space' is not as severe as in purely nonparametric models. Robinson (1988a)'s simulation results provided limited support of this view in finite sample using higher–order kernels. But this requires further investigation.

In empirical work, we seldom work with less than 4 dimensions, especially for the estimation of rational expectations condition on a large information set. In fact, our information set consists of no less than 7 variables of 3 lags each.

Therefore, in the absence of firm theoretical evidence supporting the use of q larger than those suggested by nonparametric density literature, it may seem desirable to reduce the dimension q. This can be done by some linear multivariate methods such as principal components (PC) or projection pursuit (PP) (see e.g. Huber (1985) for some discussion on the PP version of PC); or nonlinear methods such as generalized principal components (Gnanadesikan (1977)). There is also an enormous amount of literature on PC including techniques for time series and nonlinear projections (see Jolliffi (1986)).

We have decided to use the linear and generalized principal component (GPC) to reduce the dimension. The reasons are that GPC is easy to understand and has practical advantage. Gnanadesikan's GPC method is an extension of the method of linear PC to nonlinear projection. Consider the simple case of $q = 2$, so $z = (z_1, z_2)'$, we can add any functions $g(z_1, z_2)$ to z and then apply the usual method of finding the PCs which give the maximum variance. Gnanadesikan has concentrated on the quadratic functions, $z^* = (z_1, z_2, z_1^2, z_2^2, z_1z_2)$. Thus his procedures has amounted to finding quadratic functions of z which maximize variance rather than the usual linear function of z. This method requires the application of the regular PCs analysis on q original variables with $q^* = q + q(q-1)/2$ derived variables which can be expensive (for 10 variables, one needs to apply regular PCA to 65 variables). One way of getting over this problem is to apply linear PCA to the original variables and then polynomial PCA on the first few PCs with the highest variance. Alternatively, one can augment the method to only searching for quadratic, cubic or indeed any other polynomial functions over a limited number of variables. The linear PCs will be a special case of the nonlinear PCs and each eigenvector provides a nonlinear coordinate of the original space.

It is well known that PCA based on covariance matrix suffers from the sensitivity of the PCs to the unit of measurements of z. Therefore we define the principal components as principal components from minimizing the correlation matrix, i.e.,

$$W = A'Z^{**}$$

where W is Nxk matrix and the kth column is the kth PC with the kth largest variance. A is an Nxk matrix whose kth column is the kth eigenvector of the correlation matrix of Z^{**}. Z^{**} is a Nxq matrix of standardized variables $(z_1^{**},\ldots,z_p^{**})$ each possesses the property of unit variance, i.e., $z_i^{**} = z_i/\sigma(z_i)$. Singular Value Decomposition (SVD) is used to obtain the eigenvalues and the eigenvectors.

We have implicitly assumed that W admits a pdf f and have to ensure that the W_i's are stationary after such a preliminary step of dimensionality reduction by PCA.

In the application that follows, the original information set includes 22 variables, namely, a constant (strictly speaking, constant should not be included but it makes no difference in practice since there are no variations in a vector of ones), three lags each of the consumer expenditure (c), nominal interest rates (r^*), average per capita hours worked (H), Government expenditure (g), income (y), inflation (I) and stock prices (S). The results of PCA on these variables are reported in Table 7.4. All the variables are in logarithm except H, I and S. Using linear PCA, the first PC is able to capture 98.07% of the variations. However, we feel that linear projection may be inadequate in summarizing the information and that a nonlinear projection may be more appropriate.

The first two PCs from the linear PCA are used to form quadratic functions before another PCA are applied to the five resulting variables. To capture more than 90% of the variations, we need at least 4 PCs.

A number of other transformations and combinations were tried. We have tried using lag values of Δc_t, Δr_t^*, Δy_t, ΔH_t, Δg_t, ΔI_t and ΔS_t as our information set and have applied PCA to reduce the dimension. However, we were unable to reduce q to less than 6 in order to capture at least 90% of the variations. We have also reported the case where we included 3 lags each of 5 variables with logarithmic transformation and quadratic functions of S and I in our PCA. There are therefore 15 variables from the variables in log and 27 variables from the quadratic functions of 3 lags each of S and I. Together with the constant, we have a total of 43 variables. In this case, we are able to capture 95.94% of the variations with just the first 2 PCs. An examination of the correlogram of the first two PCs (not reported) up to 20 lags revealed that there were no signs of nonstationarity. We have therefore decided to use these two PCs for all the semiparametric analysis that follows.

The use of a single smoothing parameter implies that PCs have to be individually scaled by its standard deviation. The PCs are prescaled to avoid extreme differences and spread in the various coordinate directions. So, after the scaling, we have the two W_i's which are orthogonal to each other and also possess the property of unit variance.

From the practical point of view, our procedure for nonparametric estimation has introduced an element of parametric modelling. By making use of this additional information, we have introduced a more general version of conditional expectation estimation than that of a purely parametric one. Our procedure is closely related to the class of additive models. The following procedures can also be viewed as a nonparametric approach to the "generated regressor" problem.

7.5 ESTIMATION OF ANTICIPATED TERMS

The anticipated terms in (2) can be consistently estimated by instrumental variable. It is well known that the consistency of the instrumental variable estimations is not affected by the specification of the instruments, but efficiency may be affected. Our main concern here is efficiency improvement. We employ an augmented asymptotically efficient estimator proposed by Newey (1987) using the conditional moment restrictions of the model for the expectation terms. Consider (2) again, if we replace the expectation terms by their realized values, we may yield the errors–in–variables model of the form :

$$y_t = b_0 + b_1{}'z_t^* + \beta'x_t + (\epsilon_t + (\pi - \beta)'\,\xi_t)$$
$$= \theta'X_t + \nu_t \qquad t = 1, 2, \ldots, N$$

Let us assume that the following conditional moment restriction is satisfied:

$$E[h(X_t,\theta)\,|\,z_t] = 0 \tag{3}$$

where $h(X_t,\theta) = y_t - \theta'X_t$. The conditional moment restriction (3) implies that the disturbance is uncorrelated with the instruments z_t, i.e.,

$$E[h(X_t,\theta)z_t] = 0.$$

This in turn implies that any function of the instruments, i.e. $T(z_t)$, has to satisfy the orthogonality condition

$$E[T(z_t)h(X_t,\theta)] = 0.$$

In this case, $\hat{\theta}$ is obtained by minimizing the objective function, i.e.,:

$$\hat{\theta} = \operatorname{argmin}_{\theta\in\otimes} S(\theta)'P_N S(\theta)$$

where P_N is a positive definite matrix, $\otimes$ is some set of feasible values for θ and $S(\theta) = N^{-1}[\Sigma_t T(z_t)h(X_t,\theta)]$. The feasible approach is to let $P_N = I_N$ and

$$\hat{X} = \hat{T}(z_t) = \hat{D}(z_t)'[\hat{v}(z_t)]^{-1}$$

with

$$D(z_t) \equiv E[\partial h(X_t,\theta)/\partial\theta\,|\,z_t] \tag{4}$$
$$v(z_t) \equiv E[h(X_t,\theta)^2\,|\,z_t] \tag{5}$$

Notice that (4) and (5) are conditional expectations and can be estimated by nonparametric regression. In this case the optimal instruments $\hat{X}$ are obtained via the nonparametric kernel regression estimates. Under the conditional moment restrictions and the assumption that the variables are i.i.d., Chamberlain (1987) has shown that the infeasible optimal instrument $T^*(z_t) \equiv C\ D(z_t)'[v(z_t)]^{-1}$, where C is a nonsingular constant matrix, attains the semiparametric asymptotic variance bound

$$\Lambda^* = [D(z_t)'D(z_t)/v(z_t)]^{-1}.$$

Furthermore, it is shown in proposition 1 of Chamberlain (1987, pp.324) that the asymptotic covariance matrix of $\hat{\theta}$, $\hat{\Lambda}$, is no smaller than Λ^*. However, Newey (1987) noted that if $T(z_t) \equiv D(z_t)'[v(z_t)]^{-1}$, then $\Lambda^* = \hat{\Lambda}$, and thus $\hat{\theta}$ attains the semiparametric efficiency bound.

One may choose to work with the linearized version of the feasible IV estimator which only requires a one step update from an initial 2SLS estimator, θ^{2SLS}. For equation (2) with homoscedasticity, $v(z_t)$ is independent of z_t and equal to a constant σ^2. We have

$$\hat{\theta} = \theta^{2SLS} - [\Sigma_t \hat{X}_t X_t']^{-1}\ \Sigma_t \hat{X}_t h(X_t, \theta^{2SLS})$$

We can also replace θ^{2SLS} by any other consistent estimates. The asymptotic covariance matrix can be estimated as

$$\hat{\Lambda} = \hat{\sigma}^2[N^{-1}\Sigma_t \hat{X}_t \hat{X}_t']^{-1}.$$

where

$$\hat{\sigma}^2 = N^{-1}\Sigma_t(y_t - \hat{\theta}'X_t)^2$$

Notice that we have not imposed any functional form on the conditional expectations. We have therefore relaxed the linearity assumption of the expectation formation of X_t using the nonparametric regression of X_t on z_t. It should also be obvious that the conditional expectation and therefore the optimal instrument of the constant and the z^* are the variables themselves. Newey (1987) proves the efficiency of the estimator $\hat{\theta}$ using the method of nearest neighbours for the estimation of conditional expectations for the case with i.i.d. observations. His procedures involve a preliminary step of "trend removal" suggested by Stone (1977), to improve the finite sample performance of the nonparametric estimates. Newey's estimator is as efficient as the class of instrumental variable estimators under homoscedasticity with the restriction that the disturbance terms have mean zero conditional on exogenous variables.

Our estimator is "augmented" because it differs from Newey's original proposal on two counts: we use kernel regression instead of nearest neighbours; higher–order kernels are also used to deal with high dimension common in time series application. We have resorted to using the higher–order kernel estimator with trimming device for various reasons. In particular, the properties of kernel estimator for time series are better known and also because of its facilities to deal with multivariate observations in the technical sense.

7.6 ESTIMATION OF THE SURPRISE TERMS

We focus our attention on the estimation of the surprise terms in this section. The estimator proposed in Robinson (1988a) for the estimation of the surprise parameters are used in our empirical work. Consider equation (2) again, rearranging the model, we get

$$y_t = \pi' x_t + [b_0 + b_1' z_t^* + (\beta - \pi)' E_t x_t] + \epsilon_t$$

$$y_t = \pi' x_t + g(z_t; b^*) + \epsilon_t \tag{6}$$

where g is an unknown function of z_t and b^* are some unknown parameters. π is the parameter of interest. For our purpose, we have taken expectation as an unknown function of z_t. Thus, we have relaxed the assumption made in most parametric models that the expectation formation is linear. The proposed augmented Robinson root–N consistent estimator (R estimator) is

$$\hat{\pi} = [\Sigma_t \hat{x}_t^* \hat{x}_t^{*\prime} I(|f_t| \geq \zeta)]^{-1} \Sigma_t \hat{x}_t^* \hat{y}_t^{*\prime} I(|f_t| \geq \zeta)$$

where

$$\hat{x}_t^* = x_t - \hat{x}_t$$

$$\hat{y}_t^* = y_t - \hat{y}_t$$

f_t = the nonparametric kernel estimates of the density $f(z_t)$

ζ = a trimming constant chosen by the practitioner

The $\hat{x}_t$ and $\hat{y}_t$ are kernel estimates of the respective conditional expectations on z_t. Consider the following

$$\hat{y}_t^* = \pi'\hat{x}_t^* + [(g_t - \hat{g}_t) + (\epsilon_t - \hat{\epsilon}_t)]$$
$$= \pi'\hat{x}_t^* + e_t$$

The terms e_t can be considered as the residuals. Therefore, the R estimator has the interpretation of a no–intercept ols estimator. The main idea is to ensure that e_t is orthogonal to $\hat{x}_t^*$. However, in order to achieve $\sqrt{N}$ consistency for π, we have to ensure that plim $N^{-1/2}\Sigma_t\hat{x}_t^* \hat{g}_t^*{}'I(|f_t| \geq \zeta) = 0$. The presence of the biased term $(g_t-\hat{g}_t)$ poses a technical problem. However, we know from Chapter 2 that the bias of $\hat{g}_t$ can be reduced using higher–order kernels provided g possesses some smoothness properties.

One can resort to the bias–reduction device of Barlett (1963). Although we have used GPC to reduce the dimensionality q to a manageable number for kernel estimation, the use of higher–order kernels may still have some advantage in finite sample. It should be mentioned that the theorem is independent of the kernel k and suggests that the higher the order of kernel, the wider the band of admissible a and ζ sequence. Although the simulation results of Robinson (1988a) suggest that there may be improvement when using higher order kernels in finite sample, we should be aware of the side effects of increased variance from using too large an order. It is clear that the R estimator is a linear estimator and it can easily be shown to have the MM and GLS interpretations as well.

The asymptotic covariance matrix can be estimated as

$$\hat{\Gamma} = \hat{\sigma}^2 [N^{-1}\Sigma_t\hat{x}_t^* \hat{x}_t^*{}'I(|f_t| \geq \zeta)]^{-1}$$

Where

$$\hat{\sigma}^2 = [N^{-1}\Sigma_t(\hat{y}_t^* - \hat{\pi}'\hat{x}_t^*)^2 I(|f_t| \geq \zeta)]^{-1}$$

The main advantage of the R estimator is that although it is $\sqrt{N}$ consistent, it is in fact as efficient as OLS computed under the correct assumption that g is in fact constant. Thus, if the expectation formulation is linear, as assumed in the empirical work of many authors such as Hansen and Singleton (1983) and Bean (1986), the estimator is just as efficient as the parametric two stage estimators. However, if the ϵ_t is serially dependent, then the estimator is inconsistent. But, this is also true for OLS estimator.

Newey's estimator is as efficient as other estimators which only made use of the conditional restrictions. The R estimator, however, is $\sqrt{N}$ consistent in the sense that the limiting distribution is normally distributed, but it is not clear whether the estimator achieves the semiparametric efficient bound. Of course, these results hold under the assumptions of i.i.d. observations.

As for the time series case, it is known that the simple nonparametric regression estimates are $\sqrt{(Na^q)}$ consistent and asymptotically normal (Robinson (1983, 1986a)). We conjectured that under some strengthening of the regularity conditions such as weak dependence and smoothness conditions, the asymptotic distributions for both estimators should be the same as in the i.i.d. case. In this spirit, we apply both the augmented Newey and Robinson's estimators to our time series data in the second part of our paper.

7.7 HAUSMAN AND ROBINSON TEST STATISTICS

Before we present the results, we will discuss two useful test statistics: the Hausman–type specification test and the Robinson's test for significance of π and its subvector of parameters. In the case of (6), if g is indeed linear, then we can write it as

$$y_t = b_0 + \pi' x_t + b_3 z_t + u_t$$

We can rearrange the equation when $E[x|z] = C_1 + C_2 z$ and yield

$$y_t = (b_0 + \pi' C_1) + \pi'(x_t - E_t x_t) + (\pi' C_2 + b_3')z_t + u_t$$

We have $E[x_t - E_t x_t] = 0$ and $E[(x_t - E_t x_t)z_t] = 0$ by construction. OLS estimator for π, π^{OLS}, and indeed the R estimator is asymptotically efficient in the Cramer–Rao sense if $E[u_t|x_t,z_t] = 0$ and u_t is normally distributed. Under the null hypothesis of no misspecification (linear expectation), OLS and R estimators are both consistent and asymptotically efficient. Under the alternative hypothesis (nonlinear expectation), OLS is inconsistent but R estimator is consistent. A Hausman's type test can therefore be constructed.

Define $\hat{V}$ to be the consistent estimator for the asymptotic variance of $(\hat{\pi} - \pi^{OLS})$, then according to Lemma 2.1 of Hausman (1978), under the null hypothesis that there is no misspecification, an asymptotically valid level–α statistic for rejecting the linear formation of rational expectations is

$$H = N((\hat{\pi} - \pi^{OLS})'\ \hat{V}^{-1}\ (\hat{\pi} - \pi^{OLS}) \sim \chi^2_p$$

where

$$\hat{V} = \hat{\sigma}^2 N[(\Sigma_t \hat{x}^*_t \hat{x}^*_t{}' I_t)^{-1} - (\Sigma_t x_t x_t{}' - \Sigma_t x_t \tilde{z}_t{}'(\Sigma_t \tilde{z}_t \tilde{z}_t{}')^{-1} \Sigma_t \tilde{z}_t x_t{}')^{-1}]$$

$$\tilde{z}_t = (1, z_t{}')'$$

The decision rule is to reject the linear specification of expectation if H exceeds 100(1–α)th percentile of the χ^2_p distribution.

We also introduce the test statistic of Robinson (1988c), which we should refer to as R test, for testing zero–type restrictions (e.g. $\pi = 0$ or $\pi_i = 0$) on our semiparametric model. The test will be useful for testing the the significance of a vector or the individual surprise term in the rational expectations models.

The motivation of employing the R test is that if $E_t x_t$ is parameterized as linear when in fact it is nonlinear, then the associated test statistics will give us incorrect rejection probabilities. The conclusion may then be invalid, though it may not necessarily be so.

Many authors do not pay much attention to the fit of these rational expectation terms. In fact, in many applied studies, the authors do not bother to report the fit of the linear formation of expectations and test whether the parametric assumptions in these models (for example, the model of Hansen and Singleton (1983)) are satisfied. If the parametric assumptions are not satisfied, the expectations may be misspecified and may well be nonlinear. The R test has the advantage, in that the practitioner does not have to specify a parametric form for the expectations.

Let $w_t = w(z_t)$ be a scalar function. The null hypothesis depends crucially on the understanding of the following Lemma:

Lemma 1: If $E[(x_t - E_t x_t)(x_t - E_t x_t)' w_t]$ is positive definite, $\pi = 0$ if and only if

$$E[(y_t - E_t y_t) x_t w_t] = 0. \qquad (7)$$

Lemma 1 is crucial in the construction of the test statistics. The hypothesis states that $y_t - E_t y_t$ and x_t are uncorrelated.

Lemma 2: A sufficient condition for (7) is

$$E[y_t | x_t, z_t] = E[y_t | z_t], \qquad \text{almost surely.} \qquad (8)$$

We can see that if (8) is rejected, we can conclude that in a parametric or nonparametric regression of y on x and z, we cannot omit x. Lemma 2 states that (8) implies (7). So if (7) is rejected, (8) is also rejected. But if (7) is not rejected, we cannot say much about (8).

The following discussions may be helpful in understanding the general structure of the test statistics. We deliberately leave out the technical discussions in order to avoid introducing further notations.

We need to express the restriction or null hypothesis in the form $\tau = 0$. τ is a px1 vector if we are interested in testing the whole vector π, and a scalar if we are interested in only one of the element of π. Let us introduce the notation

$$d_{ts} = (y_t - y_s) x_t v_t K_{ts} a^{-q}$$

and Robinson (1988c) has shown that we can in fact construct

$$\hat{\tau} = N^{-2} \Sigma_t \Sigma_s (y_t - y_s) x_t v_t K_{ts} a^{-q}$$

Compare d_{ts} with (7) and one may see some similarities. We have set the scalar function $v_t = v(z_t) = 1$ and so $w_t = v_t f(z_t) = f(z_t)$. f(z) is probability density function of z which is of an unknown form. Under some regularity conditions, $\sqrt{N}\ \hat{\tau}$ is asymptotically normally distributed with zero mean under the null hypothesis of $\tau = 0$. In order to form a correctly scaled statistic, one has resorted to spectral estimation for the autocovariance of $\hat{\tau}$. The test statistic is

$$\hat{R} = N/4\ \hat{\tau}'\ \hat{\Omega}^{-1}\ \hat{\tau}$$

where

$$\hat{\Omega} = \Sigma_{i=-m}^{m} H(i/m)\hat{\Psi}_i \qquad (9)$$

$$\hat{\Psi}_i = N^{-1}\Sigma_{t=1}^{N-i} c_t c_{t+1}, \qquad \hat{\Psi}_{-i} = \hat{\Psi}_i, \qquad i = 0.1,..,m$$

$$c_t = N^{-1}\Sigma_s c_{ts}$$

$$C_{ts} = (d_{ts} + d_{st})/2$$

(9) is the weighted–autocovariance nonparametric estimator. The weight, *H*, is known as the lag window and has to be chosen by the practitioner to satisfy the conditions of the theorem. The lag number m should be larger, the larger N is. There are a variety of lag windows to choose from and the properties of these lag windows have been studied in the time series literature. Our preferred choice for lag window *H* is the Hanning window (see the next section), i.e., $H(u) = (1 + \cos \pi u)/2$. Finally, under some regularity conditions,

$$\hat{R} \sim \chi^2_p, \qquad \text{as } N \to \infty$$

$\hat{R}$ has the usual asymptotic chi–squared distribution and it is easy to adopt the decision rule for an α–level test of significance, i.e., to reject the null hypothesis that $\pi = 0$ if $\hat{R}$ exceeds the upper 100α percentile of the χ^2 distribution, given a probability level α.

