Inference in Nonparametric Instrumental Variables with Partial Identification

Andres Santos*
Department of Economics
University of California, San Diego
e-mail: a2santos@ucsd.edu
October 11, 2007

Abstract

This paper develops methods for hypothesis testing in a nonparametric instrumental variables (IV) setting within a partial identification framework. I construct and derive the asymptotic distribution of a test statistic for the hypothesis that at least one element of the identified set satisfies a conjectured restriction. This procedure can be used to test for features of the model that may be identified even when the true model is not. This framework can also be employed to construct confidence regions for functionals of the elements of the identified set, such as consumer surplus and price elasticity of demand at a point. I apply this procedure to study Engel curves for gasoline and ethanol in Brazil. For both ethanol and gasoline I fail to reject that there are log-linear Engel curves in the identified set. In addition, I derive confidence regions for the level and slope of the Engel curves at the sample average as well as for the compensated variation associated with a price change in gasoline.

KEYWORDS: Instrumental variables, nonparametric hypothesis testing, partial identification.

*I would like to thank my advisors Peter Hansen, Aprajit Mahajan and Frank Wolak for their generous and helpful guidance. I am especially grateful to Azeem Shaikh and Frank Wolak for their suggestions. I am also indebted to Liran Einav, Sriniketh Nagavarapu, Luigi Pistaferri, Peter Reiss and seminar participants where this paper was presented for their helpful comments. The financial support of the Stanford Institute for Economic and Policy Research in the form of a doctoral dissertation fellowship is gratefully acknowledged.
1 Introduction

Empirical work in economics is often concerned with the estimation and analysis of econometric models that are derived from the behavior of optimizing agents. The underlying structural relations typically imply some of the regressors are endogenous, so that these models do not fit the classical regression framework, but are instead of the form:

\[ Y = \theta_0(X) + \epsilon \]  

(1)

where \( E[\epsilon|X] \neq 0 \). Instrumental variables (IV) methods have become immensely popular in econometrics as they allow for the estimation of the unknown function \( \theta_0(x) \). Under the parametric assumption that \( \theta_0(x) \) is of the form \( \theta(x, \beta_0) \) for some vector \( \beta_0 \), and the appropriate identifying conditions, IV estimators are consistent and asymptotically normally distributed.

The potential misspecification of parametric methods makes it desirable to extend the IV approach to a more flexible nonparametric framework. Unfortunately, the theoretical study and empirical implementation of nonparametric IV methods has faced two important challenges. First, as originally pointed out in Newey & Powell (2003), nonparametric identification is hard to attain. Without parametric assumptions limiting the possible forms of \( \theta_0(x) \), the burden of identification falls entirely on the instrument. Nonparametric identification of \( \theta_0(x) \) requires the availability of an instrument satisfying conditions far stronger than the usual covariance restrictions needed in the parametric case. Second, few methods are available for hypothesis testing in a nonparametric IV setting. Ai & Chen (2003) establish the asymptotic normality of the parametric component in a semiparametric specification, which allows for hypothesis testing on the parametric component. Horowitz (2007) shows the asymptotic normality of the fully nonparametric estimator proposed in Hall & Horowitz (2005), and is in this way able to build confidence intervals for the level of \( \theta_0(x) \).

In the present paper I extend the existing results in the nonparametric IV literature by constructing a family of test statistics for a wider set of hypotheses than was previously available. In addition, since nonparametric identification is not only hard to attain but also to test, my analysis does not assume identification. Instead I apply many of the prevalent ideas in the partial identification literature, see Manski (1990, 2003), to the nonparametric IV setting. For any dependent variable, endogenous regressor and instrument triplet \( (Y, X, Z) \), I study the nonparametric IV problem as one of partial identification. The identified set is defined to be:

\[ \Theta_0 = \{ \theta(x) \in \Theta : E[Y - \theta(X)|Z] = 0 \} \]  

(2)

This is the relevant definition of endogeneity in this context, in contrast to \( E[X\epsilon] \neq 0 \), because \( E[\epsilon|X] = 0 \) implies \( \theta_0(x) \) is nonparametrically identified in (1) while \( E[\epsilon X] = 0 \) does not.
where $\Theta$ is a nonparametric set of functions that by assumption contains $\theta_0(x)$. The elements of $\Theta_0$ are those functions in $\Theta$ that are consistent with the conditional moment assumption $E[\epsilon|Z] = 0$, which is the instrument exclusion restriction in a nonparametric setting.

For a family of restrictions on functions $\theta(x)$, I show how to construct a test statistic for the null hypothesis that these restrictions are satisfied by at least one element of $\Theta_0$. Special cases of the restrictions we can test for include point values of a function and its derivatives, shape restrictions such as monotonicity, concavity, economies of scale, economies of scope and parametric specification tests against a nonparametric alternative. In the special case where $\theta_0(x)$ is identified, these tests reduce to tests on $\theta_0(x)$. The testing framework developed in this paper can be used to build informative confidence intervals for functionals on $\theta_0(x)$, even when such functionals are not identified. For example, suppose $\theta_0(x)$ is a demand function and we are interested in its consumer surplus $CS(\theta_0)$. We can construct a set $\hat{\Theta}$ such that if $\theta(x) \in \Theta_0$ has consumer surplus $CS(\theta)$, then $P(CS(\theta) \in \hat{\Theta}) \rightarrow 1 - \alpha$ for $\alpha$ the size of the test. This concept is referred to as confidence regions for functions of the identifiable parameters in Romano & Shaikh (2006a), which elaborates on the concept of confidence regions for identifiable parameters, as originally explored in Imbens & Manski (2004) for the case of the mean.

Another advantage of not requiring identification is that a test for whether at least one element of $\Theta_0$ satisfies the hypothesized restrictions may still be consistent for the null hypothesis that $\theta_0(x)$ satisfies the tested restrictions and correctly reject even when $\theta_0(x)$ is not identified. For example, in a parametric specification test against a nonparametric alternative, if no element of the identified set is a member of the hypothesized parametric family, then we can conclude $\theta_0(x)$ does not fit the parametric specification either. On the other hand, without identification the test may not be consistent for the null hypothesis that $\theta_0(x)$ satisfies the hypothesized restrictions and asymptotically fail to reject because some other element of $\Theta_0$ satisfies the restrictions we are testing for. This is not necessarily a weakness of the procedure, but the result of the instrument employed being unable to nonparametrically identify not only $\theta_0(x)$ but also the features we are testing for.

I apply the test statistics derived in this paper to study Engel curves for gasoline and ethanol in Brazil with endogenous total expenditures. For both ethanol and gasoline I fail to reject the null hypothesis that there is at least one log-linear Engel curve in the identified set, which is a commonly used parametric specification. I build log-linear and nonparametric confidence regions for the level and derivative of the Engel curves at the sample mean. The nonparametric confidence regions for the level and slope of the ethanol Engel curve are considerably different from those obtained assuming log-linearity. This discrepancy, however, might be explained by a small sample
size. In contrast, for the gasoline Engel curve I find a log-linear parametrization implies confidence regions for the level that are similar to the nonparametric ones. The analogous confidence regions for the slope of the gasoline Engel curve, however, differ substantially. While the upper bounds of the confidence regions are almost identical, their lower bounds are significantly different. I find that these differences do not translate into dissimilar confidence regions for the compensated variation associated with a price change in gasoline. The results illustrate that the impact of a particular parametrization are dependent on what the hypothesis of interest is.

The literature on nonparametric IV is fairly recent. Newey & Powell (2003) propose a nonparametric estimator for $\theta_0(x)$ and show its consistency. The authors solve the ill-posed inverse problem by obtaining compactness through smoothness assumptions on $\theta_0(x)$, an insight I rely upon in this paper. Darolles, Florens & Renault (2003) and Hall & Horowitz (2005) propose alternative consistent estimators and obtain rates of convergence. Ai & Chen (2003) analyze semiparametric specifications and establish efficient estimators for the parametric component. Blundell, Chen & Kristensen (2004) use a semiparametric specification and provide the first empirical application of semiparametric IV together with “low level” identification conditions and new rates of convergence results. Severini & Tripathi (2006) explore the identification of linear functionals of $\theta_0(x)$ when $\theta_0(x)$ is not identified, while Severini & Tripathi (2007) obtain an efficiency bound for the estimation of these functionals. Santos (2007) constructs $\sqrt{N}$ consistent estimators for such continuous linear functionals on $\theta_0(x)$. Horowitz (2006) derives a parametric specification test against nonparametric alternatives for model (1). In related work, Newey, Powell & Vella (1999), Chesher (2003, 2005, 2007) and Imbens & Newey (2006) explore estimation and identification in triangular systems. Work estimating informative bounds when instruments fail to provide identification has also been done in the treatment effects literature. See for example Manski & Pepper (2000), Heckman & Vytlacil (2001), Manski (2003), Shaikh & Vytlacil (2005) and references therein.

The remainder of the paper is organized as follows. Section 2 expands on the partial identification problem. Section 3 develops the test statistic that allows us to test whether at least one element of the identified set satisfies a set of hypothesized restrictions. Section 4 addresses two aspects of the implementation of the test: the utilization of sieves to approximate the functional space and the use of subsampling if the statistic is not pivotal. In Section 5 I apply these methods to study the Brazilian market for ethanol and gasoline. Section 6 briefly concludes. All proofs are contained in a mathematical appendix.
2 Partial Identification

This section elaborates on the problem of partial identification in the nonparametric IV setting. I first review the requirements for identification and then proceed to motivate why it may be advantageous to adopt a framework that does not assume $\theta_0(x)$ is nonparametrically identified.

2.1 The Identified Set

Identification in a nonparametric IV setting was originally discussed in Newey & Powell (2003). In this section I review their results before discussing the identified set. For expositional purposes the discussion of nonparametric identification will proceed by drawing analogies to the familiar linear setting. The two models are:

$$Y = X'\beta_0 + \epsilon \quad \quad Y = \theta_0(X) + \epsilon$$  \hspace{1cm} (3)

where $Y \in \mathbb{R}$, $X \in \mathbb{R}^k$, $\epsilon$ is unobservable and $\beta_0 \in \mathbb{R}^k$ is an unknown vector, while $\theta_0(x) : \mathbb{R}^k \rightarrow \mathbb{R}$ is an unknown function. The appropriate definition of endogeneity in the linear model is $E[X\epsilon] \neq 0$, because if $E[X\epsilon] = 0$, then $\beta_0$ can be consistently estimated by OLS. In contrast, the pertinent definition of endogeneity in the nonparametric model is $E[\epsilon|X] \neq 0$, because if $E[\epsilon|X] = 0$ then $\theta_0(x)$ can be consistently estimated by a nonparametric regression. In order to simplify the discussion of the linear model I will also assume, in this section only, that $Z$ and $X$ have the same dimension. The strategy to obtain identification in the linear case is to multiply both sides of the first equation in (3) by $Z$ and take expectations. For the nonparametric case we instead take conditional expectations on both sides of the second equation in (3). In this way we obtain:

$$E[ZY] = E[ZX']\beta_0 + E[Z\epsilon] \quad \quad E[Y|Z] = E[\theta_0(X)|Z] + E[\epsilon|Z]$$ \hspace{1cm} (4)

For the linear model, obtaining covariances with $Z$ is a useful approach as it yields $k$ equations to solve for $\beta_0$, which is $k$ dimensional. For the nonparametric model, however, this approach is unsatisfactory as it yields $k$ equations to identify a whole function, which is an infinite dimensional object. Instead, we take conditional expectations to derive a functional equation. In this way we obtain an infinite number of equations, one for every possible value $z$, to solve for $\theta_0(x)$. Under the exogeneity assumption on the instrument, we simplify the identifying equations from (4) to:

$$E[ZY] = E[ZX']\beta_0 \quad \quad E[Y|Z] = E[\theta_0(X)|Z]$$ \hspace{1cm} (5)
In order to identify $\beta_0$ and $\theta_0(x)$ from the equations in (5) we need the existence of unique solutions. Uniqueness follows from the rank conditions on the instruments, which are:

$$E[ZX']\beta = 0 \Rightarrow \beta = 0 \quad E[\theta(X)|Z] = 0 \Rightarrow \theta(X) = 0$$

(6)

If the rank condition is not satisfied in the linear model, then $\beta_0$ will not be identified. For any $\beta \neq 0$ satisfying $E[ZX']\beta = 0$, the model $Y = X'(\beta_0 + \beta) + \hat{\epsilon}$ with $\hat{\epsilon} = \epsilon - X'\beta$ is observationally equivalent to the true model and also satisfies $E[Z\hat{\epsilon}] = 0$. This rank condition will hold if and only if the matrix $E[ZX']$ is invertible. Similarly, for the nonparametric model, if there exists a function $\theta(x) \neq 0$ satisfying $E[\theta(X)|Z] = 0$, then the model $Y = (\theta_0(X) + \theta(X)) + \hat{\epsilon}$ with $\hat{\epsilon} = \epsilon - \theta(X)$ is observationally equivalent to the true model and also satisfies $E[\hat{\epsilon}|Z] = 0$. The requirement that there be no function $\theta(x) \neq 0$ satisfying $E[\theta(X)|Z] = 0$ was aptly named a nonparametric rank condition by Newey & Powell (2003). If $Z$ and $X$ satisfy the nonparametric rank condition, then the distribution of $X$ conditional on $Z$ is said to be complete in $Z$. Sufficient conditions for completeness are only known for a few distributions, most notably for the exponential families.

