TRANSFORMATION FUNCTION TESTS

BY KAVEH AKRAM

A dissertation submitted to the

Graduate School - New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Economics

Written under the direction of

Roger Klein

and approved by

_____

_____

_____

_____

New Brunswick, New Jersey

October, 2011

ABSTRACT OF THE DISSERTATION

Transformation Function Tests

by Kaveh Akram

Dissertation Director: Roger Klein

Econometric models usually relate a known function of a dependent variable, Y, with some observable covariates X. A misspecified transformation function, however, can cause inconsistency. The estimated parameters will be biased and inconsistent. The marginal effects and the elasticities calculated using these estimated parameters can be far from the truth and any inference based on them will be misleading.

The problem of misspecification can be resolved simply by estimating the transformation function. There are different methods of estimating the model based on the parametric assumptions on the transformation function, or the distribution of the error term, or both. In a completely parametric setup, both the transformation function and the distribution of the error term are known up to a vector of parameters. The Box-Cox transformation is a good example of using parametric transformation functions. It is also possible, however, to estimate the transformation function without parametric assumptions.

In this dissertation two Hausman tests for transformation functions are proposed where validity does not depend on distributional assumptions. These tests compare estimators that remain consistent regardless of the transformation function to an

estimator whose consistency depends on the transformation function. The properties of these test statistics are studied in finite sample and under different designs. The behavior of these test statistics is studied both when the adopted transformation function is correct and when the true transformation function deviates from the hypothesized one.

This dissertation applies the semiparametric transformation function test to study reported crimes in the U.S. metropolitan areas. Most studies in this literature adopt a logarithmic transformation of reported crimes. There is no theoretical justification for this specific function. In addition, it is likely that city level crime is misreported (underreported). Therefore, testing is particularly relevant. I show that, although for particular types of crimes the log function is appropriate, the same log function cannot be used for broadly defined categories of crimes.

Acknowledgements

This dissertation would not have been written without the help and support of many people. I would like to express my gratitude to all those people who made this work possible.

I would like to express my sincere gratitude to my advisor, Roger Klein, for his patience, guidance, support, and encouragement. He was always available whenever I needed his advice.

I would like to thank Anne Piehl for her suggestions and support. I have benefited greatly from her invaluable guidance.

I would like to thank John Landon-Lane and Chan Shen for their insightful comments.

I would also like to thank my friends and colleagues for their help and support during my graduate studies. I specially want to thank Costanza Biavaschi.

I would like to thank Dorothy Rinaldi for her caring, kindness, and support.

Last, but certainly not least, I would like to thank my parents for their unconditional love and support. They are the ones who taught me to be a critical thinker.

Dedication

To my parents

Table of Contents

List of Tables

List of Figures

Chapter 1
Introduction

It is often of interest to estimate the price elasticity of demand or the elasticity of crime with respect to police force. It would be tempting to take the logarithm of quantity or the logarithm of crime, and use it as the dependent variable in a regression on the logarithm of price or the logarithm of police force. The estimates from these regressions could be interpreted as the elasticities of interest, and the use of the logarithmic transformation would simplify an otherwise more cumbersome calculation. In other settings, for example in wage regressions, the use of the logarithm of the dependent variable has been justified to obtain the normality of the error term. Such transformations of the dependent variable also have been justified to produce a better fit of the model to the data. In all these examples, for various reasons, the researchers have used a transformation function (e.g. the logarithm) of the dependent variable.

If, however, the transformation is incorrect, the estimators will be inconsistent. Suppose the true transformation function is T_o. i.e.:

$$T_o(\boldsymbol{Y}) = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}$$

but instead, the parameters are estimated in a misspecified model. i.e.:

$$T(\boldsymbol{Y}) = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where T is not a linear transformation of T_o.

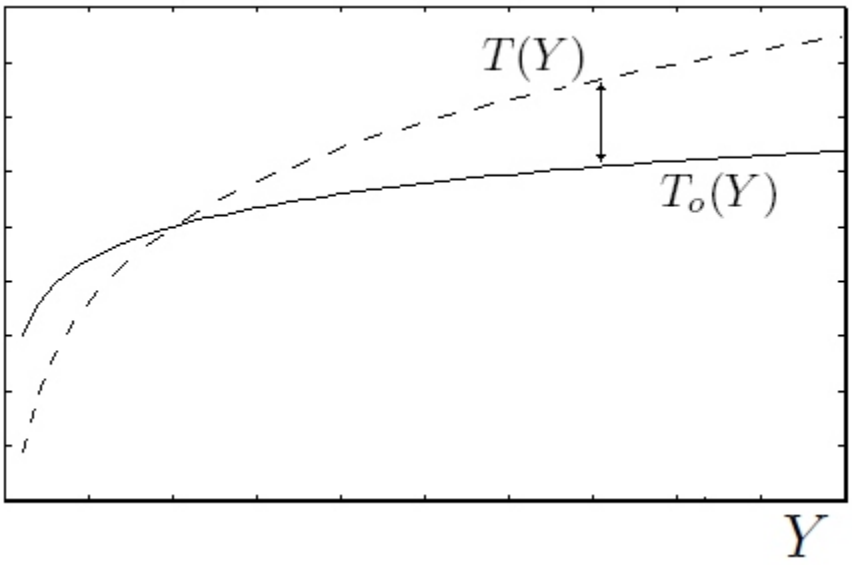


Figure 1.1: True Transformation Function Vs. Used Transformation Function

Then the error term in the misspecified model is:

$$\boldsymbol{\epsilon} = \boldsymbol{u} + \left[T(\boldsymbol{Y}) - T_o(\boldsymbol{Y})\right]$$

The difference between the two functions $T(\boldsymbol{Y})$ and $T_o(\boldsymbol{Y})$ is now part of this error term. This part of the error term depends on Y and consequently on X. The induced endogeneity is the reason why using an incorrect transformation can lead to inconsistent estimators.

To illustrate the magnitude of the bias, assume that the true model is given as:

$$ln(\boldsymbol{Y}) = 0.5\boldsymbol{X} + 1 + \boldsymbol{\epsilon},$$

However, suppose that we misspecify the transformation function and estimate:

$$\boldsymbol{Y} = \boldsymbol{X}b + c + \boldsymbol{u},$$

With $\boldsymbol{X}$ generated as a normal random variable centered at zero and with the

variance of 9 and $\boldsymbol{\epsilon}$ generated as a standard normal random variable , The above model was estimated in a Monte Carlo with N = 1000 observations and R = 1000 replications. For the parameter estimates, the results were as follows:

Monte-Carlo Results	*Bias*	*Std.Error*
Estimated Slope	6.32	2.12
Estimated Intercept	12.69	2.00

The bias and the standard error when the true transformation function is used are shown in the following table.

Monte-Carlo Results	*Bias*	*Std.Error*
Estimated Slope	-0.0002	0.0107
Estimated Intercept	-0.0006	0.0308

As seen the estimated bias and standard error are larger in the misspecified model; However, in comparing results across models, it is difficult to compare coefficients. In the true model, the coefficient on $\boldsymbol{X}$ describe the percentage change in $\boldsymbol{Y}$ for a one unit change in $\boldsymbol{X}$, while in the misspecified model, the coefficient describes the change in $\boldsymbol{Y}$ for a one unit change in $\boldsymbol{X}$. Accordingly, rather than comparing coefficients, we turn to a comparison of marginal effects. Namely, in both models we report the percentage change in $\boldsymbol{Y}$ for a one unit change in $\boldsymbol{X}$. In the true model, this percentage change is constant and is equal to the estimated slope; However, this percentage change is not constant in the misspecified model and depends on the value of $\boldsymbol{X}$. The percentage changes of $\boldsymbol{Y}$ when $\boldsymbol{X}$ is increased by one unit are presented in Table 1.1. These marginal effects are calculated using the average estimates for the intercept and the slope in the Monte Carlo study. As seen in Table 1.1, the marginal effects can be very different in the misspecified model compared to the true model. The reason the marginal effect is very different at $X = -2$ is that the marginal effect

in the misspecified model is equal to $\frac{\hat{b}}{X*\hat{b}+\hat{c}}$ and in our Monte Carlo $\hat{c}$ happens to be very close to $2\hat{b}$, thus the denominator of the marginal effect is very close to zero.

Table 1.1: Marginal effects in the true and the misspecified models

$\boldsymbol{X}$	-3	-2	-1	0	1	2	3
The marginal effects in the misspecified model	-1.0059	169.77	0.9941	0.4985	0.3327	0.2496	0.1998
The marginal effects in the true model	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997

Transformation functions are very common in many different fields of study. There are different justifications why the dependent variable is transformed. For example, a logarithmic transformation is usually used when the dependent variable is positive and skewed to the right. Logarithmic transformation also reduces the effect of outliers. Such a transformation also tends to diminish the effect of heteroskedasticity. Probably the most famous example of using logarithmic transformation is Mincer's wage equation (Mincer, 1974). In such models, the logarithm of wage or earnings is used as the left-hand-side variables with explanatory variables like education, experience, gender, and race.

The logarithmic transformation is also used in health economics literature. In many cases where the dependent variable is health care expenditure, health care utilization, or number of hospital days, the log transformation is used. For example in Zweifel et al. (1999), the logarithm of health care expenditure is used to reduce the skewness of this variable. This paper studies the relationship between age and health care expenditure.

In Bamezai et al. (1999), the logarithm of operating hospital costs is used as the dependent variable in a study that analyzes the effect of health maintenance organizations and preferred provider organizations on the cost of hospitals in different

market structures.

In addition, the logarithmic transformation is used for variables like the length of stay in hospital. The log transformation for such a variable is usually justified because it is right skewed and can have a long right tail. For example in Robinson and Luft (1985), where the impact of market structure on average hospital cost is studied, the log transformation is used for the following dependent variables: length of stay, number of outpatient visits, and inpatient admissions.

The logarithmic transformation is used in crime literature as well. This literature will be discussed in more detail later in this dissertation. In most studies regarding city level crime rate, the logarithm of crime or logarithm of crime rate are used as left-hand-side variables. Some examples are: Glaeser and Sacerdote (1999), Levitt (1997), McCrary (2002), Kelly (2000), and Lott and Mustard (1997).

The misspecification problems arisen by using a wrong transformation function can be resolved by estimating the transformation function. There are different estimation methods that are based on parametric assumptions on the transformation function, on the distribution of the error term, or on both. In a completely parametric setup, both the transformation function and the distribution of the error term are known up to a vector of parameters. In these models, a "flexible function" is used as the left-hand-side variable. In these setups, the shape of the transformation function is assumed to be known up to a vector of parameters. The Box-Cox transformation is a good example of using a parametric transformation function (Box and Cox, 1964). The transformation functions in these models are as follow.

$$T(\boldsymbol{Y}) = \begin{cases} \frac{\boldsymbol{Y}^{\lambda}-1}{\lambda} & \text{if} \quad \lambda \neq 0 \\ ln(\boldsymbol{Y}) & \text{if} \quad \lambda = 0 \end{cases} \tag{1.1}$$

for $\boldsymbol{Y} > 0$

or

$$T(\boldsymbol{Y}) = \begin{cases} \frac{(\boldsymbol{Y}+\lambda_2)^{\lambda_1}-1}{\lambda_1} & \text{if} \quad \lambda_1 \neq 0 \\ ln(\boldsymbol{Y}+\lambda_2) & \text{if} \quad \lambda_1 = 0 \end{cases} \tag{1.2}$$

for $\boldsymbol{Y} > -\lambda_2$

Box-Cox transformation is widely used because this family of functions includes both the logarithmic transformation and the linear transformation (no transformation). For example Box and Cox themselves studied the survival time of animals in a 3×4 factorial experiment where the factors are 3 poisons and 4 treatment. Although the Box-Cox transformation is probably the most common flexible function, there are many other flexible functions in the literature. Most of these flexible functions are modifications of the Box-Cox transformation. Some of these flexible functions are presented in John and Draper (1980), Bickel and Doksum (1981), and MacKinnon and Magee (1990).

Bickel and Doksum argued that Box-Cox transformation requires $\boldsymbol{Y}$ to be positive as $\boldsymbol{Y}^{\lambda}$ is not real for negative values of $\boldsymbol{Y}$, unless λ is an integer. Also the left-hand-side variable in such a model is bounded from below (when $\lambda > 0$). One way to extend the model in (1.1) is a model in which the dependent variable can also take on negative values. Bickel and Doksum suggested the following transformation function:

$$T(\boldsymbol{Y}) = \frac{|\boldsymbol{Y}|^{\lambda} sign(\boldsymbol{Y}) - 1}{\lambda} \tag{1.3}$$

where $\lambda > 0$ and $Y \in \mathbb{R}$.

John and Draper argued that a power transformation such as Box-Cox transformation eliminates the skewness even when such a correction is not necessary. They suggest the alternative "modulus" transformation:

$$T(\boldsymbol{Y}) = \begin{cases} sign(\boldsymbol{Y})\frac{(|\boldsymbol{Y}|+1)^{\lambda}-1}{\lambda} & \text{if} \quad \lambda \neq 0 \\ sign(\boldsymbol{Y})ln(|\boldsymbol{Y}|+1) & \text{if} \quad \lambda = 0 \end{cases} \tag{1.4}$$

where $Y \in \mathbb{R}$.

They argued that such a transformation is more appropriate for distributions that are not normal but almost symmetric. They also applied this transformation in a 4×5 factorial experiment of comparing the performance of expert inspectors in assessing the thickness of certain types of piping. The experiment was done with 4 different locations on pipes and 5 equally qualified inspectors.

MacKinnon and Magee argued that the original Box-Cox transformation cannot be used when the dependent variable takes on negative values or zero. Also, the assumption of normality of the error term contradicts the fact that the left-hand-side variable is bounded from below when $\lambda > 0$ and is bounded from above when $\lambda < 0$. To overcome these problems, they suggested the following transformation function:

$$T(\boldsymbol{Y}) = \frac{H(\alpha \boldsymbol{Y})}{\alpha}$$

Where the function H satisfies the following conditions:

$$H(0) = 0,$$

$$H^{'}(0) = 1, \text{and}$$

$$H^{''}(0) \neq 0$$

However, such a family of functions does not include logarithm. The proposed function by MacKinnon and Magee was:

$$H(\boldsymbol{Y}) = sinh^{-1}(\boldsymbol{Y}) = ln(\boldsymbol{Y} + \sqrt{Y^2 + 1})$$

Using a flexible function can reduce the problem of misspecification. By allowing the transformation function to take on different shapes, we can choose the closest function to the true transformation function in the flexible function family. While a flexible function would appear to reduce the problem, it requires the distribution of the error term to be correctly specified. If the distribution of the error term is not correctly specified, the estimators are typically inconsistent. Since the distribution of the error term is seldom known, it is problematic to employ an estimator that depends on a correct error distribution.

In order to see why in a flexible function model, both the error distribution and the transformation function may need to be correctly specified, first consider the following model without a flexible function:

$$T_0(\boldsymbol{Y}) = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

If there are no unknown parameters in T_0, then we can estimate the above model by OLS and obtain consistent estimates provided that we correctly specify the transformation function. The distribution of the error is irrelevant in this case, provided that it does not depend on X; However, in this case there is no flexibility in the transformation that would allow it to adapt to the data. Now, consider a wider class of transformation functions that would allow their shape to adapt to the data:

$$T_0(\boldsymbol{Y}; \boldsymbol{\lambda}) = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

Here, $\boldsymbol{\lambda}$ is a parameter that alows the shape of the transformation to vary. While this formulation provides needed flexibility in the transformation function, the distribution of the error term now becomes important. If the above model is estimated by OLS, it can be shown that the resulting estimator is not consistent.

To deal with the above inconsistency, one could employ maximum likelihood.

However, to obtain the likelihood, we require the distribution of the error. We should note that even if the error is normally distributed, the maximum likelihood estimator will not be equivalent to OLS. When the value of $\boldsymbol{\lambda}$ is unknown, we construct the likelihood from the conditional density of $\boldsymbol{Y}$ conditioned on $\boldsymbol{X} = \boldsymbol{x}$: $f(y|X = x)$. This distribution will not be normal even when the error term is normally distributed. More formally, this density depends not only on that for the error, but also on the jacobian of the transformation, which is not constant and generally depends on $\boldsymbol{\lambda}$.

To overcome this limitation, the models with parametric transformation functions can also be estimated by allowing both the error distribution and the transformation to have flexible functional forms. However, such flexibility can result in numerical problems in estimating the model. Moreover, the true transfromation and error distribution may not fall in the class of flexible forms. In other words, the "flexible functions" are not flexible enough to encompass all transformations and error distributions. Therefore, it would be desirable to be able to estimate the parameters with less assumptions on the transformation function.