Let us partition $\pi = (\pi_1, \pi_2)$ and $x = (x_1, x_2)$, where π_1 is a scalar and π_2 is (p–1)x1. Then we may test the hypothesis that $\pi_1 = 0$ by replacing E_t as expectation conditional on both z as well as x_2. $\hat{R}$ is then distributed asymptotically as χ^2_1. It should be mentioned that we have not employed the trimming device and the higher–order kernels in the computation of the R statistics, instead we have resorted to the simple density estimator.

Some further remarks are given below. The asymptotic relative efficiency (ARE(R;b)) of the R test is non zero. In other words, test statistic R based on a N^b –consistent estimate compared to one based on a $\sqrt{N}$ consistent and fully efficient estimates is non zero, i.e. ARE(R;b) $= 0$ if $b < 1/2$ and ARE(R;b) >0 if $b = 1/2$.

The ARE for the R statistic can be calculated based on a true parametric model (see, e.g. Stock (1985)). Although the R statistic is positive, it may be much less than one in value. Nevertheless, it may be better than some parametric statistics based on $\sqrt{N}$ consistent but inefficient estimates. The R test is not confined to just the surprise model, in fact, it has wider applications than that. The proof of the asymptotic distribution of the test statistic is set up with time series in mind. Therefore the test allows direct application to time series data.

Further discussions and interpretations of the regularity conditions can be found in Robinson (1988c). In general, the conditions on the class of kernels is the same as the R estimator. We require some weak dependence conditions (absolute regularity), smoothness/moment conditions similar to those for the R estimator and some conditions on the lag window. Finally, some conditions need to be imposed on the rate of convergence of the bandwidth a and lag number m.

7.8 CHOICE OF BANDWIDTH, KERNELS, LAG WINDOW AND NUMBER

We discuss various intuitions regarding the choice of the bandwidth a, the choice of order ℓ of kernels H_ℓ, the choice of kernels k, the choice of lag window and the choice of lag number m. Some of these choices are guided by the theorems but others are not. We have to limit the scope of our discussion to the following and describe the course of action that we have taken in applied work. The first two points apply to both estimators and test statistics while the last two apply only to the test. The user has to decide on the choice of bandwidth, kernel, lag window, and lag number, in either estimation or statistical inference:

(i) Bandwidth a

One of our intentions, as in Chapter 4, is to examine the sensitivity of our estimates to the choice of a and ζ after a preliminary step of dimensional reduction. We have estimated the models under various bandwidth. The strategy is that if in this case, the estimates are not very sensitive to bandwidth choice, we may avoid using the expensive "automatic" estimates.

(ii) Order of kernel k, H_ℓ, and the choice of ψ

Although we have applied principal component analysis to z and reduced the dimension to q = 2, we report the results obtained under H_2, H_4 and H_6 to have an idea of the sensitivity of these estimates to the choice of ℓ.

(iii) Lag Window *H*

Like the kernels, the choice of lag window is less crucial than the choice of lag number. The advantages and disadvantages of using certain lag windows have been discussed extensively in the spectral density literature. The theorem of Robinson (1988c) has emphasized the Hanning window because of its ease of computations and its ability to protect against the influence of spectral peaks at distant frequencies. We will therefore adopt the Hanning window.

(iv) Lag number m

In using the R test, we have to decide on the choice of the lag number. We have conducted a very simple simulation (not reported here) on the set of real data that we have. The simplest model possible was used for the simulations. We took p = q = 1, $x_t = r_t$, $z_t = r_{t-1}$, and $E_t r_t = r_{t-1}/(1+r_{t-1})$. We generated u_t as first order Gaussian autoregressions with mean zero and variance 1, and the lag–1 autocorrelation coefficient ρ = 0.0, 0.5, 0.7, 0.9, 0.95 and 0.99. The number of replications were 5000. Similar conclusions were reached as those from the simulations of Robinson (1988c). The results suggested the maximum desirable value for m was 6 for our sample size of 133. We will report the results for m = 6 in our empirical work.

7.9 RESULTS FROM SEMIPARAMETRIC METHODS

We are interested in building a surprise consumption function and Model 1a and Model 2a are two simple models to start with. The results for Model 1a and 2a are reported in Table 7.5. The trimming constant ζ is set equal to 0.000001 and the bandwidth a is selected to be 0.30. The results are in fact insensitive to the trimming constant for the data set on hand for a wide range of values. We have estimated the models using many different values of a. However, we have decided not to report all the results.

The estimates for expected interest rates are not statistically significant, providing very little support for intertemporal substitution. The estimates vary widely with the bandwidth, but the general conclusion is that consumption is sensitive to current income but not inflation. This is consistent with the results that we have obtained in the last chapter.

According to the t–ratios, significant news come from income and inflation respectively in Model 1a and 2a. But this is not confirmed by the R test statistics. The Hausman statistics rejected the null hypothesis of linear rational expectation formation in Model 1a but not Model 2a.

The most interesting results come from Model 3 reported in Table 7.6a, 7.6b and 7.6c. The trimming constant is set at 0.001 and three different bandwidths are used. All the coefficients for the anticipated terms are of the "correct" signs and mostly significant for estimates obtained using H_4 and H_6 in all the tables. In particular, the estimates obtained with a = 0.3 and H_6 in Table 7.6a are most promising. All the anticipated variables are significant at the 1% level, except inflation.

From the t–ratios and R test statistics in the three tables, there is little doubt that significant news are from hours, income, and inflation. judging from the t–ratios. The most significant news is from hours followed by income. The anticipation–only hypothesis can be taken as rejected because the test statistics for a = 0.6 and 0.9 exceed the critical value.

Comparing the results in Table 7.2a and 7.6a, we observe that semiparametric estimates are all larger in magnitude. The long run steady state income elasticity are 2.8 and 3.4 for parametric and semiparametric models respectively, which appear to be far in excess of the findings of Blinder and Deaton (1985)'s 0.78.

We summarized our results in the following discussions. First of all, the semiparametric estimates are sensitive to the bandwidth a but not over sensitive. In some cases, the range of estimates we obtained is relatively small compared to the range of bandwidth used. The larger the bandwidth, the lower the t–ratios.

Secondly, the results are not sensitive to the trimming constant ζ as far as the set of data that we used are concerned but the t–ratios are smaller as we increase ζ.

Thirdly, the estimates for the anticipated terms for both methods are fairly similar despite these differences, with the semiparametric estimates mostly larger in magnitude. This supports the view that the linear instruments are closed to efficient and therefore nonparametric instruments may not have gained much efficiency. This perhaps suggests that in future application, one need not use the computational more expensive kernel estimates, but this requires further investigation. However, the results for the surprise terms cannot claim to be similar for both methods. This reinforces the point that expectation formation is crucial for consistent estimates.

One interesting question is whether most of the nonlinearity was generated by the GPC analysis. If this is the case, then we can also obtain the results simply using the PCs without resorting to kernel estimation. While this may be the case for the anticipated terms, the Hausman statistics do not support this claim for the unanticipated terms. We would like to believe that we have captured the nonlinearity in the formation of rational expectations in the latter case while the parametric methods are clearly inappropriate.

The fourth observation is that the (statistical significant) estimates for lagged consumption and income seem to suggest that they are cointegrated. Blinder and Deaton (1985)'s have rejected this hypothesis. However, all the estimates reported here for lagged consumption and income are almost identical in magnitude and opposite in sign which suggests that we may take the difference rather than treat these two variables as separate.

7.10 CONCLUSION

In some rational expectations models, the conditional expectations are only nuisance functions. Unfortunately, misspecification of these nuisance parameters may still lead to loss of efficiency in estimating some parameters of interest and consistency in others.

A semiparametric method has been proposed to estimate the anticipated and unanticipated terms. The augmented Newey's estimator for expected terms are as efficient as our parametric estimates. However, Robinson's estimator for the unanticipated terms only achieves the same efficiency as parametric two stage methods under the restrictive assumption that the error term is normally distributed and the expectation formation is linear. A Hausman–type misspecification test and Robinson's test of zero–type restrictions are introduced and applied to the surprise consumption function.

We have used semiparametric methods to analyze the "surprise" consumption model. We have found that the results using parametric methods are fairly similar to those using semiparametric methods for the anticipated terms. But, there are differences for the unanticipated terms from both methods. The differences cannot be entirely due to the fact that we have omitted the time trend and time squared in our information set or to the nonlinear projection that we used to reduce the dimension for nonparametric estimations. Furthermore, the linear vector autoregression used in most of the parametric studies may not be appropriate when additional variables such as inflation and stock prices are added to the system of equations. The breakdown of the normality and other parametric assumptions are confirmed by the diagnostic tests on the autoregression and the semiparametric test statistics. It would seem that we have captured some of the nonlinearity in the formation of expectations.

The traditional belief that lagged variables of income and consumption expenditure are always significant in the consumption regressions is once again confirmed by both parametric and semiparametric methods. We have also found significant roles for real rates of interest which is generally hard to establish empirically in parametric models. This is perhaps one of our main contributions.

We have also found that inflation variables matter in the consumption function thus confirming the findings of Koskela and Viren (1987). There may be roles for inflation because of price confusion effect of Deaton (1977) (see Blinder and Deaton (1985)) and because of imperfect capital market (see Jackman and Sutton (1982)). It is not clear why expected inflation and stock prices with positive coefficients should matter in the regression function, though Hall (1978) did find it difficult to reject stock prices' significant roles in his regression. Perhaps they simply reflect specific symptoms of the US economy and the importance of the underlying demand and supply conditions. Having obtained these results, more can be done in the future. The role of wealth is important and it is interesting to see what effect it has on consumption. The regularity conditions of the estimators have to be worked out though it is reasonable to assume that the asymptotic distribution of both the augmented estimators should hold in the time series case.

Table 7.1

Regression results using standard two stage methods

Model

(1): Δc_t = constant + $\beta_1 E_t r_t$ + $\beta_2 E_t y_t$ + surprise terms + ϵ_{1t}

(2): Δc_t = constant + $\beta_1 E_t r_t$ + $\beta_4 E_t y_t$ + surprise terms + ϵ_{2t}

variable	(1a)	(1b)	(2a)	(2b)		
constant	0.0035**			0.0050**		
	(0.0005)			(0.0005)		
anticipated terms						
Int. rates	0.091			–0.145		
	(0.98)			(0.11)		
Income	0.322					
	(0.58)					
Inflation				0.0007**		
				(0.0002)		
surprise terms						
Int. rates	–0.094	–0.070	–0.203	–0.070		
	(0.12)	(0.19)	(0.21)	(0.21)		
Hours		0.25*		0.25*		
		(0.10)		(0.10)		
Govt. expend.		0.028		0.028		
		(0.05)		(0.05)		
Income	0.13**	0.08*		0.08*		
	(0.05)	(0.05)		(0.05)		
Inflation		–0.0002	–0.0003	–0.0002		
		(0.0004)	(0.0004)	(0.0004)		
Stock Prices		0.00008		0.0008		
		(0.00009)		(0.00010)		
					critical value	
RSS	0.00278	0.00260	0.00346	0.00310	5%	1%
Serial corr.						
χ^2_4	6.08	9.37	9.07	9.04	9.49	13.28
Functional form						
χ^2_1	11.35**	6.93*	1.47	0.12	3.84	6.64
Normality						
χ^2_2	53.29**	54.64**	3.86	1.23	5.99	9.21
Heteroscedasticity						
χ^2_1	3.65	2.41	0.36	3.29	3.84	6.64
Sargan's test						
χ^2_{21}	54.27**		65.26**		32.6	38.9

Note:

* indicates that the coefficient is significant at 5% level.

** indicates that the coefficient is significant at 1% level.

Asymptotic standard errors are reported in the parentheses.

Table 7.2a

Regression results using standard two stage methods

(1): $\Delta c_t = \text{constant} + \beta' E_t x_t + \pi'(x_t - E_t x_t) + \epsilon_{4t}$

(2): $\Delta c_t = \text{constant} + b_{21} c_{t-1} + b_{22} y_{t-1} + \beta' E_t x_t + \pi'(x_t - E_t x_t) + \epsilon_{4t}$

variable	(1)	(2)	(3)	(4)
no.of obs n	133	124	133	124
constant	0.0047**	0.0071**	0.0028	0.0036
	(0.0007)	(0.0010)	(0.0028)	(0.0029)
lag c.			−0.088**	−0.080*
			(0.025)	(0.032)
lag y			0.090**	0.082**
			(0.023)	(0.032)
anticipated terms				
Int. rates	0.040**	0.333*	0.284*	0.402**
	(0.10)	(0.14)	(0.12)	(0.14)
Hours	0.310**	0.376**	0.285**	0.340**
	(0.12)	(0.13)	(0.11)	(0.13)
Govt. expend.	−0.119*	−0.301**	−0.146*	−0.214*
	(0.06)	(0.09)	(0.08)	(0.10)
Income	0.177*	0.066	0.210**	0.144
	(0.082)	(0.093)	(0.078)	(0.095)
Inflation	0.0006**	0.00046*	0.0005**	0.00042*
	(0.00021)	(0.00022)	(0.00019)	(0.00021)
Stock prices	0.00027*	0.00019	0.0004**	0.0004*
	(0.00015)	(0.00016)	(0.00015)	(0.00017)
surprise terms				
Int. rates	−0.070	0.018	−0.070	−0.012
	(0.109)	(0.179)	(0.159)	(0.172)
Hours	0.249**	0.231**	0.249**	0.339**
	(0.09)	(0.09)	(0.08)	(0.08)
Govt. expend.	0.028**	0.014	0.028	0.023
	(0.041)	(0.042)	(0.039)	(0.040)
Income	0.083*	0.082*	0.083*	0.080*
	(0.041)	(0.042)	(0.039)	(0.041)
Inflation	−0.00024	−0.00015	−0.00024	−0.00014
	(0.00032)	(0.00032)	(0.00030)	(0.00030)
Stock prices	0.00007**	0.00012	0.000078**	−0.00010
	(0.00008)	(0.00008)	(0.00007)	(0.00008)

Note:
* indicates that the coefficient is significant at 5% level.
** indicates that the coefficient is significant at 1% level.
Asymptotic standard errors are reported in the parentheses.

Table 7.2b

Diagnostic tests results

	(1)	(2)	(3)	(4)	critical values 5%	1%
RSS	0.00201	0.00182	0.00173	0.00165		
Serial corr. χ_4^2	3.21	5.07	8.37	8.17	9.49	13.28
Functional form χ_1^2	0.66	0.76	0.62	0.29	3.84	6.64
Normality χ_2^2	1.64	0.15	0.78	0.87	5.99	6.64
Heteroscedasticity χ_1^2	1.40	1.36	0.66	0.88	3.84	6.64
Predictive failure χ_9^2		11.35		5.72	16.9	21.6
Sargan's Test χ_{15}^2	34.59**	29.06*	22.37	20.94	25.0	30.5

Tests of "surprises only" and "anticipated only" hypothesis

					critical values 5%	1%
Surprises only						
LM χ_6^2	53.62**	51.29**	60.56**	52.13**	12.5	16.8
LR χ_6^2	68.64**	66.20**	80.80**	67.63**	12.5	16.81
F(6,120)	13.51**				3.6	6.88
F(6,111)		13.05**			3.6	6.88
F(6,118)			16.44**		3.67	6.88
F(6,109)				13.18**	3.67	6.88
Anticipation only						
LM χ_6^2	21.26**	19.52**	24.02**	21.21**	12.59	16.81
LR χ_6^2	23.17**	21.24**	26.50**	23.27**	12.59	16.81
F(6,120)	3.81*				3.67	6.88
F(6,111)		3.47				
3.67 6.88						
F(6,118)			4.34*		3.67	6.88
F(6,109)				3.75*	3.67	6.88

Note:
* indicates that the coefficient is significant at 5% level.
** indicates that the coefficient is significant at 1% level.

Table 7.3

Specification tests for supplementary regression

$$x_t = \text{constant} + a(L)z_{t-1} + b_1t + b_2t^2 + \xi_t$$

(I) 3 LAGS

	r	ΔH	Δg	Δy	ΔI	ΔS
Serial corr. χ_4^2	51.36**	15.23**	27.07**	5.93	1.71	1.67
Functional form χ_1^2	0.20	4.84*	6.71**	3.42	10.83**	0.42
Normality χ_2^2	5.66	1.56	13.63**	181.4**	188.6**	65.54**
Heteroscedasticity χ_1^2	9.63**	0.34	1.17	13.47**	6.97**	0.20

(II) 2 LAGS

	r	ΔH	Δg	Δy	ΔI	ΔS
Serial corr. χ_4^2	42.36**	13.95**	8.10	6.62	4.24	1.74
Functional form χ_1^2	0.09	5.03*	5.85*	1.02	10.13**	0.35
Normality χ_2^2	16.75**	1.65	8.62*	229.2**	246.6**	82.73**
Heteroscedasticity χ_1^2	19.78**	0.28	2.46	7.03**	5.21*	0.80

Note:
* indicates that the coefficient is significant at 5% level.
** indicates that the coefficient is significant at 1% level.

Table 7.4

Principal component analysis

(a) Linear PCA (3 lags of each variable and constant).
Total number of variables = 22.

Principal Components	Variance Explained	Total Variance
1	98.07	98.07
2	0.22	99.29
3	0.44	99.73
4	0.08	99.81

(b) Nonlinear PCA

(i) quadratic functions of first two PCs obtained from (a).
Total number of variables = 5.

Principal Components	Variance Explained	Total Variance
1	42.89	42.89
2	20.86	63.75
3	19.73	83.48
4	10.95	94.43

(ii) 22 variables as in (a) in addition to the quadratic functions of S and I.
Total number of variables = 43.

Principal Components	Variance Explained	Total Variance
1	89.63	89.63
2	6.31	95.94
3	2.32	98.26
4	0.46	98.72

Correlation matrix is used for the PCA.

Table 7.5

Semiparametric regression results for Model 1a and 2a

Model

(1a): $\Delta c_t = \text{constant} + b_{11}c_{t-1} + b_{12}\, y_{t-1} + \beta_1 E_t r_t + \beta_2 E_t \Delta y_t + \text{surprise terms} + \epsilon_{5t}$

(2a): $\Delta c_t = \text{constant} + b_{11}c_{t-1} + b_{12}\, y_{t-1} + \beta_1 E_t r_t + \beta_2 E_t \Delta I_t + \text{surprise terms} + \epsilon_{6t}$

variable	H_2	H_4	H_6	H_2	H_4	H_6
a = 0.30		b = 0.000001				
constant	0.0062	0.0062	0.0062	0.0000	0.0007	0.0012
	(0.0038)	(0.0038)	(0.0038)	(0.0450)	(0.0449)	(0.0450)
lag c	–0.078*	–0.081*	–0.084*	–0.1351	–0.1178	–0.1031
	(0.039)	(0.039)	(0.039)	(0.4621)	(0.4601)	(0.4633)
lag y	0.081*	0.084*	0.087*	0.0455	0.0274	0.0120
	(0.039)	(0.039)	(0.039)	(0.4599)	(0.4572)	(0.4603)
anticipated terms						
Interest	0.283	0.301	0.317	1.257	0.317	–0.102
	(0.384)	(0.268)	(0.238)	(5.433)	(3.826)	(3.212)
Income	0.507**	0.493**	0.481**			
	(0.155)	(0.110)	(0.094)			
Inflation				0.0026	0.0014	0.0009
				(0.0099)	(0.0067)	(0.0057)
surprise terms						
Interest	–0.045	–0.007	0.034	–0.019	0.028	0.071
	(0.076)	(0.076)	(0.077)	(0.085)	(0.082)	(0.082)
R test χ_1^2		1.147			1.736	
Income	0.201**	0.187**	0.193**			
	(0.037)	(0.037)	(0.038)			
R test χ_1^2		3.367				
Inflation				0.0003*	0.0005**	0.0006**
				(0.0001)	(0.0001)	(0.0001)
R test χ_1^2					0.488	
Hausman's Test						
χ_2^2	12.30**	10.18**	10.49**	1.17	1.33	1.04
R test Anticipation Hypo.						
χ_2^2		5.045			1.151	

Note:
* indicates that the coefficient is significant at 5% level.
** indicates that the coefficient is significant at 1% level.
Asymptotic standard errors are reported in the parentheses.