On the one hand, the nonparametric rank condition is quite demanding, and one that may not be met in many empirical applications. On the other hand, the nonparametric rank condition is precisely what is required for nonparametric identification and hence, without further assumptions, doing away with it is not possible without losing identification. Without the rank condition, the model is partially identified. The identified set is composed of the models that are observationally equivalent to the true model and are consistent with the exogeneity assumption on $Z$. Hence, the identified sets are given by the solutions to the equations in (5), which can be characterized as:

$$\beta_0 + \mathcal{N}(E[ZX']) \quad \theta_0(x) + \mathcal{N}(E[\cdot|Z])$$

(7)

where $\mathcal{N}(E[ZX'])$ is the null space of $E[ZX']$ and $\mathcal{N}(E[\cdot|Z])$ is the null space of $E[\cdot|Z]$. That is, $\mathcal{N}(E[\cdot|Z])$ is the set of functions with finite second moment such that $E[\theta(X)|Z] = 0$. The rank conditions are equivalent to assuming $\mathcal{N}(E[ZX'])$ and $\mathcal{N}(E[\cdot|Z])$ are equal to $\{0\}$. Without further assumptions the identified sets can be quite large. The sets $\mathcal{N}(E[ZX'])$ and $\mathcal{N}(E[\cdot|Z])$ are both vector spaces. Furthermore, while the dimension of $\mathcal{N}(E[ZX'])$ is less than or equal to $k$, the dimension of $\mathcal{N}(E[\cdot|Z])$ can be infinite.

2.2 Limiting the Identified Set

In order to derive the asymptotic distribution of the test statistic I need to assume the true model $\theta_0(x)$ is differentiable a specified number of times and that some of its derivatives are bounded.
Newey & Powell (2003) pioneered the use of these technical assumptions in the nonparametric IV literature to obtain compactness of $\Theta$ and in this manner solve the ill-posed inverse problem. These assumptions also imply the statistics in our analysis behave uniformly on $\Theta$.

Besides their technical advantages, these regularity assumptions also aid in identification. The use of known or assumed properties of $\theta_0(x)$ to aid in identification has been previously used in the partial identification literature. Manski (1997), for example, explores the identifying power of monotonicity, semi-monotonicity and semi-concavity restrictions. In a similar vein, Blundell, Chen & Kristensen (2004) exploit the boundedness of Engel curves to aid in identification. The authors require a unique bounded solution to the equation $E[Y|Z] = E[\theta_0(X)|Z]$, because any unbounded solutions cannot be an Engel curve. This is equivalent to assuming that there exist no bounded functions $\theta(x)$ such that $E[\theta(X)|Z] = 0$. This condition is known as bounded completeness and it is weaker than completeness.

Since $\theta_0(x)$ is assumed to satisfy a set of regularity conditions, functions that do not satisfy them should be excluded from the identified set. In some instances, these conditions may be enough to attain identification. Thus, before precisely defining the identified set, it is necessary to first formally state the regularity conditions I assume for the true model $\theta_0(x)$. Let $X \in \mathbb{R}^k$ and define $\lambda$ to be a $k$ dimensional vector of nonnegative integers, also known as a multi-index. In addition, define $|\lambda| = \sum_{i=1}^{k} \lambda_k$ and let $D^\lambda \theta(x) = \partial^{\lambda_1} \theta(x)/\partial x_1^{\lambda_1} \ldots \partial x_k^{\lambda_k}$. For $m, m_0$ and $\delta_0$ positive integers satisfying $m > k/2$, $\delta_0 > k/2$ and $(k/m_0 + k/\delta_0) < 1/2$ define the norm:

$$||\theta||_S = \left\{ \sum_{|\lambda| \leq m+m_0} \int \left[ D^\lambda \theta(x) \right]^2 (1 + x'x)^{\delta_0} \, dx \right\}^{1/2}$$

(8)

where the function $\theta(x)$ is assumed to be $m + m_0$ times differentiable with respect to its arguments. Weighting the integrand by $(1 + x'x)^{\delta_0}$ implies that if $||\theta||_S < \infty$, then the tails of the function $\theta(x)$ and its derivatives decay at a rate of at least $o((1 + x'x)^{-\delta_0})$. This tail condition allows us to carry out the analysis for $X$ with full support. If we assume $X$ has compact support, then the tail condition is unnecessary and $\delta_0$ can be set equal to zero. I assume the true model $\theta_0(x)$ satisfies $||\theta_0||_S \leq B$ for some known constant $B$. The functional space $\Theta$ is defined to be:

$$\Theta = \{ \theta(x) : ||\theta||_S \leq B \}$$

(9)

Gallant & Nychka (1987) show the $||\theta||_S \leq B$ implies $\max_{|\lambda| \leq m_0} \sup_x |D^\lambda \theta(x)|$ is uniformly bounded in $\theta(x) \in \Theta$. Therefore, in the definition of $\Theta$ I am implicitly assuming the true model $\theta_0(x)$ and its derivatives up to order $m_0$ are bounded. These assumptions, which are made for
technical reasons, aid the instrument $Z$ in identifying $\theta_0(x)$ as well. By assumption, it is only necessary to consider alternatives within $\Theta$, and hence the identified set is defined to be:

$$\Theta_0 = \{\theta(x) \in \Theta : E[Y - \theta(X)|Z] = 0\}$$

(10)

In particular, since the functions $\theta(x) \in \Theta$ are uniformly bounded, similarly as in Blundell, Chen & Kristensen (2004), bounded completeness is sufficient to attain identification.

### 2.3 Analysis Without Identification

Even with the restrictions imposed on $\Theta$, it may still be a strong requirement that $\theta_0(x)$ be identified. In addition, there are currently not tests for the hypothesis that $\theta_0(x)$ is identified. It is therefore prudent to utilize a testing framework that is robust to $\theta_0(x)$ not being identified. Instead of performing hypothesis tests on a function, the test statistic developed in this paper will allow us to test restrictions on the identified set $\Theta_0$. The kind of hypothesis tests I allow for are of the form:

$$H_0 : \Theta_0 \cap R \neq \emptyset \quad \quad H_1 : \Theta_0 \cap R = \emptyset$$

(11)

where $R$ is a set of functions that satisfy a property we wish to test for. Some restrictions are imposed on $R$, and these will be discussed in detail in Section 3. The null hypothesis in (11) is that at least one element of $\Theta_0$ satisfies the restrictions imposed in $R$. For example, $R$ can be a parametric set of functions, in which case the test in (11) is a parametric specification test against a nonparametric alternative. $R$ can also be the set of demand functions that are inelastic at a point $x_0$. In this context, the null hypothesis in (11) is that at least demand function in $\Theta_0$ is inelastic at $x_0$. Shape restrictions can also be included in this framework. For example, $R$ can be the set of cost functions with economics of scope or the set of productions functions with economies of scale.

When $\theta_0(x)$ is identified, the null hypothesis and the alternative in (11) simplify to $H_0 : \theta_0(x) \in R$ and $H_1 : \theta_0(x) \notin R$. If nonparametric identification is attained, then the present testing framework reduces to tests on the true model $\theta_0(x)$. Therefore, there are no clear disadvantages to adopting a testing framework that does not assume $\theta_0(x)$ is identified. There are, however, three important advantages. First, when we are interested in a functional of $\theta_0(x)$, we may be able to construct informative confidence intervals even when the functional is not identified. The type of coverage requirement we adopt was originally explored in Imbens & Manski (2004). I will refer to these confidence regions as confidence regions for identifiable functionals, which is equivalent to the concept of confidence regions for functions of the identifiable parameters in Romano & Shaikh (2006a).
Second, even though \( Z \) may not be able to identify \( \theta_0(x) \), it may still be able to answer interesting questions about it, such as whether it satisfies certain shape restrictions. Third, the inability to reach compelling conclusions using this approach points out the limitations of the instrument to nonparametrically identify parameters of interest. This is not a weakness of the procedure, but rather its virtue of reflecting the limits of what can be learned without parametric assumptions. The following sections elaborate on the first two points.

2.3.1 Confidence Regions for Identifiable Functionals

In many instances, we will not be interested in \( \theta_0(x) \) but in a functional of it. Often, \( \theta_0(x) \) not being identified will imply certain functionals of it will not be identified either. It is still possible, however, to use hypotheses like (11) to construct a confidence region that asymptotically covers the value of the functional at every element of \( \Theta_0 \) with a prespecified probability that controls the size of the test. Suppose, for example, that the functions \( \theta(x) \) are Walrasian inverse demand functions, \( X \) is quantity, and we are interested in Walrasian consumer surplus with endogenous quantity at some level \( q^* \). Every \( \theta(x) \in \Theta_0 \) has a corresponding consumer surplus \( CS(\theta) = \int_0^{q^*} \theta(x)dx - \theta(q^*)q^* \).

The identified set for Walrasian consumer surplus is given by:

\[
S_0 = \left\{ CS(\theta) = \int_0^{q^*} \theta(x)dx - \theta(q^*)q^* : \theta(x) \in \Theta_0 \right\}
\]  

(12)

Consumer surplus will not be identified unless every \( \theta(x) \in \Theta_0 \) has the same consumer surplus. If we do not restrict our attention to the functional space \( \Theta \), then we might not be able to learn anything at all about consumer surplus unless \( \theta_0(x) \) is identified. To see this, recall from (7) that the identified set without smoothness restrictions is given by \( \theta_0(x) + \mathcal{N}(E[\cdot|Z]) \). Suppose \( \theta_0(x) \) is not identified, and that there is a function \( \theta_1(x) \in \mathcal{N}(E[\cdot|Z]) \) such that \( CS(\theta_1) \neq 0 \). For all scalars \( \lambda \), the function \( \theta_0(x) + \lambda \theta_1(x) \) is in the identified set as well because \( \lambda \theta_1(x) \in \mathcal{N}(E[\cdot|Z]) \) for all \( \lambda \). The consumer surplus of \( \theta_0(x) + \lambda \theta_1(x) \) is given by \( CS(\theta_0 + \lambda \theta_1) = CS(\theta_0) + \lambda CS(\theta_1) \), and hence, by choosing \( \lambda \) appropriately we can find an inverse demand function in the identified set with an arbitrary Walrasian consumer surplus.

The assumption \( \theta_0(x) \in \Theta \) helps in the partial identification problem by limiting the possible inverse demand functions in \( \Theta_0 \) and therefore also the possible consumer surpluses in \( S_0 \). As pointed out in Manski (1997), however, economic restrictions should also be exploited to limit \( S_0 \). For example, by imposing the constraint that demand functions must be weakly positive we can avoid \( S_0 \) being the whole real line. If warranted, additional assumptions such as monotonicity or convexity of the demand function can also be used to further limit \( \Theta \) and hence \( \Theta_0 \) and \( S_0 \).
The set $S_0$ is the limit of what can be nonparametrically learned about consumer surplus without further assumptions. A large set $S_0$ reflects that the instrument $Z$ is not able to provide much information about consumer surplus for the true inverse demand function $\theta_0(x)$. Under such circumstances, it may be advisable to adopt a parametric model for $\theta_0(x)$ in order to perform inference on consumer surplus. In this case, precise parametric estimates of consumer surplus reflect not the ability of the instrument $Z$ to estimate it, but the importance of the parametric assumption in identifying it. A large identified set $S_0$ puts emphasis on the crucial role the parametric assumptions play not only in estimation but also in identification. It also shows how far from the truth the estimates might be if the parametric model is misspecified.

Using a family of hypotheses of the form $H_0 : \Theta_0 \cap \mathcal{R} \neq \emptyset$, it is possible to construct confidence regions for the identifiable functionals. This is a weaker coverage requirement than building a confidence region for the whole set $S_0$, as discussed in Chernozhukov, Hong & Tamer (2007) and Romano & Shaikh (2006b). We aim to construct a set $\hat{S}_0$ such that if $CS(\theta) \in S_0$, then:

$$\lim_{N \to \infty} P \left( CS(\theta) \in \hat{S}_0 \right) \geq 1 - \alpha$$

(13)

where $\alpha$ is the chosen level of control for a Type I error. The construction of this type of confidence regions is a special case of our framework. To see this, consider the family of sets $\mathcal{R}(\lambda) = \{ \theta(x) \in \Theta : CS(\theta) = \lambda \}$ and the corresponding family of null hypotheses $H_0(\lambda) : \Theta_0 \cap \mathcal{R}(\lambda) \neq \emptyset$. In Section 3 I will show how to construct statistics to test this kind of hypothesis while controlling the probability of a Type I error. Therefore, if $\hat{S}_0$ is the set of $\lambda$ such that $H_0(\lambda)$ is not rejected, then $\hat{S}_0$ will have the desired property (13) by the duality of confidence intervals and hypothesis testing.