It is also possible to estimate (up to location and scale) the transformation function with no parametric assumptions either on the transformation function or the distribution of the error term (see Horowitz (1996), Klein and Sherman (2002)). In these models, the consistency of the estimators neither depends on the form of the transformation nor on the distribution of the error term.

However, because in most studies the transformation function is not estimated, and a known function of Y is used as the left-hand side variable, this dissertation tries to test the validity of the employed transformation function without estimating it. Moreover, typically the transformation function is not of direct interest. Rather, we are only interested in it to the extent that it enables us to correctly estimate marginal effects. In proposing a test below, one of the estimators that we will study will provide correct marginal effects without having to even specify or estimate the

transformation function.

The tests constructed in this dissertation are Hausman tests which compare two estimators: one is consistent only if the used transformation function is the true transformation function, and one is consistent regardless of the form of the true transformation function (provided it is monotone). Two different tests are proposed in this dissertation based on two different ways of deriving the latter estimator. The test is constructed once using an ordered estimator and once again using a semiparametric least squares estimator. While the ordered test presented in this dissertation uses a semiparametric estimator, it is important to note that this test can also be done parametrically where the distrubution of the error term is specified. For example under normality or logistic assumptions, such tests would use the ordered probit or the ordered logit estimators as estimators whose consistencies do not depend on the employed transformation function. In applications, it is important to have good parametric tests of transformation function in addition to ones based on semiparametric estimators. Accordingly, while the focus of this dissertation will be on tests based on semiparametric methods for which consistency does not depend on the form of the transformation function, we will also discuss in detail how to perform such tests using parametric ordered models.

Although these tests can be used in various models, testing the transformation function is particularly relevant in some specific cases. First, such tests are particularly important in models where there is no theoretical justification for the transformation function that is used. Almost all examples in economics suffer from this problem. Economic theory usually has little to say about the functional form of the relationship between the dependent variable and the explanatory variables. Economic theory sometimes suggests that some variables might be related and in some other cases also suggests the direction of the relationship, but almost never specifies the functional form of the relationship. One example can be the wage equation where the

logarithm of wage is used as the dependent variable. The theory might suggest that human capital and consequently education, are related to wage. The theory can also suggest that there is a positive relationship between education and wage; However, the functional form of the relationship is not specified through economic theory.

Second, transformation models are important where the dependent variable is underreported or overreported. In these models, the reported value of the dependent variable is used instead of the actual value of the dependent variable. In most cases, it is plausible to assume that the reported value of the dependent variable is a monotone transformation of the actual value of the dependent variable. By using the reported value rather than the actual value, the transformation function which relates the actual value to the reported value is neglected and these models can suffer from the misspecification of the transformation function. These models can be written in the following way:

$$\boldsymbol{Y}^{actual} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

but since $\boldsymbol{Y}^{actual}$ is not known, the reported value of $\boldsymbol{Y}$, $\boldsymbol{Y}^{reported}$, is used.

$$\boldsymbol{Y}^{reported} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{1.5}$$

if $\boldsymbol{Y}^{actual} = T(\boldsymbol{Y}^{reported})$, then (1.5) suffers from the misspecification of the transformation function.

Some examples where variables are underreported are: crime, traffic accidents with minor damages, and health status. Self-reported data would seem to be particularly susceptible to misreporting isues. In the case of crime, which forms the basis for the application employed here, the degree of missreporting is probably not the same for all categories of crime. For example one expects that homicide suffers less from underreporting compared to larceny theft. An example of an overreported variable is

reported demand for future products from survey data.

The ordered test developed in this dissertation is applied to study the logarithmic transformation of city level crimes. Most studies in this area use the logarithm of crime as the left-hand-side variable, and there is no theoretical justification for this specific function. Also as mentioned above, it is likely that city level crime is misreported (underreported). This study shows that the logarithmic transformation function is not rejected for most of the specific crime categories but is more likely to be rejected for more aggregate categories.

This dissertation is organized as follows. In Chapter 2, the estimators needed to construct the tests are presented. The estimators that remain consistent independent of the shape of the transformation function are the semiparametric ordered estimator and the semiparametric least squares estimator. Some of the properties of these estimators are also discussed in this section. Furthermore, the test statistics are constructed and their asymptotic properties are discussed. In Chapter 3, the finite sample properties of these test statistics are presented using some Monte Carlo experiments. In Chapter 4, an empirical application is presented. In this chapter the validity of the logarithmic transformation function which is widely used in the study of city level crime is tested. Concluding remarks are presented in Chapter 5.

Chapter 2

Semiparametric Tests for Transformation Functions

Economists often transform the dependent variable in the regression model. A common example is the log transformation of the wage equation

$$\ln(\boldsymbol{Wage}) = \boldsymbol{X\beta} + \boldsymbol{u}$$

or the log-log transformation imposed to calculate elasticities in a demand function.

$$\ln(\boldsymbol{Quantity}) = \boldsymbol{X\beta} + \eta \ln(\boldsymbol{Price}) + \boldsymbol{u}$$

In this example, if the logarithmic transformation is the true transformation of the dependent variable, there is no misspecification and the estimators are consistent. For example, $\hat{\eta}$ can be safely interpreted as the price elasticity of demand in the second model; However, if the true transformation in the demand equation is not logarithm, it means that the demand curve does not have constant elasticity and $\hat{\eta}$ can no longer represent "the elasticity " along such a demand curve.

In general, a transformation function model is a model where the expectation of a monotone function of the dependent variable is linearly related to the explanatory variables. In other words, in models where the dependent variable itself is not linearly related to the explanatory variables, a transformation function can help us to find a model in which a transformed version of the dependent variable has the desired linear

relationship with the explanatory variables.

As mentioned before, in this chapter three test statistics are developed to test the validity of the transformation function. One of these test statistics is based entirely on existing and widely used parametric estimators in the literature: OLS and the parametric ordered probit. The other two estimators avoid parametric assumptions on error distributions by employing semiparametric methods.

A Transformation function model has the following representation:

$$T(\boldsymbol{Y}) = \beta_0 + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}, \tag{2.1}$$

where $\boldsymbol{Y}$ is the dependent variable, T is a monotone function[1], $\boldsymbol{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_k)$ is a full rank matrix of independent variables, $\boldsymbol{\beta}$ is a vector of unknown parameters, and $\boldsymbol{u}$ is an error term with distribution F.[2] Observations are i.i.d., and $\boldsymbol{X}$ and $\boldsymbol{u}$ are independent.

As explained in the introduction, although T can be approximated using a flexible function parametrically or estimated up to location and scale semiparametrically, most researchers often use a known transformation of $\boldsymbol{Y}$ as their left-hand-side variable. The main objective of this paper is to test whether or not the used transformation is correct. Thus, the following hypothesis is tested:

$$H_0 : T(\boldsymbol{Y}) = T_0(\boldsymbol{Y}) \qquad \text{vs} \qquad H_1 : T(\boldsymbol{Y}) \neq T_0(\boldsymbol{Y})$$

where $T_0(\boldsymbol{Y})$ is the hypothesized transformation function.

As discussed in the introduction, logarithmic transformation function is used in many different studies. By setting $T_0(\boldsymbol{Y})$ equal to $ln(\boldsymbol{Y})$ for example, one can test if

---

[1]Note that T does not have to be smooth or even continuous. Note also that the left-hand-side of equation (2.1) is the vector $[T(Y_1), \ldots, T(Y_n)]'$.

[2]This distribution is unknown in a semiparametric framework.

such a transformation is appropriate. Even in models where the dependent variable itself is used as the left-hand-variable, there exists a transformation function which is the identity function. By setting $T_0(\boldsymbol{Y})$ equal to $\boldsymbol{Y}$, one can test whether or not the model needs a transformation function which is not equal to the identity function. In other words, we will know if $\boldsymbol{Y}$ itself is linearly related to the explanatory variables or if a transformation of $\boldsymbol{Y}$ has such a linear representation.

As mentioned before, the transformation function can be estimated semiparametrically up to location and scale. Since the transformation function can be estimated, it is tempting to construct a test statistic based on a measure of difference between the estimated transformation function and the hypothesized transformation function. In this work, however, this method is not used. The test proposed in this dissertation is a Hausman test. To construct a test statistic using Hausman's method, two estimators are needed: one is consistent only if the used transformation function is the true transformation function, and one is consistent regardless of what the true transformation function is. The Hausman test is better than a test which is based on a measure of difference between an estimated transformation function and the hypothesized one, in the sense that in a Hausman test the comparison of interest is between two k dimensional vectors of parameters whereas in the other test, two functions should be compared with each other. Also, the convergence rate is faster for a k dimensional vector of estimated parameters than for a function. When testing the validity of a transformation function using a Hausman test, the transformation function in fact is not estimated. This can also save computation time.

In this chapter, two semiparametric estimators and one parametric estimator that remain consistent if an incorrect transformation function is used are presented. Let $\hat{\boldsymbol{\theta}}^*$ be an estimator that is consistent whether or not the null hypothesis holds (i.e. whether or not the employed transformation function is the true transformation function). Let $\hat{\boldsymbol{\theta}}$ be an estimator for the same parameter vector of interest under the null

hypothesis. Both of these estimators are close to the truth under the null and a distance measure of $(\hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}})$ is close to zero; However, while $\hat{\boldsymbol{\theta}}^*$ remains to be close to the truth under the alternative, $\hat{\boldsymbol{\theta}}$ deviates from the truth and a distance measure of $(\hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}})$ deviates from zero; Hence based on how different $\hat{\boldsymbol{\theta}}$ is from $\hat{\boldsymbol{\theta}}^*$, one can judge the validity of the employed transformation function. The tests presented in this dissertation are based on a distance measure between an estimator that remains to be consistent under the alternative and an estimator that is consistent under the null.

2.1 The estimator that is consistent under the null

This estimator can be chosen among all consistent estimators under the null. One possibility would be to make no assumptions on the distribution of the error term and employ OLS under the assumption that the transformation function is correct. Another possibility is to assume a distribution for the error term and employ maximum likelihood. For example, if we assume that the error is i.i.d. distributed as normal and that $\boldsymbol{X}$ is exogenous, then the following log-likelihood can be used to estimate the unknown parameters:

$$Q(\beta_0, \boldsymbol{\beta}, \sigma) = -\ln \sigma^2 - \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{T_0(\boldsymbol{Y}_i) - \beta_0 - \boldsymbol{X_i}\boldsymbol{\beta}}{\sigma}\right)^2. \tag{2.2}$$

Notice that if we maximize the above likelihood, then we must minimize the sum of squared residuals. In other words, and as is well known in the literature, under normality and with no unknown parameters in the transformation function, the OLS estimator and the maximum likelihood estimator coincide. If the error term is not normally distributed, then we can still employ the OLS estimator in constructing our test, but we need to be careful not to rely on this estimator being fully efficient (as would be the case if it were a maximum likelihood estimator). As discussed before, if

T_0 is not the true transformation function, this misspecification causes the error terms in the model to be related to the explanatory variables. Such an induced endogeneity is the source of inconsistency of the estimators when the transformation function is misspecified.

The inconsistency of the estimators in the presence of misspecified transformation function is desirable in the sense that in order to construct a Hausman test, one also needs an estimator that becomes inconsistent if the null hypothesis is not true. By comparing this inconsistent estimator (when the null is false) with a consistent one, one has a natural test for the transformation function.

We must emphasize again that this particular likelihood also assumes the normality of the error term; However, even if the error term is not normally distributed, this estimator is still consistent and asymptotically normal provided that the transformation function is equal to the hypothesized transformation function. The reason is although (2.2) is not the true likelihood function when the errors are not normally distributed, the function in (2.2) still converges uniformly to its expectation, and the expectation has a unique maximizer at the truth. These are the sufficient conditions to prove the consistency of the estimator.

In the rest of this dissertation, this model is referred to as the "linear model", since one who uses $T_0(\boldsymbol{Y})$ as the left-hand-side variable, believes that $T_0(\boldsymbol{Y})$ is linearly related to the explanatory variables.

2.2 Estimators that remain consistent under the alternative

In this work, we present three estimators whose consistency does not depend ont he assumed form of the transformation function. One of these estimators is the parametric ordered model which assumes the error distribution is normal. The other estimators are semiparametric and remain consistent regardless of the form of the

transformation function or the form of the error distribution. One of them is a semiparametric ordered estimator and the other one is a Semiparametric Least Squares (SLS) estimator. Although I focused more on these semiparametric estimators, it is important to note that the ordered estimation can also be done parametrically and a Hausman test can be constructed by comparing the parametric ordered estimator to the linear estimator. This is important because most researchers are more familiar with parametric estimators and also many computer packages provide these estimators. Another advantage of a parametric estimator is that the computation time is much shorter than of a semiparametic estimator. Therefore, in this section three consistent estimators under the alternative are presented: the parametric ordered estimator, the semiparametric ordered estimator, and the semiparametric least squares estimator.

2.2.1 The parametric ordered estimator

An ordered model can provide a consistent estimator of the parameters even if the employed transformation function is not correct. The reason is that in an ordered model, the value of the transformation function is not important. What is important is the monotonicity of the function and the fact that the transformation function preserves the order of the dependent variable. Intuitively, because T is a monotone function, data can be sorted based on $\boldsymbol{Y}$ into some ranked categories, and the unknown parameters can be estimated using an ordered model. This estimator is consistent with any monotone transformation function, and its consistency does not depend on whether or not the hypothesized transformation function is correct. The reason is that any monotone transformation of the dependent variable preserves the same ordinality of the sorted categories. It is not important whether or not this monotone transformation is actually the true transformation function. As long as one knows that the true transformation function is increasing (decreasing) and sorts

the dependent variable, $\boldsymbol{Y}$, ascendingly (descendingly), the estimator remains consistent. The next few paragraphs formalize this intuition, which will apply to both parametric and semiparametric ordered models.

To estimate the unknown parameters in the ordered model, first $\boldsymbol{Y}$ is categorized into q quantiles. Let $t_1 = \min\{Y_i\}$, $t_2, \ldots, t_q$ be the cutpoints of quantiles of $\boldsymbol{Y}$, and $t_{q+1} = \max\{Y_i\}$. Then, the monotonicity of T (in this case T is increasing) implies that:

$$t_{j-1} < Y_i \leq t_j \quad \text{iff} \quad T(t_{j-1}) < T(Y_i) \leq T(t_j) \quad \forall j = 2, \ldots, q+1. \tag{2.3}$$

Using equation (2.1) , equation (2.3) can be written as:

$$t_{j-1} < Y_i \leq t_j \quad \text{iff} \quad T(t_{j-1}) < \beta_0 + \boldsymbol{X}_i\boldsymbol{\beta} + u_i \leq T(t_j) \quad \forall j = 2, \ldots, q+1. \tag{2.4}$$

Notice that (2.4) is the description of an ordered model with unknown cutpoints. The dependent variable, Y_i, falls in some ordered categories based on the value of a linear index, $\beta_0 + \boldsymbol{X}_i\boldsymbol{\beta}$. The reason why the cutpoints are unknown is that although t_js are known, T, the true transformation function, is not known. Since it is impossible to seperate β_0 from the unknown cutpoints, β_0 is not identified in these ordered models. For example adding the same constant to both the unknown cutpoints and β_0 will keep the model the same, thus β_0 is not separately identified. σ is not identified either. For example if u_i has a normal distribution with mean zero and variance σ^2, then dividing the inequalities in (2.3) by σ will not change the probabilities of being in different categories since the inequalities contain unknown cutpoints. Although σ is not identified, $\boldsymbol{\beta}/\sigma$ is identified.

Let $\boldsymbol{\theta}^* = \boldsymbol{\beta}/\sigma$. $\boldsymbol{\theta}^*$ can be estimated consistently using an ordered model. The properties of this estimator are independent of the shape of the transformation function, since in this ordered model only the monotonicity of T is important; Thus the ordered estimators remain consistent whether or not the employed transformation

function is the true transformation function. If one assumes that the error terms are normally distributed, then this model is ordered probit with unknown cutpoints, and $\boldsymbol{\theta}^*$ can be estimated using a maximum likelihood estimator. In order to construct the Hausman test, $\boldsymbol{\theta}^*$ should be compared to its counterpart in the linear model. Therefore the likelihood in (2.2) is rewritten as follows:

$$Q(\beta_0, \boldsymbol{\theta}, \sigma) = -\ln \sigma^2 - \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{T_0(\boldsymbol{Y}_i)}{\sigma} - \frac{\beta_0}{\sigma} - \boldsymbol{X_i\theta}\right)^2, \tag{2.5}$$

where $\boldsymbol{\theta} = \boldsymbol{\beta}/\sigma$.

Although both β_0 and σ are identified in the linear model, since they are not identified in ordered probit, the estimators for these parameters cannot be used to construct the test statistic.