Table 7.6a
Semiparametric regression results for Model 3a

Model (3a): Δc_t = constant + $b_{21}c_{t-1}$ + $b_{22}y_{t-1}$ +
$\beta' E_t x_t + \pi'(x_t - E_t x_t) + \epsilon_{4t}$

variable	H_2	H_4	H_6
a = 0.30	b = 0.001		
constant	0.0052	0.0054	0.0053
	(0.0069)	(0.0048)	(0.0041)
lag c	–0.125*	–0.125**	–0.124**
	(0.062)	(0.047)	(0.041)
lag y	0.125*	0.126**	0.125**
	(0.062)	(0.046)	(0.041)
anticipated terms			
Interest	0.565	0.559*	0.575**
	(0.373)	(0.250)	(0.212)
Hours	0.327	0.344*	0.353**
	(0.244)	(0.163)	(0.129)
Govt	–0.287*	–0.313**	–0.320**
	(0.145)	(0.092)	(0.071)
Income	0.277*	0.240**	0.228**
	(0.123)	(0.087)	(0.075)
Inflation	0.00086	0.00077	0.00080*
	(0.00063)	(0.00044)	(0.00038)
Stock prices	0.00070*	0.00075**	0.00077**
	(0.00027)	(0.00017)	(0.00015)
surprise terms			
Interest rates	0.164**	0.222**	0.258**
	(0.074)	(0.072)	(0.072)
R Test $\chi^2(1)$		1.132	
Hours	0.337**	0.370**	0.350**
	(0.070)	(0.073)	(0.074)
R test $\chi^2(1)$		11.973**	
Govt.	–0.024	–0.043	–0.047
	(0.037)	(0.038)	(0.037)
R test $\chi^2(1)$		0.016	
Income	0.112**	0.094**	0.104**
	(0.041)	(0.040)	(0.041)
R test $\chi^2(1)$		5.515*	
Inflation	0.00043**	0.00046**	0.00052**
	(0.00015)	(0.00015)	(0.00014)
R test $\chi^2(1)$		3.717	
Stock prices	0.00010	0.00011	0.00010
	(0.00007)	(0.00007)	(0.00007)
R test $\chi^2(1)$		0.910	
R Test: Anticipation Hypo. $\chi^2(6)$		12.264	

Note: * indicates that the coefficient is significant at 5% level.
** indicates that the coefficient is significant at 1% level.
Asymptotic standard errors are reported in the parentheses.

Table 7.6b
Semiparametric regression results for Model 3a

Model (3a): $\Delta c_t = \text{constant} + b_{21}c_{t-1} + b_{22}y_{t-1} + \beta' E_t x_t + \pi'(x_t - E_t x_t) + \epsilon_{4t}$

variable	H_2	H_4	H_6
a = 0.60	b = 0.001		
constant	0.0073	0.0055	0.0055
	(0.0218)	(0.0073)	(0.0058)
lag c	−0.112	−0.122*	−0.123*
	(0.159)	(0.063)	(0.052)
lag y	0.110	0.122*	0.123*
	(0.170)	(0.063)	(0.051)
anticipated terms			
Interest	0.636	0.565	0.539*
	(1.325)	(0.404)	(0.301)
Hours	0.309	0.334	0.332*
	(0.575)	(0.241)	(0.199)
Govt.	−0.252	−0.275*	−0.286*
	(0.494)	(0.155)	(0.117)
Income	0.330	0.276*	0.260*
	(0.380)	(0.128)	(0.102)
Inflation	0.00137	0.00091	0.00082
	(0.00182)	(0.00066)	(0.00051)
Stock prices	0.00060	0.00069	0.00069**
	(0.00105)	(0.00029)	(0.00020)
surprise terms			
Interest	0.125*	0.131*	0.133*
	(0.076)	(0.076)	(0.076)
R test $\chi^2(1)$		0.883	
Hours	0.316**	0.317**	0.319**
	(0.067)	(0.068)	(0.068)
R test $\chi^2(1)$		11.941**	
Govt.	−0.013	−0.013	−0.018
	(0.035)	(0.035)	(0.036)
R test $\chi^2(1)$		0.069	
Income	0.132**	0.125**	0.123**
	(0.038)	(0.039)	(0.039)
R test $\chi^2(1)$		5.954*	
Inflation	0.00033**	0.00035**	0.00035**
	(0.00015)	(0.00015)	(0.00015)
R test $\chi^2(1)$		4.719*	
Stock prices	0.00014*	0.00012*	0.00011**
	(0.00007)	(0.00007)	(0.00007)
R test $\chi^2(1)$		0.892	
R test Anticipation Hypo. $\chi^2(6)$		12.694*	

Note:
* indicates that the coefficient is significant at 5% level.
** indicates that the coefficient is significant at 1% level.
Asymptotic standard errors are reported in the parentheses.

Table 7.6c
Semiparametric regression results for Model 3a

Model (3a): $\Delta c_t = \text{constant} + b_{21}c_{t-1} + b_{22}y_{t-1} + \beta' E_t x_t + \pi'(x_t - E_t x_t) + \epsilon_{4t}$

variable	H_2	H_4	H_6
a = 0.30	b = 0.001		
constant	0.0091	0.0058	0.0055
	(0.0918)	(0.0125)	(0.0082
lag c	–0.089	–0.120	–0.118*
	(0.485)	(0.096)	(0.068)
lag y	0.085	0.119	0.119*
	(0.555)	(0.100)	(0.068)
anticipated terms			
Interest	0.795	0.650	0.578
	(5.045)	(0.719)	(0.467)
Hours	0.319	0.340	0.331
	(2.312)	(0.334)	(0.250)
Govt.	–0.162	–0.269	–0.268
	(1.645)	(0.253)	(0.173)
Income	0.297	0.296	0.280*
	(1.530)	(0.206)	(0.139)
Inflation	0.00185	0.00114	0.00099
	(0.00727)	(0.00103)	(0.00074)
Stock prices	0.00032	0.00068	0.00068*
	(0.00281)	(0.00055)	(0.00034)
surprise terms			
Interest	0.113	0.125*	0.127*
	(0.076)	(0.076)	(0.076)
R test $\chi^2(1)$		0.997	
Hours	0.309**	0.316**	0.315**
	(0.066)	(0.067)	(0.067)
R test $\chi^2(1)$		12.141**	
Govt.	–0.015	–0.014	–0.011
	(0.035)	(0.035)	(0.035)
R test $\chi^2(1)$		0.117	
Income	0.136**	0.133**	0.128**
	(0.036)	(0.037)	(0.038)
R test $\chi^2(1)$		6.583*	
Interest	0.00031**	0.00031**	0.00031**
	(0.00015)	(0.00015)	(0.00015)
R test $\chi^2(1)$		5.558*	
Stock prices	0.00015**	0.00015**	0.00014**
	(0.00007)	(0.00007)	(0.00007)
R test $\chi^2(1)$		0.882	
R test Anticipation Hypo. $\chi^2(6)$		12.946*	

Note:
* indicates that the coefficient is significant at 5% level.
** indicates that the coefficient is significant at 1% level.
Asymptotic standard errors are reported in the parentheses.

CHAPTER 8

BANDWIDTH SELECTION:
SOME MONTE CARLO RESULTS

8.1 INTRODUCTION

The semiparametric estimators that we have employed in this thesis have desirable properties, such as being efficient or consistent when the competing parametric ones are not. However, in order to implement the semiparametric procedures, decisions have to be made about how to calculate the nonparametric components. The most important of these choices that an empiricist has to make is the choice of bandwidth, or equivalently, the choice regarding the size of the neighbourhood.

The asymptotic properties of the estimators that we have quoted in the text usually assert that the 'optimal' bandwidth should vary directly with the sample size. Besides this, there is little useful information regarding the choice for a practitioner. Most researchers are content with a subjective choice, e.g., by plotting graphs and examining the smoothness of the nonparametric estimates, or setting a = constant x $N^{-1/5}$. These subjective rules are usually ad hoc, and is equivalent to choosing a particular parametric form to work with. Although the theorems for consistency or asymptotic normality only require that a has a suitable rate of convergence as $N \to \infty$, it should be fair to say that in finite sample, the selection of the bandwidth in a semiparametric setting should be as crucial as the selection of a parametric from in a parametric model.

As we have mentioned in Chapter 2, various automatic methods have been suggested for bandwidth selection in nonparametric methods and some of these have been adopted for semiparametric models. Since in this case the bandwidth is dependent on the data, the resulting estimator obtained by substituting the selected bandwidth into the formula will be data–dependent too. The bandwidth being data dependent has several consequences. In the case where the semiparametric estimator can be regarded as a linear estimator without data–independent bandwidth, it has now become nonlinear with the data–dependent bandwidth. Furthermore, the consistency and asymptotic normality of the parameters of interest, denoted by $\hat{\beta}$, is difficult to show in the presence of data–dependent bandwidth. This is because there is now an element of "endogeneity" in the estimator. Of course, it is not impossible to prove consistency and normality and indeed a few results exist, e.g., Robinson (1988d)'s results on the Hannan's GLS estimator in the frequency domain are derived under data–dependent bandwidth. His results are to be exploited in this study here.

Although there are few asymptotic results for semiparametric models, we can perhaps use the methods by appealing to the knowledge that we have from those purely nonparametric models. In fact, some of the literature on automatic bandwidth selection for semiparametric models depends heavily on the results from purely nonparametric problems.

In view of the importance of bandwidth selection in semiparametric models, this chapter presents some Monte Carlo results on various models of interest. In particular, we are interested in the finite sample properties of various so called rules–of–the–thumb and automatic bandwidth selection methods. We have only reported the results for the i.i.d. case corresponding to different dimensions of smoothing.

The basic cost function used in previous studies is usually the quadratic loss or expected loss that we have mentioned. Some other approaches have also been used as stopping rule for selection of regressors, fitting autoregressions when the true model is unknown. These approaches are related and have been adopted for different problems and discussed. But the criteria that we are to speculate are closely related to Robinson (1988d) and based on the principle of cross–validation.

8.2 THE PSEUDO LOG–LIKELIHOOD AND CROSS–VALIDATION CRITERIA

The first criterion that is being proposed is that of Gaussian pseudo log–likelihood criterion of Robinson (1988d):

$$\hat{\theta} = \text{argmax}_{\theta}\, L(\theta)$$
$$= \text{argmax}_{\theta} -(2N)^{-1}\Sigma_i\{\log 2\pi + \log \hat{g}_{-i}(a) + RSS_{-i}(\beta,a)/g_{-i}(a)\}$$

where $\hat{\theta} = (\beta,a)$, g_i is the conditional variance, RSS_{-i} is the residual sum of squares obtained using the leave–one–out estimates, and the subscript –i refers to the leave–one–out estimates. Although the general approach in the literature is to study the properties of the 'automatic' selected bandwidth, the eventual interest lies in a semiparametric estimator β. We are interested in finding a suitable bandwidth and we have suggested maximizing $L(\theta)$ simultaneously with respect to β and a. This likelihood criterion has the advantage of being able to take into account heteroscedasticity and is useful for models with nonparametric variance estimates. From the computational point of view, we may follow the time series problem of Robinson (1988d) by concentrating the σ out of the likelihood function. Indeed, if we can replace β by any root N consistent estimator, $\tilde{\beta}$, we need only to optimize the simplified function over some subset of a–values, such that a maximum exists:

$$\hat{a} = \text{argmax}_{a} -(2N)^{-1}\Sigma_i\{\log 2\pi + \log \hat{g}_{-i}(a) + RSS_{-i}(\tilde{\beta},a)/g_{-i}(a)\}$$

Although the two procedures are asymptotically equivalent in Robinson (1988d)'s problem, the latter has enormous computation advantage as we have reduced a multi–dimensional optimization problem to a mere one–dimensional problem.

In the event of homoscedasticity, it appears that we may concentrate σ out of the likelihood. This reduces the function to the cross–validation (CV) criterion. The CV criterion is computationally more expensive than some other criteria but has the advantage that it can be easily adapted to many semiparametric problems that we are interested in. Three models of particular interest are investigated here:

(i) Partial and partly linear model

There has been some work on the partly linear model. One of the more popular criteria is the generalized cross–validation (GCV), introduced by Craven and Wahba (1979). It has been used in some semiparametric studies of the partial linear model (Speckman (1988)) with kernel estimation. Engle et al (1986) have also suggested GCV for partial liner model using smoothing splines. Green (1985) has also addressed the issue in the context of field trials. Most of this work take g to be non–stochastic. The main problem with g being stochastic is that the MSE may not exist. Consider the model

$$y_i = \beta x_i + g(z_i) + \epsilon_i$$

with $E[\epsilon_i | z_i] = 0$. In matrix terms, we have

$$y = X\beta + g(z) + \epsilon \qquad (1)$$

and denote

$$\hat{y} = X\hat{\beta} + \hat{g}$$

In the case of purely nonparametric regression, i.e. β=0, we can show that the optimum bandwidth from minimizing the mse (call it mse_0) is proportional to $N^{-(1/2\ell+d)}$, where ℓ is the order of kernels and d is the dimension of smoothing. This gives us the optimum rate of $E(mse_0)$ being proportion to $N^{-(2\ell/[2\ell+d])}$. Speckman (1988), in his study of the relationship between the average mse of the model (1) (call it mse_1) with that of a nested one with $\beta = 0$, has shown that under some regularity conditions

$$E(mse_1(a)) = E(mse_0(a))[1+o(1)]$$

where $mse_0(a) = N^{-1}\|E(y)-Ky\|^2$,

$$mse_1(a) = N^{-1}\|E(y)-X\hat{\beta}-\hat{g}\|^2$$

and K matrix consist of the ij–th element $K_{ij}/\Sigma_j K_{ij}$. This shows that the optimum rate for the $E(mse_0)$ is also optimum for $E(mse_1)$. With this result, one can then apply the powerful GCV theorem of Craven and Wahba (1979) to prove the following:

$$E(GCV(a)) = \sigma^2 + E(mse_1(a))\{1+o(1)\}$$

given that

$$GCV = RSS/[1-N^{-1}tr(A(a))]$$

$$\mathrm{RSS} = N^{-1}\|[I-A(a)]y\|^2$$
$$A(a) = K + \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'(I-K)$$
$$\tilde{X} = (I-K)\,X$$

Of course, if one is able to express $\hat{y}$ as a linear estimator of y, such as the case of Speckman (1988) and Robinson (1988a)'s estimator with higher–order and trimming, the idea of GCV can easily be applied. His results suggest that the bandwidth that minimizes the GCV criterion is asymptotically equivalent to the minimizer of $E[mse_1(a)]$. By similar argument and provided that the MSE exists, we can apply the theorem to Robinson (1988a)'s estimator which is the estimator that we are interested in.

But we have taken the suggestions by Allen (1974) and apply the PRESS criterion or simply known as cross–validation. In this case, as noticed by Robinson (1988d), the problem with the cross–validation in partial linear model is that minimizing the $E[mse_1(a)]$ wrt a alone will yield the optimum bandwidth of zero as that is the value which gives $E[mse_1(a)] = 0$. Therefore, an alternative method has to be devised in this case. That is to minimize the criterion wrt to β and a together with a. In fact, if there is no heteroscedasticity, maximizing the likelihood function wrt θ should be equivalent to minimizing the cross–validation criterion. Robinson (1988a)'s estimator can be expressed as

$$\hat{\beta}^R = (\tilde{X}^{*\prime}\tilde{X}^{*})^{-1}\tilde{X}^{*\prime}\tilde{y}^{*}$$

where

$$\tilde{X}^{*} = (I-K)\ \mathrm{Diag}(I_1, I_2,\ldots, I_N)X$$
$$\tilde{y}^{*} = (I-K)\ \mathrm{Diag}(I_1, I_2,\ldots, I_N)y$$
$$I_i = I(\hat{f} > \zeta)$$

ζ is the trimming constant and I is the indicator function. We obtain the bandwidth by minimizing the criterion function

$$\mathrm{CV} = \Sigma_j(\tilde{y}_{-j} - \beta^R_{-j}\tilde{x}_{-j})^2$$

wrt a. $\tilde{x}_{-j}$ is the nonparametric regression leave–one–out estimates of x on the instruments z. This bandwidth is then used to estimate β.

(ii) Errors–in–variables model

In this case, we can suggest the cross–validation (CV) criterion in the absence of heteroscedasticity. Define the model to be

$$y_i = x_i\beta + \eta_i \qquad (2)$$
$$x_i = E[x_i | z_i] + \xi_i$$

where $E[\eta_i\xi_i | z_i] = 0$ and $E[\eta_i | z_i] = 0$. Newey (1987)'s estimator can be expressed as

$$\hat{\beta}^{NEW} = \beta_{IV} + (\hat{X}'X)^{-1}\hat{X}'\eta$$

β_{IV} is obtained by 2SLS estimation and $\hat{X}$ is the nonparametric estimate for $E[x_i | z_i]$. We obtain the bandwidth by minimizing the criterion function

$$CV = \Sigma_j(y_j - \beta_{IV}\hat{x}_{-j})^2$$

wrt a. $\hat{x}_{-j}$ is the nonparametric regression leave–one–out estimates of x on the instruments z. This bandwidth is then used to estimate β.

(iii) Generalized least squares model

Let us define the model as

$$y_i = \beta x_i + \nu_i$$

with $E[\nu_i | x] = g_i$. The log–likelihood criterion can be straightforwardly implemented in this case as suggested by Robinson (1988d), i.e.,

$$\hat{a} = \text{argmax}_a \; -(2N)^{-1}\Sigma_i\{\log 2\pi + \log \hat{g}_{-i}(a) + RSS_{-i}(\tilde{\beta}^{OLS},a)/g_{-i}(a)\}$$

and the semiparametric GLS estimator is

$$\hat{\beta}^{SGLS} = (X'\Omega X)^{-1}X'\Omega y$$

where Ω is a diagonal matrix with i–th element $\hat{g}_i^{*-1}$, and $\hat{g}_i^{*-1}$'s are the cross–validated variance estimates. In all the cases, we have only taken a one–step procedure. Further iterations in some models may be desired (Matloff, Rose and Tai (1984)) . For example, we may want to obtain the residuals from the one step semiparametric GLS estimates and repeat the procedures of finding the optimum bandwidth. The back–and–forth process can be repeated until the GLS estimates converge. We offer no evidence of the performance of these estimates as it may be computationally too expensive for Monte Carlo simulations.

8.3 MEAN SQUARE ERROR CRITERION

Throughout the analysis, the bias, standard deviation and mse are reported for comparison. Normal density is used for construction of kernels. Higher–order product kernels are used where appropriate. As we can see from Chapter 2 that the optimum bandwidths for nonparametric problems in the mean square error sense are usually of the order $N^{-(1/(2\ell+d))}$. Several so called rules–of–the–thumb (RT) for bandwidth selection are used in the literature because they are easy to understand. These RT are especially popular in the applied literature:

(i) OPT criterion: since $a_{opt} = CN^{-(1/(2\ell+d))}$, this criterion set the bandwidth at $N^{-(1/(4+d))}$, i.e., C = 1.0. Some empirical work employed this rule for bandwidth selection. One example is Pagan and Ullah (1989)

(ii) SD criterion: one can also take the constant C to be the standard deviation of the data to be smoothed, e.g. Whistler (1989) sets her bandwidth based on this idea.

(iii) UNIT criterion: this criterion corresponds to using a diagonal bandwidth matrix h but taking $a_i = sd(z_i) \times N^{-(1/[d+2\ell])}$. In other words, dividing the z_i by the individual bandwidth and then using a single bandwidth as in (i). The division of the z by the individual sd is essential in applied work to take into account the spread of the data. For an example, see very recent work of Sentana and Wadhwani (1989)'s choice of bandwidth $a = c \times h_i$ where c corresponds to a different constant.