This construction can be applied to a wide array of functionals such as diverse kinds of elasticities and values of the function $\theta_0(x)$ and its derivatives. The outlined procedure can also be used to construct joint confidence regions for multiple functionals. For example, if in addition to consumer surplus $CS(\theta)$ we also care about price elasticity $E(\theta)$, then we can construct a joint confidence region for $CS(\theta)$ and $E(\theta)$ by using the sets $\mathcal{R}(\lambda) = \{ \theta(x) \in \Theta : CS(\theta) = \lambda_1, E(\theta) = \lambda_2 \}$.

### 2.3.2 Identified Features Without Identification of $\theta_0(x)$

In some instances, depending on the distribution of $Y, X$ and $Z$, it will still be possible to infer properties of $\theta_0(x)$ without it being identified. I illustrate this possibility using a parametric specification test as an example.

Suppose we wish to test whether $\theta_0(x)$ belongs to some parametric family $\theta(x, \beta)$. If $\theta_0(x)$ is
identified, then we can use a hypothesis test of the form \( H_0 : \Theta_0 \cap R \neq \emptyset \) to test whether \( \theta_0(x) \) belongs to a specified parametric family by letting \( R = \{ \theta(x) : \theta(x) = \theta(x, \beta) \text{ for some } \beta \} \). Even without identification, however, we may still be able to conclude that \( \theta_0(x) \) does not belong to the specified parametric family when this is indeed the case. Suppose that for all \( \beta \):

\[
E[\theta_0(X)|Z] \neq E[\theta(X, \beta)|Z] \tag{14}
\]

with positive probability. If inequality (14) is satisfied, then it follows that \( \theta_0(x) \) does not belong to the specified parametric family, and in addition by (5) and (10) that neither do any functions in \( \Theta_0 \). Therefore, we will asymptotically reject \( H_0 : \Theta_0 \cap R \neq \emptyset \) with probability tending to one. Hence, since \( \theta_0(x) \in \Theta_0 \) we can infer \( \theta_0(x) \) does not belong to the specified parametric family either. On the other hand, if \( \theta_0(x) \) does not belong to the specified parametric family, but some other function \( \theta(x) \in \Theta_0 \) does, then \( H_0 : \Theta_0 \cap R \neq \emptyset \) will be true. We will consequently fail to reject \( H_0 : \Theta_0 \cap R \neq \emptyset \) with asymptotic probability \( 1 - \alpha \), where \( \alpha \) is the chosen size of the test. Upon failing to reject this null hypothesis it is not possible to reach any conclusions about \( \theta_0(x) \). This is not a weakness of the procedure. On the contrary, this reflects that without identification it is impossible to conclude whether the function in the identified set belonging to the parametric family is really \( \theta_0(x) \) or not.

In general, it will be possible to conclude \( \theta_0(x) \notin R \), even without identification, as long as \( \Theta_0 \cap R = \emptyset \). This requirement on the instrument \( Z \) can still be fairly restrictive, though it is always weaker than assuming \( \theta_0(x) \) is identified. The derivation of “low level” conditions that ensure \( \Theta_0 \cap R = \emptyset \) when \( \theta_0(x) \notin R \) is a challenging exercise beyond the scope of this paper.

### 3 Test Statistics

In this section I develop a test statistic for the null hypothesis \( H_0 : \Theta_0 \cap R \neq \emptyset \). Since \( \Theta_0 \subseteq \Theta \), the null hypothesis \( H_0 : \Theta_0 \cap R \neq \emptyset \) is equivalent to asking whether there is a function \( \theta(x) \in \Theta \cap R \) satisfying \( E[Y - \theta(X)|Z] = 0 \). Therefore, for certain sets \( R \), \( H_0 : \Theta_0 \cap R \neq \emptyset \) is equivalent to:

\[
H_0 : \inf_{\theta(x) \in \Theta \cap R} E \left[ (E[Y - \theta(X)|Z])^2 f_Z^2(Z) \right] = 0 \tag{15}
\]

If the set \( R \) is such that the infimum in (15) is attained, then the null hypothesis in (15) is equivalent to \( H_0 : \Theta_0 \cap R \neq \emptyset \). This type of equivalence was originally used in Romano & Shaikh (2006a) to derive confidence regions for functions of the identifiable parameters and is similar in spirit to the Anderson & Rubin test (1949).
The advantage of using (15) over $H_0 : \Theta_0 \cap R \neq \emptyset$ is that no estimation of $\Theta_0$ is required. As a preliminary step towards being able to test (15) I construct a test statistic for $H_0 : \theta(x) \in \Theta_0$ for some arbitrary $\theta(x)$ in $\Theta$. Adapting arguments from Hall (1984) I develop a test statistic $T_N(\theta)$ that is asymptotically normally distributed when evaluated at a $\theta(x) \in \Theta_0$ and diverges to infinity otherwise. Following (15), I then use this result to show it is possible to test $H_0 : \Theta_0 \cap R \neq \emptyset$ by using the test statistic $I_N(R) = \min_{\Theta \cap R} T_N(\theta)$. If $\Theta_0 \cap R = \emptyset$, then when computing $I_N(R)$ we will minimize $T_N(\theta)$ over values for which it diverges to infinity and hence $I_N(R)$ will diverge to infinity as well. On the other hand, if $\Theta_0 \cap R \neq \emptyset$ then the minimizer of $T_N(\theta)$ over $\Theta \cap R$ should be close to $\Theta_0 \cap R$ because $T_N(\theta)$ diverges to infinity for all other functions $\theta(x)$. In Section 3.3 I formalize this intuition and show $I_N(R)$ converges in probability to $\min_{\Theta \cap R} T_N(\theta)$ when $H_0 : \Theta_0 \cap R \neq \emptyset$ is true. This result in turn allows us to obtain the asymptotic distribution of $I_N(R)$ under the null hypothesis.

### 3.1 Testing Strategy and Assumptions for $H_0 : \theta(x) \in \Theta_0$

A function $\theta(x)$ is in $\Theta_0$ if and only if it is consistent with the exogeneity assumption on $Z$. Hence, the null hypothesis $H_0 : \theta(x) \in \Theta_0$ is equivalent to $H_0 : E[Y - \theta(X)|Z] = 0$. Once $\theta(x)$ is fixed, the second hypothesis can be viewed as a specification test for whether the nonparametric regression function $E[Y - \theta(X)|Z]$ is equal to zero. In order to implement this specification test, I employ the Nadaraya-Watson kernel estimator for $E[Y - \theta(X)|Z]$. Assume $Z \in \mathbb{R}^d$ and define:

$$
\hat{f}_Z(z_n) = \frac{1}{Nh^d} \sum_{i \neq n} K\left(\frac{z_i - z_n}{h}\right)
$$

where the kernel $K(u)$ is a symmetric density function and $h$ is the chosen bandwidth. The Nadaraya-Watson estimator is then given by:

$$
\hat{E}[Y - \theta(X)|z_n] = \frac{1}{Nh^d} \sum_{i \neq n} K\left(\frac{z_i - z_n}{h}\right) (y_i - \theta(x_i)) \frac{1}{\hat{f}_Z(z_n)}
$$

Hall (1984) derives the asymptotic behavior of the integrated square error of the Nadaraya-Watson estimator, which is given by $\int \left(\hat{E}[Y - \theta(X)|Z = z] - E[Y - \theta(X)|Z = z]\right)^2 dz$. The specification test for $H_0 : E[Y - \theta(X)|Z] = 0$ can be implemented by examining the integrated square error under the null hypothesis. The results from Hall (1984) can be readily applied to derive the asymptotic behavior of $\int \left(\hat{E}[Y - \theta(X)|Z = z]\right)^2 dz$ and in this way obtain a test statistic for the null hypothesis $H_0 : \theta(x) \in \Theta_0$. In order to avoid the computationally intensive calculation of
integrating over the support of $Z$, I adapt the arguments in Hall (1984) to instead examine the asymptotic behavior of:

$$Q_N(\theta) = \frac{1}{N} \sum_{n=1}^{N} \left( \hat{E}[Y - \theta(X)|z_n] \right)^2 \hat{f}_Z^2(z_n)$$

Using $\hat{f}_Z^2(z_n)$ as a weight function allows us to avoid the possible irregular behavior of the Nadaraya-Watson estimator when $f_Z(z_n)$ is close to zero. In addition, using $\hat{f}_Z^2(z_n)$ as a weight function simplifies the derivation of the asymptotic behavior of $Q_N(\theta)$ as it no longer is the ratio of two random variables.

The following assumptions are sufficient for establishing the asymptotic behavior of the statistic $Q_N(\theta)$ for any $\theta(x)$ in the functional space $\Theta$:

**ASSUMPTION 1:** The observations $\{y_n, x_n, z_n\}_{n=1}^{N}$ are i.i.d. with $Y \in \mathbb{R}$, $X \in \mathbb{R}^k$ and $Z \in \mathbb{R}^d$. They are distributed with density $f_{ZXY}(z, x, y)$ and generated by the model specified in (1) with $\theta_0(x) \in \Theta$ and $\epsilon$ not a deterministic function of $X$.

**ASSUMPTION 2:** The marginal distribution $f_Z(z)$ is bounded and continuous almost everywhere. The marginal distribution $f_{ZX}(z, x)$ is also bounded.

**ASSUMPTION 3:** There exists a $\delta > 0$ such that $E[Y^{4+\delta}] < \infty$.

**ASSUMPTION 4:** The moments $E[(Y - \theta(X))^J|Z = z]$ are well defined for $J \in \{1, 2, 4\}$ and continuous and bounded in $z$ for $J \in \{1, 2\}$ uniformly in $\theta(x) \in \Theta$. The moments $E[|Y|^J|Z]$ are bounded for $J \in \{1, 2, 3, 4\}$.

**ASSUMPTION 5:** The kernel $K(u)$ is a bounded, symmetric density with full support on $\mathbb{R}^d$. The bandwidth $h$ satisfies $h \rightarrow 0$ and $Nh^d \rightarrow \infty$.

Assumptions 1-5 are stronger than what is needed to establish the asymptotic behavior of $Q_N(\theta)$ for any particular $\theta(x) \in \Theta$. In Section 3.4, however, I need to analyze the behavior of $Q_N(\theta)$ as a stochastic process defined on the space of bounded functionals on $\Theta_0$. In this instance the uniformity in $\Theta$ aspects of Assumptions 1-5, which are unnecessary for the present analysis, will be crucial. In order to avoid the introduction of multiple sets of assumptions I adopt the stronger than necessary Assumptions 1-5 that are applicable throughout the paper. The assumption that $K(u)$ has full support in $\mathbb{R}^d$ is also unnecessary. It is made to simplify the proofs and notation when changes of variables in integration are necessary.
3.2 Test Statistic for \( H_0 : \theta(x) \in \Theta_0 \)

Under Assumptions 1-5, if \( Q_N(\theta) \) is properly centered and scaled, then it converges in distribution to a standard normal random variable when evaluated at a function \( \theta(x) \in \Theta_0 \) and diverges to infinity otherwise. In order to simplify notation we define:

\[
v_n(\theta) = y_n - \theta(x_n)
\]

(19)

where the notation \( v_n(\theta) \) is meant to emphasize that the residual depends on the function \( \theta(x) \) that implies it. The U-Statistic that will provide the proper centering for \( Q_N(\theta) \) instead use the also consistent estimator \( \hat{\theta} \). Furthermore, if Assumptions 1-5 hold and \( \hat{\theta} \) is properly centered and scaled, then it converges in distribution to a standard normal random variable when evaluated at a function \( \theta(x) \in \Theta_0 \) and diverges to infinity otherwise. In order to simplify notation we define:

\[
\hat{B}_N(\theta) = \frac{1}{N^3h^{2d}} \sum_{n=1}^{N} \sum_{i<n} K^2 \left( \frac{z_i - z_n}{h} \right) \left( v_n^2(\theta) + v_i^2(\theta) \right)
\]

(20)

\( \hat{B}_N(\theta) \) is the sum of the squares that are generated upon expanding the terms \( \left( \hat{E}[Y - \theta(X)|z_n] \right)^2 \) in \( Q_N(\theta) \). It is also necessary to define the estimator for the asymptotic variance of \( Q_N(\theta) \). We let \( \hat{\sigma}^2_C(\theta) = \hat{\sigma}^2_{CI}(\theta) \hat{\sigma}^2_{CI} \), where \( \hat{\sigma}^2_{CI} = 2 \int \left[ \int K(u)K(u+w)du \right]^2 dw \) and \( \hat{\sigma}^2_C \) is defined by:

\[
\hat{\sigma}^2_C(\theta) = 24(N-4)! \sum_{n=1}^{N} \sum_{i<j<k<l} \int K \left( \frac{z_n - z}{h} \right) K \left( \frac{z_i - z}{h} \right) K \left( \frac{z_j - z}{h} \right) v_n^2(\theta) v_i^2(\theta) v_j^2(\theta) v_l^2(\theta) dz
\]

(21)

The estimator \( \hat{\sigma}^2_C(\theta) \) can be computationally intensive. For this reason it may be desirable to instead use the also consistent estimator \( \hat{\sigma}^2_{CI}(\theta) = \int \left( \hat{E}[v_n^2(\theta)|Z = z] \right)^4 \hat{f}_z(z)dz \). The asymptotic behavior of \( \hat{\sigma}^2_{CI}(\theta) \) is dominated by the cross terms that arise from expanding the fourth power. The sum of these cross terms is identical to \( \hat{\sigma}^2_{CI}(\theta) \). I use \( \hat{\sigma}^2_{CI}(\theta) \) instead of \( \hat{\sigma}^2_{CI}(\theta) \) because it is less cumbersome to show its uniform consistency over \( \Theta \). We now state the theorem that will allow us to test \( H_0 : \theta(x) \in \Theta_0 \).