The test statistic for the transformation function is a measure of the difference between $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}^*$. If the true transformation function is employed, both are close to the truth and the difference is close to zero and while $\hat{\boldsymbol{\theta}}^*$ remains close to the truth with an incorrect transformation function, $\hat{\boldsymbol{\theta}}$ deviates from the truth and the difference between $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}^*$ becomes greater. Thus a measure of difference between $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}^*$ can be helpful in identifying an incorrect transformation function.

A Hausman test could now be described as follows:

Under the null hypothesis, both $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}^*$ are consistent estimators of $\boldsymbol{\theta_0}$ and both are normally distributed.

$$\sqrt{n}(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta_0}) \xrightarrow{d} \boldsymbol{Z}^* \sim N(\mathbf{0}, \boldsymbol{\Sigma}^*),$$

and

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta_0}) \xrightarrow{d} \boldsymbol{Z} \sim N(\mathbf{0}, \boldsymbol{\Sigma}).$$

Thus under the null, $(\hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}}) \xrightarrow{p} \mathbf{0}$, and it can also be proven that $\sqrt{n}(\hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}})$ is convergent in distribution to a random variable that is normally distributed. In fact, these estimators have the following linear structure that guarantees their normality:

$$\sqrt{n}(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta_0}) = -\boldsymbol{H}^{*-1}\boldsymbol{G}^*,$$

and

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta_0}) = -\boldsymbol{H}^{-1}\boldsymbol{G},$$

where $\boldsymbol{H}$ and $\boldsymbol{H}^*$ are the hessians of these estimators, and $\boldsymbol{G}$ and $\boldsymbol{G}^*$ are the gradients. It can also be shown that the difference between these two estimators follows this linear form and it is also normally distributed.

It should be noted that under the null, and with correct specification of the distribution of the error term (normality), the linear estimator is more efficient than the ordered estimator. In other words, the difference between $\boldsymbol{\Sigma}^*$ and $\boldsymbol{\Sigma}$ is positive definite.

Hausman (1978) proved that once you have a consistent and normally distributed estimator like $\hat{\boldsymbol{\theta}}^*$ whose consistency does not depend on the specification, and another estimator like $\hat{\boldsymbol{\theta}}$ which is normally distributed and consistent under the null, then the difference between these two estimators is normally distributed and centered at zero under the null. Furthermore, he proved that if one of the estimators is more efficient than the other one, then the covariance matrix of the difference between the estimators is equal to the difference between the covariance matrices of the estimators. In other words:

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}^*) \xrightarrow{d} \boldsymbol{W}^{\sim} N(\mathbf{0}, \boldsymbol{\Sigma} - \boldsymbol{\Sigma}^*),$$

and the test statistic has χ^2_k distribution where k is the dimension of $\boldsymbol{\theta_0}$.

$$Test = \sqrt{n}(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}^*)'(\hat{\Sigma} - \hat{\Sigma}^*)^{-1}(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}^*)\sqrt{n}$$

It is important to notice that constructing such a test statistic is very easy. Most of the computer packages are equipped with both the OLS estimator and the ordered probit estimator. All that is needed to construct this test statistic are these two estimators and their covariance matrices.

This test is done under the normality for null and alternative hypotheses; However, the error term may not be normally distributed and the incorrect assumption of normality may lead to inconsistent estimators. Furtunately is it possible to relax the normality assumption in two respects. First, while the OLS estimator is not fully efficient under non-normality, it can still be employed in the test as it will be consistent when the transfromation function is correct irrespective of the error distribution. As to an estimator that is consistent regardless of of whether the assumption transformation function is correct, we note that the ordered model can be estimated semiparametrically without any parametric assumptions on the distribution of the error term. The semiparametric ordered estimator provides a consistent estimator that depends neither on the shape of the transformation function, nor on the distribution of the error term; Thus, the test statistic which employes the semiparametric ordered estimator is independent of the distribution of the error term. Below, we describe the semiparametric ordered estimator and the resulting test statistic.

2.2.2 The semiparametric ordered estimator

The Model, Assumptions, and Identification

As mentioned before, the monotonicity of the transformation function implies that by sorting the dependent variable into some ordered categories, the transformation

function of the dependent variable is also sorted into the same ordered categories (This happens if the transformation function is increasing, and if the transformation function is decreasing they are sorted inversely). For example, if $\boldsymbol{Y}$ is sorted into q quantiles, the fact that the true transformation function is linearly related to the explanatory variables, implies that:

$$t_{j-1} < Y_i \leq t_j \quad \text{iff} \quad T(t_{j-1}) < \beta_0 + \boldsymbol{X}_i\boldsymbol{\beta} + u_i \leq T(t_j) \quad \forall j = 2, \ldots, q+1.$$

In the semiparametric model the distribution of u_i is not known and at least one of the explanatory variables, let's say $\boldsymbol{X_1}$, is continuous. Just like before, T is a monotone function, $\boldsymbol{X}$ is full rank, the observations are i.i.d., and the explanatory variables are independent from $\boldsymbol{u}$.

This ordered model is an example of single-index models. Here:

$$P(Y_i \text{ being in the } j\text{th category}|\boldsymbol{X_i}) = G(\beta_0 + \boldsymbol{X_i}\boldsymbol{\beta}) \tag{2.6}$$

where G is an unknown function, and $v_i = \beta_0 + \boldsymbol{X_i}\boldsymbol{\beta}$ is the index. In this model, the interaction between $\boldsymbol{X_i}$ and the probability of Y_i being in different categories is through this linear index. This means that the single-dimensional vector of $\boldsymbol{v}$ contains the same information as the k dimientional vector of $\boldsymbol{X}$ when it comes to calculate the probability of Y_i being in the jth category. In other words:

$$P(Y_i \text{ being in the } j\text{th category}|\boldsymbol{X_i}) = P(Y_i \text{ being in the } j\text{th category}|v_i = \beta_0 + \boldsymbol{X_i}\boldsymbol{\beta}) \tag{2.7}$$

Under the single-index assumption, any linear transformation of the index contains

the same information as the original index.

$$\begin{aligned} &P(Y_i \text{ being in the } j\text{th category}|v_i = \beta_0 + \boldsymbol{X_i\beta}) \\ =&P(Y_i \text{ being in the } j\text{th category}|av_i + b = a(\beta_0 + \boldsymbol{X_i\beta}) + b) \end{aligned} \tag{2.8}$$

for any two constants $a \neq 0$ and b.

No information is gained or lost by linearly transforming the index in a single-index model, since the function G in (2.6) is unknown and can adjust accordingly. For this reason, the unknown parameters in these single-index models are only identified up to location and scale. The constant, β_0, is not identified and other parameters are only identified up to a multiplicative constant. There are different methods for scale normalization. One of these methods enforces the norm of β to be equal to a constant, for example one, while the other method enforces the coefficient on the first explanatory variable to be equal to a constant, for example one. In this work, the second method is chosen. The index, v_i, is equal to $X_{1i} + \boldsymbol{Z_i\theta}$. Where $\boldsymbol{\theta}' = [\beta_2, \ldots, \beta_k]/\beta_1$ and $\boldsymbol{Z} = (\boldsymbol{X}_2, \ldots, \boldsymbol{X}_k)$.

Although βs are not identified, $\boldsymbol{\theta}$ is identified. Since the β's are not identified, one might wonder how to calculate the conditional probabilities of Y_i falling in different categories. Fortunately, $\boldsymbol{\theta}$ is all we need to calculate such probabilities. According to (2.8), knowing $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are equivalent when it comes to calculating these probabilities. In other words, the same reason that causes $\boldsymbol{\beta}$ not to be identified enables us to use $\boldsymbol{\theta}$ instead to calculate the probabilities. Any question regarding these probabilities (e.g. marginal effects) that can be answered through $\beta_0, \beta_1, \ldots, \beta_k$, can also be answered through $\boldsymbol{\theta}$.

To estimate the identified parameter, the ordered model in equation (2.4) can be

reparametrized as follows:

$$t_{j-1} < Y_i \leq t_j \quad \text{iff} \quad T(t_{j-1}) < \beta_0 + \beta_1(X_{1i} + \boldsymbol{Z}_i\boldsymbol{\theta}) + u_i \leq T(t_j) \quad \forall j = 2, \ldots, q+1. \tag{2.9}$$

To estimate the vector of identified parameters, a likelihood function is maximized. This likelihood is based on the reparametrized model in (2.9).

Estimation

In order to estimate $\boldsymbol{\theta}$ in (2.9), one can no longer rely on knowing the distribution of the error term; However, the probabilty of Y_i falling in different categories can still be calculated without distributional assumptions on the error term. To do so, Bayes' rule can be used.

$$P(Y \text{ being in category } j|v) = \frac{P(Y \text{ being in category } j) * g_1(v|Y \text{ being in category } j)}{g(v)}, \tag{2.10}$$

where g is the density of the index, and g_1 is conditional density of the index based on the category in which Y_i falls in.

The probability of $\boldsymbol{Y}$ being in the jth category can be estimated by sample frequency. g and g_1 can also be estimated by kernel density estimators. Let $P_{ij}(\boldsymbol{\theta})$ be the semiparametric conditional probability of Y_i being in the jth quantile. Then $P_{ij}(\boldsymbol{\theta})$ can be estimated as follows:

$$\hat{P}_{ij}(\boldsymbol{\theta}) = \frac{\sum_{j \neq i}\{t_{j-1} < Y_i \leq t_j\}K_{ij}(\boldsymbol{\theta})}{\sum_{j \neq i} K_{ij}(\boldsymbol{\theta})},$$

where $K_{ij}(\boldsymbol{\theta}) = \frac{1}{h}\phi\Big(\frac{v_i(\boldsymbol{\theta}) - v_j(\boldsymbol{\theta})}{h}\Big)$, with ϕ being a symmetric density around 0. In this dissertation, ϕ is the standard normal density, and h is the bandwidth.

Using these semiparametric probabilities, $\boldsymbol{\theta}$ can be estimated. The estimation method used in this work is that in Klein and Sherman (2002), but employs the bias reduction devices in Klein and Shen (2010).[3] The conditional probability in (2.10) is estimated as the ratio of two estimated densities. In order to consistently estimate such a probability, the density in the denominator should be kept away from zero. This is done by using a trimming function. The estimation is done in two stages. In the first stage, the trimming is on the continuous Xs. In order to prove the normality of the estimator, the gradient of the likelihood at the truth should have zero expectation. This can be achieved utilizing a property the conditional expectations in single-index models proven by Whiteney Newey. Newey proved that the gradient of the conditional expectation of the dependent variable on the linear index has zero expectation when evaluated at the truth. This conditional expectation however is not necessarily equal to zero if the trimming function is based on the explanatory variables. With a trimming function based on the index, Newey's result holds and one can prove the normality of the estimator. Thus a second stage likelihood is maximized where the trimming is on the estimated index using the first stage estimates.

The second stage quasi-log-likelihood is:

$$\hat{Q}_2(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\tau(\hat{v}_i)\sum_{j=2}^{q+1}\{t_{j-1} < Y_i \leq t_j\}\ln(\hat{P}_{ij}(\boldsymbol{\theta})) \tag{2.11}$$

and

$$\hat{\boldsymbol{\theta}}_{ac1} = \arg\max_{\boldsymbol{\theta}} \hat{Q}_2(\boldsymbol{\theta})$$

Where τ is a simple trimming function on the first stage estimated index. Since this trimming function is very commonly used, it is also used in this dissertation. τ

[3] A brief overview of the estimation technique can be found in the appendix.

is zero on the tails of the index and is equal to one at other places. This trimming function is particularly useful if the density of the index is only close to zero on the tails of the distribution. A more complex trimming function could also be used which does not require the density of the index to be close to zero only on the tails of the index. Such a trimming function would be close to zero where the estimated density of the index is close to zero and the trimming function is close to one otherwise. Using such a trimming function, one does not need to worry if the density of the index is close to zero only at extreme values of the index.

It is important to first point out that the above estimator requires a modification so that it is technically correct. In the second stage above, trimming is based on an estimating index. View such trimming as depending on true index, as that is what one proves in establishing the properties of the estimator. Accordingly, such trimming provides protection against small denominators when parameter values are evaluated at the truth. In establishing normality of the estimator, the derivative of the objective function is evaluated at the truth, and such trimming is all that one requires. However, to show that the parameter estimates are consistent, one needs to show that the objective function converges to a fixed function "essentially" for all values of the parameters. Without some adjustment or modification, it is then not possible to establish consistency as trimming only provides "protection" at the true parameter values. To circumvent this difficulty, in the second stage the semiparametric probabilities need to be adjusted as in Klein and Shen (2010) or in Klein et al. (2010).

It is also important to notice that this estimator is independent of the shape of the transformation function. In fact, the transformation function does not even appear in the likelihood (2.11). The presence of a monotone transformation is only reflected in this likelihood through Y_i falling in some ordered categories. Therefore, all properties of this estimator are independent from the transformation function. In particular,

this estimator is consistent whether or not the hypothesized transformation function is the true transformation function.

Here, this estimator, $\hat{\boldsymbol{\theta}}_{ac1}$, is referred to as the "always consistent estimator1 "to emphasize that it is consistent both under the null and under the alternative. It is called $\hat{\boldsymbol{\theta}}_{ac1}$ because later, another estimator will be presented which also stays consistent with any transformation function. That estimator, $\hat{\boldsymbol{\theta}}_{ac2}$, will be called "always consistent estimator2 "

Since the unknown parameters of the semiparametric model are only identified up tp location and scale, and since this semiparametric estimator is going to be compared to the consistent estimator under the null in order to construct the test statistic, throughout this study the model in equation (2.1) is reparametrized in the following way:

$$T(\boldsymbol{Y}) = \beta_0 + \beta_1(\boldsymbol{X}_1 + \boldsymbol{Z}\boldsymbol{\theta}) + \boldsymbol{u}, \tag{2.12}$$

where $\boldsymbol{\theta}' = [\beta_2, \ldots, \beta_k]/\beta_1$ and $\boldsymbol{Z} = (\boldsymbol{X}_2, \ldots, \boldsymbol{X}_k)$

In the rest of the dissertation, the comparison of interest will be between two different estimators for $\boldsymbol{\theta}$, the identified vector of parameters. Under the null hypothesis that the transformation is correct, we employ OLS. This estimator can be viewed as the maximum likelihood estimator, in which case we rewrite the likelihood in (2.2) as:

$$Q(\boldsymbol{\theta}, \beta_0, \beta_1, \sigma) = -\ln \sigma^2 - \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\Big(\frac{T_0(\boldsymbol{Y}_i)}{\sigma} - \frac{\beta_0 + \beta_1(\boldsymbol{X}_{1i} + \boldsymbol{Z}_i\boldsymbol{\theta})}{\sigma}\Big)^2.$$

Notice that the presence of $T_0(Y)$ in this likelihood implies that the consistency of this estimator depends on whether or not $T_0(Y)$ is the true transformation function and unlike the ordered estimator, the linear estimator suffers from inconsistency when the transformation function is not correctly specified.

Let $\hat{\boldsymbol{\theta}}_{nc}$ be the Maximum Likelihood Estimator of $\boldsymbol{\theta}$ in equation (2.12), i.e. :

$$\{\hat{\boldsymbol{\theta}}_{nc}, \hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}\} = \arg \max_{(\boldsymbol{\theta},\beta_0,\beta_1,\sigma)} Q(\boldsymbol{\theta}, \beta_0, \beta_1, \sigma)$$

In this dissertation, this estimator, $\hat{\boldsymbol{\theta}}_{nc}$, is referred to as the "null consistent estimator" to emphasize that its consistency is only guaranteed under the null. When we do not want to assume that the error term is normally distributed, we still can employ OLS but note that it is not fully efficient.

2.2.3 The Semiparametric Least Squared (SLS) estimator

As mentioned before, in order to construct this Hausman test, two estimators are needed. One of them is only consistent under the null while the other one remains consistent even if the null is not true. In the previous section an ordered estimator was used as the latter. However, the ordered estimator is not the only estimator that remains consistent under the alternative. In this section another consistent estimator is proposed that remains consistent under the alternative. Another estimator of equation (2.12) is a semiparametric least squares estimator. This estimator takes advantage of the single-index assumption. In such a model, as it will be explained in more detail later, the relationship between the dependent variable and the explanatory variables is through a linear index. Unlike the objective function of the ordered estimator which does not change with different transformation functions, the SLS objective function changes with different transformation functions. However, as explained in the next few paragraphs, the SLS estimator remains consistent with different transformation functions. The SLS estimator unlike the ordered estimator can take advantage of the curvature of the transformation function whereas in the ordered estimation method, only the ranking of the dependent variables is important. The ordered estimator thus does not use all the information that is available in the

data.