There are reasons to believe that the OPT criterion will not work in practice for $q > 1$ if the z's have large standard deviations and a non–zero mean. One of the main reasons is the empty space phenomenon caused by the grouping of the z and therefore the product kernels are not operational with a single a. Furthermore, it is difficult to justify the use of product kernels when the z's are of different units of measurement. Nevertheless, we produce results for the OPT criterion bearing in mind that we have adopted a design that enables it to work.

8.4 THE SIMULATION RESULTS

All the programs are written in double precision FORTRAN–77. The results are obtained using the AMDHALL–5890/300 computer at the University of London Computer Centre (ULCC). The subroutines for generation of random variables are from NAG–13 library and the minimization routines are taken from Numerical Recipes (Press, Flannery, Teukolsky and Vetterling (1987), we shall refer to this as PFTV). The bandwidths reported are the average for CV and SD criteria.

(i) Partly Linear model

In order to study the finite sample property of the partly linear model, we use similar design as that of Robinson (1988a). This will enable us to compare the results obtained. The model is

$$y_i = \beta_0 + \beta_1 x_i + g(z_i) + \epsilon_i \qquad\qquad i = 1,\ldots, N.$$
$$g(z_i) = \Sigma_{j=1}^{d} \gamma_j z_j^{\,2},$$
$$\beta_0 = \beta_1 = \gamma_1 = \gamma_2 = \ldots = \gamma_d = 1,$$

where ϵ_i is N(0,1), x, $z_{(j)}$'s are equicorrelated and identically distributed N(1,1) variables with correlation $\surd 2$. Notice that we have to minimize the CV criterion simultaneously wrt both β_1 and a, as we cannot minimize β_1 holding a constant. This suggests that any minimization routines which use the line minimization will not work for the same reason. One routine that will work in this case will be the method of downhill simplex of Nelda and Mead (1965). The AMOEBA routine given in PFTV (Section 10.4) has performed reasonably well in this case. However, global minimum is not guaranteed and therefore we have to try various initial starting points in order to make sure that we have the minimum over a certain interval. The larger the number of restarts for each replication, the better the chances of getting the minimum. At the same time, the higher the cpu time charges, the fewer replications allowed. Since the maximum of the cpu time allowed on the ULCC AMDHALL is 60 minutes, we stretches this to the limit whenever possible. The number of restarts ranges from 3 to just none depending on d. In other words, for N greater than 300, we have no restart.

One important point to bear in mind is the use of large d is highly susceptible to computational rounding off errors. To see this, recall that when multiplicative kernels are used as in all the cases here, the individual normal kernel $k(u) = (2\pi)^{-1/2} \exp(u^2)$ is between zero and one. Therefore successive multiplication of k for $d > 1$ will give a very small number. It is strongly advisable to do the computation in higher order as d increases. The rounding off errors have considerable effects over the estimates from some simulation results not reported.

In Table 8.1.a, the bias, standard deviation (std dev) and mean square error (mse) are reported. Directly below the results, we have the ratios of the values of the first three columns to the fourth.

There is little doubt that the OPT criteria gives considerably better results. This suggests that if a single a is operational with the data, it is perhaps good news to the empiricist. If it is not operational, then the CV criterion is desirable. Indeed, there is an enormous difference between the mse of the CV criteria and that of the SD or UNIT Z.

In Table 8.1.b, we have reported the results for $d = 2$ and the difference in the performance of CV with the last two criteria is substantial. Further justification for the use of CV is provided by the evidence in Table 8.1.c and Table 8.1.d. In both Tables, with $N = 400$, the results clearly indicate the performance of SD and UNIT criterion deteriorates with increased d.

Table 8.1.a Results for Partly Linear Model: $y = \beta_0 + \beta_1 x + g(z) + \epsilon$

Replications = 5000, N = 100, d = 1

	1.CV	2.OPT	3.SD	4.UNIT Z
$\ell = 2$				
a	2.89811	0.39811	2.13579	0.39811
bias	0.05302	0.00173	0.05168	0.05168
std dev	0.07256	0.07169	0.07390	0.07390
mse	0.00808	0.00514	0.00813	0.00813
bias	1.02590	0.03351	1.00000	1.00000
std dev	0.98180	0.96999	1.00000	1.00000
mse	0.99301	0.63225	1.00000	1.00000
$\ell = 4$				
a	3.09948	0.59948	3.21615	0.59948
bias	0.06302	0.00198	0.06059	0.06059
std dev	0.09625	0.07170	0.09954	0.09954
mse	0.01324	0.00515	0.01358	0.01358
bias	1.04018	0.03261	1.00000	1.00000
std dev	0.96701	0.72038	1.00000	1.00000
mse	0.97481	0.37895	1.00000	1.00000
$\ell = 6$				
a	3.20170	0.70170	3.76455	0.70170
bias	0.08785	0.00209	0.08690	0.08690
std dev	0.15506	0.07196	0.20062	0.20062
mse	0.03176	0.00518	0.04780	0.04780
bias	1.01097	0.02402	1.00000	1.00000
std dev	0.77288	0.35868	1.00000	1.00000
mse	0.66443	0.10842	1.00000	1.00000

Table 8.1.b Results for Partly Linear Model: $y = \beta_0 + \beta_1 x + g(z) + \epsilon$

Replications = 5000, N = 100, d = 2

	1.CV	2.OPT	3.SD	4.UNIT Z
$\ell = 2$				
a	2.73208	0.46416	2.49660	0.46416
bias	0.02817	0.00402	0.11327	0.11295
std dev	0.08741	0.10446	0.10161	0.10167
mse	0.00843	0.01093	0.02316	0.02309
bias	0.24942	0.03560	1.00279	1.00000
std dev	0.85973	1.02749	0.99950	1.00000
mse	0.36516	0.47319	1.00264	1.00000
$\ell = 4$				
a	2.81548	0.63096	3.39377	0.63096
bias	0.03192	0.00462	0.20679	0.21525
std dev	0.09917	0.10725	1.09309	0.62566
mse	0.01085	0.01152	1.23760	0.43778
bias	0.14828	0.02145	0.96071	1.00000
std dev	0.15851	0.17143	1.74710	1.00000
mse	0.02479	0.02633	2.82701	1.00000
$\ell = 6$				
a	2.85984	0.71969	3.87101	0.71969
bias	0.04754	0.00685	0.48317	0.53778
std dev	0.16375	0.12788	2.23694	2.42992
mse	0.02907	0.01640	5.23737	6.19370
bias	0.08840	0.01273	0.89846	1.00000
std dev	0.06739	0.05263	0.92058	1.00000
mse	0.00469	0.00265	0.84560	1.00000

Table 8.1.c Results for Partly Linear Model: $y = \beta_0 + \beta_1 x + g(z) + \epsilon$

Replications = 100, N = 400, d = 4

	1.CV	2.OPT	3.SD	4.UNIT Z
$\ell = 2$				
a	2.73644	0.47287	2.58340	0.47287
bias	0.04233	0.00961	0.16785	0.16718
std dev	0.04555	0.08227	0.05623	0.05591
mse	0.00387	0.00686	0.03134	0.03108
bias	0.25322	0.05748	1.00400	1.00000
std dev	0.81471	1.47148	1.00566	1.00000
mse	0.12444	0.22079	1.00836	1.00000
$\ell = 4$				
a	3.10696	0.60696	3.31598	0.60696
bias	0.80166	0.01033	0.46334	1.13667
std dev	1.50886	0.09019	3.64520	7.57469
mse	2.91931	0.00824	13.50214	58.6680
bias	0.70527	0.00909	0.40763	1.00000
std dev	0.19920	0.01191	0.48123	1.00000
mse	0.04976	0.00014	0.23014	1.00000
$\ell = 6$				
a	3.18766	0.68766	3.75682	0.68766
bias	2.90043	0.01151	2.92993	3.25039
std dev	3.62257	0.09418	6.19412	8.09106
mse	21.5354	0.00900	46.9515	76.0302
bias	0.89233	0.00354	0.90141	1.00000
std dev	0.44773	0.01164	0.76555	1.00000
mse	0.28325	0.00012	0.61754	1.00000

Table 8.1.d: Results for Partly Linear Model: $y = \beta_0 + \beta_1 x + g(z) + \epsilon$

Replications = 100, N = 400, d = 5,

	1.CV	2.OPT	3.SD	4.UNIT Z
$\ell = 2$				
a	2.75695	0.51390	2.81250	0.51390
bias	0.06069	0.01849	0.23416	0.23251
std dev	0.06392	0.10527	0.08213	0.08214
mse	0.00777	0.01142	0.06158	0.06081
bias	0.26101	0.07952	1.00709	1.00000
std dev	0.77816	1.28154	0.99981	1.00000
mse	0.12776	0.18787	1.01262	1.00000
$\ell = 4$				
a	2.81536	0.63073	3.45185	0.63073
bias	0.05449	0.01376	0.63361	2.27120
std dev	0.07244	0.12174	9.62024	6.68173
mse	0.00822	0.01501	92.9505	49.8038
bias	0.02399	0.00606	0.27898	1.00000
std dev	0.01084	0.01822	1.43978	1.00000
mse	0.00016	0.00030	1.86633	1.00000
$\ell = 6$				
a	2.85149	0.70297	3.84723	0.70297
bias	0.06879	0.01171	3.84792	6.11724
std dev	0.52846	0.13025	8.05438	24.0566
mse	0.28400	0.01710	79.6794	616.142
bias	0.01124	0.00191	0.62903	1.00000
std dev	0.02197	0.00541	0.33481	1.00000
mse	0.00046	0.00003	0.12932	1.00000

ii) Errors–in–variables model

Consider the following model

$$y_i = \alpha + \beta\, E(x_i | z_i) + \eta_i \qquad i=1,2,\ldots,N$$

$$x_i = \Sigma_j\, \delta_j z_{ij}^{\;2} + \xi_i$$

$$\alpha = \beta = \delta = 1$$

η_i is NID(0,1), ξ_i NID(0,0.5), and z_i is NID(0,2.0). η_i and ξ_i are independent. The results are reported in Table 8.2a and Table 8.2b. We have reported the ratio of the estimates in the first three columns to that of the last. The estimates in the last column is the actual values and not ratios. In other words, for $\ell = 4$, we have the bias for the last column as –0.02706 and the bias for the CV criteria as ratio(first column) x bias(UNIT Z) = 0.13612 x–0.02706 = –0.00368.

From the results in Table 8.2, we can see that there is very little difference between the CV and the OPT criterion. However, the performance of the SD and UNIT criterion is not reliable at all. We must remind the readers again that the OPT criteria may not work in practice and there are no justifications for its use from the theory. The main point that we want to illustrate is the difference in performance of the different criteria in response to the use of rules–of–the thumb method in applied work.

Table 8.2.a Results for Error–In–Variable Model: y = α + βE[x|z] + η

Replications = 100, N = 400, d = 1,

	1.CV	2.OPT	3.SD	4.UNIT Z
$\ell = 2$				
	0.06608	0.30171	0.60245	0.30171
β coefficient				
bias	0.06711	0.24836	1.00000	–0.0523
std dev	0.81882	0.80134	1.00000	0.01178
mse	0.03666	0.08971	1.00000	0.00288
α coefficient				
bias	0.06495	0.26377	1.00000	0.47355
std dev	0.96972	0.86728	1.00000	0.07586
mse	0.02764	0.08665	1.00000	0.23001
$\ell = 4$				
a	0.14678	0.51390	1.02615	0.51390
β coefficient				
bias	–0.0797	–0.00264	1.00000	0.00874
std dev	0.16827	0.16183	1.00000	0.05659
mse	0.02780	0.02558	1.00000	0.00328
α coefficient				
bias	0.07831	–0.14216	1.00000	0.02268
std dev	0.11365	0.10987	1.00000	0.59921
mse	0.01291	0.01208	1.00000	0.35957
$\ell = 6$				
a	0.15928	0.63073	1.25942	0.63073
β coefficient				
bias	–0.0365	0.09044	1.00000	0.01123
std dev	0.11499	0.11309	1.00000	0.08117
mse	0.01300	0.01270	1.00000	0.00671
α coefficient				
bias	0.01860	0.11597	1.00000	–0.1178
std dev	0.14153	0.14260	1.00000	0.46840
mse	0.01886	0.01992	1.00000	0.23328

Table 8.2.b Results for Error–In–Variable Model: y = α + βE[x|z] + η

Replications = 100, N = 400, d = 2.

	1.CV	2.OPT	3.SD	4.UNIT Z
$\ell = 4$				
a	0.19981	0.54928	1.09688	0.54928
β coefficient				
bias	0.09383	–0.06496	0.98138	–0.0129
std dev	0.63944	0.63400	0.98638	0.01073
mse	0.17218	0.16663	0.96713	0.00028
α coefficient				
bias	0.03053	–0.10639	0.98082	0.18694
std dev	0.59493	0.57163	0.98941	0.13819
mse	0.12567	0.12278	0.96798	0.05404
$\ell = 6$				
a	0.17651	0.65184	1.30168	0.65184
β coefficient				
bias	–0.1283	0.48109	1.11402	0.00802
std dev	0.28118	0.26230	0.90678	0.02661
mse	0.07385	0.08233	0.85708	0.00077
α coefficient				
bias	0.00634	0.44804	1.08308	–0.1457
std dev	0.24567	0.24760	0.92420	0.33902
mse	0.05095	0.08305	0.90390	0.13617

(iii) Generalized least squares model

$$y_i = \alpha + \beta x_i + \epsilon_i \sigma(x_i) \qquad i=1,\dots,N$$

with $E[\epsilon_i | z_i] = 0$. This model is perhaps the most studied semiparametric models in terms of simulations. Therefore, the design is identical to many previous studies on the GLS estimators (e.g. Cragg (1983), Delgado (1987, 1988)). We have $x_i = \exp(z_i)$, z_i is NID(0,1) and ϵ_i is NID(0,1) and both ϵ_i and z_i are independent. Besides reporting the results for criterion CV, OPT (identical to UNIT Z as d=1) and SD, we have also reported the results for OLS and Aitken's GLS estimator (where the weights are the true variance). In the following tables, the results in the first four columns are ratios to the last GLS. The GLS results are the the actual values.

The following variance model is considered:

$$\sigma^2(x_i) = | \alpha_1 + \alpha_2 x_i + \alpha_3 x_i^2 |^\gamma \qquad i=1,2,\dots,N.$$

The results reported below are all replicated using $\alpha_1 = 1.0$, $\alpha_2 = 0.3$ and $\alpha_3 = 0.5$ with γ taking values between 0.5 and 3.

Table 8.3.a Results for GLS Model: $y_i = \alpha + \beta x_i + \epsilon_i \sigma_i$

Replications = 100, N = 300, d = 1

$\alpha_1 = 1.0$, $\alpha_2 = 0.3$, $\alpha_3 = 0.5$;

	CV	OPT	SD	OLS	GLS
$\gamma = 1.0$					
a	0.10848	0.31958	0.72184		
bias	−11.325	−14.689	−12.404	1.04390	0.01817
std dev	1.98776	2.02622	2.07766	1.54117	0.19238
mse	5.05050	5.97740	5.63906	2.36383	0.03734
bias	−8.6793	−9.5362	−9.4887	1.05945	−0.0201
std dev	1.66812	1.59662	1.74115	1.91111	0.21262
mse	3.43032	3.33834	3.80835	3.62977	0.04561
$\gamma = 2.0$					
a	0.12596	0.31958	4.46143		
bias	−1.9513	−4.2348	2.89794	7.69189	0.02597
std dev	1.06021	1.04478	2.28089	9.50548	0.35698
mse	1.13819	1.18026	5.21931	90.1899	0.12811
bias	−1.1749	−2.2646	2.90373	11.3529	−0.0246
std dev	1.06662	1.01650	2.67573	16.9378	0.29824
mse	1.13933	1.06108	7.16819	285.816	0.08955
$\gamma = 2.5$					
a	0.20048	0.31958	14.4231		
bias	0.00767	−2.3851	5.71154	33.28519	0.02269
std dev	1.18483	0.95703	20.2228	29.26604	0.42050
mse	1.39975	0.92975	407.871	857.2309	0.17734
bias	0.54312	−0.70889	6.16638	49.63785	−0.0227
std dev	1.25942	0.95958	37.4914	57.80669	0.32302
mse	1.57978	0.91873	1398.86	3337.281	0.10486
$\gamma = 3.0$					
a	0.05522	0.31958	50.14846		
bias	4.89216	−1.61716	153.0128	193.0087	0.01532
std dev	1.45632	0.93527	92.22610	95.71482	0.47431
mse	2.14358	0.87654	8521.186	9190.596	0.22520
bias	4.03927	0.15555	181.8416	237.1477	−0.0192
std dev	1.62472	0.94945	195.3171	204.0701	0.34083
mse	2.68319	0.89867	38132.63	41691.01	0.11653

Table 8.3.b Results for GLS Model: $y_i = \alpha + \beta x_i + \epsilon_i \sigma_i$

Replications = 100, N = 400, d = 1

$\alpha_1 = 1.0$, $\alpha_2 = 0.3$, $\alpha_3 = 0.5$;

	CV	OPT	SD	OLS	GLS
$\gamma = 0.5$					
a	0.34292	0.30171	0.40557		
bias	0.51533	–1.3013	–0.5652	0.85725	0.00910
std dev	1.35891	1.49582	1.52649	0.94047	0.10757
mse	1.83541	2.23362	2.31589	0.88342	0.01165
bias	0.70990	–0.1139	0.17492	0.74173	–0.0089
std dev	1.23184	1.31481	1.31402	0.98032	0.14718
mse	1.51368	1.72239	1.72038	0.95951	0.02174
$\gamma = 1.0$					
a	0.05828	0.30171	0.66277		
bias	–12.370	–15.683	–13.392	1.25344	0.01547
std dev	2.01561	2.03370	2.08446	1.45617	0.18607
mse	5.08518	5.79595	5.54621	2.11666	0.03486
bias	–11.793	–13.163	–13.039	1.38806	–0.0135
std dev	1.69555	1.62920	1.76357	1.78595	0.20317
mse	3.48075	3.41315	3.85253	3.18401	0.04146
$\gamma = 1.5$					
a	0.08942	0.30171	1.41565		
bias	–5.1837	–8.1719	–3.1800	2.84674	0.02004
std dev	1.20796	1.22422	1.19039	3.20561	0.26644
mse	1.60208	1.86587	1.46594	10.2637	0.07139
bias	–4.9522	–6.6979	–3.6870	4.16701	–0.0159
std dev	1.18120	1.15145	1.21860	4.79149	0.24935
mse	1.48958	1.50342	1.53438	22.9356	0.06243
$\gamma = 2.0$					
a	0.00800	0.30171	3.84941		
bias	–2.1183	–4.7617	3.33550	9.25769	0.02065
std dev	0.98846	1.04068	2.13306	8.77247	0.33575
mse	0.99029	1.16440	4.57473	76.9892	0.11316
bias	–1.8955	–3.5151	3.85717	16.2288	–0.01566
std dev	1.00469	1.01048	2.39801	15.1404	0.28251
mse	1.01732	1.05577	5.77838	229.338	0.08005
$\gamma = 2.5$					
a	0.00000	0.30171	12.2281		
bias	–0.4942	–2.8691	7.65564	40.92894	0.01702
std dev	1.07382	0.94772	16.5291	27.10506	0.39141
mse	1.15137	0.91201	272.806	736.4586	0.15349
bias	–0.2347	–1.5196	10.0253	75.35350	–0.0132
std dev	1.11222	0.94668	28.9602	51.54245	0.30498
mse	1.23482	0.89885	837.307	2662.302	0.09319

We can see that the CV criteria is the best criteria based on the results. However, the OPT criteria seems to be better than that of the CV and indeed GLS when γ is greater than or equal to 2.5. We see that the OLS and SD estimates deteriorate rapidly with γ.

We have also plotted the bias, std dev and mse for the various ratios in Figure 8.1 to 8.9 for N in between 80 and 180. R1, R2 and R3 are the ratios of CV to UNIT, OPT and GLS respectively. Look at the ratios for bias, CV is always better than UNIT and OPT criteria. As for standard deviation, CV is only better than UNIT criterion when N is greater than 160, and CV is always better than OPT. As for MSE, marked improvement over UNIT and OPT is registered only after N = 160. It is also true that the ratio of CV to GLS is never better than 3.6605 but the performance of CV becomes better when N is greater than 120.