**Theorem 3.1.** Let \( T_N(\theta) = Nh^{\frac{d}{2}} \left( Q_N(\theta) - \hat{B}_N(\theta) \right) \). If Assumptions 1-5 hold and \( \theta(x) \in \Theta_0 \), then:

\[
\frac{1}{\hat{\sigma}_C(\theta)} T_N(\theta) = \frac{Nh^{\frac{d}{2}}}{\hat{\sigma}_C(\theta)} \left( Q_N(\theta) - \hat{B}_N(\theta) \right) \overset{L}{\longrightarrow} N(0, 1)
\]

Furthermore, if Assumptions 1-5 hold and \( \theta(x) \notin \Theta_0 \), then:

\[
\frac{1}{\hat{\sigma}_C(\theta)} T_N(\theta) = \frac{Nh^{\frac{d}{2}}}{\hat{\sigma}_C(\theta)} \left( Q_N(\theta) - \hat{B}_N(\theta) \right) \overset{p}{\longrightarrow} +\infty
\]

Since \( \hat{B}_N(\theta) \) is the sum of the squares that result from expanding the terms \( \left( \hat{E}[Y - \theta(X)|z_n] \right)^2 \) in \( Q_N(\theta) \), the statistic \( T_N(\theta) = Nh^{\frac{d}{2}} \left( Q_N(\theta) - \hat{B}_N(\theta) \right) \) is equal to a sum of cross terms of the form:

\[
h^{-2d} K \left( \frac{z_i - z_n}{h} \right) K \left( \frac{z_j - z_n}{h} \right) v_i(\theta)v_j(\theta)
\]

(22)
Grouped in threes, terms like (22) form a symmetric kernel that shows \( T_N(\theta) \) is a symmetric U-Statistic of order three. The expectation of (22) is equal to \( h^{-2d}E \left[ \left( E \left\{ \frac{1}{h} (z_n - \theta) \right\} v_i(\theta) \right| z_n \right] \), which converges to \( E \left[ (E[(Y - \theta(X))|Z])^2 f_Z^2(Z) \right] \). Hence, for \( h \) sufficiently small the expectation of (22) will be strictly positive causing \( T_N(\theta) \) to be improperly centered and diverge to positive infinity. On the other hand, when \( \theta(x) \in \Theta_0 \) the expectation of (22) conditional on any pair \((z_k, v_k(\theta))\) for \( k \in \{n, i, j\} \) is equal to zero, which implies \( T_N(\theta) \) is not only properly centered but also degenerate of order one. Due to its degeneracy, the asymptotic behavior of \( T_N(\theta) \) is governed by the second term in its Hoeffding decomposition when \( \theta(x) \in \Theta_0 \). The scaling by \( Nh^\frac{d}{2} \) therefore corresponds to the standard deviation of the second term in the Hoeffding decomposition of \((Nh^\frac{d}{2})^{-1}T_N(\theta)\). When \( \theta(x) \notin \Theta_0 \), \( T_N(\theta) \) is no longer degenerate for \( h \) sufficiently small. The asymptotic behavior of \((Nh^\frac{d}{2})^{-1}T_N(\theta)\) is then governed by the first term in its Hoeffding decomposition, which has a standard deviation of order \( O(N^{-\frac{1}{2}}) \). Consequently, \( T_N(\theta) \) diverges to positive infinity when \( \theta(x) \notin \Theta_0 \) not only because it is improperly centered, but also because it is being scaled by a factor of \( Nh^\frac{d}{2} \) that is too large. While the improper centering of \( T_N(\theta) \) ensures the divergence is to positive infinity and not negative infinity, the scaling by \( Nh^\frac{d}{2} \) causes the divergence to be at a rate faster than \( \sqrt{N} \).

3.3 Testing Strategy and Assumptions for \( H_0 : \Theta_0 \cap R \neq \emptyset \)

I construct a test statistic for the null hypothesis \( H_0 : \Theta_0 \cap R \neq \emptyset \) by exploiting its equivalence with \( H_0 : \inf_{\theta(x) \in \Theta \cap R} E \left[ (E[Y - \theta(X)|Z])^2 f_Z^2(Z) \right] = 0 \). For this equivalence to hold we need to ensure that the infimum is attained. Attainment will be implied by compactness of \( \Theta \cap R \) under the norm:

\[
||\theta||_{C^\delta} = \max_{|\lambda| \leq m_0} \sup_{x} |D^\lambda \theta(x)|(1 + x'x)^{\delta}
\]

(23)

where \( m_0 \) was specified in the definition of the norm \( ||\theta||_{S} \) in (8) and \( k/2 < \delta < \delta_0 \). The functional \( E \left[ (E[Y - \theta(X)|Z])^2 f_Z^2(Z) \right] \) is continuous under \( ||\theta||_{C^\delta} \) and hence compactness of \( \Theta \cap R \) under \( ||\theta||_{C^\delta} \) will imply the infimum in (15) is attained. Gallant & Nychka (1987) show \( \Theta \) is compact under \( ||\theta||_{C^\delta} \). Therefore, since closed subsets of compact sets are compact, if \( R \) is closed under \( ||\theta||_{C^\delta} \), then \( \Theta \cap R \) will be compact under \( ||\theta||_{C^\delta} \). Due to the norm \( ||\theta||_{C^\delta} \) being very strong, it is straightforward to show the set \( R \) will be closed for a wide array of interesting hypotheses. For example, for any \( M \in \mathbb{R}^m \) and transformation \( F : \Theta \rightarrow \mathbb{R}^m \) that is continuous under \( ||\theta||_{C^\delta} \), the sets \( R = \{ \theta(x) : F(\theta) \leq M \} \) and \( R = \{ \theta(x) : F(\theta) = M \} \) will be closed under \( ||\theta||_{C^\delta} \). Since \( ||\theta||_{C^\delta} \) is a strong norm, \( F : \Theta \rightarrow \mathbb{R}^m \) being continuous under it is a weak requirement. Examples
of functionals that are continuous under $||\theta||_{C^5}$ include levels of $\theta(x)$ and its derivatives, consumer surplus and price elasticities.

Replacing $H_0 : \Theta_0 \cap R \neq \emptyset$ by (15) and assuming $R$ is closed under $||\theta||_{C^5}$ has two limitations. The first limitation is that requiring $R$ to be closed is not just a technical condition but actually restricts the economic content of the hypotheses that can be tested. As an example consider testing whether an Engel curve is weakly decreasing. By letting $F(\theta) = \inf_x \theta'(x)$ and $R = \{\theta(x) : F(\theta) \leq 0\}$, we see that $R$ is closed under $||\theta||_{C^5}$ because $F(\theta)$ is continuous under $||\theta||_{C^5}$. Therefore, a test for weak monotonicity fits the present framework. Suppose, however, that we wish to test for strict monotonicity instead, which corresponds to $R = \{\theta(x) : F(\theta) < 0\}$. The set $R$ is now open, which implies $H_0 : \Theta_0 \cap R \neq \emptyset$ and $H_0 : \inf_{\theta(x) \in \Theta \cap R} E \left[ (E[Y - \theta(X)|Z])^2 f_Z^2(Z) \right] = 0$ may no longer be equivalent. Assume, for simplicity, that $\theta_0(x)$ is identified and in addition that it is not strictly monotonic but only weakly so. In this case $H_0 : \Theta_0 \cap R \neq \emptyset$ is false. On the other hand, the hypothesis $H_0 : \inf_{\theta(x) \in \Theta \cap R} E \left[ (E[Y - \theta(X)|Z])^2 f_Z^2(Z) \right] = 0$ is actually true. By evaluating $E \left[ (E[Y - \theta(X)|Z])^2 f_Z^2(Z) \right]$ at a sequence of strictly monotonic Engel curves approaching the true model $\theta_0(x)$ we can make $E \left[ (E[Y - \theta(X)|Z])^2 f_Z^2(Z) \right]$ arbitrarily close to zero. If we utilize the present framework regardless, then we are likely to incorrectly infer $\theta_0(x)$ is strictly monotonic when in reality it is only weakly so.

A second limitation of the present testing framework is that hypotheses are restricted to the functional space $\Theta$. The smoothness assumptions on $\Theta$ apriori limit the set of hypotheses that can be tested. For example, for many parametric families $\theta(x, \beta)$, the set $R = \{\theta(x) : \theta(x, \beta)\}$ for some $\beta$ will be closed under $||\theta||_{C^5}$. Therefore, the present framework includes a parametric specification test as a special case. There is an implicit assumption, however, that the parametric family $\theta(x, \beta)$ itself is sufficiently smooth and hence included in $\Theta$. It is not possible, for example, to utilize the present framework to do a specification test for a parametric family $\theta(x, \beta)$ that is not differentiable.

The equivalence of $H_0 : \Theta_0 \cap R \neq \emptyset$ and $H_0 : \inf_{\theta(x) \in \Theta \cap R} E \left[ (E[Y - \theta(X)|Z])^2 f_Z^2(Z) \right] = 0$ for a large class of sets $R$ allows us to base our test on the latter null hypothesis. The advantage of using the second null hypothesis is that it does not require estimation of the identified set $\Theta_0$. To test this hypothesis we replace $E \left[ (E[Y - \theta(X)|Z])^2 f_Z^2(Z) \right]$ with the statistic $T_N(\theta)$ from Theorem 3.1 and define:

$$I_N(R) = \inf_{\Theta \cap R} T_N(\theta)$$  \hspace{1cm} (24)

From Theorem 3.1 we know $T_N(\theta)$ diverges to infinity for all $\theta(x) \notin \Theta_0$. Therefore, $I_N(R)$ diverges to infinity when $\Theta_0 \cap R = \emptyset$, because in this case the infimum in (24) is taken over functions that
are not in $\Theta_0$. On the other hand, when $\Theta_0 \cap R \neq \emptyset$ there are functions $\theta(x) \in \Theta \cap R$ that are also in $\Theta_0$. Since $T_N(\theta)$ converges to a normal distribution for such functions and it diverges to infinity for the rest, $T_N(\theta)$ should be minimized on functions that are close to the identified set. We therefore expect $I_N(R)$ to behave like $\inf_{\Theta_0 \cap R} T_N(\theta)$ when $\Theta_0 \cap R \neq \emptyset$. In principle, this rationale can be used to construct similar test statistics that utilize alternatives to $T_N(\theta)$ as a first step. An excellent reference to the large number of admissible tests for the null hypothesis $H_0 : E[Y - \theta(X)|Z] = 0$ can be found in Hart (1997). While there is an extensive literature examining the power of these tests for the null hypothesis $H_0 : E[Y - \theta(X)|Z]$, it is unclear how the choice of a particular test statistic affects the power in the second step for the null hypothesis $H_0 : \Theta_0 \cap R \neq \emptyset$. This is a complex question beyond the scope of this paper.

I need the following assumptions to formalize the intuition developed in this section:

**ASSUMPTION 6:** The set $R$ is closed under the norm $||\theta||_{C_k}$.

**ASSUMPTION 7:** For some $l \leq 1$, the bandwidth $h$ satisfies $\sqrt{Nh^l} \to \infty$ and $Nh^{d+l}\left(1-\frac{(m_0+k)\delta}{m_0\delta_0}\right) \to 0$.

**ASSUMPTION 8:** The density $f_Z(z)$ is differentiable, with bounded gradient, and for $h$ small enough there exists a function $g(u, z, y, x)$ satisfying $h^{-1}K(u)||f_{ZYX}(hu+z, y, x) - f_{ZYX}(z, y, x)||y - \theta(x)||^i \leq g(u, z, y, x)$ for all $\theta, i \in \{1, 2\}$ and $\int g(u, z, y, x)dudzdydx < \infty$.

The role of Assumption 6 has already been extensively discussed. Assumption 7 is necessary to ensure that certain biases converge to zero sufficiently fast. The assumption that $k/m_0 + k/\delta_0 < 1/2$, which is necessary to control the covering numbers of $\Theta$, also guarantees Assumption 7 is compatible with Assumption 5, which requires $Nh^d \to \infty$. Alternatively, if higher order kernels are used for $K(u)$ to reduce these biases, then it is possible to allow for a higher range of rates for the bandwidth $h$. Assumption 8 is necessary to allow us to exchange the order of integration and differentiation. It also implies certain error terms in Taylor expansions will behave uniformly in $\Theta$.