Here is the reason why the single-index assumption can be exploited in transformation function models and why an SLS estimator can be used to estimate the parameters consistently:

Since T is a monotone function, it has an inverse and the equation in (2.1) can be rewritten as follows:

$$\boldsymbol{Y} = T^{-1}(\beta_0 + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}),$$

which means

$$E(\boldsymbol{Y}|\boldsymbol{X}) = G(\beta_0 + \boldsymbol{X}\boldsymbol{\beta}),$$

or

$$\boldsymbol{Y} = G(\beta_0 + \boldsymbol{X}\boldsymbol{\beta}) + \boldsymbol{\epsilon}, \tag{2.13}$$

where G is an unknown function, and the explanatory variables are related to $\boldsymbol{Y}$ through a linear index, $\boldsymbol{v} = \beta_0 + \boldsymbol{X}\boldsymbol{\beta}$.

The single-index assumption is used in this model as well. Therefore, the identification is the same as the ordered model. Since $E[Y|\beta_0+\boldsymbol{X}\boldsymbol{\beta}] = E[Y|a(\beta_0+\boldsymbol{X}\boldsymbol{\beta})+c]$, the parameters in this model are also only identified up to location and scale. For example, if we assume the single-index assumption holds in a wage equation with the explanatory variables of education and experience, it means that in order to calculate the expectation of wage, knowing the exact value of education and experience is not necessary. Knowing the index is enough to calculate such an expectation since the index encapsulates all the necessary information; However, any linear transformation of the index provides the same information as the index itself. For example:

$$E\big[Wage_i|\beta_1 educ_i+\beta_2 exper_i = 3\big] = E\big[Wage_i|2(\beta_1 educ_i+\beta_2 exper_i)-1 = 2*3-1 = 5\big]$$

Another way to explain the identification of the parameters in a single-index model is as follows.

The Single-index assumption implies:

$$E(\boldsymbol{Y}|\boldsymbol{X}) = E(\boldsymbol{Y}|\beta_0 + \boldsymbol{X}\beta) = G(\beta_0 + \boldsymbol{X}\beta) \tag{2.14}$$

Notice that G is an unknown function. Let M be another function which is related to G in the following way:

$$G(x) = M(ax + b)$$

for all values of x and some $a \neq 0$ and b.

In other words:

$$\begin{aligned} E(\boldsymbol{Y}|\boldsymbol{X}) &= G(\beta_0 + \boldsymbol{X}\beta) \\ &= M(a(\beta_0 + \boldsymbol{X}\beta) + b) = M(a\beta_0 + ab + a\boldsymbol{X}\beta) \end{aligned}$$

The functions M and G are both unknown and are indistinguishable from one another. Any linear transformation of the index in (2.14) keeps the model the same since the unknown function can also change accordingly. In fact, the model when the function is G and the unknown parameters are β_0 and $\boldsymbol{\beta}$ is the same as when the function is M and the unknown parameters are $a\beta_0 + b$ and $a\boldsymbol{\beta}$. Thus the parameters in single-index models are only identified up to location and scale.

Using Ichimura's unweighted Semiparametric Least Squares (SLS) estimator (Ichimura, 1993), the normalized unknown parameter, $\boldsymbol{\theta}$, can be estimated. Here, we employ Shen's variant of SLS for two reasons (Shen, 2011). First, the finite sample performance of the estimator can be improved by employing regular kernels and

making a bias correction based on Newey's result as was done above. Second, even if $\boldsymbol{u}$ is homoskedastic in (2.1), the nonlinearity of the transformation function causes $\boldsymbol{\epsilon}$ to be heteroskedatic in equation (2.13). The reason is $E[\boldsymbol{Y}|\boldsymbol{X}]$ is equal to $E[T^{-1}(\beta_0 + \boldsymbol{X\beta} + \boldsymbol{u})|\boldsymbol{X}]$ and unless T is linear, this conditional expectation is not equal to $E[T^{-1}(\beta_0 + \boldsymbol{X\beta})|\boldsymbol{X}] + E[T^{-1}(\boldsymbol{u})|\boldsymbol{X}]$. In other words, the new error term has some interaction with the explanatory variables. Such an interaction will cause the variance of $\boldsymbol{\epsilon}$ not to be constant and be dependent on the explanatory variables. The following example will clarify why a nonlinear transformation causes $\boldsymbol{\epsilon}$ to be heteroskedastic.

Consider the following model:

$$ln(\boldsymbol{Y}) = \beta_0 + \boldsymbol{X}\beta + \boldsymbol{u}$$

where $\boldsymbol{u}$ is homoskedastic with standard normal distribution, $\boldsymbol{u}$ and $\boldsymbol{X}$ are independent, and observations are i.i.d.

Then:

$$\begin{aligned}\boldsymbol{Y} &= e^{\beta_0+\boldsymbol{X}\beta+\boldsymbol{u}} \\ &= e^{\beta_0+\boldsymbol{X}\beta}.e^{\boldsymbol{u}}\end{aligned}$$

Thus:

$$\begin{aligned}E(\boldsymbol{Y}|\boldsymbol{X}) &= E(e^{\beta_0+\boldsymbol{X}\beta}.e^{\boldsymbol{u}}|\boldsymbol{X}) \\ &= e^{\beta_0+\boldsymbol{X}\beta}E(e^{\boldsymbol{u}}|\boldsymbol{X}) \\ &= e^{\beta_0+\boldsymbol{X}\beta}E(e^{\boldsymbol{u}})\end{aligned}$$

Since $e^{\boldsymbol{u}}$ has log normal distribution, its expectation is equal to $e^{\frac{1}{2}}$. Thus:

$$E(\boldsymbol{Y}|\boldsymbol{X}) = G(\beta_0 + \boldsymbol{X}\beta) = e^{\beta_0 + \boldsymbol{X}\beta + \frac{1}{2}}$$

and since

$$\boldsymbol{Y} = G(\beta_0 + \boldsymbol{X}\beta) + \boldsymbol{\epsilon},$$

$\boldsymbol{\epsilon}$ is equal to

$$e^{\beta_0 + \boldsymbol{X}\beta}(e^{\boldsymbol{u}} - e^{\frac{1}{2}}).$$

Variance of such an error term clearly depends on the explanatory variables and even when $\boldsymbol{u}$ is homoskedastic, $\boldsymbol{\epsilon}$ is heteroskedastic.

To deal with the issues raised above, Shen's method of estimating the parameters in transformation and retransformation models is employed (Shen, 2011). First, $\sigma_i^2 = Var(\epsilon_i)$, is estimated. Then the following objective function is minimized:

$$\hat{Q}_2(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \tau(\hat{v}_i)\frac{(\boldsymbol{Y} - \hat{E}[Y_i|v_i = X_{1i} + \boldsymbol{Z_i\theta}])^2}{\hat{\sigma}_i^2} \tag{2.15}$$

Where $\hat{E}[Y_i|v_i = X_{1i} + \boldsymbol{Z_i\theta}]$ is the semiparametric conditional expectation of Y_i on $v_i = X_{1i} + \boldsymbol{Z_i\theta}$, and τ is a trimming function on the first stage estimated index. $\hat{E}[Y_i|v_i = X_{1i} + \boldsymbol{Z_i\theta}]$ has the following structure:

$$\hat{E}[Y_i|v_i = X_{1i} + \boldsymbol{Z_i\theta}] = \frac{\sum_{j\neq i} Y_i K_{ij}(\boldsymbol{\theta})}{\sum_{j\neq i} K_{ij}(\boldsymbol{\theta})},$$

where $K_{ij}(\boldsymbol{\theta}) = \frac{1}{h}\phi\Big(\frac{v_i(\boldsymbol{\theta}) - v_j(\boldsymbol{\theta})}{h}\Big)$, with $\boldsymbol{v} = \boldsymbol{X}_1 + \boldsymbol{Z\theta}$ being the index and ϕ being a symmetric density around 0. Just like the previous estimator, ϕ is chosen to be the standard normal density, and h is the bandwidth.

In order to have an estimate for σ_i^2, first the following objective function is minimized:

$$\frac{1}{n}\sum_{i=1}^{n}\tau(\hat{v}_i)(\boldsymbol{Y}-\hat{E}[Y_i|v_i=X_{1i}+\boldsymbol{Z_i\theta}])^2 \tag{2.16}$$

Although the estimator for $\boldsymbol{\theta}$ is not efficient because of heteroskedastisity, it is consistent. Using this estimator, $\hat{\sigma}_i^2$ can be estimated consistently as follows:

$\hat{\sigma}_i^2 = \hat{E}(\hat{\epsilon}_i^2|X_{1i}+\boldsymbol{Z}_i\hat{\theta})$

where $\hat{\epsilon}_i = Y_i - \hat{E}[Y_i|X_{1i}+\boldsymbol{Z}_i\hat{\boldsymbol{\theta}}]$

There are several different semiparametric estimators for $\boldsymbol{\theta}$ that are consistent and asymptotically normal. The estimator employed here is the same that used for the other semiparametric estimator. This estimator is the one proposed by Klein and Shen (2010).

It is important to notice that this estimator remains consistent as long as the transformation function satisfies the monotonicity condition. Unlike the ordered estimator, the objective function of this estimator presented in equation (2.15) is not independent of the shape of the transformation function. In fact, $\hat{E}[Y_i|v_i=X_{1i}+\boldsymbol{Z_i\theta}]$ can change with different transformation functions. What keeps this estimator consistent is the single-index assumption. In this paper, this estimator, $\hat{\boldsymbol{\theta}}_{ac2}$, is referred to as the "always consistent estimator2" to emphasize that it is consistent both under the null and under the alternative.

Just like the previous estimator, it can be shown that this estimator also has the consistency and the asymptotic normality properties (Klein and Shen, 2010). Thus, to construct the Hausman test, the "always consistent estimator2 " and the "null consistent estimator" can also be used.

2.3 The Test Statistics

Recall that a Hausman specification test is based on a comparison between two estimators. One of them must be consistent even if an incorrect specification is used, while the consistency of the other estimator is guaranteed only if the correct specification is used. The previous section introduced one estimator that is consistent under the null, and two semiparametric and one parametric estimators that remain consistent even when an incorrect transformation function is used. The consistency and normality of these estimators have been proven elsewhere.

Theorem. *Let $\boldsymbol{\theta}_{ac}$ be either of the two "always consistent" estimators discussed previously. Under the null, the difference of the "always consistent" estimator and the "null consistent" estimator is normally distributed and $Test = (\hat{\boldsymbol{\theta}}_{ac} - \hat{\boldsymbol{\theta}}_{nc})' \hat{\boldsymbol{\Omega}}^{-1} (\hat{\boldsymbol{\theta}}_{ac} - \hat{\boldsymbol{\theta}}_{nc})$ has χ^2_k distribution, where $\hat{\boldsymbol{\Omega}}$ is the covariance matrix of the difference of the estimators.*

In fact, both of the "always consistent" estimators and the "null consistent" estimators have the following structure:

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = -\boldsymbol{H}^{-1}\sqrt{n}\boldsymbol{G},$$

where $\boldsymbol{H}$ is the hessian matrix, $\boldsymbol{G}$ is the gradient, and $\sqrt{n}\boldsymbol{G}$ has a normal distribution centered at $\mathbf{0}$. Thus under the null,

$$\sqrt{n}\Big(\hat{\boldsymbol{\theta}}_{ac} - \hat{\boldsymbol{\theta}}_{nc}\Big) \xrightarrow{d} \boldsymbol{Z} \sim N(\mathbf{0}, \boldsymbol{V}),$$

where $\boldsymbol{V}$ is the covariance matrix of $\sqrt{n}(\hat{\boldsymbol{\theta}}_{ac} - \hat{\boldsymbol{\theta}}_{nc})$.

More generally, let $\hat{\boldsymbol{\theta}}_a$ and $\hat{\boldsymbol{\theta}}_b$ be any two estimators for $\boldsymbol{\theta}_0$ with the following structure:

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_a - \boldsymbol{\theta}_0) = \boldsymbol{A}_a \sum_{i=1}^{n} \frac{\boldsymbol{\omega}_{ai}}{n}\sqrt{n} + o_p(1)$$

and

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_b - \boldsymbol{\theta}_0) = \boldsymbol{A}_b \sum_{i=1}^{n} \frac{\boldsymbol{\omega}_{bi}}{n}\sqrt{n} + o_p(1)$$

where $\boldsymbol{A}_a$ and $\boldsymbol{A}_b$ are fixed $k \times k$ matrices with bounded elements, $\boldsymbol{\omega}_{ai}$ and $\boldsymbol{\omega}_{bi}$ are $k \times 1$ vectors, and $\boldsymbol{\omega}_{ai}$ and $\boldsymbol{\omega}_{bi}$ are i.i.d.

Also

$$E(\boldsymbol{\omega}_{ai}) = 0 \quad , \quad E(\boldsymbol{\omega}_{bi}) = 0$$

and

$$Var(\boldsymbol{\omega}_{aj}) < \infty \quad , \quad Var(\boldsymbol{\omega}_{bj}) < \infty \quad \forall j = 1, ..., k$$

In our case, $\boldsymbol{A}_a$ and $\boldsymbol{A}_b$ are the negative inverse hessians and $\boldsymbol{\omega}_{ai}$ and $\boldsymbol{\omega}_{bi}$ are the gradients.

Given the linear structure of $\hat{\boldsymbol{\theta}}_a$ and $\hat{\boldsymbol{\theta}}_b$, the difference between these two estimators has the following structure:

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_a - \hat{\boldsymbol{\theta}}_b) = \sum_{i=1}^{n} \frac{\boldsymbol{A}_a\boldsymbol{\omega}_{ai} - \boldsymbol{A}_b\boldsymbol{\omega}_{bi}}{n}\sqrt{n} + o_p(1)$$

Let $\boldsymbol{\gamma}_i = \boldsymbol{A}_a\boldsymbol{\omega}_{ai} - \boldsymbol{A}_b\boldsymbol{\omega}_{bi}$

$\boldsymbol{\gamma}_i$ is i.i.d. because $\boldsymbol{\omega}_{ai}$ and $\boldsymbol{\omega}_{bi}$ are i.i.d. Also:

$$E(\boldsymbol{\gamma}_{ij}) = 0 \quad \forall j = 1, ..., k$$

because $E(\boldsymbol{\omega}_{ai}) = E(\boldsymbol{\omega}_{bi}) = 0$.

Also $Var(\boldsymbol{\gamma}_j)$ is bounded for all $j = 1, ..., k$ since the variance of $\boldsymbol{\omega}_{aj}$ and $\boldsymbol{\omega}_{bj}$ are bounded and $\boldsymbol{A}_a$ and $\boldsymbol{A}_b$ are matrices with bounded elements.

Thus the difference between these two estimators also has a linear structure. This difference therefore is normally distributed exactly for the same reason $\hat{\boldsymbol{\theta}}_a$ and $\hat{\boldsymbol{\theta}}_b$ are normally distributed.

Using the "null consistent" estimator and the "always consistent" estimator, a Hausman test for the transformation function can be constructed as:

$$Test = (\hat{\boldsymbol{\theta}}_{ac} - \hat{\boldsymbol{\theta}}_{nc})' \hat{\boldsymbol{\Omega}}^{-1} (\hat{\boldsymbol{\theta}}_{ac} - \hat{\boldsymbol{\theta}}_{nc}), \qquad \boldsymbol{\Omega} \equiv \frac{\boldsymbol{V}}{n}, \quad \hat{\boldsymbol{\Omega}} \equiv \frac{\hat{\boldsymbol{V}}}{n} \tag{2.17}$$

$Test$ has a standard quadratic form. Thus, under the null $Test \sim \chi^2_k$, where k is the dimension of $\boldsymbol{\theta}$, which equals the number of identified parameters in the semiparametric models.

Hausman proved that in cases where one of the estimators is efficient, the covariance matrix of the difference of the estimators, $\boldsymbol{\Omega}$, has the following structure under the null hypothesis (Hausman, 1978):

$$\boldsymbol{\Omega} = \boldsymbol{\Sigma_{ac}} - \boldsymbol{\Sigma_{nc}},$$

where $\boldsymbol{\Sigma}_{ac}$ is the covariance matrix of the "always consistent" estimator, and $\boldsymbol{\Sigma}_{nc}$ is the covariance matrix of the "null consistent" estimator. This structure of the covariance matrix enables the researcher to calculate the test statistic very easily. All that is needed to construct the test statistic are the estimators and their covariance matrices. The test statistic then can be calculated as follows.

$$Test = (\hat{\boldsymbol{\theta}}_{ac} - \hat{\boldsymbol{\theta}}_{nc})' (\hat{\boldsymbol{\Sigma}}_{ac} - \hat{\boldsymbol{\Sigma}}_{nc})^{-1} (\hat{\boldsymbol{\theta}}_{ac} - \hat{\boldsymbol{\theta}}_{nc}) \tag{2.18}$$

Calculating the test statistic as done in (2.18) can cause a problem. Even if the

"null consistent" estimator is efficient under the null, its efficiency is not guaranteed if the null is not true. In other words, $\hat{\boldsymbol{\Sigma}}_{ac} - \hat{\boldsymbol{\Sigma}}_{nc}$ need not be positive semidefinite, and the test may return a negative value under the alternative. Another problem is that calculating the covariance matrix of the difference using Hausman's method, relies on the asymptotic efficiency of one of the estimators. Although one of the estimators might be asymptotically more efficient compared to the other one, its variance is not necessarily smaller in a finite sample. If so, the difference between $\hat{\boldsymbol{\Sigma}}_{ac}$ and $\hat{\boldsymbol{\Sigma}}_{nc}$ is not necessarily positive definite even under the null.