Figure 8.1 Plot of Ratio of Bias of CV Estimates to UNIT for Model 3

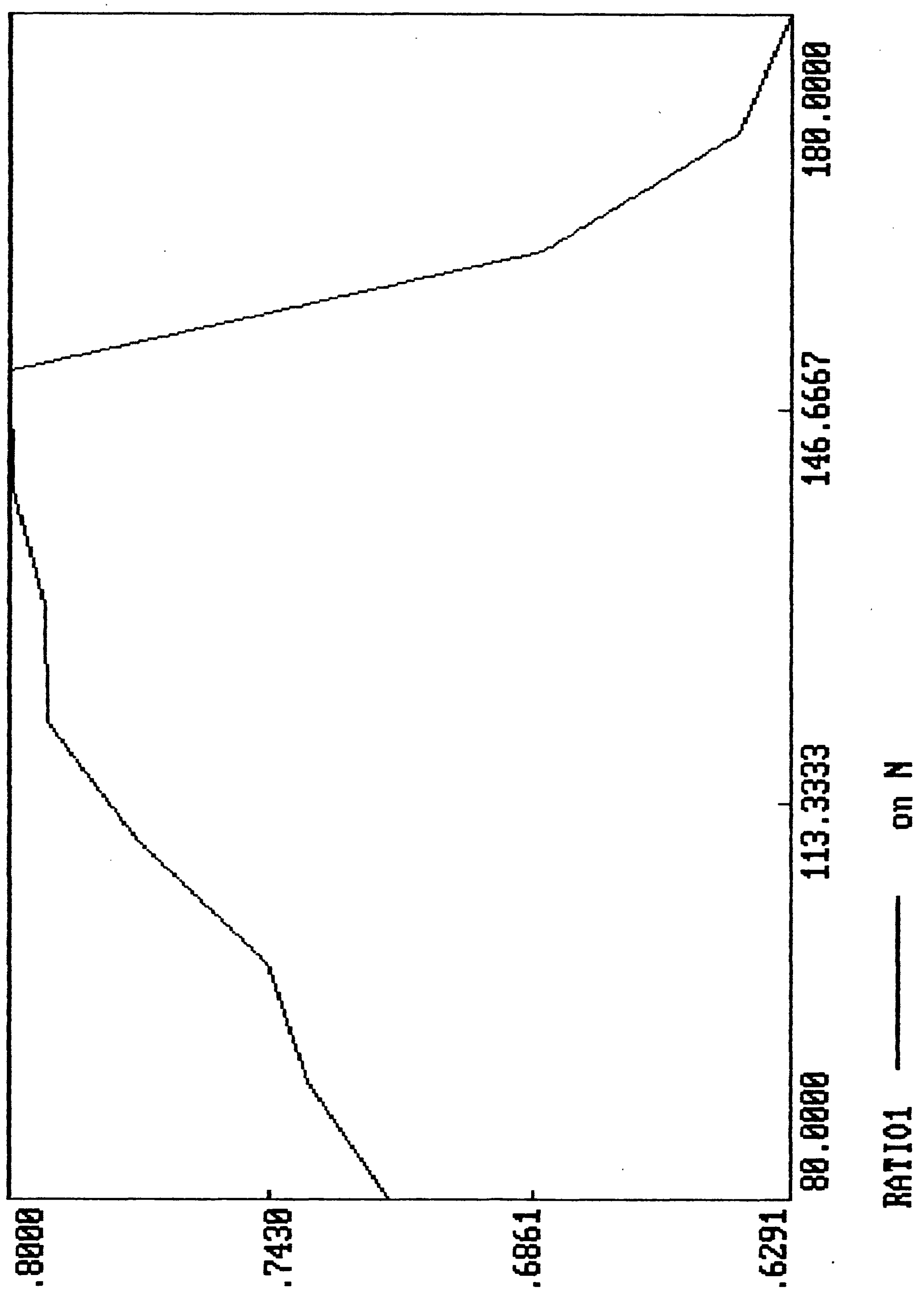


Figure 8.2 Plot of Ratio of Bias of CV Estimates to OPT for Model 3

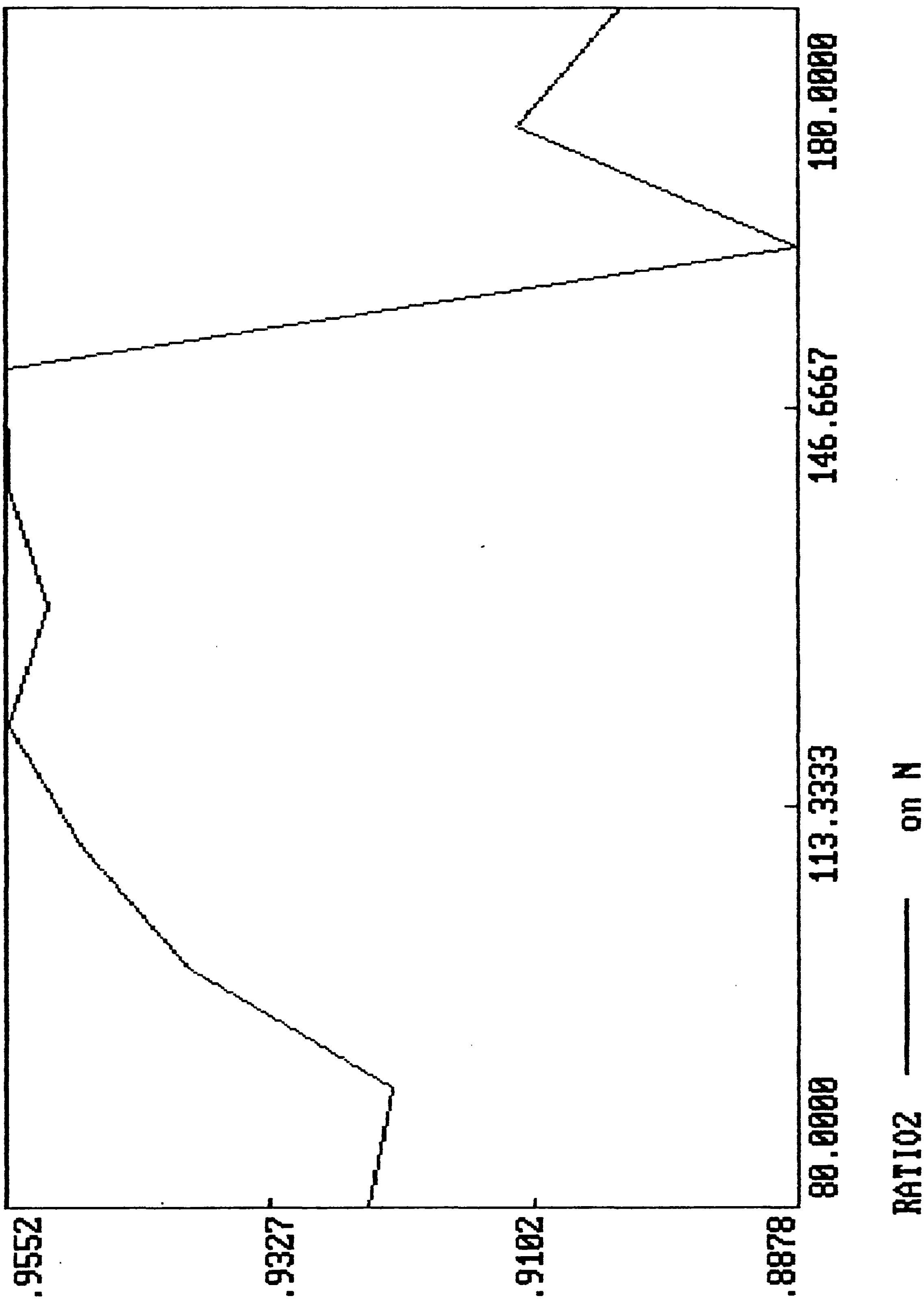


Figure 8.3 Plot of Ratio of Bias of CV Estimates to GLS for Model 3

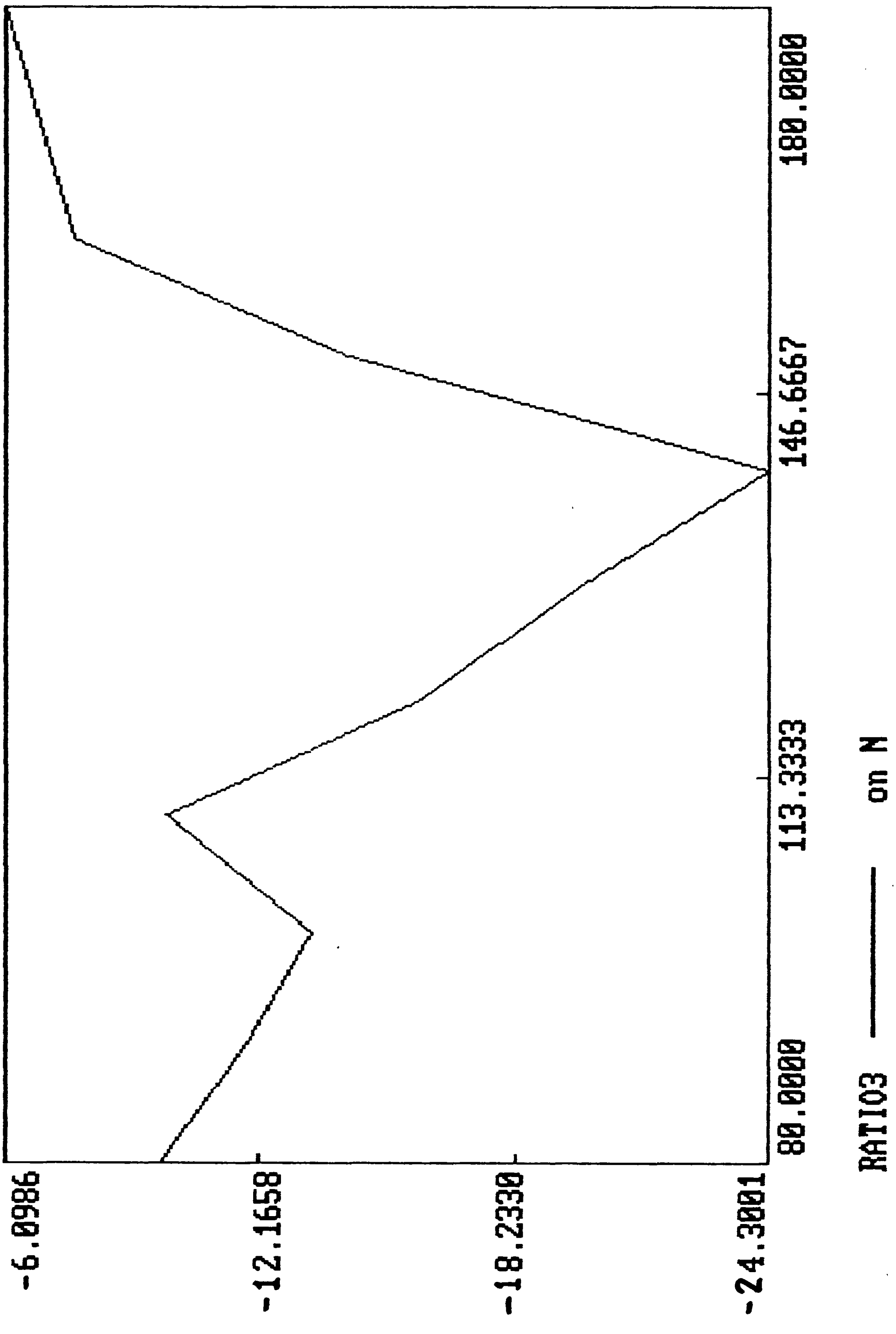


Figure 8.4 Plot of Ratio of Standard Deviation of CV Estimates to UNIT for Model 3

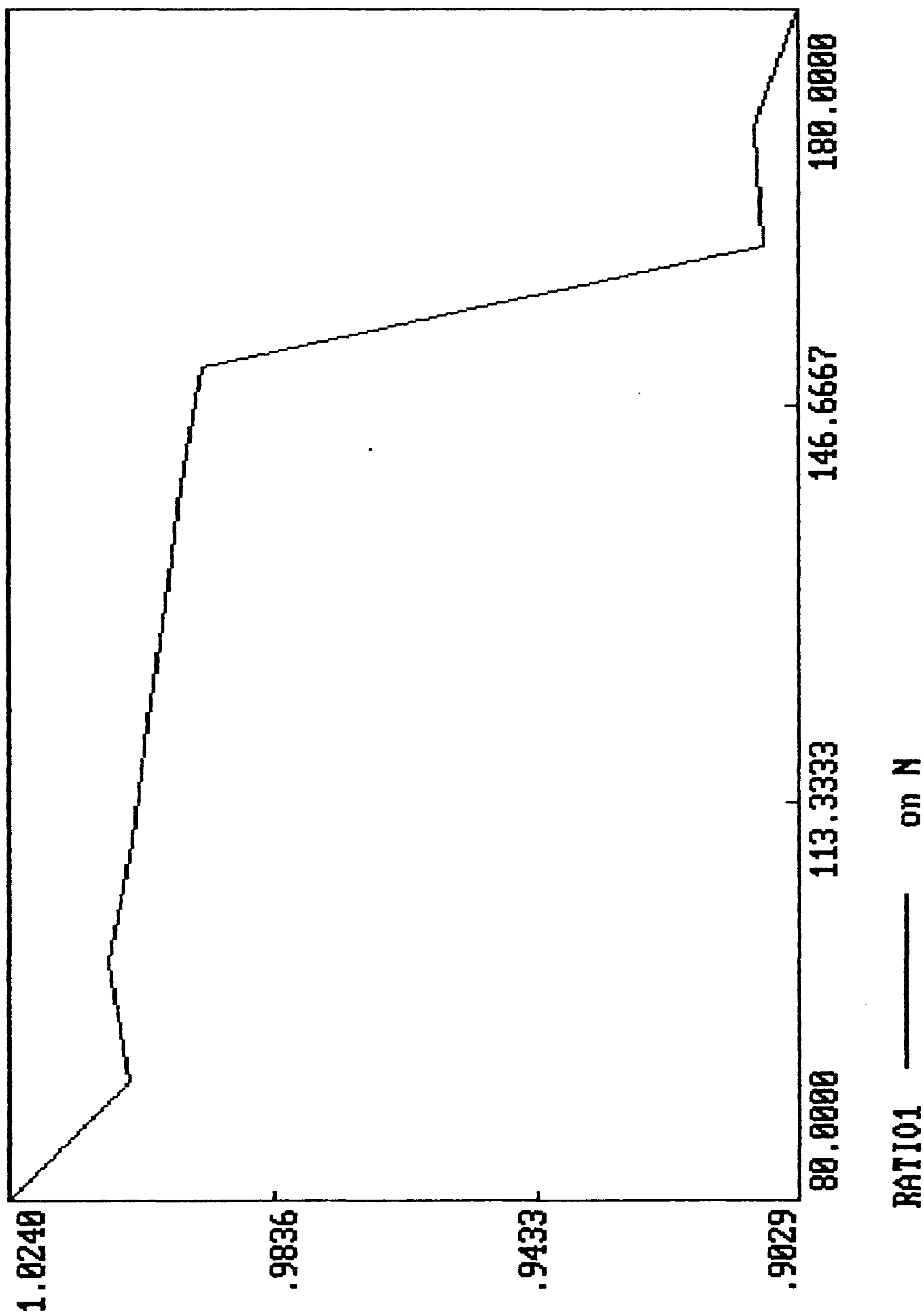


Figure 8.5 Plot of Ratio of Standard Deviation of CV Estimates to OPT for Model 3

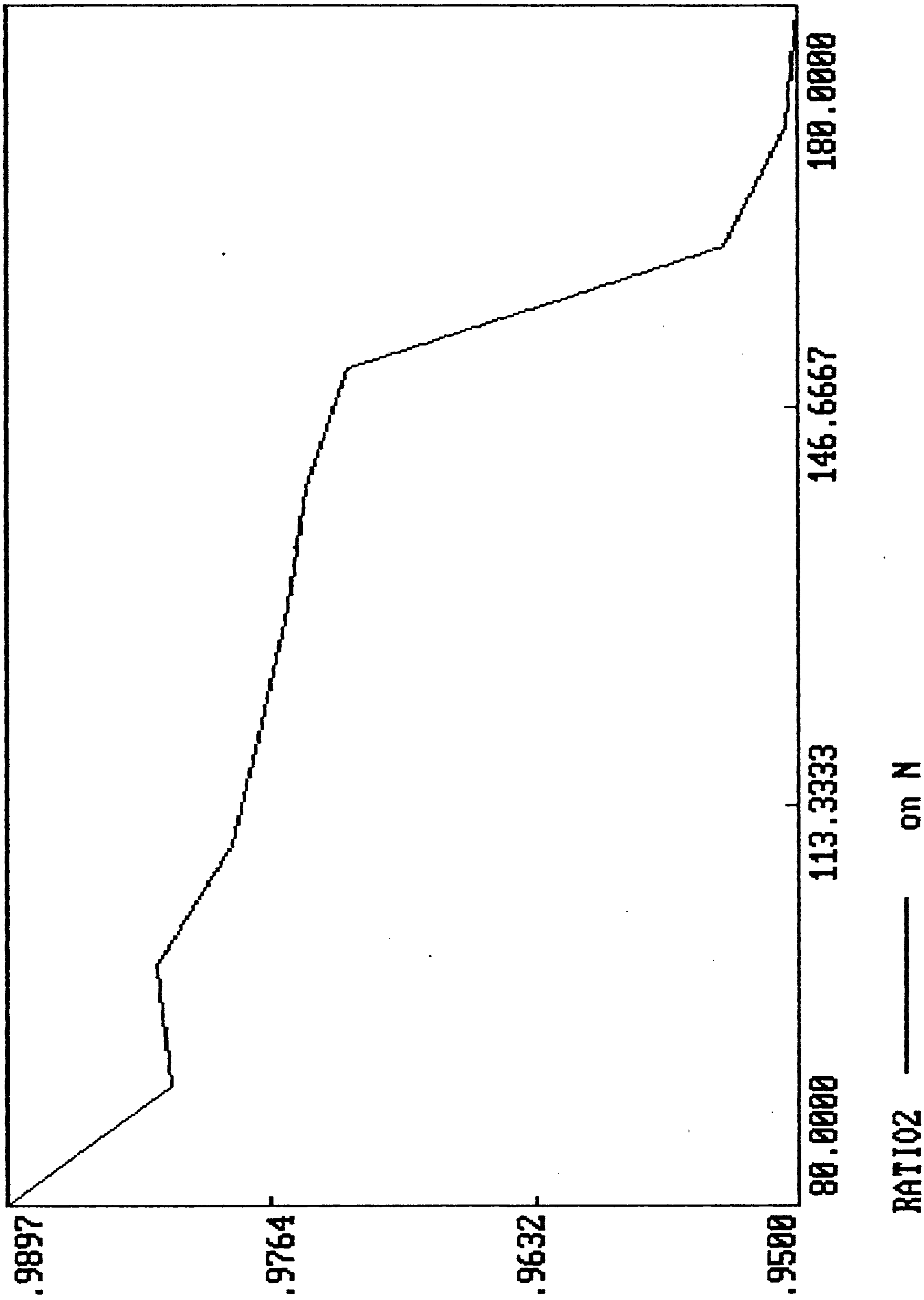


Figure 8.6 Plot of Ratio of Standard Deviation of CV Estimates to GLS for Model 3