### 3.4 Test Statistic for $H_0 : \Theta_0 \cap R \neq \emptyset$

Before stating the asymptotic distribution of $I_N(R)$ it will be helpful to provide a basic overview of convergence in distribution in the space of bounded functionals. For a detailed discussion please refer to Chapter 1.5 in van der Vaart & Wellner (1998). The set of functionals on $\Theta_0$, $F : \Theta_0 \to \mathbb{R}$ satisfying $\sup_{\theta(x) \in \Theta_0} |F(\theta)| < \infty$ form a metric space denominated $\mathcal{L}\infty(\Theta_0)$. In this space the distance between two functionals $F_1$ and $F_2$ is measured by $\sup_{\theta(x) \in \Theta_0} |F_1(\theta) - F_2(\theta)|$. It is helpful to think of $T_N(\theta)$ as an element of $\mathcal{L}\infty(\Theta_0)$. The statistic $T_N(\theta)$ maps every $\theta(x) \in \Theta_0$ to $\mathbb{R}$. In
addition, because $\theta(x) \in \Theta$ are uniformly bounded, we have $\sup_{\theta(x) \in \Theta_0} |T_N(\theta)| < \infty$, which implies $T_N(\theta)$ is indeed an element of $L^\infty(\Theta_0)$. The value $T_N(\theta)$ assigns to any particular $\theta(x)$, however, is random as it depends on the realization of the data. As such, $T_N(\theta)$ is a random variable defined on $L^\infty(\Theta_0)$ because every realization of the data generates a different mapping $T_N(\theta)$ from $\Theta_0$ to $\mathbb{R}$.

Just as for random variables defined on the real line, it is also possible to think of $T_N(\theta)$ converging in distribution to a random variable defined on $L^\infty(\Theta_0)$. In fact, $T_N(\theta)$ converges in distribution to a Gaussian process $G(\theta)$ in $L^\infty(\Theta_0)$. A Gaussian process in $L^\infty(\Theta_0)$ is a random mapping from $\Theta_0$ to $\mathbb{R}$ with the property that $(G(\theta_1), \ldots, G(\theta_M))$ is jointly normally distributed for any finite vector of functions $(\theta_1(x), \ldots, \theta_M(x))$ in $\Theta_0$. We now proceed to the statement of the main theorem:

**Theorem 3.2.** If Assumptions 1-8 hold and $\Theta_0 \cap R \neq \emptyset$, then:

$$I_N(R) = \inf_{\Theta \cap R} T_N(\theta) = \inf_{\Theta_0 \cap R} N h^2 \left( Q_N(\theta) - \bar{B}_N(\theta) \right) + o_p(1) \xrightarrow{L} \inf_{\Theta_0 \cap R} G(\theta)$$

where $G(\theta)$ is a Gaussian process in $L^\infty(\Theta_0)$. If Assumptions 1-8 hold and $\Theta_0 \cap R = \emptyset$, then:

$$I_N(R) = \inf_{\Theta \cap R} T_N(\theta) \xrightarrow{p} +\infty$$

Theorem 3.2 implies that if $\Theta_0 \cap R$ is a singleton with unique element $\theta^U(x)$, then the test statistic $I_N(R)$ will converge in distribution to $G(\theta^U)$, which is a normal random variable. For example, if $\theta_0(x)$ is identified, then $\Theta_0 \cap R$ will be either a singleton, when $\theta_0(x) \in R$, or empty, when $\theta_0(x) \notin R$. Corollary 3.1 establishes that it is possible to estimate the asymptotic variance of $G(\theta^U)$ and construct a pivotal statistic that will converge in distribution to a standard normal random variable when $\Theta_0 \cap R = \{\theta^U(x)\}$.

**Corollary 3.1.** Suppose Assumptions 1-8 hold and $\theta^*(x) \in \arg\min_{\Theta \cap R} T_N(\theta)$. If the set $\Theta_0 \cap R = \{\theta^U(x)\}$ is a singleton, then $\sigma^{-1}_C(\theta^*)I_N(R) \xrightarrow{L} N(0,1)$. Furthermore, if $\Theta_0 \cap R = \emptyset$ then $\sigma^{-1}_C(\theta^*)I_N(R) \xrightarrow{p} +\infty$.

There are a number of interesting technical challenges present in the proof of Theorem 3.2. The statistic $T_N(\theta)$ does not converge to an asymptotically tight random variable in $L^\infty(\Theta)$ and hence the asymptotic behavior of $I_N(R)$ cannot be derived through a direct use of the continuous mapping theorem. Even though for every sample of size $N$ the U-Statistic $T_N(\theta)$ has continuous sample paths in all of $\Theta$, it is only asymptotically uniformly equicontinuous with respect to $||\theta||_{L^2(X)}$ on $\Theta_0$, where $||\theta||_{L^2(X)}^2 = E[|\theta^2(X)|]$. The lack of asymptotic uniform equicontinuity follows immediately from Theorem 3.1, as $T_N(\theta)$ converges to a normal random variable for every $\theta(x) \in \Theta_0$ but diverges to infinity for every $\theta(x) \notin \Theta_0$. The identified set $\Theta_0$, however, is small relative to $\Theta$; in fact its
interior relative to Θ is empty. Therefore, \( T_N(\theta) \) actually fails to be asymptotically uniformly equicontinuous in almost all of Θ, which implies it does not converge in distribution in \( \mathcal{L}^\infty(\Theta) \). I address this challenge by examining the behavior of the Hoeffding decomposition of \( T_N(\theta) \) within a shrinking neighborhood \( \Theta_0^{c_N} \) of the identified set \( \Theta_0 \), and outside this neighborhood.

Since \( T_N(\theta) \) is a U-Statistic of order three, its Hoeffding decomposition generates four fully degenerate U-Statistics \( P_N^0(\theta), P_N^1(\theta), P_N^2(\theta) \) and \( P_N^3(\theta) \), where \( P_N^0(\theta) \) is of order \( i \) and \( P_N^i(\theta) = E[T_N(\theta)] \). For \( \theta(x) \in \Theta_0 \), \( T_N(\theta) \) is degenerate and therefore \( P_N^1(\theta) = 0 \). In addition, since \( P_N^0(\theta) \geq 0 \) and \( \Theta_0 \subseteq \Theta_0^{c_N} \), the following inequality follows:

\[
\sum_{i=1}^{3} \inf_{\theta \in \Theta_0^{c_N} \cap R} P_N^i(\theta) \leq \inf_{\theta \in \Theta_0^{c_N} \cap R} T_N(\theta) \leq \sum_{i=2}^{3} \inf_{\theta \in \Theta_0 \cap R} P_N^i(\theta) \tag{25}
\]

The triangular array nature of the problem and \( T_N(\theta) \) not being linear in \( \theta(x) \) prevent the direct use of the results in Arcones & Gine (1993). Instead, I apply the maximal inequalities from Arcones & Gine (1993) to changing classes of functions that are Lipschitz in Θ. With these maximal inequalities it is possible to show \( P_N^3(\theta) \) converges in probability to zero uniformly in Θ and \( P_N^2(\theta) \) is uniformly asymptotically equicontinuous with respect to \( ||\theta||_{\mathcal{L}^2(X)} \) on all of Θ. From Theorem 3.1 we know the marginals of \( P_N^2(\theta) \) on \( \Theta_0 \) are normally distributed and therefore \( P_N^2(\theta) \) converges in distribution to a Gaussian process in \( \mathcal{L}^\infty(\Theta_0) \), though probably to Gaussian chaos in \( \mathcal{L}^\infty(\Theta) \). For \( \theta(x) \notin \Theta_0 \), the standard deviation of \((Nh^{\frac{2}{3}})^{-1}P_N^1(\theta)\) is of order \( O(N^{-\frac{1}{2}}) \), but its level decreases to zero as \( \theta(x) \) gets closer to \( \Theta_0 \). If \( \epsilon_N \to 0 \) fast enough, then the variance of \( P_N^1(\theta) \) converges uniformly to zero within \( \Theta_0^{c_N} \). This result follows by letting the rate at which the level of the variance of \((Nh^{\frac{2}{3}})^{-1}P_N^1(\theta)\) decreases because of \( \theta(x) \) being close to \( \Theta_0 \) compensate for \( T_N(\theta) \) being scaled by the factor \( Nh^{\frac{2}{3}} \) instead of \( \sqrt{N} \). Therefore, since \( P_N^1(\theta) \) and \( P_N^3(\theta) \) converge in probability to zero uniformly within \( \Theta_0^{c_N} \) and \( P_N^2(\theta) \) is uniformly asymptotically equicontinuous with respect to \( ||\theta||_{\mathcal{L}^2(X)} \) in Θ, inequality (25) implies \( \inf_{\Theta_0^{c_N} \cap R} T_N(\theta) = \inf_{\Theta_0^{c_N} \cap R} P_N^2(\theta) + o_p(1) \xrightarrow{L} \inf_{\theta \in \Theta_0 \cap R} G(\theta) \). If \( \epsilon_N \to 0 \) slowly enough, then \( E[T_N(\theta)] + P_N^1(\theta) \), and hence \( T_N(\theta) \), will diverge to infinity uniformly in \( (\Theta_0^{c_N})^c \) because of the improper centering. Therefore, the infimum of \( T_N(\theta) \) over \( \Theta \cap R \) will with high probability be attained on \( \Theta_0^{c_N} \cap R \), which implies \( I_N(R) = \inf_{\Theta_0^{c_N} \cap R} T_N(\theta) + o_p(1) \) and hence establishes the theorem.

## 4 Implementation

There are two challenges that need to be addressed before being able to implement the test specified in Theorem 3.2. First, the test statistic \( I_N(R) \) is the solution to an optimization problem over the
possibly nonparametric set of functions $\Theta \cap R$, which is computationally problematic. Second, the appropriate critical values are known only when $\Theta_0 \cap R$ is a singleton, in which case Corollary 3.1 can be used to construct a pivotal test statistic. In order to address the first problem, I show that it is possible to compute $I_N(R)$ by optimizing over spaces that approximate $\Theta \cap R$ without losing the asymptotic results. For the second problem, I discuss when we might expect Corollary 3.1 to apply and show that subsampling can be used to estimate the appropriate critical values when this is not the case.

### 4.1 Approximations to $\Theta \cap R$

For certain hypotheses, such as a parametric specification test, the set $\Theta \cap R$ is parametric and the computation of $I_N(R)$ is straightforward. For most hypotheses, however, the computation of $I_N(R)$ will require solving a minimization problem over a nonparametric set of functions. I address this challenge by showing in Theorem 4.1 that it is possible to solve the minimization problem over an approximating sieve without losing the asymptotic results of Theorem 3.2 and Corollary 3.1.

**Theorem 4.1.** Suppose Assumptions 1-8 hold, let $\{\Theta_J\} \subseteq \Theta$ be a sequence of closed sets under the norm $||\theta||_{\ell^2(X)} = [E[\theta^2(X)]]^{1/2}$ and define $\theta^*_J(x) \in \arg\min_{\Theta_J \cap R} T_N(\theta)$.

1. If $\Theta_J \cap R \subseteq \Theta \cap R$ and $\Theta_0 \cap R = \emptyset$, then $\min_{\Theta_J \cap R} T_N(\theta) \overset{p}{\rightarrow} +\infty$ and $\hat{\sigma}^{-1}(\theta^*_J) T_N(\theta^*_J) \overset{p}{\rightarrow} +\infty$.

2. If $\Theta_0 \cap R \neq \emptyset$, then $\sup_{\theta \in \Theta \cap R} \inf_{\theta_J \in \Theta_J \cap R} ||\theta - \theta_J||_{\ell^2(X)} = o(h^{l})$ for $l \leq 1$ with $\sqrt{Nh} \rightarrow \infty$ and $Nh^{l+d}(1-\frac{\log h}{\log \log h}) \rightarrow 0$, then $\min_{\Theta_J \cap R} T_N(\theta) = \min_{\Theta \cap R} T_N(\theta) + o_P(1)$. If in addition the set $\Theta_0 \cap R$ is a singleton, then $\hat{\sigma}^{-1}(\theta^*_J) T_N(\theta^*_J) \overset{d}{\rightarrow} N(0,1)$.

Theorem 4.1 requires that $\{\Theta_J \cap R\}$ be able to approximate $\Theta \cap R$ uniformly well, which is different from the typical requirement that $\{\Theta_J\}$ be able to approximate $\Theta$. If we wish to test whether demand is inelastic at a point $x_0$ using splines, for example, then the splines must satisfy this constraint and still be able to approximate those demand functions in $\Theta$ that are inelastic at $x_0$. For pointwise restrictions this is a straightforward requirement. For global restrictions, such as non-negativity monotonicity or concavity, we will need to choose shape-preserving splines; see Chen (2006) for examples.

The sieve $\{\Theta_J \cap R\}$ is required to approximate $\Theta \cap R$ under $||\theta||_{\ell^2(X)} = E[\theta^2(X)]$, which is weaker than the norm $||\theta||_{C^0}$ under which $R$ is required to be closed. The approximation error must decrease to zero at a rate $o(h^{l})$, which is governed by the bandwidth. Despite depending only on
the bandwidth, this rate cannot be made arbitrarily slow because Assumption 7 provides a lower bound on how slow $h$ can decrease to zero. The use of higher order kernels would increase the rate at which biases vanish, allowing for a slower rate requirement on the bandwidth $h$ and therefore also a slower rate requirement on the approximation error from using $\{\Theta_J \cap R\}$ instead of $\Theta \cap R$.