To overcome these problems, the efficiency property of the "null consistent" estimator is not exploited in this work. $\hat{\boldsymbol{\Omega}}$ is not calculated as the difference between the estimated covariance matrices, but it is calculated in a way that assures the estimated covariance matrix is positive semidefinite.

Let $\boldsymbol{G}_{ac}$ and $\boldsymbol{G}_{nc}$ be the gradients of the objective functions of the "always consistent" and the "null consistent" estimators; let $\boldsymbol{H}_{ac}$ and $\boldsymbol{H}_{nc}$ be the hessians of the objective functions of these estimators. Using a Taylor expansion, the difference between each estimator and the truth can be written as:

$$\hat{\boldsymbol{\theta}}_{ac} - \boldsymbol{\theta}_0 = -\boldsymbol{H}_{ac}^{-1}(\boldsymbol{\theta}^+)\boldsymbol{G}_{ac}(\boldsymbol{\theta}_0) \qquad \text{where } \boldsymbol{\theta}^+ \in [\hat{\boldsymbol{\theta}}_{ac}, \boldsymbol{\theta}_0],$$
$$\hat{\boldsymbol{\theta}}_{nc} - \boldsymbol{\theta}_0 = -\boldsymbol{H}_{nc}^{-1}(\boldsymbol{\theta}^+)\boldsymbol{G}_{nc}(\boldsymbol{\theta}_0) \qquad \text{where } \boldsymbol{\theta}^+ \in [\hat{\boldsymbol{\theta}}_{nc}, \boldsymbol{\theta}_0],$$

where $\boldsymbol{\theta}^+ \xrightarrow{p} \boldsymbol{\theta}_0$ because both $\hat{\boldsymbol{\theta}}_{ac}$ and $\hat{\boldsymbol{\theta}}_{nc}$ are consistent under the null.

Therefore, the difference in the estimators can be written as:

$$\hat{\boldsymbol{\theta}}_{ac} - \hat{\boldsymbol{\theta}}_{nc} = -\boldsymbol{H}_{ac}^{-1}(\boldsymbol{\theta}^+)\boldsymbol{G}_{ac}(\boldsymbol{\theta}_0) - (-\boldsymbol{H}_{nc}^{-1}(\boldsymbol{\theta}^+)\boldsymbol{G}_{nc}(\boldsymbol{\theta}_0))$$

As a consequence, the variance of $\sqrt{n}(\hat{\boldsymbol{\theta}}_{ac} - \hat{\boldsymbol{\theta}}_{nc})$ has the following structure:

$$\boldsymbol{V} \equiv n\boldsymbol{\Omega} = Var(\sqrt{n}(\hat{\boldsymbol{\theta}}_{ac} - \hat{\boldsymbol{\theta}}_{nc})) =$$

$$
\begin{aligned}
=&E\Big[\sqrt{n}\Big(- \boldsymbol{H}_{ac}^{-1}(\boldsymbol{\theta}_0)\boldsymbol{G}_{ac}(\boldsymbol{\theta}_0) + \boldsymbol{H}_{nc}^{-1}(\boldsymbol{\theta}_0)\boldsymbol{G}_{nc}(\boldsymbol{\theta}_0)\Big) \\
&\sqrt{n}\Big(- \boldsymbol{H}_{ac}^{-1}(\boldsymbol{\theta}_0)\boldsymbol{G}_{ac}(\boldsymbol{\theta}_0) + \boldsymbol{H}_{nc}^{-1}(\boldsymbol{\theta}_0)\boldsymbol{G}_{nc}(\boldsymbol{\theta}_0)\Big)'\Big], \qquad (2.19)
\end{aligned}
$$

To estimate $\boldsymbol{\Omega}$, the estimated gradients and hessians evaluated at the estimates can be used:

$$
\begin{aligned}
\hat{\boldsymbol{V}} \equiv n\hat{\boldsymbol{\Omega}} =&\Big(- \hat{\boldsymbol{G}}_{ac}^{*}(\hat{\boldsymbol{\theta}}_{ac})\hat{\boldsymbol{H}}_{ac}^{-1}(\hat{\boldsymbol{\theta}}_{ac}) + \hat{\boldsymbol{G}}_{nc}^{*}(\hat{\boldsymbol{\theta}}_{nc})\hat{\boldsymbol{H}}_{nc}^{-1}(\hat{\boldsymbol{\theta}}_{nc})\Big)' \qquad (2.20) \\
&\Big(- \hat{\boldsymbol{G}}_{ac}^{*}(\hat{\boldsymbol{\theta}}_{ac})\hat{\boldsymbol{H}}_{ac}^{-1}(\hat{\boldsymbol{\theta}}_{ac}) + \hat{\boldsymbol{G}}_{nc}^{*}(\hat{\boldsymbol{\theta}}_{nc})\hat{\boldsymbol{H}}_{nc}^{-1}(\hat{\boldsymbol{\theta}}_{nc})\Big).
\end{aligned}
$$

Where $\hat{\boldsymbol{G}}^{*}$ is an $n \times k$ matrix of gradients. $\hat{\boldsymbol{G}}_{ij}^{*}$ is the gradient of the ith observation with respect to the jth parameter. $\hat{\boldsymbol{\Omega}}$ is a quadratic form and is therefore positive semidefinite.

The proof that (2.20) is a consistent estimator for (2.19), uses the uniform convergence of the estimated hessians and gradients, and the consistency of both estimators.

To sum up, these test statistics follow χ_k^2 distribution under the null, and never return a negative value. In the next section the performance of these tests in finite samples is presented.

Chapter 3
Finite Sample Properties of the Test Statistics

3.1 The Data Generating Process

To examine the finite sample properties of the test, the behavior of test statistic is studied in Monte Carlo experiments under different designs. First, the distribution of the test statistic is studied when the null is true. In these designs, the transformation function is natural logarithm, and the data are generated in the following way:

$$Y = e^{\boldsymbol{X\beta}+\boldsymbol{u}}.$$

Thus:

$$ln(\boldsymbol{Y}) = \boldsymbol{X\beta} + \boldsymbol{u},$$

where $\boldsymbol{u}$ follows standard normal distribution, $\boldsymbol{X} = (\boldsymbol{X1}, \boldsymbol{X2}, \boldsymbol{X3}, \boldsymbol{X4}, \boldsymbol{X5})$ has five columns. $\boldsymbol{X1}$ through $\boldsymbol{X4}$ are explanatory variables and $\boldsymbol{X5}$ is a vector of ones. In design 1, all the explanatory variables are continuous and follow standard normal distributions. In design 2, all the explanatory variables are continuous except $\boldsymbol{X4}$ which is a standardized binary variable that is either 1 or -1. The probabilities of being 1 or -1 are $\frac{1}{2}$. In design 3, $\boldsymbol{X1}$ and $\boldsymbol{X2}$ are standard normal and $\boldsymbol{X3}$ and $\boldsymbol{X4}$ are standardized binary variables. In design 4 all explanatory variables are standardized binary variables except $\boldsymbol{X1}$ which is standard normal. $\boldsymbol{\beta}' = [-2, 2, 1, -2, 1]$. In these designs $n = 1000$ and number of Monte Carlo replications is 500. These designs were chosen to check the robustness of the test statistic and the estimators in the

presence of discontinuous variables. As mentioned before, in order to estimate the semiparametric probabilities consistently, one need to make sure that these probabilities are calculated where the density of the index is not going to zero. This is usually done through trimming of the continuous index. Although the consistency and normality of these estimators are proven asymptotically, and although the presence of one continuous variable guarantees a continuous index, in finite sample, the absence of enough continuous variables can be problematic. Furthermore, in some empirical analysis researchers do not have many continuous explanatory variables.

3.2 Performance of the Tests Under the Null

To observe the performance of the test under the null, the following hypothesis is tested:

$$H_0 : T(\boldsymbol{Y}) = ln(\boldsymbol{Y}) \quad \text{vs} \quad H_1 : T(\boldsymbol{Y}) \neq ln(\boldsymbol{Y}).$$

Since only three of the parameters are identified in semiparametric estimations, the test statistic should have χ^2_3 distribution under H_0. The results of these Monte Carlo experiments when the ordered estimator is used are shown in Figure 3.1(a) to Figure 3.1(d).

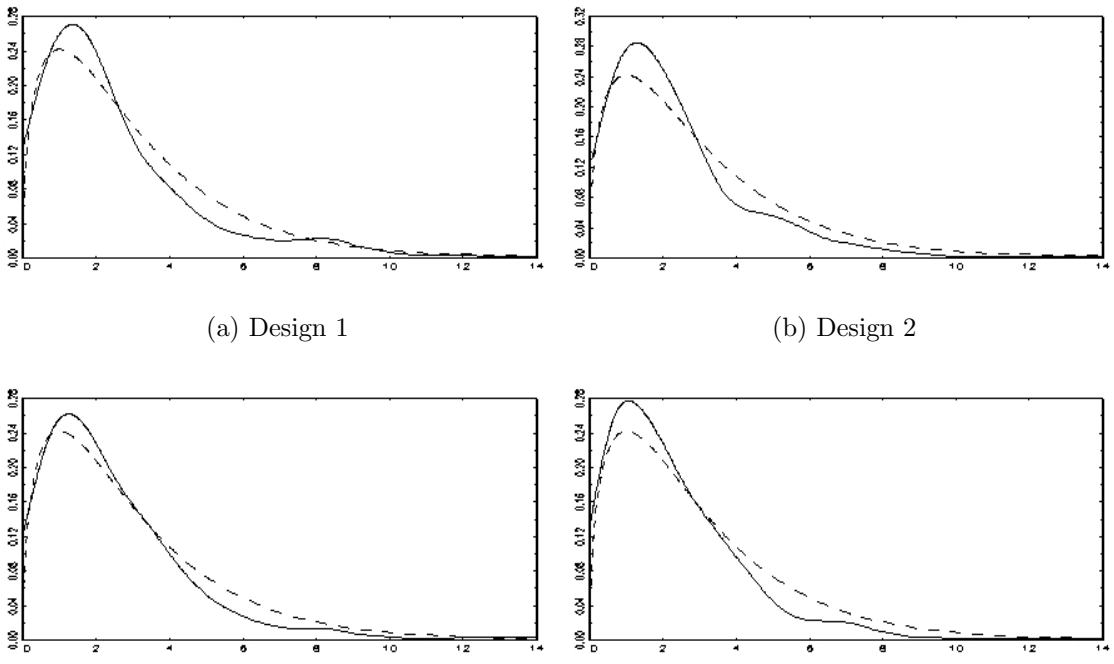
The solid line is the estimated density of the test statistic and the dashed line is the density of χ^2_3.[1]

As seen, the estimated density is close to χ^2_3. By increasing n, the estimated density of the test statistic becomes closer to χ^2_3.

With exactly the same data generating process, the behavior of the test statistic that uses the SLS estimator is studied. The results of these Monte Carlo experiments are shown in Figure 3.2(a) to Figure 3.2(d).

---

[1]Results are robust to changes in the distribution of the error term. These results are available upon request. More designs will be presented in future versions of this paper.



(a) Design 1

(b) Design 2

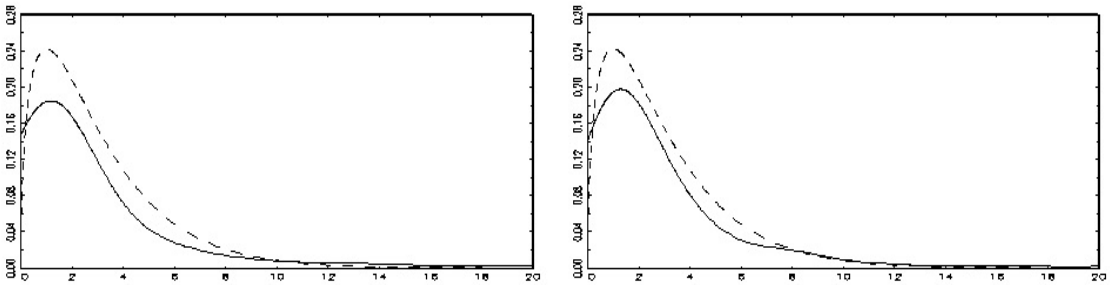
(c) Design 3

(d) Design 4

Figure 3.1: Estimated density of the test statistic using the ordered estimator in different designs when the hypothesized transformation function is logarithm and the null is true

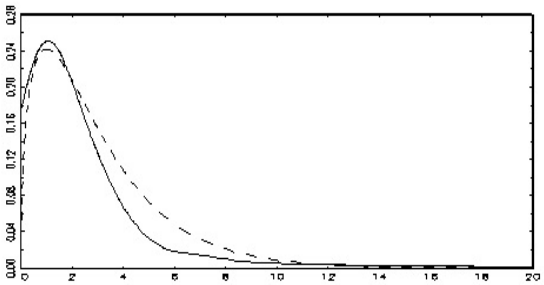
The solid line is the estimated density of the test statistic and the dashed line is the density of χ^2_3. As seen, the estimated density is close to χ^2_3.

To compare these tests under the null, the estimated densities of both test statistics along with the density of χ^2_3 are presented in Figure B.1(a) to Figure B.1(d) in the appendix. The solid line is the estimated density of the test statistic which uses the semiparametric ordered estimator, the dashed line is the estimated density of the test statistic which uses the SLS estimator, and the dotted line is the density of χ^2_3. Both tests perform better in designs number three and four compared to designs number one and two as the estimated densities of the test statistics are closer to the true density in these designs. The test that uses the SLS estimator has a better size except in design number one. Since the comparison of these test statistics vary by design, we cannot state that one of them is better than the other one.

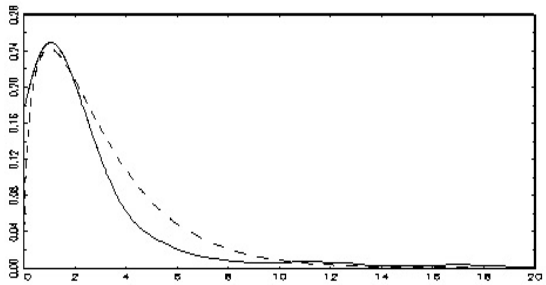


(a) Design 1

(b) Design 2



(c) Design 3



(d) Design 4

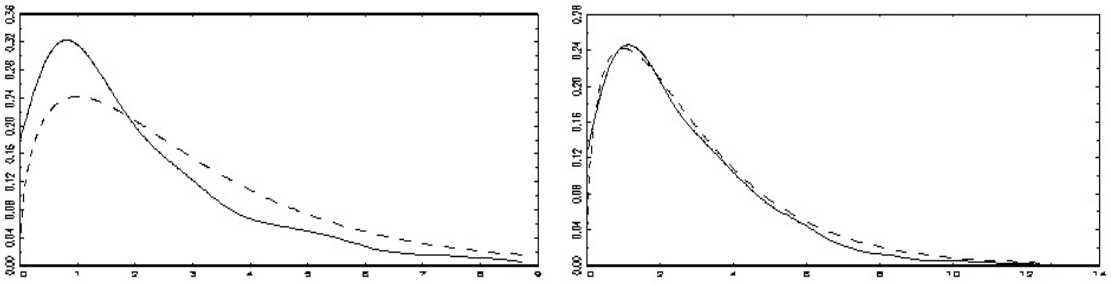
Figure 3.2: Estimated density of the test statistic using the SLS estimator in different designs when the hypothesized transformation function is logarithm and the null is true

Unlike the test that uses the ordered estimator where the "always consistent estimator" is independent of the transformation function, the SLS estimator depends on the transformation function. Although the SLS estimator remains consistent with different transformation functions (as long as they are monotone), to observe the behavior of the test statistic in finite sample with another transformation function, the following Monte Carlo experiment is implemented.