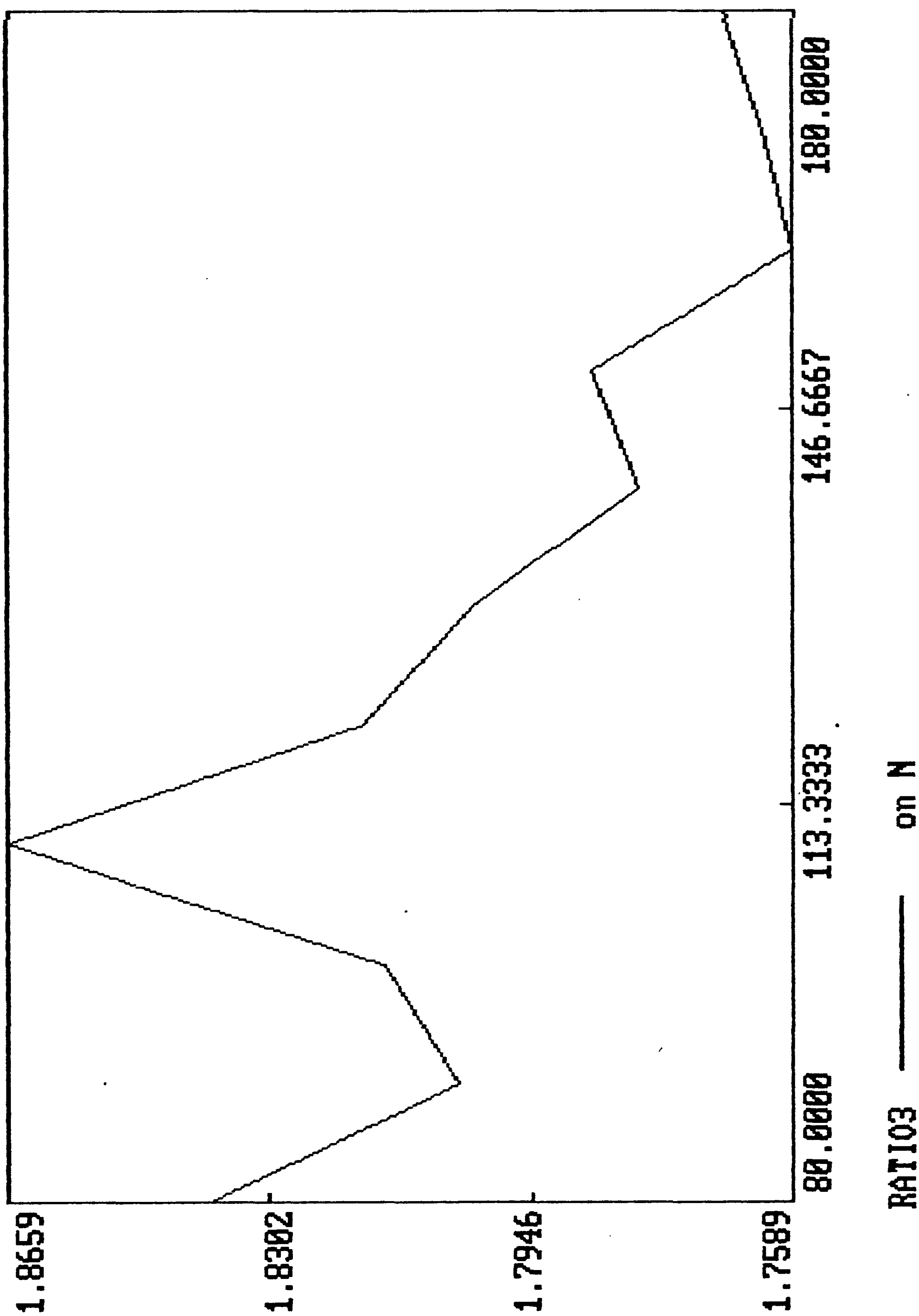


Figure 8.7 Plot of Ratio of MSE of CV Estimates to UNIT for Model 3

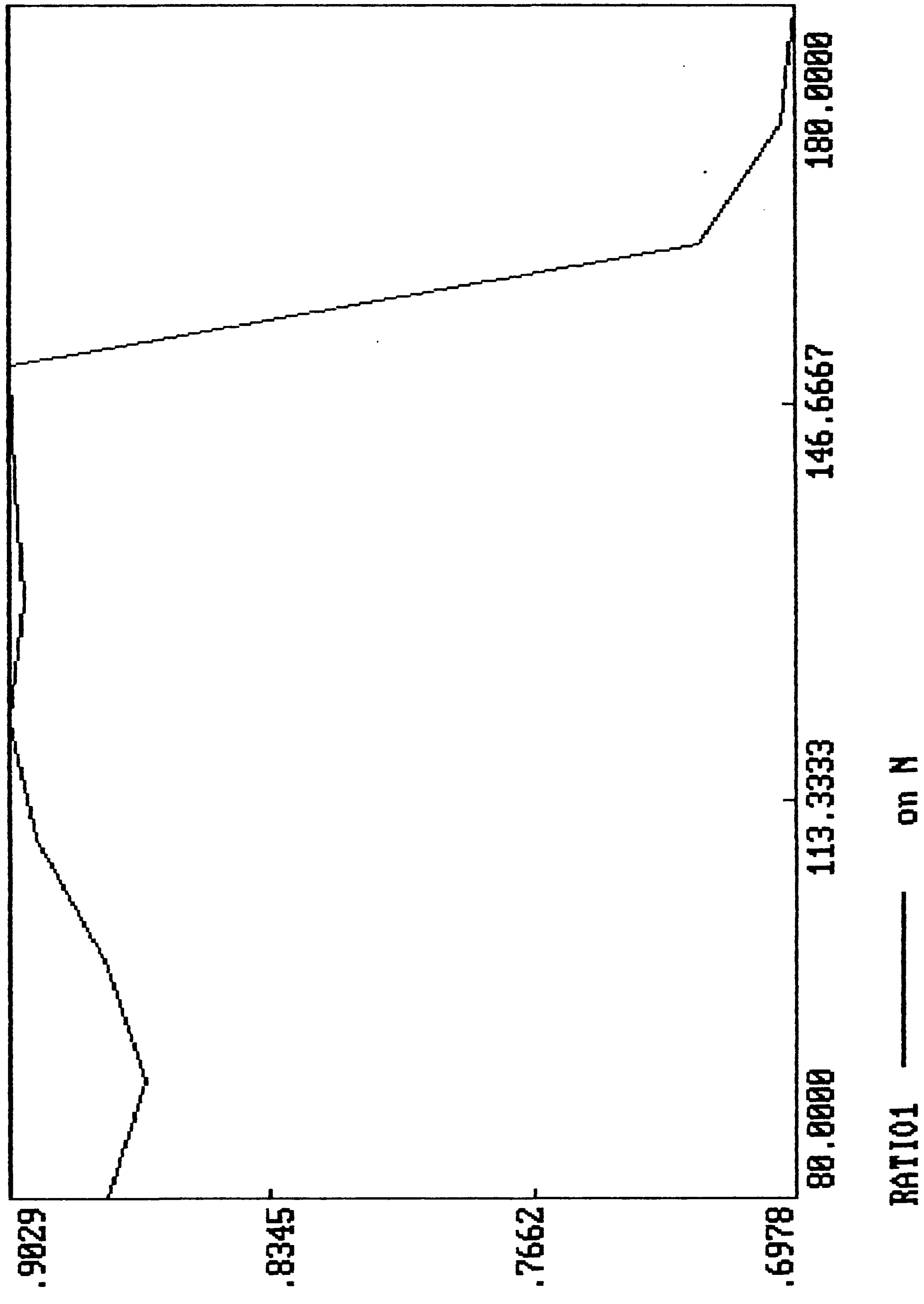


Figure 8.8 Plot of Ratio of MSE of CV Estimates to OPT for Model 3

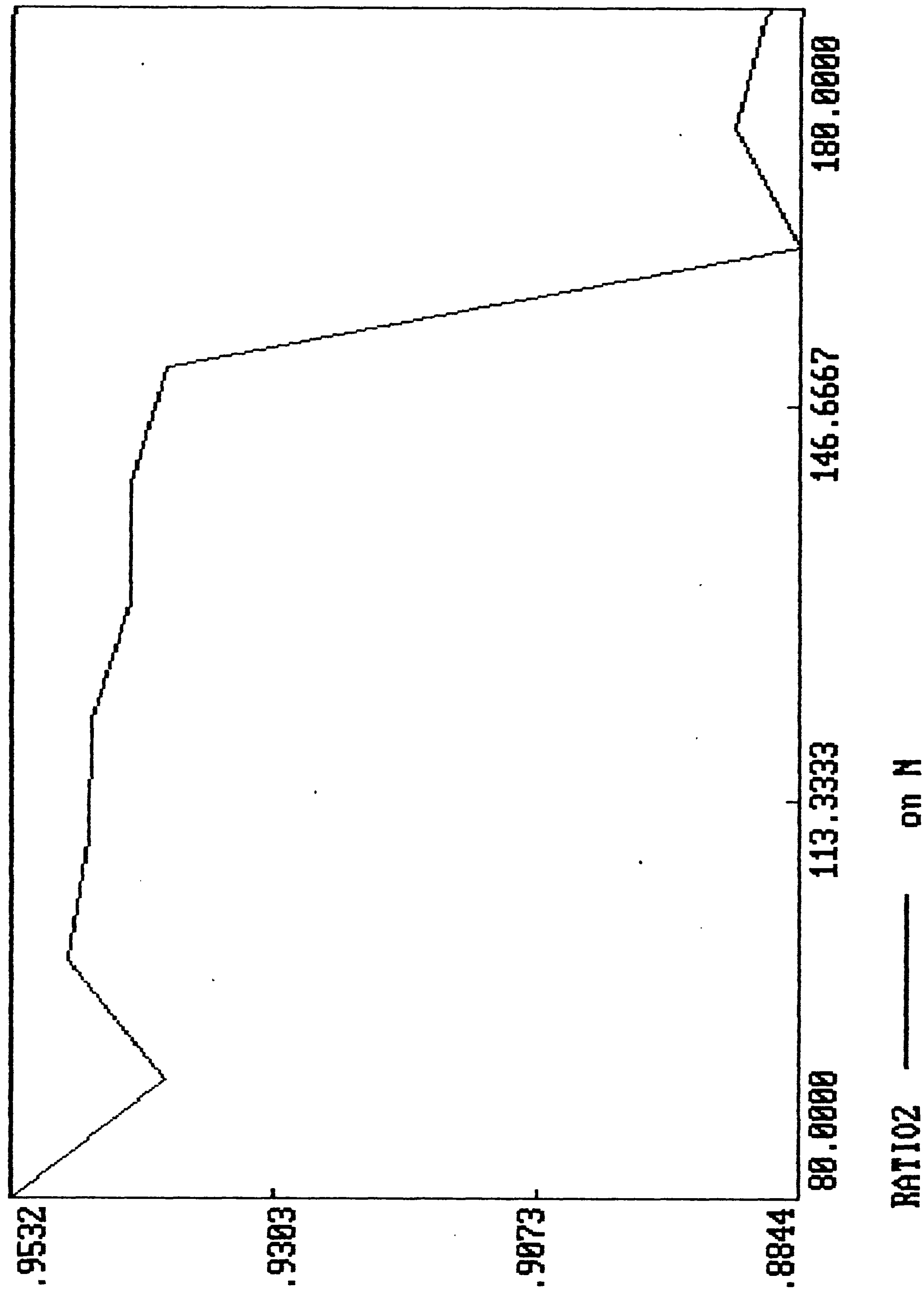
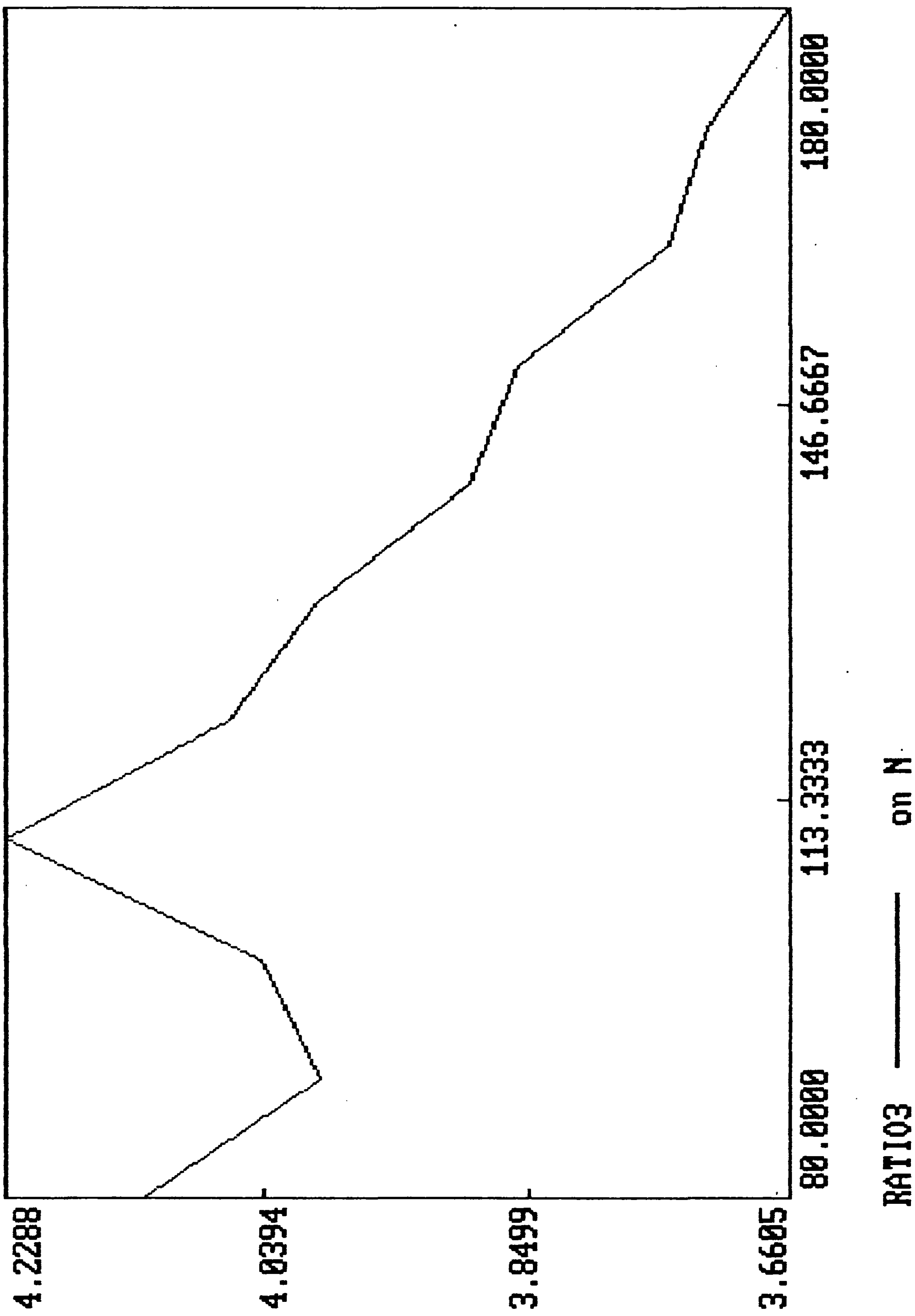


Figure 8.9 Plot of Ratio of MSE of CV Estimates to GLS for Model 3



CHAPTER 9

SOME OPEN PROBLEMS

In this concluding chapter, we briefly list some general problems related to the techniques used as well as specific problems in relation to our empirical work, to be investigated and suggest some possible course of action for future research.

GENERAL PROBLEMS

(1) Bandwidth Selection

The issue of bandwidth selection is perhaps the most important of all the issues that we have raised in this thesis. First of all, the asymptotic results of the semiparametric estimators with data–dependent bandwidth need to be worked out. In particular, we have to work out the regularity conditions under which the "automatic" bandwidth is consistent and possibly the rate of convergence.

Secondly, the performance of the different criteria needs to be compared in order to select a satisfactory criterion for practical applications with finite sample. In particular, most of the automatic bandwidth criteria are computationally expensive. One solution may be to split the sample into subsample and minimize each subsample criterion function with respect to the bandwidth a. An average of these subsample automatic bandwidths is used for the semiparametric estimator. Although this approach may have computational advantage when N is large, it introduces another source of problem discussed below.

There may also be reasons to believe that some of these criteria are subjected to sample noise. In order to overcome this across sample variability, partitioning the sample into subsample may also be desirable. Recently, Marron (1988) has suggested partitioned cross–validation for nonparametric curve estimation. The procedure involves working out the value of the CV function for each subsample and then minimizing the average of these subsample CV values.

Of course, this introduces another source of problem, i.e., the choice of the number of subsamples, m. We have a similar dilemma as that of bandwidth selection: having too larger a m induces large bias and having too small a m induces large variance.

Define $\hat{a}_{CV}$ as the minimizer of the CV criterion and a_{MISE} as the minimizer of the Mean Integrated Squared Error, i.e., MISE = E[ISE] = $E(\int[\hat{f}(x)-f(x)]^2dx)$. Hall and Marron (1987) have shown that the rate of convergence of $\hat{a}_{CV}$ to a_{MISE} is fairly slow and subjected to a good deal of sample noise. In particular, if the density function f is no more than twice differentiable and under some other regularity conditions,

$$N^{-1/10}(\hat{a}_{CV}-a_{MISE})/a_{MISE} \sim N(0,\sigma_1^2) \qquad (1)$$

Define the partitioned CV bandwidth as $\hat{a}_{PCV}=m^{-1/5}\hat{a}^*_{CV}$ where $\hat{a}^*_{CV}$ is the a which minimizes $CV^* = m^{-1}\Sigma_{j=1}^{m} CV_j$. Then under an appropriate choice of m and regularity conditions,

$$N^{-1/4}(\hat{a}_{PCV}-a_{MISE})/a_{MISE} \sim N(0,\sigma_2^2)$$

which is an improvement over (1). This idea of partitioning the sample is also very useful for time–series problem and semiparametric models. It appears that if we can partition the serially correlated data sample so that these subsamples are independent, then we can simultaneously reduce across sample variability and overcome the potential problems of serial correlation in the sample.

(2) Nonparametric smoothing and Dimension Reduction

Another pressing issue is the problem of the curse of dimensionality in nonparametric smoothing. There are suggestions which can be employed in future empirical work.

We have only used the kernel estimators in our empirical work. However, this is not the only nonparametric smoother that one can use. Two other favoured competitors are the nearest neighbour and spline smoothers. While spline smoothing has been applied in various areas, the method of nearest neighbour is relatively new in economics. The first semiparametric model which combines the use of nearest neighbour nonparametric component and parametric component is the GLS estimator of Robinson (1988). A nearest neighbour estimators for $E[y_i|x_i]$ can be expressed as:

$$\hat{E}[y_i|x_i] = \Sigma_{j=1}^{N} y_i W_{ij}$$

where $W_{ij} = I(i\neq j)q_{ij}^{-1} \Sigma_{\ell=P_{ij}}^{p_{ij}+q_{ij}-1} c_\ell + I(i=j)\, 0$

$$P_{ij} = 1 + \Sigma_\ell I(r_{\ell j} < r_{ij}) \qquad i\neq j$$

$$q_{ij} = 1 + \Sigma_\ell I(r_{\ell j} = r_{ij}) \qquad i\neq j$$

$$r_{ij} = \Sigma_{m=1}^{d} \frac{(x_{im}-x_{jm})^2}{s_m} \qquad i,j = 1,\dots,N,\ i\neq j.$$

and $s_m = (N-1)^{-1}\Sigma_i(x_{im}-(\Sigma_i x_{im}/N))^2, \qquad 1 \leq m \leq d.$

We can see that nearest neighbour estimator is also a linear estimator. The important differences between kernel and nearest neighbour estimators are that (i) we are averaging over a same k number of nearest neighbour rather than over a fixed size of neighbourhood as in kernel estimation; (ii) the x_i's are of unit variance in this nearest neighbour setting. The first will correspond to using a kernel with different bandwidth and the second will overcome the problem of having different unit of measurement for different explanatory variables in economic applications. For a given constant $k < N$, the weight c_ℓ is a non–negative constant chosen by the practitioner if $1 \leq \ell \leq k$ and equal to zero if $\ell > k$. Of course, this weight is equal to one if it is summed from 1 to k. Popular choice of weights are Uniform, Triangular and Quadratic.

As for the choice of dimensional reduction technique, we have not worked out the asymptotic properties in our case. In fact, the asymptotic properties of many additive models have not been worked out and is a current area of research. There is very few useful results for projection pursuit regression which is very closely related to the approach that we have suggested.