### 4.2 Choice of Critical Values

In this section I discuss how to choose the appropriate critical values for the test statistic. Since for most hypotheses the set $\Theta \cap R$ will be nonparametric, we will need to compute our test statistic over an approximating space $\{\Theta_J \cap R\}$. Therefore, I will present the results in this section in terms of the test statistics from Theorem 4.1:

$$
\tilde{I}_N(R) = \inf_{\Theta_J \cap R} T_N(\theta) \quad \hat{\sigma}_C^{-1}(\theta^*) T_N(\theta^*)
$$

(26)

where $\theta^*_J(x) \in \arg\min_{\Theta_J \cap R} T_N(\theta)$. As implied by Theorem 4.1, the results in this section will also apply to the statistics $I_N(R)$ and $\hat{\sigma}_C^{-1}(\theta^*) T_N(\theta^*)$, where $\theta^*(x) \in \arg\min_{\Theta \cap R} T_N(\theta)$.

#### 4.2.1 Critical Values From the Normal Approximation

Corollary 3.1 and Theorem 4.1 imply that if the set $\Theta_0 \cap R$ is a singleton, then $\hat{\sigma}_C^{-1}(\theta^*_J) T_N(\theta^*_J)$ $\xrightarrow{L} N(0,1)$. Therefore, if we expect $\Theta_0 \cap R$ to be a singleton whenever $\Theta_0 \cap R \neq \emptyset$, then we can use the quantiles of the standard normal distribution as critical values. This will be the case, for example, if $\theta_0(x)$ is identified. Nonparametric identification of $\theta_0(x)$, however, is not a necessary condition for $\hat{\sigma}_C^{-1}(\theta^*_J) T_N(\theta^*_J)$ to be pivotal. As an illustrative example suppose that $X, Z \in \mathbb{R}$ and we wish to test whether there is a linear function $\theta(x) = \alpha + x\beta$ in $\Theta_0$. The constraint set $R_l$ corresponding to this hypothesis is:

$$
R_l = \{\theta(x) = \alpha + x\beta : \alpha, \beta \in \mathbb{R}\}
$$

(27)

Suppose there exist two linear functions, $\theta_1(x) = \alpha_1(x) + \beta_1 x$ and $\theta_2(x) = \alpha_2(x) + \beta_2 x$ in $\Theta_0$. It follows that $E[Y|Z] = \alpha_i + E[X|Z]\beta_i$ for both $i = 1$ and $i = 2$, and therefore $(\alpha_1 - \alpha_2) = E[X|Z](\beta_2 - \beta_1)$. This equality can hold for $(\alpha_1, \beta_1) \neq (\alpha_2, \beta_2)$ if and only if $E[X|Z] = E[X]$. Therefore, if $X$ is not mean independent of $Z$ and the set $\Theta_0 \cap R_l$ is not empty, then $\Theta_0 \cap R_l$ will be a singleton. Hence, $\hat{\sigma}_C^{-1}(\theta^*_J) T_N(\theta^*_J)$ will be pivotal as long as $E[X|Z] \neq E[X]$, even if $\theta_0(x)$ is not nonparametrically identified.

In general, assuming $\Theta_0 \cap R$ is a singleton when it is not empty is equivalent to assuming the instrument can identify a unique model within $R$ that is consistent with the exogeneity assumption
on $Z$. In the linear specification test example, $E[X|Z] \neq E[X]$ is exactly what is needed to identify the linear model when $E[\varepsilon|Z] = 0$. For certain hypotheses, the set $R$ will provide enough structure for the instrument to identify a unique model within $R$. In these cases, $\Theta_0 \cap R$ will be a singleton if it is not empty and the statistic $\hat{\sigma}_C^{-1}(\theta^*_j)T_N(\theta^*_j)$ will be pivotal. This discussion, however, is secondary to whether $\theta_0(x)$ is nonparametrically identified or not. For example, even though in a linear specification test we expect $\hat{\sigma}_C^{-1}(\theta^*_j)T_N(\theta^*_j)$ to be pivotal, if $\theta_0(x)$ is not nonparametrically identified, then upon failing to reject $H_0 : \Theta_0 \cap R_l \neq \emptyset$ we cannot conclude that $\theta_0(x)$ is indeed linear. Furthermore, for certain hypotheses, such as shape restrictions or levels of a function at a point, the requirements for $\Theta_0 \cap R$ to be a singleton may be either hard to satisfy or interpret, in which case it might be advisable to use the statistic $\tilde{I}_N(R)$ instead of $\hat{\sigma}_C^{-1}(\theta^*_j)T_N(\theta^*_j)$.

If the set $\Theta_0 \cap R$ is a singleton when it is not empty, then we can use the quantiles of the standard normal distribution as critical values. We can construct a test with asymptotic size $\alpha$ for $H_0 : \Theta_0 \cap R \neq \emptyset$ by rejecting whenever:

$$\hat{\sigma}_C^{-1}(\theta^*_j)T_N(\theta^*_j) > q_{1-\alpha}$$  \hspace{1cm} (28)

where $q_{1-\alpha}$ is the $1-\alpha$ quantile of the standard normal distribution. When the null hypothesis is true and $\Theta_0 \cap R$ is a singleton, Corollary 3.1 and Theorem 4.1 imply $P \left( \hat{\sigma}_C^{-1}(\theta^*_j)T_N(\theta^*_j) > q_{1-\alpha} \right) \longrightarrow \alpha$, which controls the probability of a Type I error. The test is also consistent, since when the null is not true $P \left( \hat{\sigma}_C^{-1}(\theta^*_j)T_N(\theta^*_j) > q_{1-\alpha} \right) \longrightarrow 1$ because $\hat{\sigma}_C^{-1}(\theta^*_j)T_N(\theta^*_j)$ diverges to positive infinity.

### 4.2.2 Critical Values From Subsampling

If we do not expect $\hat{\sigma}_C^{-1}(\theta^*_j)T_N(\theta^*_j)$ to be pivotal, then we need to instead use $\tilde{I}_N(R)$ as our test statistic. Theorem 3.2 and Theorem 4.1 imply that under the null hypothesis $\tilde{I}_N(R) \overset{L}{\longrightarrow} \min_{\Theta_0 \cap R} G(\theta)$, where $G(\theta)$ is a Gaussian process defined on $L^\infty(\Theta_0)$. Since the distribution of $\min_{\Theta_0 \cap R} G(\theta)$ depends on the infinite dimensional set $\Theta_0$, it is unclear how to obtain critical values through simulation. I have also not yet been able to show the consistency, or lack thereof, of the bootstrap in the present context. Since the bootstrap can sometimes be inconsistent, see Andrews (2000) for examples, it seems imprudent to utilize the bootstrap without first establishing its validity. For this reason I resort to using subsampling, which is consistent under very general conditions.

The subsampling construction I use is from Politis, Romano & Wolf (1999). Let $b_N$ be a sequence of positive integers satisfying $b_N \longrightarrow \infty$ and $b_N/N \longrightarrow 0$. Define $B_N$ to be the set of all possible subsets of size $b_N$ of a dataset with $N$ observations and $\tilde{B}_N$ a random subset of $B_N$ satisfying
Let $|\hat{B}_N| \to \infty$, where $|\hat{B}_N|$ is the cardinality of $\hat{B}_N$. Let $\hat{I}_{b_N,V}(R)$ denote the statistic $\hat{I}_N(R)$ evaluated at a subset $V$ of the data of size $b_N$. The estimated $1 - \alpha$ quantile is then given by:

$$\hat{c}_{1-\alpha} = \inf \left\{ x : \frac{1}{|\hat{B}_N|} \sum_{V \in \hat{B}_N} 1 \left\{ \hat{I}_{b_N,V}(R) \leq x \right\} \geq 1 - \alpha \right\}$$

(29)

Given this definition of $\hat{c}_{1-\alpha}$, the following Theorem follows from Theorem 2.6.1 in Politis, Romano & Wolf (1999), Theorem 3.2 and Theorem 4.1:

**Theorem 4.2.** Suppose Assumptions 1-8 hold, $\{\Theta_J\} \subseteq \Theta$ are closed under $||\theta||_{L^2(X)} = \sqrt{\mathbb{E}[\theta^2(X)]}$ with $\Theta_J \subseteq \Theta_{J+1}$ and $\sup_{\theta \in \Theta \cap R} \inf_{\theta_J \in \Theta_J \cap R} ||\theta - \theta_J||_{L^2(X)} = o(h^{\frac{1}{2}})$ for $l \leq 1$ with $\sqrt{N}h^l \to \infty$ and $N h^{d+1} \left(1 - \frac{(m_0 + k_0)}{m_0^2}\right) \to 0$. In addition, assume $b_N \to \infty$, $b_N/N \to 0$, the $1 - \alpha$ quantile of $\min_{\theta \in \Theta \cap R} G(\theta)$ is continuous and $\hat{c}_{1-\alpha}$ is as defined in (29). If $\Theta_0 \cap R \neq \emptyset$, then $P \left( \hat{I}_N(R) > \hat{c}_{1-\alpha} \right) \to \alpha$. Furthermore, if $\Theta_0 \cap R = \emptyset$, then $P \left( \hat{I}_N(R) > \hat{c}_{1-\alpha} \right) \to 1$.

Theorem 4.2 establishes that a test that rejects whenever $\hat{I}_N(R) > \hat{c}_{1-\alpha}$ asymptotically controls the probability of a Type I error, as $P(\hat{I}_N(R) > \hat{c}_{1-\alpha}) \to \alpha$ under the null hypothesis. In addition, Theorem 4.2 also establishes that such a test is consistent, since when the null hypothesis is not satisfied $P \left( \hat{I}_N(R) > \hat{c}_{1-\alpha} \right) \to 1$.

### 5 Brazilian Fuel Engel Curves

In response to the oil shocks of the 1970s, Brazil embarked in 1975 on a national program to substitute gasoline consumption with ethanol processed from sugar cane. Originally aided by heavy subsidies, ethanol production quickly expanded, ethanol powered cars became widely available and most gas stations began selling ethanol. During the 1990s, the end of most subsidies to the ethanol industry and the low price of oil caused a sharp decline in ethanol consumption. The increase in the price of oil over the past few years, however, and the development of cars that run on both ethanol and gasoline has caused a resurgence in ethanol consumption. Today, ethanol accounts for as much as 20% of Brazil’s transport fuel market.

In this section I study the Engel curves for ethanol and gasoline in Brazil using the data from “Pesquisa de Orçamentos Familiares 2002-2003” (POF). I first test and fail to reject the hypothesis that there exist log-linear Engel curves in the identified set, a specification that corresponds to Price Independent Generalized Logarithmic (PIGLOG) preferences. I then construct confidence regions

\[\text{Source: Luhnow, D. and Samor, G. (2006)}\]

\[\text{A special case of the PIGLOG specification is the Almost Ideal Demand System of Deaton & Muellbauer (1980)}\]
for the levels of the Engel curves and their derivatives at the sample mean using both a log-linear and a fully nonparametric specification. The nonparametric confidence regions for the level of the ethanol Engel curve and its derivative are significantly larger than their parametric counterparts. In contrast, the nonparametric and log-linear confidence regions for the level of the gasoline Engel curve are similar, while their confidence regions for the derivative of the Engel curve are different. Despite the latter discrepancy, a log-linear specification provides estimates of the compensated variation associated with a price change in gasoline that are similar to the nonparametric ones.

The goal of the POF survey is to characterize Brazilian household income and consumption. It is similar to the United States Bureau of Labor Statistics Consumer Expenditure Survey (CEX). Unlike the CEX, however, the POF is done more sporadically (previous study was in 1995-1996) but more extensively (total of 48,470 households). I will use the notation $S_n^E$ and $S_n^G$ for the share of total nondurable expenditures spent by household $n$ on ethanol and gasoline respectively and $X_n$ for the log of total nondurable expenditures of household $n$. I assume the Engel curves for ethanol and gasoline fit the specification:

$$S_n^M = \phi_M(X_n) + \epsilon_n$$

(30)

where $M \in \{E, G\}$ and $\epsilon_n$ is unobserved heterogeneity. Assumption 1 requires $S_n^M$ to be continuously distributed, and hence it is not compatible with households consuming zero ethanol or gasoline with positive probability. Consequently, I carry out the analysis conditional on positive consumption, which unfortunately exposes us to an infrequency of purchase problem. Conditioning on positive consumption also implies the analysis is done only for households that own a vehicle. Therefore, the estimated Engel curves will not reflect increments in consumption due to households purchasing their first vehicle as they become wealthier.