$Y = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}$

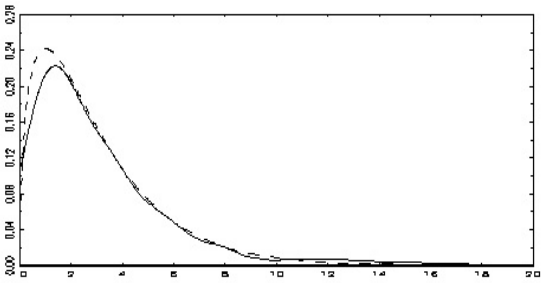
In this case the true transformation function is the identity function. The following hypothesis is tested.

$$H_0 : T(\boldsymbol{Y}) = \boldsymbol{Y} \quad \text{vs} \quad H_1 : T(\boldsymbol{Y}) \neq \boldsymbol{Y}.$$

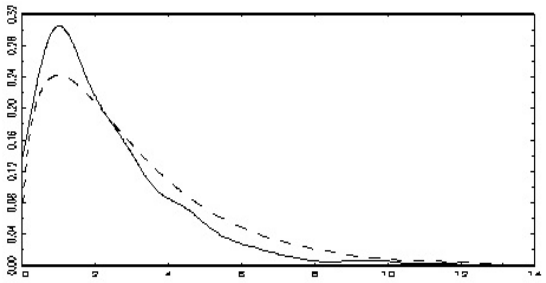


(a) Design 1

(b) Design 2



(c) Design 3



(d) Design 4

Figure 3.3: Estimated density of the test statistic using the SLS estimator in different designs when the hypothesized transformation function is the identity and the null is true

As seen in Figure 3.3(a) to Figure 3.3(d), the estimated density of the test statistic is very close to χ^2_3. In this example, since the transformation function is the identity function, the estimator does not suffer the heteroskedasticity that the nonlinear transformation functions induce. If one wants to test the hypothesis that the transformation function is the identity function with the assumption that the error terms are homoskedastic, the test can also be done without calculating $\hat{\sigma}_i^2$. One can use the minimizer of the (2.16) as the "always consistent estimator".

3.3 Performance of the Tests Under the Alternative

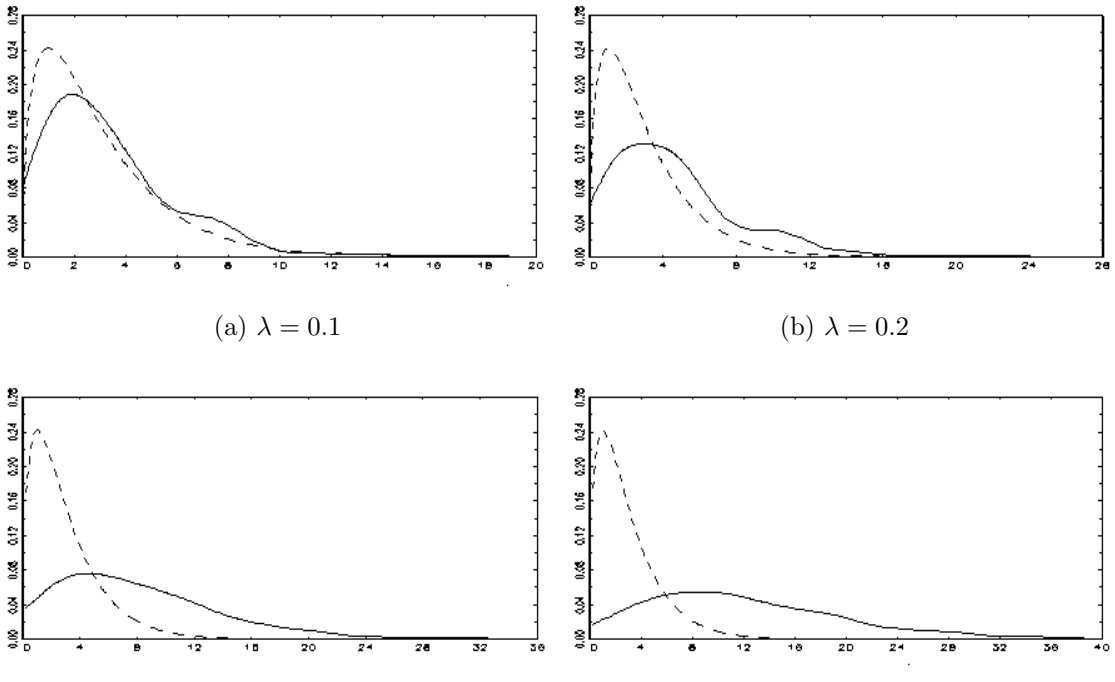
To observe the power of the test, the data is generated under the alternative hypothesis. In the first example, the hypothesized transformation function is natural

logarithm but the data is generated as follows:

$$\boldsymbol{Y} = \begin{cases} e^{\boldsymbol{X\beta}+\boldsymbol{u}} & \text{if} \quad \boldsymbol{X\beta}+\boldsymbol{u} \leq 0 \\ (\lambda(\boldsymbol{X\beta}+\boldsymbol{u})+1)^{\frac{1}{\lambda}} & \text{if} \quad \boldsymbol{X\beta}+\boldsymbol{u} > 0, \end{cases} \tag{3.1}$$

where $(\lambda(\boldsymbol{X\beta}+\boldsymbol{u})+1)^{\frac{1}{\lambda}}$ is the inverse of the Box-Cox transformation for positive values of $\boldsymbol{X\beta}+\boldsymbol{u}$. So

$$\boldsymbol{X\beta}+\boldsymbol{u} = \begin{cases} ln(\boldsymbol{Y}) & \text{if} \quad \boldsymbol{X\beta}+\boldsymbol{u} \leq 0 \\ \frac{\boldsymbol{Y}^{\lambda}-1}{\lambda} & \text{if} \quad \boldsymbol{X\beta}+\boldsymbol{u} > 0 \end{cases} \tag{3.2}$$



(a) $\lambda = 0.1$ (b) $\lambda = 0.2$

(c) $\lambda = 0.5$ (d) $\lambda = 1.0$

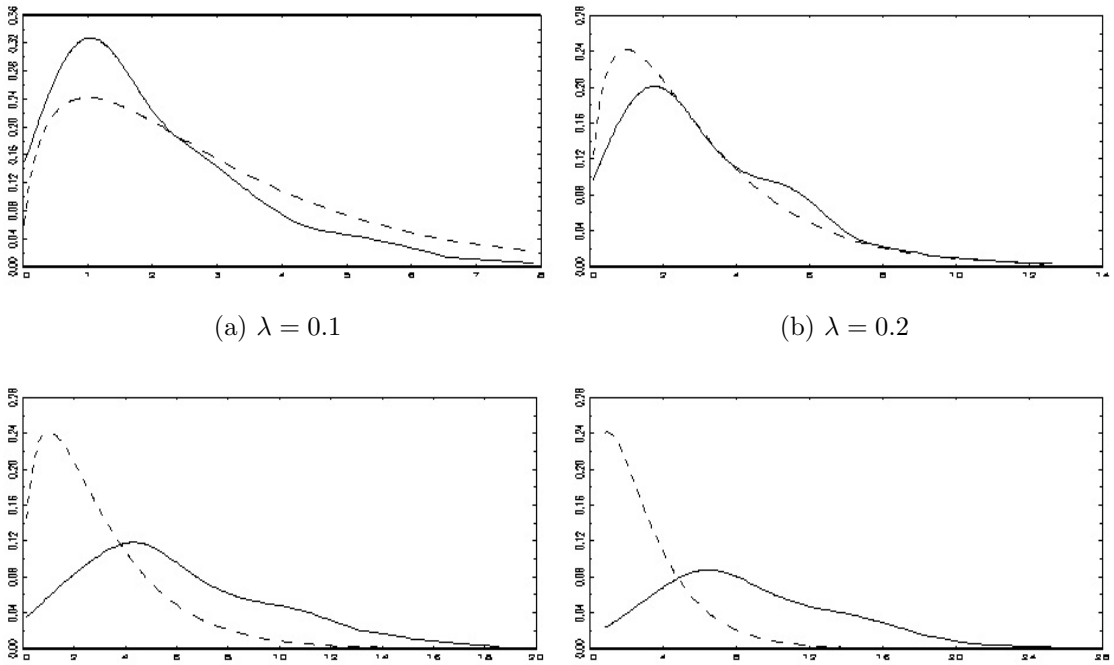
Figure 3.4: Estimated density of the test statistic using the ordered estimator when the true transformation function is Box-Cox but the hypothesized function is logarithm

For small values of λ, $\frac{\boldsymbol{Y}^{\lambda}-1}{\lambda}$ is close to $ln(\boldsymbol{Y})$, and the null is not expected to be rejected very frequently. On the other hand, as λ gets large, the true transformation

function deviates from the hypothesized one. Therefore, for high values of λ, the null hypothesis is expected to be rejected frequently.

The results of these Monte Carlo studies using the ordered estimator are shown in Figure 3.4(a) to Figure 3.4(d). The solid line is the estimated density of the test statistic and the dashed line is the density of χ^2_3.

As seen in these figures, the solid line and the dashed line are close when λ is small. As λ increases, the true transformation function deviates from the hypothesized transformation function, the density of the test statistic deviates from χ^2_3, and the null hypothesis is more likely to be rejected.



(a) $\lambda = 0.1$ (b) $\lambda = 0.2$

(c) $\lambda = 0.5$ (d) $\lambda = 1.0$

Figure 3.5: Estimated density of the test statistic using the SLS estimator when the true transformation function is Box-Cox but the hypothesized function is logarithm

This Monte Carlo experiment is repeated using the SLS estimator. The results of these Monte Carlo studies are shown in Figure 3.5(a) to Figure 3.5(d). The solid line is the estimated density of the test statistic and the dashed line is the density of χ^2_3.

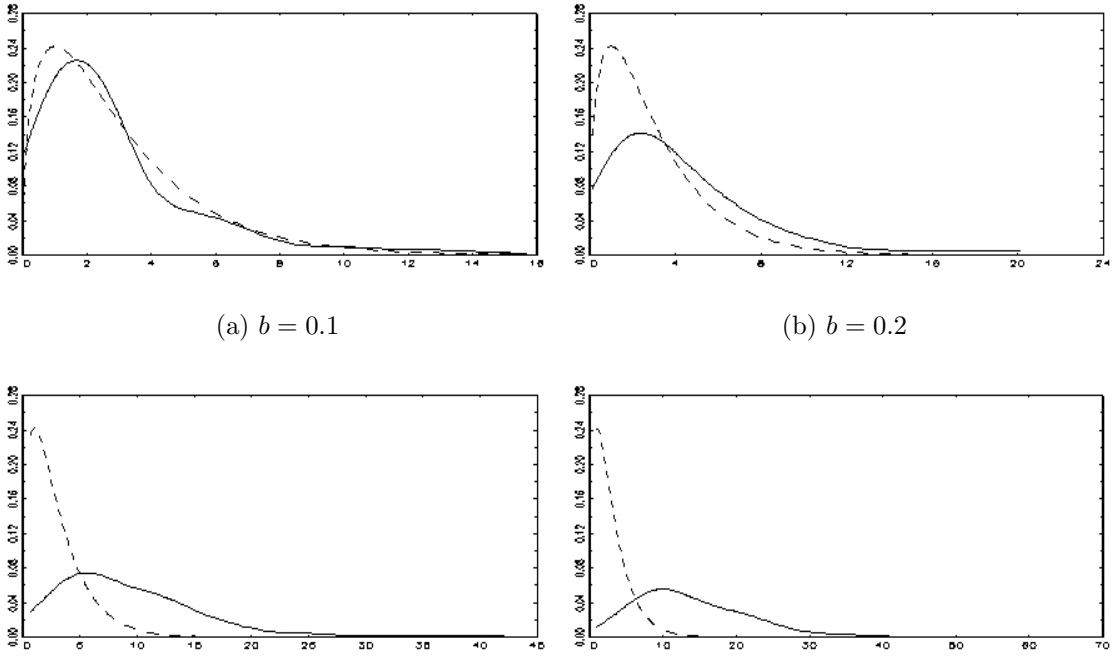
When the true transformation function is close to logarithm, the estimated density

of the test statistic is close to χ^2_3 but as λ increases and the true transformation function deviates from logarithm, the null hypothesis is rejected more frequently and the estimated density of the test statistic deviates from χ^2_3.

To see the behavior of the test when the hypothesized transformation function is the identity function, data is generated as follows:

$$\boldsymbol{Y} = a(\boldsymbol{X\beta} + \boldsymbol{u})^3 + b(\boldsymbol{X\beta} + \boldsymbol{u})^2 + (\boldsymbol{X\beta} + \boldsymbol{u}), \tag{3.3}$$

and the following hypothesis is tested: $H_0 : T(\boldsymbol{Y}) = \boldsymbol{Y}$ vs $H_1 : T(\boldsymbol{Y}) \neq \boldsymbol{Y}$.



(a) $b = 0.1$ (b) $b = 0.2$

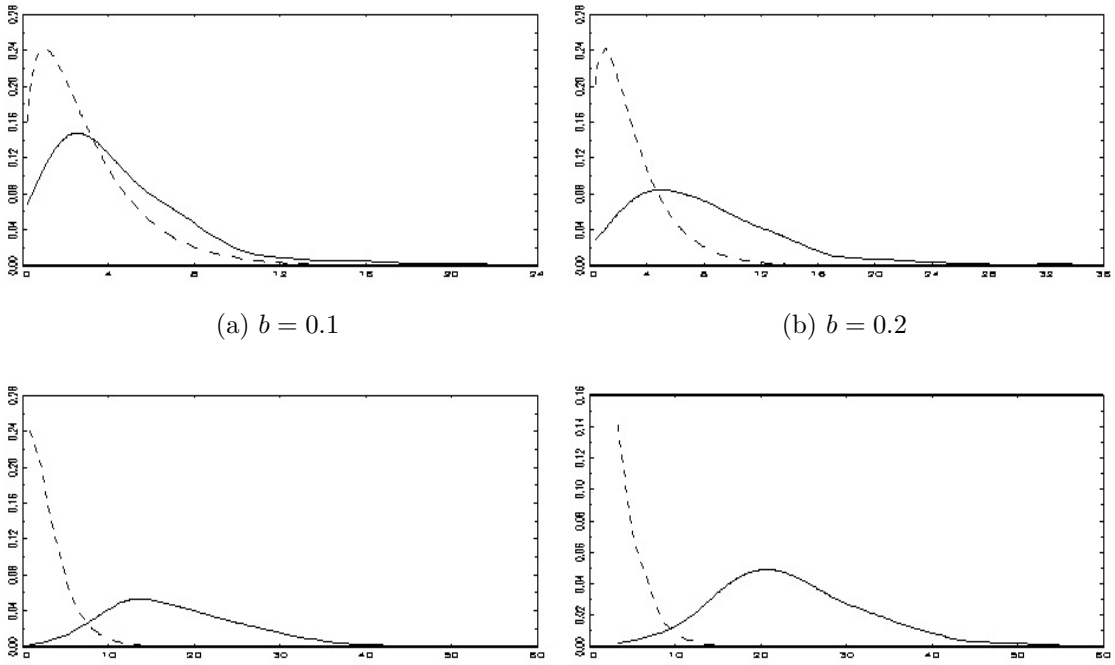
(c) $b = 0.5$ (d) $b = 1.0$

Figure 3.6: Estimated density of the test statistic using the ordered estimator when the true transformation function is an inverse cubic function but the hypothesized function is the identity

As $a, b \to 0$, the transformation function is getting closer to the hypothesized transformation function and as a and b deviate from zero, the true transformation function deviates from the hypothesized one. To make sure that the transformation function is monotone, a is chosen to be greater than $\frac{b^2}{3}$. In this study, $a = \frac{1.1b^2}{3}$.

The results are shown in Figure 3.6(a) to Figure 3.6(d). As the true transformation function deviates from the hypothesized one, the null is more likely to be rejected.

The behavior of the test statistic that uses the SLS estimator is also studied in some Monte Carlo experiments. The results are shown in Figure 3.7(a) to Figure 3.7(d). As the true transformation function deviates from the hypothesized one, the null is more likely to be rejected.



(a) $b = 0.1$ (b) $b = 0.2$

(c) $b = 0.5$ (d) $b = 1.0$

Figure 3.7: Estimated density of the test statistic using the SLS estimator when the true transformation function is an inverse cubic function but the hypothesized function is the identity

To compare these tests, we show the size adjusted power results for these examples. First, the size adjusted powers of the example where the true transformation function is a modified Box-Cox are presented in Tables 3.1 and 3.2.

As seen in Tables 3.1 and 3.2, the test which uses the semiparametric ordered estimator seems to have slightly higher power compared to the test that uses the SLS estimator.

Table 3.1: Size-adjusted power for the test statistic which uses the semiparametric ordered estimator when the true transformation function is the one in (3.2)

	Significance level		
	10%	5%	1%
$\lambda = 0.1$	23.0%	13.5%	2.0%
$\lambda = 0.2$	37.5%	22.5%	12.5%
$\lambda = 0.5$	66.0%	55.5%	34.0%
$\lambda = 1.0$	83.5%	75.5%	61.0%

Table 3.2: Size-adjusted power for the test statistic which uses the SLS estimator when the true transformation function is the one in (3.2)

	Significance level		
	10%	5%	1%
$\lambda = 0.1$	8.5%	0.5%	0.0%
$\lambda = 0.2$	28.0%	6.0%	0.0%
$\lambda = 0.5$	60.5%	31.0%	1.5%
$\lambda = 1.0$	84.0%	69.0%	25.0%

The size adjusted powers are also presented in Tables 3.3 and 3.4 for the example where the true transformation function is the inverse cubic function.