(3) Transformation models

We have suggested parametric transformation models. However, it may also be of interest to look at some of the economic relationships that we have investigated using the alternating conditional expectations (ACE) of Breiman and Friedman (1985). Their procedures are useful for preliminary data analysis when one is seeking "optimal" transformations. In particular, the semiparametric model

$$T_y(y_i) = \Sigma_j T_j(x_{ij}) + \epsilon_i$$

which seeks the optimal transformation for T_y and T_j may be use as a preliminary step for subsequent parametric transformation. The procedures involve minimizing the objective function

$$E[\epsilon_i^2]/E[T_y(y_i)]^2$$

subject to the constraint of zero expectations:

$$E[T_y(y_i)] = E[T_1(x_1)] = ... = E[T_d(x_d)] = 0.$$

There is no doubt that by plotting graphs, this method will lead to better understanding of the nonlinear relationships between the dependent and independent variables and it is possible to detect any nonlinear singularity of the data as in the case of the fuel efficiency and rental cost function.

SPECIFIC PROBLEMS

We outline some further specific problems for Chapters 3 to 8 to be investigated in future empirical studies.

CHAPTER 3

In Chapter 3, we have constructed an intertemporal model with the assumption that the miles driven, M, is exogenously given. It is plausible that both E(z) and z will affect M, and thus M = M(E(z),z). If data on individual motorists can be collected, then a distance travel function can also be estimated. Such a function may perhaps be of interest to policy analysts, e.g., in relation to electronic road pricing.

The constructed model does not take into account technological improvement.

It is also plausible that fuel efficiency improvement can be brought about by technical factors independent of the attributes. While it is easy to include such a variable in the model, it is not clear how we can proceed with the estimation.

We have not provided the standard errors of the elasticity estimates. It is worthwhile considering the options of bootstrapping or computing the standard errors conditional on some of the parameters. However, these methods are also likely to be computationally very expensive.

Although our work has concentrated on the hedonic approach, it may be fruitful to proceed along other lines of investigation. One such approach is to use of survey methods (see e.g. Goodman (1989)). Many studies in housing economics have addressed the issue.

CHAPTER 4

Despite some of the criticisms of the two–stage method of Rosen (1974), it is still worthwhile to compare the semiparametric and nonparametric estimates with the parametric estimates obtained under appropriate instruments. It is still worthwhile to estimate the benefits using the TS method if individual household data from multiple markets are available,.

Besides the usual problem of bandwidth selection, one may want to conduct hypothesis testing to select appropriate variables to be included in the semiparametric hedonic price function. This can be done by, say, Robinson's test statistics.

CHAPTER 6

While the relaxation of the distribution assumptions has been adopted in some studies, the relationship between consumption growth and real interest rates should be investigated further along the line of Harvey (1988). It would be desirable to confirm the results that there are relationship between these two macroeconomic variables using data on returns of more than a quarter.

It is of also of interest to work out the regularity conditions under which the SARCH estimates are asymptotically efficient. Furthermore, the asymptotic properties of the automatic bandwidth need to be investigated.

CHAPTER 7

In Chapter 7, we have investigated the statistical relationship between macroeconomic variables and consumption. The most important task is to suggest a economic framework which is consistent with the data. In particular, we have to explain why inflation and stock prices should matter. The explanatory power and the magnitude of lagged consumption and income variables pose some further questions from the theoretical point of view.

CHAPTER 8

There are two very important observations from Chapter 8. First of all, while it is relatively cheap to compute some of the semiparametric estimators, it is computationally expensive to conduct cross–validation or other automatic methods for bandwidth selection in multidimensional problems. This is clearly reflected in the number of replications that we have. A search for computationally efficient algorithms will be important for future multivariate semiparametric applications. Secondly, the simulation results using finite sample produced some insights into the interesting behaviour of the automatic semiparametric estimators. While there are few asymptotic results on the automatic semiparametric estimators, it is obvious that further investigations are fruitful. This lack of asymptotic results also suggests that investigation into the finite properties of the semiparametirc estimators will no doubt be more difficult but highly desirable.

DATA APPENDICES

Appendix A: Cross Sectional Data

The same set of cross sectional data is used in Chapter 3 and 4. There are various known sources for the required data but the one which has been widely used in empirical studies and considered reliable is collected by Haymarket Magazines Limited (reported in WHAT CAR?). The number of attributes that can be included in our model is enormous. Many of these variables are highly collinear. However, there is some consensus of what variables should be included in the model which are the main determinants of consumer choice and fuel efficiency among the models and makes. Basically, we can classify these attributes into three classes, namely, performance, specification and dimension variables. The attributes that we used are general representations of these three classes and are constructed from these variables to be consistent with the more recent studies on car efficiency in the United States. The variables used in the study are reported below and can be found in WHAT CAR? unless otherwise stated.

(a) Price, P

The actual prices used are the list price of new cars inclusive of car tax, Value Added Tax and delivery costs. Standard equipment is included in the list price. Ideally, one would prefer a measure of the price which excludes these standard equipment but lack of data prevent us from doing so. There have been attempts by other authors to subtract these costs from the prices by treating standard equipment in different makes at a fixed price. This may not be a good practice because it is unlikely that prices of these standard equipment are the same across different models, and one would end up with severe measurement errors. The alternative approach is to obtain the information from the dealers. We left the prices as they were. This is measured in UK sterling and the prices are given in June 1988 issue.

(b) Size, z_1

This is the engine capacity measured in cubic centimetres.

(c) Spaciousness, z_2

We adopt the usual approach of using the the product of length, width and height as a proxy for spaciousness. It is not intended to be a variable to capture the aesthetic qualities which will distinguish one make of car from another. This is the usual constraint that empiricalist has to face but it is believed that this variable will be a good proxy.

(d) Power, z_3

This variable is constructed by dividing the brake horse power by the kerb weight and is generally taken to reflect the performance of the car.

(e) Acceleration, z_4

This is the inverse of the timing in seconds from 0–60 miles per hour.

(f) Fuel Efficiency, E

Fuel consumption is measured by miles per gallon (MPG). Three figures are given for travelling in urban, at 56 miles per hour and at 75 miles per hour respectively. We adopt the approach of taking the weighted average of the three consumption figures shown for each model on the following basis: 2xUrban figure + 56mph figure + 75mph figure.

(g) Fuel, D_U

This is a dummy variable for car which can run on unleaded fuel without adjustments. It takes the value of 1 if unleaded and 0 otherwise.

(h) Running Cost Per Mile, R

This figure is calculated for three years over three thousands miles and quoted in pence by Emmerson Hill Associates. The original observations include fuel consumption, servicing cost, insurance payment and other funding and maintenance costs. With the information provided by Emmerson Hill Associates, we are able to exclude fuel consumption from the data.

Given in more details, the figure includes insurance, based on the average premium for the grouping of the model; depreciation over 3 years, based on past performance through Auction houses (CAP Black Book); the cost of financing the purchase of the vehicle (a flat rate of 10% per annum of the purchase price) allowing for discount where available. The cost of servicing, repair and maintenance is included in the operating cost constructed according to Manufacturers' frequency and standard, and the replacement of parts such as tyres, brake and clutch linings.

(i) market shares

We use the information provided by the Consumer Association in Car Buyer's Guide for June 1988. This variable is the firm's share of the market.

(j) British made, D_B

This is a dummy variable which takes the value of 1 if the car is manufactured in the Great Britain and zero otherwise.

(k) Luxury car, D_L

This is a dummy variable which takes the value of 1 if there are more than 6 extra features included in the standard equipment lists. The features refer to air conditioning, central locking, manual sunroof, electric sunroof, electric front windows, four electric windows, power seats, cruise control, headlamp cleaning features, electric mirrors, trip computer, split/fold rear seats, pre radio kit, radio, radio/cassette, five speed, power steering and ABS brakes.

(l) Speciality car, D_S

This is also a dummy variable which takes the value of 1 if the rating of insurance group exceeds 8. The ratings are according to the Association of British Insurers.

(m) interest rate, r

Treasury bill rate at June 1988 taken from Table 39, pp. 68, Economic trends, CSO, number 423, Jan. 89, HMSO. We have also used an r close to the market rate used by credit card companies.

(n) Miles travelled, M

This is converted to miles from 14,000 kilometers for cars and Taxis in 1987. It is taken from Table 2.5, pp. 69, Transport Statistics of Great Britain, 1977–87, Department of Transport.

(o) Price of Petrol, P , and growth rate, f

Price per gallon of 4 star petrol includes excise duties and tax. Petrol prices has been growing at a slow rate since 1980 after the second oil shock. There is also indication of price deflation in recent years. f is assumed to be 10%. We have not taken into account the price differential of leaded and unleaded petrol which is negligible during the preparation of this chapter. Table 1.28, pp. 46, Transport Statistics of Great Britain, 1977–1987, Department of Transport.

(p) Life of car, T

We have tried to evaluate the elasticities using different planning horizons. One of them is the average life time of a car. This is taken to be 13 by observing the age distribution of stock and vehicle survival rate for cars in all taxation classes in 1987. While there are 45.5% of the cars of age between 10–12 years survived, only 24.8% survived in the age group 12–14. Therefore, it is reasonable to assume that the average length of the life is around 13. Source: Table 2.25, Transport Statistics of Great Britain, 1977–1987, Department of Transport.

Our data consist of 399 observations which are only a subset of those reported in What Car?. This is because we have to exclude those models which do not have complete observations on all the variables. We have also excluded diesel engine models as well as some minor variations of the same model.

APPENDIX B: TIME SERIES DATA

The data set used in Chapter 6 is described in the text. This appendix describes only the data used in Chapter 7. The U.S. seasonally adjusted quarterly data from the period 1949Q2 to 1979Q4 can be obtained from the Survey of Current Business and the 13 raw data are taken from Bean (1986). I am grateful to Charles Bean for supplying the set of raw data. Attempts have been made to transform the data in the same way as Bean did, so that some comparisons can be made with the estimates obtained there. There are 138 observations in total, but we will use only 133 observations for estimation. The transformed data are described in the data appendix of Bean.

The expenditure on non–durables and services plus services are calculated at 1972 prices. To compute the services from durables which is taken to be equal to the value of exponential depreciation, we have to make assumptions on the depreciation rate. This is assumed to be constant at 4% per quarter. The benchmark value of stock at the beginning of 1948 is 726.7 bn. dollars at 1972 prices. The series of consumption measure is divided by the population size to get the consumption expenditure per capita. The non–property income per capita (Y) is defined as

$$Y = (\text{Personal disposable–rent–dividends–interest rate}) \times \text{retention ratio}.$$

The 3 month treasury bill rate (÷400) is the rate at the end of the quarter. Rate of Inflation (I) is calculated as $I_t = 100x(P_t/P_{t-1})^4-100$. Ex–post Real Rate of Interest (r) is nominal rate net of tax deflated by inflation. Standard and Poor's Corporation 500 index (S) is the index at end of the quarter deflated by P_t. Hours (H) is the aggregate hours of wages and salary workers in non–agriculture establishment per capita. The Government expenditure (G) is state and local spending on goods and services per capita at 1972 prices. We have also excluded Federal expenditures which are largely defense expenditures, and as Bean has argued that they are unlikely to be a substitute of private consumption.

REFERENCES

Allen, D.M. (1974): "The Relationship Between Variable Selection and Data Augmentation and A Method of Prediction", Technometrics 16, 125–127.

Amemiya, T. (1974) : "The Nonlinear Two–Stage Least–Squares Estimators", Journal of Econometrics, 2, 105–110.

Amemiya, T. (1977): "The Maximum Likelihood and the Nonlinear Three–Stage Least Squares Estimator in the General Nonlinear Simultaneous Equation Model", Econometrica, 45, 955–968.

Amemiya, T. (1985): Advance Econometrics, Oxford: Basil Blackwell.

Amemiya, T. and J.L. Powell (1981): "A Comparison of the Box–Cox Maximum Likelihood Estimator and the Nonlinear Two Stage Least Squares Estimator", Journal of Econometrics, 17, 351–381.

Atkinson, S. E. and R. Halvorsen (1984): "A New Hedonic Technique for Estimating Attribute Demand: An Application to the Demand for Automobile Fuel Efficiency", Review of Economics and Statistics, 66, 417–426.

Attanasio, O.P. and G. Weber (1989): "Intertemporal Substitution, Risk Aversion and the Euler Equation for Consumption: Evidence from Aggregate and Average Cohort Data", Economic Journal, 99, Conference 1989, 59–73.

Barlett, M.S. (1947): "The Use of Transformation", Biometrics, 3, 39–52.

Barlett, M.S. (1963),"Statistical Estimation of Density Functions", Sankhya, Ser.A, 25, 145–154.

Bartik, T.J. (1987a): "The Estimation of Demand Parameters in Hedonic Price Models", Journal of Political Economy, 95, 81–89.

Bartik, T.J. (1987b): "Estimating Hedonic Demand Parameters with Single Market Data: The Problems Caused by Unobserved Tastes", Review of Economics and Statistics, 69, 178–180.

Berndt, E.K., Hall B.H., Hall R.E. and Hausman J.A. (1974), "Estimation and Inference in Nonlinear Structural Models", Annals of Economic and Social Measurement, 4, 653–665.

Baillie, R.T. (1987) : "Inference in Dynamic Models Containing 'Surprise' Variables." Journal of Econometrics, 35, 101–117.

Baillie, R.T., Lippens, R.E. and McMahon, P.C. (1983) : "Testing Rational Expectations and Efficiency in the Foreign Exchange Market." Econometrica, 51, 553–563.

Banerjee, A. and Dolado, J. (1987a) : "Do We Reject Rational Expectation Models too often: Interpreting Evidence using Nagar Expansions." Economics Letters, 24, 27–32.

Banerjee, A. and Dolado, J. (1987b) : "Tests of the Life Cycle–Permanent Income Hypothesis in the Presence of Random Walks: Asymptotic Theory and Small–Sample Interpretations." Mimeo, Oxford.

Banerjee, A., Dolado, J. and Galbraith, J.W. (1987) : "Rejections of Orthogonality in Rational Expectations Models: Further Monte Carlo Results for an Extended Set of Regressors." Forthcoming, Economics Letters.

Beall, G. (1942): "The Transformation From Entomological Field experiments so that the Analysis of Variance Becomes Applicable", Biometrika, 32, 243–262.

Bean, C.R. (1986) : "The Estimation of "Surprise" Models and the Surprise Consumption Function.", The Review of Economic Studies, 53, 497–516.

Bickel P.J. and K.A. Docksum (1981): "An Analysis of Transformation Revisited", Journal of the American Statistical Association, 76, 296–311.

Blaylock J.R. and D.M. Smallwood (1983): "Interpreting the Effect of Binary Variables in Transformed Models, Economics Letters, 12, 255–259.

Blinder, A.S. and A. Deaton (1985) : "The Time Series Consumption Function Revisited", Brookings Paper on Economic Activity, 2, 465–511.

Blinder, A.S. and Deaton, A. (1985) : "The Time Series Consumption Function Revisited." Brookings Paper on Economic Activity, 2, 465–511.

Bollerslev, T. (1986), "A Generalized Autoregressive Conditional Heteroscedasticity", Journal of Econometrics, 31, 307–327.

Box, G.E.P. and D.R. Cox (1964): "An Analysis of Transformations", Journal of Royal Statistical Society Series B, 26, 211–252.

Boyd, H.J. and R.E. Mellman (1980): "The Effect of Fuel Economy Standards on the U.S. Automotive Market: An Hedonic Demand Analysis", Transportation Research, 14A, 367–378.

Bray, M.M. (1981) : "Futures Trading, Rational Expectations, and the Efficiency Markets Hypothesis." Econometrica, 49, 575–596.

Bray, M.M. (1985) : "Rational Expectations, Information and Asset Markets: An Introduction." Oxford Papers, 37, 161–195.

Breeden, D.T. (1979) : "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities." Journal of Financial Economics, 8, 265–296.

Breiman, L. and Friedman, J.H. (1985) 'Estimating Optimal Transformations for Multiple Regression and Correlation', Journal of the American Statistical Association, 80, 580–619.

Breusch, T.S. and A.R. Pagan (1979): "A simple Test for Heteroscedasticity and Random Coefficient Variation", Econometrica, 47, 1287–1294.

Brown, J.N. and Rosen, H.S. (1982): "On the Estimation of Structural Hedonic Price Model", Econometrica, 50, 765–768.

Buja, A., Hastie, T, and Tibshirani, R. (1989) 'Linear Smoothers and Additive Models', Annals of Statistics, 17, 453–510.

Burbidge, J.B., L. Magee and A.L. Robb (1988): "Alternative Transformations to Handle Extreme Values of the Dependent Variable", Journal of the American Statistical Association, 83, 123–127.

Cacoullos, T. (1966): "Estimation of a Multivariate Density", Annals of the Institute of Statistical Mathematics, 28, 77–89.

Campbell, J.Y. and Shiller, R.J. (1986) : "Cointegration and Tests of Present Value Models." NBER Working Paper No. 1885.

Carroll, R.J. (1982): "Adapting for Heteroscedasticity in Linear Models", Annals of Statistics, 10, 1224–1233.

Carroll, R.J. and D. Ruppert (1988): Transformation and Weighting in Regression, Monographs on Statistics and Applied Probability, London: Chapman and Hall.

Cassel, E. and R. Mendelsohn (1985): "The Choice of Functional Forms for Hedonic Price Equations: Comment", Journal of Urban Economics, 18, 135–142.

Chamberlain, G. (1987): "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions", Journal of Econometrics, 34, 305–334.

Chen, N.F., Roll, R. and Ross, S.A. (1986) : "Economic Forces and the Stock Market." Journal of Business, 59, 383–403.

Collomb, G. (1981): "Estimation Non–parametric de la Regression: Revue Bibliography", International Statistical Review, 49, 75–93.

Court, A.T. (1939): "Hedonic Price Indexes with Automotive Examples", In The dynamics of Automobile Demand, New York, General Motors Corporation, 99–117.

Cragg, J.G. (1983): "More Efficient Estimation in the Presence of Heteroscedasticity of Unknown Form", Econometrica, 51, 751–764.

Craven, P. and Wahba, G. (1979): "Smoothing Noisy Data With Spline Functions: Estimating the Correct Degree of Smoothing By the Method of Generalized Cross–Validation", Numerical Mathematics, 31, 377–403.

Curtiss, J.H. (1943): "On Transformation Used in Analysis of Variance", Annals of Mathematical Statistics, 14, 107–122.

Davidson, J., Hendry, D., Srba, F. and Yeo, S. (1978) : "Econometric Modelling of the Aggregate Time–Series Relationship Between Consumers' Expenditure and Income in the United Kingdom." Economic Journal, 88, 661–692.

Dagenais, M.G., M.J.I. Gaudry, and C.L. Tran (1987): "Urban Travel Demand: The Impact of Box–Cox Transformations with Nonspherical Residual Errors", Transportation Research, 21B, 443–477.

Deaton, A. (1977) : "Involuntary Saving Through Unanticipated Inflation", American Economic Review, 67, 899–910.

Deaton, A. (1986) : "Life–Cycle Model of Consumption: Is the Evidence Consistent With the Theory?" NBER Working Paper No. 1910.

Delgado, M.A. (1988a): "KREG and RONCO: Routines for Nonparametric and Semiparametric Kernel Regression", Preprint, LSE.

Delgado, M.A. (1988b): "An Alternative Efficient Semiparametric Generalized Least Squares Estimator", Preprint, LSE.

Dicky, D.A. and Fuller W.A. (1979) : "Distributions of the estimators for Autoregressive Time Series with a Unit Root." Journal of American Statistical Association, 74, 427–431.

Dicky, D.A. and Fuller W.A. (1981) : "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root." Economic Journal, 88, 661–692.

Draper, N.R. and D.R. Cox (1969): "On Distributions and Their Transformation to Normality", Journal of Royal Statistical Society, Series B, 31, 472–476.

Dunn, K.B. and Singleton, K.J. (1986), "Modelling the Term Structure of Interest Rates Under Non–Separable Utility and Durability of Goods", Journal of Financial Economics, 17, 27–56.

Eicker, F. (1963), "Asymptotic Normality and Consistency of the Least Squares

Estimator for Families of Linear Regression", Annals of Mathematical Statistics, 34, 447–456.

Elbadawi, I., A.R. Gallant and G. Souza (1983): "An Elasticity can be Estimated Consistently Without A Priori Knowledge of Functional Form", Econometrica, 51, 1731–1751.

Engle, R.F. (1982), "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of U.K. Inflation", Econometrica, 50, 987–1008.