A concern in Engel curve estimation is the possible endogeneity of total nondurable expenditures. Total nondurable expenditures is jointly determined with budget shares as part of the consumer’s optimization problem. Therefore, it is likely to be influenced by unobserved heterogeneity in preferences. As pointed out in Blundell (1988), if preferences are intertemporally weakly separable, then the consumer optimization problem can be thought of as a two stage problem. The consumer first decides how to divide total income into savings and total expenditures, and then he decides how to allocate total expenditures into the different budget shares. A commonly used instrument is total household income. For a detailed discussion see for example Summers (1959) or the analysis in Blundell, Pashardes & Weber (1993). Total household income is strongly related to total nondurable expenditures and if in addition unobserved heterogeneity conditional on positive consumption is mean independent of total income, then total income will also satisfy the exogeneity
assumption. Blundell, Chen & Kristensen (2004) use head of household income as their instrument and reject the null hypothesis of exogeneity of total nondurable expenditures. They also find their semiparametric IV estimator to provide more plausible estimates for the equivalence scale between couples with children and no children than regular semiparametric estimators. I will also use total household income as an instrument. Unlike Blundell, Chen & Kristensen (2004), however, I do not use a semiparametric specification that accounts for differences in demographics and I instead conduct a fully nonparametric analysis that focuses on only one demographic group without assuming the true Engel curve is identified.

In order to attain homogeneity across households the analysis is done for cohabitating couples in urban areas with children. A drawback of the POF dataset is that the survey was carried out over a period lasting one year. Engel curves, however, are only valid for a fixed set of prices. In order to account for changes in the price level the POF uses region specific price deflators. Unfortunately, I am not able to account for households facing different relative prices due to being interviewed in different time periods or dissimilar geographical regions.

### 5.1 Parameter Choices and Parametric Specification Test

In this section I test the null hypothesis that there exists a log linear Engel curve in the identified set. In order to implement this test, I first need to specify a number of parameter choices. For the sieve $\Theta_J$ I use polynomial approximations, so that if $J = j$, then $\Theta_J$ is the set of polynomials of order $j - 1$. In the analysis of the ethanol dataset, which contains 467 observations, I set $J = 4$ and for the gasoline dataset, which contains 4994 observations, I set $J = 5$. Changing $J$ in the neighborhood of the specified levels did not significantly affect the level of $\inf_{\Theta_J} T_N(\theta)$ in both the ethanol and gasoline datasets. In particular, adding higher powers to the gasoline optimization problem failed to significantly change the optimum obtained by using a linear fit. I also need to specify the parameters determining the norm $||\theta||_S$ as defined in (8). I set the range of integration to be $X \in [7.3, 12.5]$, which contains all the observations in both the ethanol and gasoline datasets. Because I integrate over a bounded set I let $\delta_0 = 0$, which is only needed to be positive to control the tails of the functions $\theta(x) \in \Theta$ when their domain is unbounded. The constants $m$ and $m_0$ are set to 1 and 4 respectively. Recall from Section 3.3 that the set $R$ needs to be closed under the norm $||\theta||_{C5}$ as defined in (23). By setting $m_0 = 4$, it is possible to test restrictions on the level of the Engel curve and its derivatives up to order four. Given these parameters, the functions $\theta^*_J(x) = \arg \min_{\Theta_J} T_N(\theta)$ had norms $||\theta^*_J||_S = 0.26$ in the ethanol dataset and $||\theta^*_J||_S = 0.07$ in the
gasoline application. In order to complete the characterization of \( \Theta = \{ \theta(x) : ||\theta||_S \leq B \} \) I set \( B = 5 \), which seems appropriately large given the reported values for \( ||\theta^*_J||_S \).

Subsampling requires the choice of the following additional parameters: the size of the subsample \( b_N \), the bandwidth for the subsample \( h_{b_N} \), and the choice of sieve for the subsample \( \Theta_{J_{b_N}} \). Assumptions 5 and 7 provide some guidance as to how to set \( h_{b_N} \) because they require \( h \) to satisfy

\[
Nh^d \rightarrow \infty, \quad \sqrt{Nh}^l \rightarrow \infty \quad \text{and} \quad Nh^{d+l\left(1-\frac{\ln m_0+\delta_0}{2m_0}\right)} \rightarrow 0 \quad \text{for some} \quad l \leq 1.
\]

When \( \delta_0 = 0 \), as in the present case, the last rate simplifies to

\[
Nh^{d+l} - \delta_0 k^2 m_0 \rightarrow 0.
\]

In order to limit the choices of \( h \) and \( h_{b_N} \), I assumed \( Nh^{1+\frac{l}{2}} = C \) for some constant \( C \), which is compatible with Assumptions 5 and 7 with \( l \in \left( \frac{6}{7}, \frac{7}{8} \right) \). With this rate constraint, a choice for \( h \) and \( b_N \) imply a particular \( h_{b_N} \). Utilizing this restriction I carried out a small sensitivity analysis of the effects of different parameter choices for \( h \) and \( b_N \) on hypothesis testing. In particular, I examined the effects of different parameter choices on a family of hypotheses of the form

\[
H_0 : \Theta_0 \cap R_u \neq \emptyset \quad \text{for} \quad R_u = \{ \theta(x) : \theta(\bar{X}) \geq u \}
\]

for different values of \( u \). A summary of the results can be found in Table 5 in Appendix E. Instead of reporting the subsampling estimate \( \hat{c}_{1-\alpha} \) for the critical value, the results are given in terms of the statistic:

\[
\hat{F} = \frac{1}{|B_N|} \sum_{V \in B_N} 1\left\{ \hat{I}_{b_N,V}(R) \leq \hat{I}_N(R) \right\}.
\]  

(31)

The definition of \( \hat{c}_{1-\alpha} \) in (29) implies \( \hat{I}_N(R) > \hat{c}_{1-\alpha} \) if and only if \( \hat{F} > 1 - \alpha \). Therefore, \( 1 - \hat{F} \) is the minimum size of the test for which we would have rejected and hence provides a measure of how far the null hypothesis is from being rejected. The sensitivity analysis shows that it is harder to reject for higher choices of \( b_N \), i.e. higher values of \( b_N \) yield lower values for \( \hat{F} \). I also find that a higher choice for \( h \) makes rejection more likely. The theory requires \( h \) to decrease quickly to zero in order to control certain biases, and thus a large choice for \( h \) seems unadvisable. Overall, however, I did not find considerable changes in the levels of \( u \) for which the hypothesis \( H_0 : \Theta_0 \cap R_u \neq \emptyset \) was rejected as \( b_N \) and \( h \) changed. All parameter choices are summarized in Table 1.

### Table 1: Parameter Values

<table>
<thead>
<tr>
<th>Market</th>
<th>( N )</th>
<th>( J )</th>
<th>( h )</th>
<th>( b_N )</th>
<th>( h_{b_N} )</th>
<th>( \hat{B}_N )</th>
<th>( B )</th>
<th>( m )</th>
<th>( m_0 )</th>
<th>( \delta_0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gasoline</td>
<td>4994</td>
<td>5</td>
<td>0.03</td>
<td>100</td>
<td>3</td>
<td>0.28</td>
<td>500</td>
<td>5</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Ethanol</td>
<td>467</td>
<td>4</td>
<td>0.1</td>
<td>30</td>
<td>2</td>
<td>0.48</td>
<td>500</td>
<td>5</td>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>

I found that a large value for \( \hat{B}_N \) was necessary to obtain precise estimates for \( \hat{F} \). The choices of \( J_{b_N} \) seemed natural given the values for \( b_N \). The value of \( \hat{F} \) for the null hypothesis \( H_0 : \Theta_0 \cap \Theta \neq \emptyset \) was 0.04 and 0.36 for gasoline and ethanol respectively. Therefore, I fail to reject the null hypothesis that there is a function \( \theta(x) \in \Theta \) satisfying \( E[Y - \theta(X)|Z] = 0 \).
I first test whether there are log-linear Engel curves in the identified set. This test was discussed in Section 4.2.1 and corresponds to the null hypothesis $H_0 : \Theta_0 \cap R_i \neq \emptyset$, where $R_i$ was defined in (27). For both the gasoline and ethanol Engel curves, I fail to reject the null hypothesis for a test size $\alpha = 0.05$. Table 2 summarizes the values of the relevant statistics:

Table 2: Linear Specification Test

<table>
<thead>
<tr>
<th>Market</th>
<th>$I_N(R_i)$</th>
<th>$F$</th>
<th>$\hat{\sigma}^{-1}(\hat{\theta}_j^* T_N(\hat{\theta}_j^*)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gasoline</td>
<td>-0.0016</td>
<td>0.04</td>
<td>-275.31</td>
</tr>
<tr>
<td>Ethanol</td>
<td>-0.0008</td>
<td>0.55</td>
<td>-48.81</td>
</tr>
</tbody>
</table>

The values of the test statistic $\hat{\sigma}^{-1}(\hat{\theta}_j^* T_N(\hat{\theta}_j^*)$ are problematic. As discussed in Section 4.2.2, the statistic $\hat{\sigma}^{-1}(\hat{\theta}_j^* T_N(\hat{\theta}_j^*) \overset{\mathcal{L}}{\rightarrow} N(0, 1)$ when the null hypothesis is satisfied and diverges to infinity otherwise. Large negative values, as reported in Table 2, are therefore not consistent with either the null hypothesis being true nor with it being false. The large negative values reflect that $\hat{\sigma}^2_C(\hat{\theta}_j^*)$ provides a poor estimator for the asymptotic variance or that the normal distribution provides a poor approximation to the finite sample distribution.

If the true Engel curves are not identified, then the failure to reject $H_0 : \Theta_0 \cap R_i \neq \emptyset$ does not imply the log-linear specification is correct. It is possible that even though there are linear functions in the identified set, the true Engel curves are actually nonlinear. Figure 1 depicts the best linear fit, i.e. $\arg \min_{\Theta \cap R_i} T_N(\theta)$, the best nonlinear fit, i.e. $\arg \min_{\Theta_j} T_N(\theta)$, and a kernel
estimator of the density of $X$. While the linear and nonlinear fits look similar for gasoline, they are quite different for ethanol. However, in both cases the linear and nonlinear fits agree where most of the data is located, as illustrated by the density estimator. Even though the linear and nonlinear ethanol Engel curves look rather different, I fail to reject the null hypothesis that the residuals they imply are mean independent of $Z$. Hence, both the linear and nonlinear fits could be in $\Theta_0$.

5.2 Hypothesis Testing on Levels and Derivatives

After failing to reject the null hypothesis that there are linear Engel curves in the identified set, we may feel comfortable using a PIGLOG specification throughout the rest of the analysis. Even if there indeed is a linear Engel curve in the identified set, however, if the true Engel curve is not identified, then there is no guarantee that it is actually linear. Therefore, confidence regions that are constructed assuming linearity may asymptotically exclude the true functional of interest when the Engel curve is not identified. In this section I explore the effects of a linearity assumption by using the framework discussed in Section 2.3.1 to build confidence regions for the level of the Engel curve and its derivative at the sample average with and without a linearity assumption. These results should not be interpreted as confidence regions for $\phi_M(E[X])$ and $\phi'_M(E[X])$ respectively, but literally for the functions $\phi_M(x)$ and $\phi'_M(x)$ evaluated at the point $\bar{X}$. Obtaining confidence regions for $\phi_M(E[X])$ and $\phi'_M(E[X])$ requires the evaluation of the statistic $\hat{I}_N(R)$ at a sequence of sets $R$ changing with sample size as $\bar{X}$ changes. Theorems 4.1 and 4.2, however, apply only to deterministic choices of $R$ and can therefore not be used to construct confidence regions for $\phi_M(E[X])$ and $\phi'_M(E[X])$.

I build a nonparametric confidence region for $\phi_M(\bar{X})$ by using a family of sets of the form $R_m = \{\theta(x) \in \Theta : \theta(\bar{X}) = m\}$ and including in the confidence region all those values of $m$ for which I fail to reject $H_0 : \Theta_0 \cap R_m \neq \emptyset$. Similarly, in order to construct a confidence region assuming linearity I include all those values of $m$ for which I fail to reject $H_0 : \Theta_0 \cap R_m \cap R_l \neq \emptyset$, for $R_l$ defined in (27). These linear confidence intervals are not equivalent to the traditional linear IV confidence regions. Figure 2 depicts the value of $\hat{F}$ for the hypotheses $H_0 : \Theta_0 \cap R_m \neq \emptyset$ and $H_0 : \Theta_0 \cap R_m \cap R_l \neq \emptyset$ as a function of $m$. Theorem 4.2 implies $H_0 : \Theta_0 \cap R_m \neq \emptyset$ should be rejected whenever $\hat{F} > 1 - \alpha$, where $\alpha$ is the size of the test. Therefore, in Figure 2, the confidence regions are determined by all the values of $m$ for which $\hat{F}$ is below the horizontal line, which signals the value 0.95. For the gasoline Engel curve, assuming linearity is an innocuous assumption as the confidence regions for the nonparametric and the linear specifications are similar. For the ethanol
Engel curve, however, the upper bounds on the linear and nonparametric confidence regions for $\phi_E(\bar{X})$ are similar, but their lower bounds are quite different.