As seen in Tables 3.3 and 3.4, the test which uses the SLS estimator has clearly higher power compared to the test that uses the ordered estimator. Thus the power comparison is inconclusive as one of the tests has slightly higher power in one example and the other test has a higher power in the other example.

Table 3.3: Size-adjusted power for the test statistic which uses the semiparametric ordered estimator when the true transformation function is the one in (3.3)

	Significance level		
	10%	5%	1%
$\lambda = 0.1$	15.0%	7.5%	2.0%
$\lambda = 0.2$	31.0%	19.0%	6.5%
$\lambda = 0.5$	71.5%	56.5%	37.5%
$\lambda = 1.0$	89.5%	85.0%	63.0%

Table 3.4: Size-adjusted power for the test statistic which uses the SLS estimator when the true transformation function is the one in (3.3)

	Significance level		
	10%	5%	1%
$\lambda = 0.1$	24.5%	12.0%	3.5%
$\lambda = 0.2$	58.0%	42.5%	18.5%
$\lambda = 0.5$	97.5%	94.5%	75.5%
$\lambda = 1.0$	98.5%	98.5%	93.0%

3.4 Conclusion

In summary, in the designs where the data are generated under the null, the distribution of the test statistics is close to their asymptotic distributions. Therefore, the hypothesized transformation function is not rejected frequently. On the other hand, when the data are generated under the alternative, the tests reject the hypothesized transformation function more frequently and specially in cases where the true transformation function is very different from the hypothesized one. Thus, in the Monte Carlo experiments presented, the tests behave as expected under the null and under the alternative. The properties of the two test statistics are very similar. While the test which uses the semiparametric ordered estimator is slightly more powerful in the example that the true transformation function is Box-Cox, the test that uses the SLS estimator is more powerful when the true transformation function is an inverse cubic function.

Chapter 4
An Application of a Transformation Function Test

As mentioned before, in many studies, a known function of the dependent variable is used as the left-hand-side variable. In many cases, there is no theoretical justification for the employed transformation function. The use of the logarithmic transformation function in particular simplifies the calculation of elasticities. This transformation is also justified in the literature in cases where the variable of study is positive and skewed to the right. Diminishing the effect of heteroskedasticity is another justification for the logarithmic transformation; However, although this particular transformation may solve some of these problems or simplifies some calculations, this does not mean that it is the true transformation function. In other words, the logarithmic transformation of the dependent variable may not be linearly related to the explanatory variables. Using this function can still lead to inconsistent estimators if logarithm is not the true transformation function. In this chapter, the validity of logarithmic transformation is tested in an example.

4.1 Example: Explaining Crime Rates

In most studies of city level crime, logarithm of crime is used as the left-hand-side variable. The fact that the theory does not say anything about the functional form of the dependent variable, and the common belief that crimes are underreported, make the study of this transformation function particularly interesting. Reviewing the literature, no discussion of the choice of the functional form of the dependent variable

was found. The logarithmic transformation therefore appears to be the preferred specification because it is the most used one, it simplifies some calculations (e.g. elasticities), and it reduces the problems of heteroskedasticity and having large outliers; However, the choice of this particular function seems arbitrary and not discussed in any study.

4.2 Data

The data used are from Metropolitan Statistical areas between 2000 and 2008. The data on crime are from Uniform Crime Report and the data on explanatory variables are from the Current Population Survey (CPS). The explanatory variables are as follows. $ln(population)$, is the natural log of the population. Most studies assume a linear relationship between $ln(crime)$ and $ln(population)$. The expected sign of the coefficient on $ln(population)$ is obviously positive. In some other studies, $ln(crime\ rate) = ln(crime/population)$ is used as the left-hand-side variable. *Age1525* represents the fraction of people in the metropolitan area that are between 15 and 25 years old. *Age2535* is the fraction of people in the metropolitan area that are between 25 and 35 years old. These two variables are generated because of the belief that different age groups may have different propensity to crime. The effect of these age groups may also be different across different categories of crime. *Median(*$ln(Wage)$) is used to capture the relationship between income and crime rate. We expect the coefficient on this variable to be negative. *Black* is the fraction of blacks, and *Hispanic* is the fraction of Hispanics in the metropolitan area. The race and ethnicity variables are generated to see if the crime rate is different in the cities that have higher percentage of blacks or hispanics. Based on the previous studies, we expect the coefficient on these two variables to be positive. *Drop out* is the fraction of people between 19 and 30 who do not have at least 12 years of schooling. *College* is the fraction of people who have at least some college education. These two variables

show the relationship between crime and the distribution of different education levels in the city. *Employed* is the fraction of the population in each metropolitan area that is employed. The expected sign of the coefficient on this variable is negative. *Central* is the fraction of people who live in central city. We expect a positive coefficient on this variable. The means and standard deviations of the dependent and explanatory variables can be found in Table C.1.

4.3 Testing the Logarithmic Transformation

The test is done using the semiparametric ordered estimator. As observed in the previous chapter the performance of these two tests are very similar so the same results are expected if the SLS estimator is used. The test is done for different groups of crimes. FBI categorizes crimes into two main categories: property crimes and violent crimes. Property crimes are burglary, larceny, and motor vehicle theft. Violent crimes are homicide, rape, robbery, and aggravated assault. These categories of crime are often separately studied since the nature of these crimes are different and the relationship between crimes and explanatory variables might vary across different categories. Researchers often use as dependent variable both the logarithm of total crime and the logarithm of a specific crime within the same analysis. In this work, first the plausibility of the logarithmic transformation function is tested in specific categories.

The test statistic and the P-values can be seen in Table 4.1.[1] As seen, the logarithmic transformation function is not rejected at 5% significance level in any of the categories of property crimes.

The test is done for different categories of violent crimes as well. The results can be seen in Table 4.2. For all categories of violent crimes except robbery, the logarithm

[1] The linear and the semiparametric ordered estimates of different categories of property crimes, violent crimes, and total crimes can be seen in Table C.2 to Table C.4.

Table 4.1: Testing the Log Transformation Function for Different Categories of Property Crime

	Burglary	Larceny	Motor Vehicle Theft
Critical Value at 5% significance level	16.9190	16.9190	16.9190
Test	10.1407	7.4642	7.1910
P-Value	0.3392	0.5889	0.6172

transformation function is not rejected at 5% significance level.

Table 4.2: Testing the Log Transformation Function for Different Cartegories of Violent Crime

	Homicide	Assault	Rape	Robbery
Critical Value at 5% significance level	16.9190	16.9190	16.9190	16.9190
Test	4.0215	5.3463	8.7787	19.9294
P-Value	0.9100	0.8031	0.4579	0.0184

It should also be noted that most of the coefficients have expected signs in both the linear and the ordered estimations. These results are shown in Table C.2 to Table C.4. The coefficients on the variables of $ln(population)$, $Median(ln(Wage))$, *Black*, *Hispanic*, *Employed*, and *Central* all have the expected signs. The coefficients on *Drop out* and *College* sometimes do not have the expected signs or they are not significant. The cities with larger fraction of people between 15 and 25, on average have less property crimes and violent crimes (except rape) and the coefficient on *Age2535* is either positive or insignificant in all categories.

Obviously, if logarithm is the transformation function for burglary, larceny, and motor vehicle theft, it cannot possibly also be the transformation function for property crimes as a whole, as by definition property crime is the sum of these three categories of crimes.

If

$$ln(\mathbf{burglary}) = \boldsymbol{X\beta_1} + \boldsymbol{\epsilon_1},$$

and

$$ln(\textbf{larceny}) = \boldsymbol{X\beta_2} + \boldsymbol{\epsilon_2},$$

and

$$ln(\textbf{motor vehicle theft}) = \boldsymbol{X\beta_3} + \boldsymbol{\epsilon_3},$$

then it is wrong to model the property crimes as:

$$ln(\textbf{property crimes}) = \boldsymbol{X\beta} + \boldsymbol{\epsilon}. \tag{4.1}$$

Similarly, if the researcher assumes that the transformation function for property crimes and violent crimes is logarithm, i.e.:

$$ln(\textbf{violent crimes}) = \boldsymbol{X\alpha} + \boldsymbol{v} \tag{4.2}$$

as in (4.1) and (4.2), then the transformation function for total crimes cannot possibly be logarithm as well.

It is, however, a common practice in this literature to use the log specification of property crimes, violent crimes, and total crimes as dependent variables. Accordingly, subject to qualifications discussed below, results for tests on the log transformation of broadly defined categories of crime are reported in Table 4.3. Before presenting the results, it should be noted that if you assume that the specification for each category of crime is a non-linear function (e.g. logarithmic), the transformation function for the aggregate category might not even exist, and if it does, it is not necessarily a monotone function. In such a case, the test presented in this dissertation is not appropriate as the consistency of the ordered estimator depends on the monotonicity of the transformation function. In general, a researcher who uses a transformation function model is implicitly making three claims about the model: That there is a function of the dependent variable which is linearly related to the explanatory variables, that

such a function is monotone, and that such a function is known. Throughout this work, the first two claims are accepted, and the third claim is tested. There are, however, examples in which one should doubt the existence or the monotonicity of the left-hand-side function. For example, if there is evidence that there is a logarithmic transformation for specific categories, then the existence of a transformation function for an aggregate category should be doubted.

From Table 4.3, the logarithm transformation function is not rejected for violent crimes but it is rejected at 10% significance level both for property crimes and total crimes. Notice that probably the logarithmic transformation for violent crimes is not rejected as most of these crimes are assaults, and two of the categories are very small (homicide and rape). Besides assault and the other two small categories, robbery is the other component of violent crimes, and it is not adequately represented with a log function.

Table 4.3: Testing the Log Transformation Function for Different Cartegories of Total Crime

	Violent	Property	Total
Critical Value at 5% significance level	16.9190	16.9190	16.9190
Test	4.4599	14.9661	16.9058
P-Value	0.8786	0.0919	0.0502

To conclude, if the transformation function exists and is monotone, more broadly defined categories of crime are less likely to have logarithmic representations. Unfortunately in many papers on city level crime rate, the same nonlinear transformation function (logarithm)is used for very specific crimes as well as very broadly defined crimes. In Glaeser and Sacerdote (1999), $\ln(crimes)$ is used as the dependent variable and similarly $\ln(assault)$ and $\ln(rape)$ are also used as dependent variables. Levitt (1997) and McCrary (2002) estimate the elasticity of crime with respect to the number of sworn police officers. Both authors use the logarithm of seven specific crimes,

the logarithm of all property crimes, and the logarithm of all violent crimes as their dependent variables. In Kelly (2000), the elasticity of crime with respect to inequality is calculated, and the logarithm of different categories of crime is used as the dependent variable as well as the logarithm of all property crimes. In Spenkuch (2010), the elasticity of crime with respect to different groups of immigrants is calculated and the logarithm of different categories of crime is used as dependent variables as well as the logarithm of all violent crimes and the logarithm of all property crimes. In Lott and Mustard (1997), the effect of carrying concealed weapons on crime rate is calculated, and again the logarithm of different categories of crime is used as dependent variables as well as the logarithm of all violent crimes and the logarithm of all property crimes. Although it is very convenient to calculate elasticities in a $log - log$ model, it does not justify the use of an incorrect transformation function. The logarithmic transformation function may be close to the true transformation function in some categories of crime, but it should not be used for more broadly defined categories since a misspecified transformation function will lead to inconsistent estimators and although calculating the elasticities are easier with this transformation function, such elasticities are not consistently estimated.

If the use of logarithmic transformation is justified in specific categories, the elasticity of a broader category can be calculated without imposing the logarithmic transformation. For example if:

$$ln(\boldsymbol{Burglary}) = ln(Median(\boldsymbol{Wage}))\beta_1 + \boldsymbol{\epsilon_1}$$

and

$$ln(\boldsymbol{Larceny}) = ln(Median(\boldsymbol{Wage}))\beta_2 + \boldsymbol{\epsilon_2}$$

and

$$ln(\boldsymbol{Motor\ Vehicle\ Theft}) = ln(Median(\boldsymbol{Wage}))\beta_3 + \boldsymbol{\epsilon_3}$$

then elasticity of $Property\ Crimes$ with respect to $Median(Wage)$ in the ith city is:

$$\frac{Burglary_i}{Property\ Crimes_i}\beta_1 + \frac{Larceny_i}{Property\ Crimes_i}\beta_2 + \frac{Motor\ Vehicle\ Theft_i}{Property\ Crimes_i}\beta_3 \qquad (4.3)$$

This method does not assume constant elasticity. In fact, elasticity is constant only if $\beta_1 = \beta_2 = \beta_3$. For example in our data, $\hat{\beta}_1 = -0.4369$, $\hat{\beta}_2 = -0.6488$, and $\hat{\beta}_3 = -0.4311$. These results suggest that elasticity of property crimes with respect to median wage is close to -0.64 for cities where larceny theft is a big fraction of property crimes and elasticity is close to -0.43 in cities where motor vehicle theft or burglary are bigger fractions of property crimes. When a constant elasticity is assumed for property crimes, the calculated elasticity is -0.5631 and the model assumes that this elasticity is the same for all cities. The method in (4.3) will deliver consistent estimates of the elasticities and does not rely on the logarithmic transformation of the property crimes. It should therefore be preferred in cases where both broad and specific categories of crimes are analyzed.

Chapter 5
Conclusion

In this dissertation two Hausman specification tests are presented where the source of the misspecification is a wrong transformation function of the dependent variable. These tests are constructed by comparing an estimator that is only consistent if the true transformation function is used to two estimators that remain consistent regardless of what the true transformation function is. One of the latter estimators is a semiparametric ordered estimator and the other one is a semiparametric least squares estimator. The ordered estimator exploits the fact that the transformation function is a monotone function thus without knowing the curvature of the function one can estimate the unknown parameters by sorting the dependent variable in ordered categories. The semiparametric least squares estimator utilizes the single-index assumption and the face that the transformation function is monotone thus invertible.

The asymptotic distribution of these test statistics is discussed and the finite sample properties of the test are presented in Monte Carlo experiments. The test statistic has the desired distrubution under the null and the alternative. In examples where the true transformation function is close to the true transformation function, the empirical distribution of the test statistic is close to its asymptotic distribution. On the other hand, as the true transformation function deviates from the true transformation function, the empirical distribution of these test statistics deviate from the asymptotic distribution and the null hypothesis of the assumed transformation function is rejected more frequently.

At last, the ordered test is used to check if the log transformation is appropriate

in models studying crime across metropolitan areas. It was shown that although for particular types of crimes the log function is appropriate, the same log function should not be used for broadly defined categories of crimes.

Appendix A

Estimation of the Semiparametric Ordered Model (Klein and Shen, 2010)

The first stage of estimator is the maximizer of the following Quasi- Log-Likelihood:

$$\hat{Q}_1(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\tau(X_i)\sum_{j=2}^{q+1}\{t_{j-1} < Y_i \leq t_j\}\ln(\hat{P}_{ij}(\theta)) \tag{A.1}$$

In order to have a normally distributed estimator, the gradient of the likelihood should have zero expectation at the truth. If the trimming function only depends on $\boldsymbol{X}$, then Newey's result about zero expectation of the derivative of the conditional expectation at the truth does not hold anymore thus the gradient of the likelihood does not necessarily have zero expectation. We can overcome this problem by maximizing a second stage likelihood function in which the trimming function depends on the estimated index rather than $\boldsymbol{X}$ (Klein, 1993, Klein and Sherman, 2002).

In this likelihood, the semiparametric probabilities are calculated slightly differently using an inside trimming function. The semiparametric probability can be written as the ratio of two estimated densities.

$$\hat{P}(\{t_{j-1} < Y_i \leq t_j\} = 1|X) = \frac{\hat{f}_i}{\hat{g}_i}$$

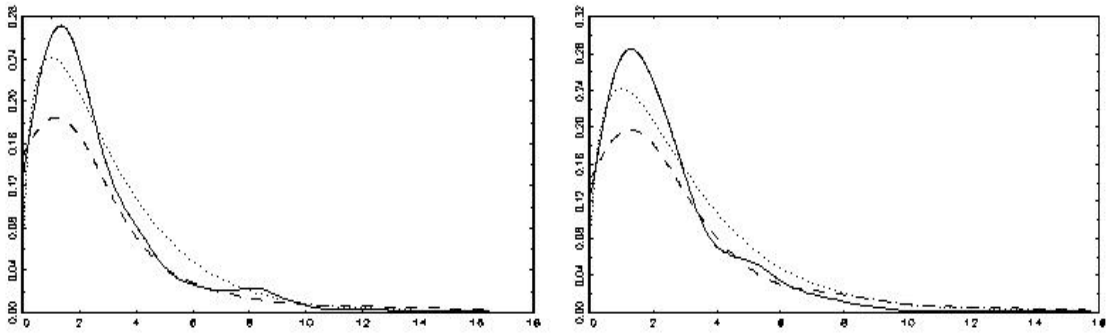
To keep the denominator away from zero, $\hat{P}$ is approximated by $\frac{\hat{f}_i+\Delta_{f_i}}{\hat{g}_i+\Delta_{g_i}}$ where Δs go to zero rapidly when the estimated densities are away from zero and go to zero slowly when the estimated densities go to zero. In other words, Δ_{f_i} and Δ_{g_i} can be chosen

such that $\frac{\hat{f}_i+\Delta_{f_i}}{\hat{g}_i+\Delta_{g_i}} \xrightarrow{p} \frac{f}{g}$.