Engle, R.F. (1983), "Estimates of the Variance of U.S. Inflation Based on the ARCH Model", Journal of Money, Credit and Banking, 15, 286–301.

Engle, R.F. and T. Bollerslev (1986), "Modelling the Persistence of Conditional Variances", Econometric Reviews, 5, 1–50.

Epanechnikov, V.A. (1969): "Nonparametric Estimation of a Multivariate Probability Density", Theory of Probability and Applications, 14, 153–158.

Epple, D. (1987): "Hedonic Prices and Implicit Markets: Estimating Demand and Supply Functions for Differentiated Products", Journal of Political Economy, 95, 59–80.

Eubank, R.L. (1988): Spline Smoothing and Nonparametric Regression, New York, Marcel Dekker.

Fama, E.F. (1984) : "Forward and Spot Exchange Rate." Journal of Monetary

Economics, 14, 47–66.

Flavin, M.A. (1981) : "The Adjustment of Consumption to Changing Expectations About Future Income." Journal of Political Economy, 89, 974–1009.

Freeman, A.M. (1979): The Benefits of Environmental Improvement, John Hopkins University Press, Baltimore.

Friedman, J.H. and Stuetzle, W. (1981): "Projection Pursuit Regression", Journal of the American Statistical Association, 76, 817–823.

Gallant, A.R. (1981): "On the Bias in Flexible Functional Forms and an Essentially Unbiased Form", Journal of Econometrics, 15, 211–245.

Gallant, A.R. and G. Tauchen (1989) : "Seminonparametric Estimation of Conditionally Constrained Heterogeneous Process: Asset Pricing Applications", Econometrica, Forthcoming.

Gaudry, M.J.I. and M.G. Dagenais (1979): "Heteroscedasticity and the Use of Box–Cox Transformation", Economics Letters, 2, 225–229.

Geweke, J. (1986), "Comment", Econometric Reviews, 5, 57–61.

Gnanadesikan, R. (1977) : Methods for Statistical Data Analysis of Multivariate Observations, New York: Wiley.

Goodman, A.C. (1983): "Willingness to Pay for Car Efficiency: The Hedonic Price Approach", Journal of Transport Economics and Policy, 17, 247–266.

Goodman, A.C. (1989): "Identifying Willingness–to–Pay for heterogeneous Goods with Factorial Survey Methods", Journal of Environmental and Management, 16, 58–79.

Green, P. (1985): "Linear Models for Field Trials, Smoothing and Cross–Validation", Biometrika, 72, 527–537.

Griffin, J.M. (1979): Energy Conservation in the OECD: 1980 to 2000, Cambridge, Mass.: Ballinger Publishing Company.

Griliches, Z. (1971): Price Indexes and Quality Change, Harvard University Press, Cambridge.

Grossman, S.J. and Shiller, R.J. (1981) "The Determinants of Variability of Stock Market Prices." American Economic Review, 71, 222–227.

Hall, P and Marron J.S. (1987): "On the Amount of Noise Inherent in Bandwidth Selection for a Kernel Density Estimator." Annals of Statistics, 15, 163–181.

Hall, R.E. (1978) : "Stochastic Implication of Life Cycle – Permanent Income Hypothesis: Theory and Evidence", Journal of Political Economy, 86, 971–987.

Hall, R.E. (1988), "Intertemporal Substitution in Consumption", Journal of Political Economy, 96, 339–357..

Hall, R.E. and Mishkin, F.S. (1982) : "The Sensitivity of Consumption to Transitory Income: Estimates from Panel Data on Household." Econometrica, 50, 461–481.

Halvorsen, R. and H.O. Pollakowski (1981): "Choice of Functional Form for Hedonic Price Equations", Journal of Urban Economics, 10, 37–49.

Hansen, L.P. (1980) : "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis." Journal of Political Economics, 88, 829–853.

Hansen, L.P. (1982) : "Large Sample Properties of Generalized Method of Moments Estimators." Econometrica, 50, 1029–1054.

Hansen, L.P. and Hodrick, R.J. (1980) : "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis." Journal of Political Economy, 88, 829–853.

Hansen, L.P. and Singleton, K.J. (1982) : "Generalized Instrumental Variables Estimation of Non linear Rational Expectations Models." Econometrica, 50, 1269–1286.

Hansen, L.P. and Singleton, K.J. (1983) : "Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns." Journal of Political Economy, 91, 249–265.

Hansen, L.P. and Singleton K.J. (1982), "Generalized Instrumental Variables

Estimation of Nonlinear Rational Expectations Models", Econometrica, 50, 1269–1285.

Hansen, L.P. and Singleton K.J. (1983), "Stochastic Consumption, Risk Aversion, and the Temporal Behaviour of Asset Returns", Journal of Political Economy, 91, 249–265.

Hansen, L.P. and Singleton K.J. (1984), Errata, Econometrica, 52, 267–268.

Harrison, D., Jr. and Stock, J.H. (1984): "Hedonic Housing Values, Local Public Goods, and the Benefits of Hazardous Waste Cleanup", Discussion Paper, John Fitzgerald Kennedy School of Government, Harvard University.

Harvey, C.R. (1988) "The Term Structure and Consumption Growth", Journal of Financial Economics, 22, 305–333.

Hausman, J.A. (1978) : "Specification Tests in Econometrics", Econometrica, 46, 1251–1271.

Hayashi, F. (1982) : "The Permanent Income Hypothesis: Estimation and Testing by Instrumental Variable." Journal of Political Economy, 90, 895–916.

Hayashi, F. (1985) : "Tests of Liquidity Constraints: A Critique." NBER working paper No. 1720.

Huang, C.J. and J.A. Kelingos (1979): "Conditional Mean Function and a General

Specification of the Disturbance in Regression Analysis", Southern Economic Journal, 45, 710–717.

Huang, C.J. and O.R. Grawe (1980): "Functional Form and the Demand for Meat in the United States: A Comment", Review of Economic and Statistics, 62, 144–146.

Huber, P.J. (1985) : "Projection Pursuit", Annals of Statistics, 12, 435–475.

Harvey, C.R. (1988), "The Term Structure and Consumption Growth", Journal of Financial Economics, 22, 305–333.

Jackman, R. and J. Sutton (1982) : "Imperfect Capital Markets and the Monetarist Black Box: Liquidity Constraints, Inflation and the Asymmetric Effects of Interest Rate Policy", Economic Journal, 92, 108–128.

Jarque, C.M. and A.K. Bera (1980): "Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals", Economics Letters, 6, 255–299.

Johnson, N.L. (1949): "Systems of Frequency Curves Generated by Methods of Translation", Biometrika, 36, 149–176.

Jolliffi, I.T. (1986): Principal Component Analysis, New York: Springer Series.

Jones, L.E. (1988): "The Characteristics Models, Hedonic Prices, and the Clientele Effect", Journal of Political Economy, 96, 551–567.

Jorgenson, D.W. Laffont,J.J. (1974) : "Efficient Estimation of Nonlinear Simultaneous Equations With Additive Disturbances." Annals of Economic and Social Measurement, 3, 615–640.

Kahn, S. and Lang, K. (1988): "Efficient Estimation of Structural Hedonic Systems", International Economic Review, 29,157–166.

King, M.A. (1985) : "The Economics of Saving: A survey of Recent Contributions." In K.J. Arrow and S. Honkapohia (eds.) Frontiers of Economics, Basil Blackwell: Oxford. 227–294.

Koskela, E. and Viren M. (1987), "Inflation and the Euler Equation Approach to Consumption Behaviour: Some Empirical Evidence", Economics Letters, 25, 233–238..

Kreps, D.M. and Porteus, E.L. (1978), "Temporal Resolution of Uncertainty and Dynamic Choice Theory", Econometrica, 46, 185–200.

Kugler, P. and Bossard, A. (1987) : "Consumption of Non–Durables and Durables." Economics Letters, 24, 219–222.

Lahiri, K. and D. Egy (1981): "Joint Estimation and Testing for Functional Form and Heteroscedasticity", Journal of Econometrics, 15, 299–307.

Lancaster, K. (1971): Consumer Demand: A New Approach, Columbia University Press, New York.

Lee, D.K.C. (1988): "Semiparametric Analysis of Surprise Consumption Function", Preprint, LSE.

Lee, D.K.C. (1989): "Inflation, Intertemporal Substitution and Consumption: Some Semiparametric Evidence", Preprint, LSE.

Lucas, R.E. (1975): "Hedonic Price Functions", Economic Inquiry, 8, 157–178.

Lucas, R.E. (1978) : "Asset Prices in an Exchange Economy" Econometrica, 46, 1429–1445.

MacKinnon, J.G. and L. Magee (1989): "Transforming the Dependent Variable in Regression Models", Forthcoming, International Economic Review.

Mankiw, N.G. (1981) : "The Permanent Income Hypothesis and the Real Interest Rates." Economics Letters, 7, 307–311.

Mankiw, N.G., Rotemberg, J.J., and Summers, L.H. (1985) : "Intertemporal Substitution in Macroeconomics." Quarterly Journal of Economics, 100, 225–251.

Mankiw, N.G. and Shapiro, M.D. (1985) : "Trends, Random Walks, and Tests of the Permanent Income Hypothesis." Journal of Monetary Economics, 16, 165–174.

Mankiw, N.G. and Shapiro, M.D. (1986) : "Do we reject too often? Small Sample Properties of Tests of Rational Expectations Models." Economics Letters, 20, 139–145.

Mankiw, N.G. and Summers, L.H. (1983) : "Do Long Term Interest Rates Overreact To Short Term Interest Rates?", Mimeo. ?

Matloff, N., Rose, R. and Tai R. (1984): "A Comparison of Two Methods for Estimating Optimal Weights in Regression Analysis", Journal of Statistical Computation and Simulation, 19, 265–274.

Marron, J.S. (1988): "Partitioned Cross–Validation", Econometric Reviews, 6, 271–283.

McCallum, B.T. (1976a) : "Rational Expectations and the Estimation of Econometric Models: An Alternative Procedure." International Economic Review, 17, 2, 484–490.

McCallum, B.T. (1976b) : "Rational Expectations and the Natural Rate Hypothesis: Some Consistent Estimates." Econometrica, 44, 1, 43–52.

McCallum, B.T. (1979) : "Topics Concerning the Formulation, Estimation, and Use of Macroeconometric Models with Rational Expectations." American Statistical Association Proceedings of the Business and Economic Statistics Section, 65–72.

Mendelsohn, R. (1984): "Estimating the Structural Equations of Implicit Markets and Household Production Functions", Review of Economics and Statistics, 66, 673–677.

Mendelsohn, R. (1985): "Identifying Structural Equations with Single Market Data", Review of Economics and Statistics, 67, 525–529.

Miron, J.A. (1986) : "Seasonal Fluctuations and the Life Cycle – Permanent Income Model of Consumption" Journal of Political Economy, 94, 1258–1279.

Muellbauer, J. (1983), "Surprises in Consumption Function", Economic Journal, 93 (supplement), 34–50.

Nadaraya, E.A. (1964): "On Estimating Regression", Theory of Probability and Its Applications, 9, 141–142.

Nelda, J.A. and Mead, R. (1965), ? , Computer Journal, 7, 308–?.

Nelson, C.R. (1975) : "Rational Expectations and the Estimation of Econometric Models." International Economic Review, 16, 555–561.

Nelson, C.R. (1987) : "A Reappraisal of Recent Tests of the Permanent Income Hypothesis." Journal of the Political Economy, 95, 641–646.

Nelson, C.R. and Plosser, C.I. (1982) : "Trends and Random Walks in Macroeconomic Time Series." Journal of Monetary Economics, 10, 139–162.

Newey, W. (1984) : "A Method of Moments Interpretation of Sequential Estimators." Economics Letters, 14, 201–206.

Newey, W.K. (1987): "Efficient Estimation of Models with Conditional Moment Restrictions", Preprint, Princeton University.

Ohta, M. and Z. Griliches (1986): "Automobile Prices and Quality: Did the Gasoline Price Increases Change Consumer Tastes in the U.S.?", Journal of Business and Economic Statistics, 4, 187–198.

Pagan, A.R. (1984), "Econometric Issues in the Analysis of Regressions with Generated Regressors", International Economic Review, 25, 221–247.

Pagan, A. (1986) : "Two Stage and Related Estimators and Their Application." The Review of Economic Studies, 53, 517–538.

Pagan, A.R. (1987), "Three Econometric Methodologies: A Critical Appraisal", Journal of Economic Surveys, 1, 3–25.

Pagan, A.R. and Hong Y. (1988), "Non–Parametric Estimation and the Risk Premium", Working Paper, University of Rochester.

Pagan, A.R. and Sabau H. (1987), "On the Consistency of the MLE in Certain Heteroscedastic Regression Models", Working Paper, University of Rochester.

Pagan, A.R. and Ullah A. (1988), "The Econometric Analysis of Models with Risk Terms", Journal of Applied Econometrics, 3, 87–105.

Parzen, , E. (1962): "On Estimation of a Probability Density Function and Mode", Annals of Mathematical Statistics, 33, 1065–1076.

Pindyck, R.S. (1979): The Structure of World Energy Demand, Cambridge, Mass.: MIT Press.

Poirier, D.J. (1978): "The Use of the Box–Cox Transformation in Limited Dependent Variable Models", Journal of the American Statistical Association, 73, 284–287.

Powell, L.P., H.S. Stock and T.M. Stoker (1986): "Semiparametric Estimation of Weighted Average Derivatives", Working Paper, University of Wisconsin at Madison.

Prakasa Rao, B.L.S. (1983): Nonparametric Functional Estimation, New York, Academic Press.

Press, W.H., Flannery, B.P., Teukolsky, S.A., and Vetterling, W.T. (1987): "Numerical Recipes", University of Cambridge: New York.

Rice, J. (1984): "Bandwidth Choice for Nonparametric Regression", Annals of Statistics, 12, 1215–1230.

Rosen, S. (1974): "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition", Journal of Political Economy, 82, 34–55.

Robinson, P.M. (1983) : " Nonparametric Estimators for Time Series", Journal of Time Series Analysis, 4, 186–207.

Robinson, P.M. (1986a) : " Nonparametric Methods in Specification", Economic

Journal, 96, 134–144.

Robinson, P.M. (1986b) : " On the Consistency and Finite Sample Properties of Time Series Regression, Autoregression and Density Estimators", Annals of the Institute of Statistical Mathematics, 38, 539–549.

Robinson, P.M. (1987a): "Adaptive Estimation of Heteroscedastic Econometric Models", Revista de Econometria, 7, 6–27.

Robinson, P.M. (1987b): "Asymptotic Efficient Estimation in the Presence of Heteroscedasticity of Unknown Form", Econometrica, 55, 875–891.

Robinson, P.M. (1988a) : "Root–N–Consistent Semiparametric Regression", Econometrica, 56, 931–954.

Robinson, P.M. (1988b) : "Semiparametric Econometrics: A Survey", Journal of Applied Econometrics, 3, 35–51.

Robinson, P.M. (1988c) : "Hypothesis Testing In Semiparametric And Nonparametric Models For Econometric Time Series", Discussion Paper, LSE Econometrics Projects.

Robinson, P.M. (1988d) : "Automatic Generalized Least Squares", Discussion Paper, LSE Econometric Projects.

Robinson, P.M. (1988e): "Best Nonlinear Three–Stage Least Squares Estimation of Certain Econometric Models", Preprint, LSE.

Robinson, P.M. (1988f): "The Stochastic Difference Between Econometric Statistics", Econometrica, 56, 531–548.

Robinson, P.M. (1989): "Nonparametric and Semiparametric Estimation", Lecture Notes, London School of Economics.

Rose, R.L. (1978): "Nonparametric Estimation of Weights in Least Squares Regression Analysis", Doctoral Dissertation, University of California at Davis.

Rosen, S. (1974): "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition", Journal of Political Economy, 82, 34–55.

Rosenblatt, M. (1956): "Remarks on Some Nonparametric Estimates of a Density Function", Annals of Mathematical Statistics, 27, 832–835.

Sargent, T.J. (1978) : "Rational Expectations, Econometric Exogeneity, and Consumption" Journal of Political Economy, 86, 673–700.

Savin, N.E. and K.J. White (1978):" Estimation and Testing for Functional Form and Autocorrelation: A Simultaneous Approach", Journal of Econometrics, 8, 1–12.

Schlesselman, J. (1971): "Power Families: A Note on the Box–Cox Transformation", Journal of Royal Statistical Society, Series B, 33, 307–311.

Schucany, W.R. and Sommers, J.P. (1977): "Improvement of Kernel Type Density Estimators", Journal of the American Statistical Association, 72, 420–423.

Schuster, E. (1972): "Joint Asymptotic Distribution of the Estimated Regression Function at a Finite Number of Distinct Points", Annals of Mathematical Statistics, 43, 84–88.

Schuster, E. and Yakowitz S. (1979): "Contributions to the Theory of Nonparametric Regression, with Application to System Identification", Annals of Statistics, 7, 139–149.

Selden L. (1978), " A New Representation of Preferences Over 'Certain x Uncertain' Consumption Pairs: The 'Ordinal Certainty Equivalent' Hypothesis", Econometrica, 46, 1045–1060.

Sentana, E. and Wadhwani, S. (1989): "Semi–Parametric Estimation and the Predictability of Stock Market Returns: Some Lessons From Japan", Preprint, LSE.

Shapiro, Mathew D (1984) "The Permanent Income Hypothesis and the Real Interest Rates: Some Evidence from Panel Data" Economics Letters, 14, 93–100.

Shiller, R. (1979) : "The Volatility of Long–Term Interest Rates and Expectations Models of the Term Structure." Journal of Political Economy, 87, 1190–1219.

Shiller, R.J. (1982) : "Consumption, Asset Markets and Macroeconomic Fluctuations." Carnegie–Rochester Conference Series on Public Policy, 17, 203–238.

Silverman, B.W. (1986): Density Estimation for Statistics and Data Analysis, London, Chapman and Hall.

Speckman, P. (1988): "Kernel Smoothing in Partial Linear Models", Journal of Royal Statistical Society, Ser. B, 50, 413–436.

Stock, J.H. (1984): "Nonparametric Estimation of the Effects of Distributional Changes in Exogenous Variables", Discussion Paper, John Fitzgerald Kennedy School of Government, Harvard University.

Stock, J.H. (1985a): "Nonparametric Policy Analysis", Discussion Paper, John Fitzgerald Kennedy School of Government, Harvard University.

Stock, J.H. (1985b): "Semiparametric Hedonics", Discussion Paper, John Fitzgerald Kennedy School of Government, Harvard University.

Stoker, T.M. (1986): "Consistent Estimation of Scaled Coefficients", Econometrica, 54, 1461–1481.

Stone, C.J. (1977) "Consistent Nonparametric Regression", Annals of Statistics, 5, 595–645.

Stone, C.J. (1985): "Additive Regression and Other Nonparametric Models", Annals of Statistics, 13, 689–705.

Tallman, E.W. (1986) : "Macroeconomic Factors in Asset Premia." Mimeo, University of Rochester.

Watson, G.S. (1964): "Smooth Regression Analysis", Sankhya, Series A, 26, 359–372.

Weiss, A.A. (1984), "ARMA Models with ARCH Errors", Journal of Time Series Analysis, 5, 129–143.

Whistler, D. (1988), "Semi–Parametric ARCH Estimation of Intra–Daily Exchange rate Volatility", Preprint, LSE.

White, H. (1980), " A Heteroscedasticity–Consistent Covariance Matrix Estimator and Direct Test for Heteroscedasticity", Econometrica, 48, 817–838.

Wickens, M.R. (1982) : "The Efficient Estimations of the Rational Expectations Hypothesis." Review of Economic Studies, 49, 55–67.

Wickens, M.R. and Molana, H. (1984) : "Stochastic Life–Cycle Theory with Varying Interest Rates and Prices." Economic Journal, 94 (supplement), 133–147.

Wertz, W. and Schneider, B. (1979): "Statistical Density Estimation: A Bibliography", International Statistical Review, 47, 155–175.

Witte, A.D., Sumka, H.J. and Erekson, H. (1979): "An Estimate of a Structural Hedonic Price Model of the Housing Market: An Application of Rosen"s Theory of Implicit Markets", Econometrica, 47, 1151–1171

Wolff, C.P. (1986) : "Forward Exchange Rates, Expected Spot Rates and Premia : A Forecasting Approach." Mimeo, London Business School.