The window that is used in (2.11) is not the optimal window to estimate the semiparametric probabilities. Hence, the bias is not going to zero at the optimal rate. To overcome this problem, Klein and Shen (2010) suggest a "bias- correction" method that is also used in this paper.

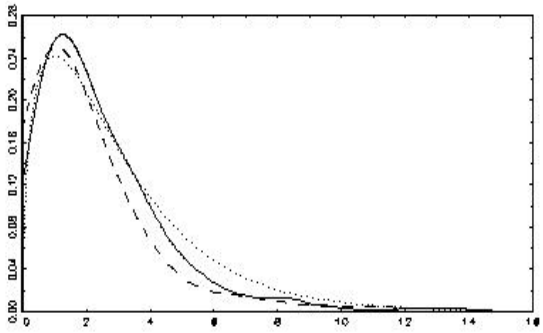
Appendix B

Figures

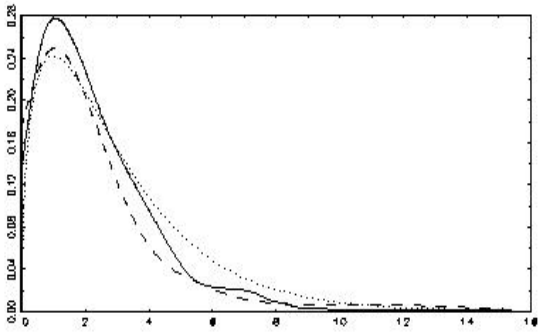


(a) Design 1

(b) Design 2



(c) Design 3



(d) Design 4

Figure B.1: Estimated density of the test statistics using both the SLS estimator and the ordered estimator in different designs when the hypothesized transformation function is logarithm and the null is true

The solid line is the estimated density of the test statistic which uses the semi-parametric ordered estimator, the dashed line is the estimated density of the test statistic which uses the SLS estimator, and the dotted line is the density of χ^2_3.

Appendix C

Tables

Table C.1: Descriptives

	Means
ln(population)	13.06
	(0.97)
Age1525	0.14
	(0.03)
Age2535	0.13
	(0.03)
Median(ln(wage))	2.62
	(0.18)
Black	0.11
	(0.12)
Hispanic	0.17
	(0.21)
Drop out	0.02
	(0.01)
College	0.37
	(0.07)
Employed	0.72
	(0.06)
Central	0.24
	(0.24)
Burglary	65,076
	(94,164)
Larceny theft	208,658
	(300,643)
Motor vehicle theft	39,169
	(78,791)
Aggravated assault	26,976
	(53,365)
Homicide	515
	(1,133)
Rape	2,664
	(3,617)
Robbery	14,348
	(34,678)
Property crime	312,903
	(464,014)
Violent crime	44,503
	(91,397)
Total crime	357,407
	(550,647)
N	1,225

Table C.2: Parametric and Semiparametric Results of Different Categories of Property Crime

	Burglary	Larceny	Motor Vehicle Theft
ln(population)	0.9544***	0.9662***	1.2241***
	(0.0100)	(0.0085)	(0.0154)
Age1525	-1.2431***	-0.9623***	-2.0375***
	(0.3544)	(0.2469)	(0.3894)
Age2535	0.8280**	1.1613***	1.0418**
	(0.3834)	(0.2911)	(0.4126)
Median(ln(wage))	-0.4369***	-0.6488***	-0.4311***
	(0.0773)	(0.0538)	(0.0831)
Black	1.0300***	0.4800***	0.5766***
	(0.1005)	(0.0827)	(0.1224)
Hispanic	0.2290***	0.1395***	0.3780***
	(0.0605)	(0.0468)	(0.0722)
Drop out	2.8267***	-0.1029	3.0441***
	(0.8452)	(0.6172)	(0.9570)
College	0.7421***	0.6190***	0.7927***
	(0.1966)	(0.1404)	(0.2112)
Employed	-1.0287***	-0.1028	-0.9671***
	(0.2190)	(0.1469)	(0.2175)
Central	0.2330***	0.2956***	0.3761***
	(0.0493)	(0.0352)	(0.0546)
Constant	-0.6560**	0.4226**	-3.8329***
	(0.2780)	(0.2019)	(0.2779)
sigma	0.3528***	0.2696***	0.4889***
	(0.0077)	(0.0057)	(0.0101)
R^2	0.884	0.928	0.871
$\bar{R}^2$	0.883	0.927	0.870
	Semiparametric Ordered Model Results		
Age1525	-1.4619**	-0.4651	-1.6503***
	(0.5838)	(0.5104)	(0.6130)
Age2535	1.5579***	1.0770**	1.2647**
	(0.6005)	(0.5258)	(0.6372)
Median(ln(wage))	-0.5334***	-0.5194***	-0.6366***
	(0.1161)	(0.1048)	(0.1272)
Black	1.2752***	0.4589***	0.6752***
	(0.1544)	(0.1377)	(0.1707)
Hispanic	0.0857	0.1792*	0.2240*
	(0.1211)	(0.0976)	(0.1178)
Drop out	0.5352	-0.1499	1.1580
	(1.2847)	(1.1927)	(1.3633)
College	0.4389	0.2465	0.6687**
	(0.2992)	(0.2584)	(0.3177)
Employed	-0.9993***	-0.1737	-1.0090***
	(0.3033)	(0.2775)	(0.3463)
Central	0.1933**	0.3430***	0.4565***
	(0.0756)	(0.0681)	(0.0844)
N	1225	1225	1225
Critical Value	16.9190	16.9190	16.9190
Test	10.1407	7.4642	7.1910
P-Value	0.3392	0.5889	0.6172

Standard errors in parentheses.
Significance levels: *: 10%, **: 5%, ***: 1%.

Table C.3: Parametric and Semiparametric Results of Different Categories of Violent Crime

	homicide	Assault	Rape	Robbery
ln(population)	1.1745***	1.0487***	0.9155***	1.3011***
	(0.0201)	(0.0172)	(0.0125)	(0.0159)
Age1525	-1.0252**	-0.9867**	0.9993**	-0.5457
	(0.4902)	(0.4663)	(0.4263)	(0.3671)
Age2535	0.6872	-0.1968	1.2700***	0.4419
	(0.4699)	(0.4726)	(0.4468)	(0.3884)
Median(ln(wage))	-0.0256	-0.3388***	-0.3590***	0.0244
	(0.1045)	(0.0977)	(0.0896)	(0.0755)
Black	2.1825***	1.0718***	0.0642	1.5289***
	(0.1457)	(0.1200)	(0.1234)	(0.0966)
Hispanic	0.4627***	0.7726***	-0.1163	0.2046***
	(0.0971)	(0.0926)	(0.0811)	(0.0768)
Drop out	1.4270	2.8181**	2.4254**	2.1263**
	(1.1776)	(1.2458)	(1.1531)	(0.9556)
College	-0.6727***	0.6679***	0.7242***	-0.5576***
	(0.2404)	(0.2464)	(0.2467)	(0.1865)
Employed	-0.8976***	-0.9278***	-0.4231*	-0.4009*
	(0.2741)	(0.2597)	(0.2386)	(0.2047)
Central	0.1717***	0.1481**	0.4516***	0.1880***
	(0.0646)	(0.0621)	(0.0624)	(0.0435)
Constant	-7.9813***	-2.9591***	-4.5132***	-6.3911***
	(0.3543)	(0.3544)	(0.3182)	(0.2584)
sigma	0.5791***	0.4864***	0.3990***	0.4744***
	(0.0099)	(0.0093)	(0.0077)	(0.0077)
R^2	0.822	0.833	0.844	0.889
$\bar{R}^2$	0.820	0.832	0.842	0.888
	Semiparametric Ordered Model Results			
Age1525	-1.4680**	-0.5147	1.0611	-0.4004
	(0.6597)	(0.5840)	(0.6580)	(0.5493)
Age2535	0.4062	0.9942	0.8522	1.1763**
	(0.6987)	(0.6541)	(0.7052)	(0.5705)
Median(ln(wage))	-0.1660	-0.2023*	-0.5672***	0.2678**
	(0.1411)	(0.1236)	(0.1461)	(0.1190)
Black	2.1346***	0.7925***	0.3469*	1.5028***
	(0.1969)	(0.1849)	(0.1881)	(0.1578)
Hispanic	0.2906**	0.6061***	-0.0246	0.1434
	(0.1309)	(0.1267)	(0.1279)	(0.1085)
Drop out	0.9422	2.0210	0.3370	0.7918
	(1.5148)	(1.3835)	(1.5417)	(1.2478)
College	-0.4534	0.3953	1.2706***	-0.5185*
	(0.3479)	(0.3203)	(0.3535)	(0.2862)
Employed	-0.8975**	-1.2711***	-0.5800	0.1026
	(0.3691)	(0.3390)	(0.3646)	(0.3102)
Central	0.2233**	0.0840	0.6175***	0.2922***
	(0.0952)	(0.0810)	(0.0916)	(0.0764)
N	1212	1225	1225	1225
Critical Value	16.9190	16.9190	16.9190	16.9190
Test	4.0215	5.3463	8.7787	19.9294
P-Value	0.9100	0.8031	0.4579	0.0184

Standard errors in parentheses.
Significance levels: *: 10%, **: 5%, ***: 1%.

Table C.4: Parametric and Semiparametric Results of Different Categories of Violent Crime, Property Crime, and Total Crime

	Violent	Property	Total
ln(population)	1.0961***	0.9861***	0.9971***
	(0.0132)	(0.0082)	(0.0081)
Age1525	-0.6587*	-1.1101***	-1.0483***
	(0.3563)	(0.2468)	(0.2405)
Age2535	0.1342	1.0489***	0.9375***
	(0.3701)	(0.2879)	(0.2810)
Median(ln(wage))	-0.1890**	-0.5631***	-0.5200***
	(0.0747)	(0.0539)	(0.0524)
Black	1.2250***	0.6060***	0.6709***
	(0.0924)	(0.0775)	(0.0739)
Hispanic	0.5516***	0.2160***	0.2516***
	(0.0699)	(0.0487)	(0.0476)
Drop out	2.4200**	0.9312	1.1020*
	(0.9562)	(0.6276)	(0.6149)
College	0.2316	0.6950***	0.6507***
	(0.1901)	(0.1448)	(0.1414)
Employed	-0.6188***	-0.4049***	-0.4349***
	(0.1992)	(0.1465)	(0.1429)
Central	0.1777***	0.2822***	0.2669***
	(0.0469)	(0.0348)	(0.0336)
Constant	-3.4377***	0.5093**	0.4059**
	(0.2672)	(0.1998)	(0.1949)
sigma	0.3934***	0.2687***	0.2641***
	(0.0071)	(0.0057)	(0.0056)
R^2	0.892	0.932	0.936
$\bar{R}^2$	0.891	0.931	0.935
	Semiparametric Ordered Model Results		
Age1525	-0.2141	-0.7581	-0.4308
	(0.5639)	(0.4747)	(0.5035)
Age2535	0.6129	1.4248***	1.3062**
	(0.5825)	(0.5104)	(0.5231)
Median(ln(wage))	-0.1769	-0.5943***	-0.5385***
	(0.1210)	(0.0974)	(0.1031)
Black	0.9956***	0.6021***	0.5438***
	(0.1631)	(0.1270)	(0.1375)
Hispanic	0.3973***	0.1158	0.1790*
	(0.1114)	(0.0901)	(0.0961)
Drop out	2.3520*	-0.6367	-0.9084
	(1.2720)	(1.0834)	(1.1236)
College	-0.0294	0.0288	-0.0747
	(0.3483)	(0.2408)	(0.2527)
Employed	-0.5638**	-0.1797	-0.2118
	(0.3105)	(0.2622)	(0.2740)
Central	0.1551**	0.3673***	0.3316***
	(0.0774)	(0.0626)	(0.0677)
N	1225	1225	1225
Critical Value	16.9190	16.9190	16.9190
Test	4.4599	14.9661	16.9058
P-Value	0.8786	0.0919	0.0502

Standard errors in parentheses.
Significance levels: *: 10%, **: 5%, ***: 1%.

References

Anil Bamezai, Jack Zwanziger, Glenn A. Melnick, and Joyce M. Mann. Price competition and hospital cost growth in the united states (1989-1994). *Health Economics*, 8(3):233–243, 1999.

Peter J. Bickel and Kjell A. Doksum. An analysis of transformations revisited. *Journal of the American Statistical Association*, 76(374):pp. 296–311, 1981.

George E.P. Box and David R. Cox. An analysis of transformations. *Journal of the Royal Statistical Society*, 26(2):211–252, 1964.

Edward L. Glaeser and Bruce Sacerdote. Why is there more crime in cities? *The Journal of Political Economy*, 107(6):pp. S225–S258, 1999.

J. A. Hausman. Specification tests in econometrics. *Econometrica*, 46(6):1251–1271, 1978.

Joel L Horowitz. Semiparametric estimation of a regression model with an unknown transformation of the dependent variable. *Econometrica*, 64(1):103–37, 1996.

Joel L. Horowitz. *Semiparametric Methods in Econometrics*. Springer-Verlag New York, Inc., 1998.

Hidehiko Ichimura. Semiparametric least squares (sls) and weighted sls estimation of single-index models. *Journal of Econometrics*, 58(1-2):71 – 120, 1993.

J. A. John and N. R. Draper. An alternative family of transformations. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 29(2):pp. 190–197, 1980.

Morgan Kelly. Inequality and crime. *The Review of Economics and Statistics*, 82(4): pp. 530–539, 2000.

Roger Klein, Chan Shen, and Francis Vella. Triangular semiparametric models featuring two dependent endogenous binary outcomes. Working Paper, 2010.

Roger W. Klein. Specification tests for binary choice models based on index quantiles. *Journal of Econometrics*, 59(3):343 – 375, 1993.

Roger W. Klein and Chan Shen. Bias Correction in Testing and Estimating Semiparametric, Single Index Models. *Econometric Theory*, 26(6):1683–1718, December 2010.

Roger W. Klein and Robert P. Sherman. Shift restrictions and semiparametric estimation in ordered response models. *Econometrica*, 70(2):663–691, 2002.

Steven D. Levitt. Using electoral cycles in police hiring to estimate the effect of police on crime. *The American Economic Review*, 87(3):pp. 270–290, 1997.

John R Jr Lott and David B Mustard. Crime, deterrence, and right-to-carry concealed handguns. *Journal of Legal Studies*, 26(1):1–68, January 1997.

James G. MacKinnon and Lonnie Magee. Transforming the dependent variable in regression models. *International Economic Review*, 31(2):pp. 315–339, 1990.

Justin McCrary. Using electoral cycles in police hiring to estimate the effect of police on crime: Comment. *The American Economic Review*, 92(4):pp. 1236–1243, 2002.

Jacob Mincer. *Schooling, experience, and earnings*. NBER New York, 1974.

James C. Robinson and Harold S. Luft. The impact of hospital market structure on patient volume, average length of stay, and the cost of care. *Journal of Health Economics*, 4(4):333 – 356, 1985.

Chan Shen. Semiparametric transformation and retransformation models. Working Paper, 2011.

Bernard W Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, 1992.

Jörg L. Spenkuch. Understanding the impact of immigration on crime. Mpra paper, May 2010.

Peter Zweifel, Stefan Felder, and Markus Meiers. Ageing of population and health care expenditure: a red herring? *Health Economics*, 8(6):485–496, 1999.

Vita

Kaveh Akram

2003-2011 Ph. D. in Economics, Rutgers University

1999-2002 M. Sc. in Socioeconomic Systems Engineering - Economics from Institute for Research on Planning and Development (IRPD) Tehran-IRAN.

1994-1999 B. Sc. in Electrical Engineering - Power from Sharif University of Technology - Tehran, Iran

2007-2011 Part Time Lecturer, Department of Economics, Rutgers University

2004-2007 Teaching assistant, Department of Economics, Rutgers University