As illustrated in Figure 2, a drawback of this procedure is that $\hat{F}$ is not necessarily a monotonic function of $m$. Even if the statistics $\hat{I}_N(R_m)$ are increasing in $m$, how they compare to their respective critical values might not be. A possible undesirable consequence is having confidence regions that are not connected even when the parameter of interest is identified. I did not run into such a problem in this study. Hence, in Table 3 the confidence regions are reported as intervals:

<table>
<thead>
<tr>
<th>Market</th>
<th>Nonparametric</th>
<th>Linear</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gasoline</td>
<td>[10.5%, 12.2%]</td>
<td>[10.8%, 11.8%]</td>
<td>[11.1%, 11.5%]</td>
</tr>
<tr>
<td>Ethanol</td>
<td>[5.0%, 14.2%]</td>
<td>[10.6%, 12.7%]</td>
<td>[11.1%, 12.9%]</td>
</tr>
</tbody>
</table>

where the IV confidence intervals are the traditional linear IV ones. The confidence regions that assume linearity are similar to those obtained through standard IV methods. The nonparametric confidence region for the level of the gasoline Engel curve at $\bar{X}$ is surprisingly narrow. This nonparametric confidence region provides conclusions similar to those that can be obtained with traditional IV methods while being a more robust procedure. In contrast, the nonparametric confidence regions for the level of the ethanol Engel curve at $\bar{X}$ is considerably wider than its IV counterpart. Given the important difference in the size of the datasets, it is unclear whether these differences can be
attributed to sample size or an inherent difference in the Engel curves for ethanol and gasoline.

These results do not imply traditional IV procedures are necessarily a reliable option. As is exemplified by the confidence regions for the slope of the Engel curves at the sample average, linearity can sometimes be a problematic assumption. In order to construct nonparametric confidence regions for $\phi'_M(\bar{X})$ I use the sets $R_s = \{ \theta(x) \in \Theta : \theta'(\bar{X}) = s \}$ and include in the confidence region all the values of $s$ for which $H_0 : \Theta_0 \cap R_s \neq \emptyset$ is not rejected. Figure 3 is analogous to Figure 2, except that the statistic $\hat{F}$ is plotted as a function of $s$ instead of $m$. The previous results suggested that linearity was an excellent parametric specification for the gasoline Engel curve. Figure 3, however, depicts how the effects of a parametric assumption depend on what the parameter of interest is. In contrast to the confidence region for $\phi_G(\bar{X})$, the confidence region for $\phi'_G(\bar{X})$ is significantly affected by whether we assume linearity or not. In Table 4 I report the linear, nonparametric and traditional IV confidence regions. The lack of monotonicity of $\hat{F}$ in $s$ generated confidence regions that were not connected. The upper/lower bounds in Table 4 correspond to the levels for which all greater/smaller values were rejected.

The linear and traditional IV confidence regions are similar. For the ethanol Engel curve, it is not possible to determine the sign of its derivative at $\bar{X}$ using the fully nonparametric confidence region. This result may be a consequence of the small sample size relative to that of the gasoline dataset and not reflective of an actual difference in the nature of both Engel curves. If there is indeed a linear Engel curve in the identified set for gasoline, then we should expect the linear confidence region
to be contained in the nonparametric one. The critical values estimated for the null hypothesis
\( H_0 : \Theta_0 \cap R_s \cap R_l \neq \emptyset \), however, will be different from those estimated for \( H_0 : \Theta_0 \cap R_s \neq \emptyset \). The former are obtained imposing linearity in the subsample as well, and hence \( J_{bN} = 2 \). In contrast, the critical values for the nonparametric confidence region were calculated using \( J_{bN} = 3 \), which explains why the upper bound on the nonparametric confidence region can be smaller than the one for the linear confidence region.

Despite agreeing on the upper bound, the lower bound of the linear and nonparametric confidence regions for the derivative of the gasoline Engel curve at \( \bar{X} \) are significantly different. I evaluate the economic significance of this discrepancy with a simplistic approximation to the compensated variation corresponding to a price change in gasoline. Using a Taylor approximation, Slutsky’s equation and duality we can derive the following approximation to compensated variation from a price change from \( p_0 \) to \( p_1 \):

\[
CV(p_0, p_1, w) \approx x(p_0, w)(p_1 - p_0) + \left[ \frac{\partial x(p_0, w)}{\partial p} + \frac{\partial x(p_0, w)}{\partial w} x(p_0, w) \right] (p_1 - p_0) \tag{32}
\]

where \( x(p, w) \) is Walrasian demand for gasoline at price \( p \) and total expenditures \( w \). It is not possible to estimate \( \frac{\partial x(p_0, w)}{\partial p} \) with the POF dataset. Short term demand for gasoline is highly inelastic, however, and I therefore assume \( \frac{\partial x(p_0, w)}{\partial p} (p_1 - p_0) \approx 0 \). In addition, by noting that \( x(p_0, w) = w\phi_G(\log(w)) \) and \( \frac{\partial x(p_0, w)}{\partial w} = \phi_G(\log(w)) + \phi'_G(\log(w)) \), we can simplify (32) to derive:

\[
CV(p_0, p_1, w) \approx \left[ 1 + \phi_G(\log(w)) + \phi'_G(\log(w)) \right] w\phi_G(\log(w))(p_1 - p_0) \tag{33}
\]

Equation (33) provides an approximation to compensated variation that is solely in terms of the Engel curve \( \phi_G(\log(w)) \). In Table 3, I found the parametric and nonparametric confidence regions for \( \phi_G(\bar{X}) \) to be similar, while Table 4 shows the confidence regions for \( \phi'_G(\bar{X}) \) are significantly different. These results and equation (33) suggest a log-linear specification may overestimate the compensated variation associated with a price change in gasoline. To examine this hypothesis, I construct confidence regions for the compensated variation associated with a price change of one Real (approximately 30 cents). In order to do so, I evaluate the series of null hypothesis \( H_0 : \)

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|}
\hline
Market & Nonparametric & Linear & IV & \\
\hline
Gasoline & [-0.089, -0.005] & [-0.031, -0.003] & [-0.026, -0.018] & \\
Ethanol & [-0.148, 0.033] & [-0.092, -0.003] & [-0.082, -0.043] & \\
\hline
\end{tabular}
\end{table}
Figure 4: Confidence Regions for $CV(p_0, p_1, w)$

$\Theta_0 \cap R_c \neq \emptyset$ for $R_c = \{ \theta(x) \in \Theta : [1 + \theta(\bar{X}) + \theta'(\bar{X})] e^{\bar{X} \theta(\bar{X})} = c \}$. Figure 4 depicts the parametric and nonparametric confidence regions for $CV(p_0, p_1, w)$. While the parametric and nonparametric confidence regions for the slope of the Engel curve differed substantially, these differences do not translate into dissimilar confidence intervals for compensated variation. Thus, the analysis suggests that while a linear specification may overestimate the derivative of the Engel curve at the sample mean, such a parametrization can still be relied upon to obtain estimates of compensated variation.

6 Conclusions

In this paper I developed methods that allow us to test a wide array of hypotheses in a nonparametric IV framework. Without parametric assumptions the conditions necessary to attain identification are very stringent. Since these requirements may not be met in many datasets, I consider the nonparametric IV problem as one of partial identification. Instead of assuming identification and testing whether the true model satisfies a certain restriction, I test whether at least one element of the identified set satisfies the hypothesized restriction.

This framework was applied to study the Engel curves for ethanol and gasoline in Brazil. I fail to reject the null hypothesis that there is at least one log-linear Engel curve in the identified set for both gasoline and ethanol. In addition, I examined parametric and nonparametric confidence
intervals for the levels of the Engel curves and the derivatives as well as for the compensated variation associated with a price change in gasoline. The results illustrate that the impact of using a particular parametrization depends on what the functional of interest is.
APPENDIX A - Notation and Definitions

The following is a table of the notation and definitions that will be used throughout the appendix, including many that go beyond the ones already introduced in the main text:

- \[ \alpha \leq \beta \] for some constant \( M \) which is universal in the context of the proof
- \[ ||\theta||_\infty \] The sup-norm \( \sup_x |\theta(x)| \)
- \[ ||\theta||_S \] The norm \( \sum_{i=1}^n |\theta(x_i)| \)
- \[ ||\theta||_{C_1} \] The norm \( \sup_x |\theta(x)| \)
- \[ ||\theta||_{C_2} \] The norm \( \sup_x |\theta(x)| \)
- \( \rho_{m,N}(f,g) \) The random semimetric \[ \sum_{i=1}^n (f(x_i) - g(x_i))^2 \]
- \[ ||f||_m \] The random norm \( \sum_{i=1}^n (f(x_i))^2 \)
- \[ C_N(\theta) \] The U-statistic \( \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n f(x_i, x_j) \)
- \[ I_{m,N} \] Set of distinct \( m \)-tuples from \( N \) observations
- \( \mathcal{L}^\infty(\Theta) \) The metric space of bounded functionals on \( \Theta \) with norm \( \sup_{\theta \in \Theta} |F_\theta(\cdot) - F_\theta(\cdot)| \)
- \( N(F, \| \cdot \|, \epsilon) \) Covering numbers of size \( \epsilon \) for \( F \) under the norm \( \| \cdot \| \)
- \( N(\mathcal{F}, \| \cdot \|, \epsilon) \) Bracketing numbers of size \( \epsilon \) for \( \mathcal{F} \) under the norm \( \| \cdot \| \)
- \( P \)-Canonical function A function \( f(X^n) \) satisfying \( E[f(X^n)X^n] = 0 \) for all \( n < m \)
- \( \sigma^2_{P}(\theta) \) The asymptotic variance of \( T_N(\theta) \) \( \theta \in \Theta_0 \)
- \( \sigma^2_{C,1}(\theta) \) Component of \( \sigma^2_{P}(\theta) \) equal to \( E[\theta - \hat{\theta}^2] \)
- \( \sigma^2_{C,1}(\theta) \) Component of \( \sigma^2_{P}(\theta) \) equal to \( 2 \int \left( E[(\theta - \hat{\theta})^2] \right) f_\theta(Z) \)
- \( X_m \) The random variable consisting of \( m \) independent copies of the random variable \( X \)
- \( w_n \) The triplet of random variables \( (z_n, x_n, y_n) \)

APPENDIX B - Proof of Theorem 3.1 and Auxiliary Theorem .1 and Lemmas .1 and .2

**Theorem .1:** If Assumptions 1-5 hold and \( \theta \in \Theta_0 \), then for \( \Phi_{C,N} \) a random variable satisfying \( \Phi_{C,N} \rightarrow \mathcal{L} N(0,1) \):

\[
Q_N(\theta) - \hat{B}_N(\theta) = \frac{\sigma_{C}(\theta)}{N h^2} \Phi_{C,N}
\]

Let \( V(z_n, v_n) = (E[v_n(\theta)]Z = z_n)^2 f^2_2(z_n) + 2E[v_n(\theta)]Z = z_n|v_n(\theta)|f_2^2(z_n) \) and \( \sigma^2_D(\theta) = \text{Var}[V(z_n, v_n)] \). Also let \( D_N(\theta) = \frac{(N-1)(N-2)}{N^2 N_{m,n}} E \left[ \left( \frac{z_n - z_n}{h} \right) \left( \frac{z_j - z_j}{h} \right) v_i(\theta) v_j(\theta) \right] \). If Assumptions 1-5 hold and in addition \( \theta \notin \Theta_0 \), then \( D_N(\theta) \rightarrow E[V(z_n, v_n)] \) and:

\[
Q_N(\theta) - \hat{B}_N(\theta) = \frac{\sigma_{D}(\theta)}{\sqrt{N}} \Phi_{D,N} + D_N(\theta)
\]

where \( \Phi_{D,N} \) is a random variable satisfying \( \Phi_{D,N} \rightarrow \mathcal{L} N(0,1) \).

**Proof of Theorem .1:** By expanding the squares in \( Q_N(\theta) \) and regrouping terms, it is possible to see that \( Q_N(\theta) - \hat{B}_N(\theta) \) is actually a symmetric U-Statistic of order 3 with kernel:

\[
H_2(w_n, w_i, w_j|\theta) = K \left( \frac{z_i - z_n}{h} \right) K \left( \frac{z_j - z_n}{h} \right) v_i(\theta) v_j(\theta)
+ K \left( \frac{z_n - z_i}{h} \right) K \left( \frac{z_j - z_i}{h} \right) v_n(\theta) v_i(\theta) + K \left( \frac{z_n - z_j}{h} \right) K \left( \frac{z_i - z_j}{h} \right) v_n(\theta) v_j(\theta)
\]

When the context is clear, the dependence on \( w_n, w_i, \) and \( w_j \) will be suppressed. The kernel \( H_2(w_n, w_i, w_j|\theta) \) yields a U-Statistic that is degenerate of order 1 if \( \theta \in \Theta_0 \) but is well behaved when \( \theta \notin \Theta_0 \). We also note that since the
kernel \( K(u) \), the \( \theta \in \Theta \) are bounded and \( E[Y^4] < \infty \), we have that \( E[H_2^2(\theta)] < \infty \). For notational simplicity, let \( C_N(\theta) = Q_N(\theta) - \hat{B}_N(\theta) \) and note that by rearranging terms we can obtain:

\[
C_N(\theta) = Q_N(\theta) - \hat{B}_N(\theta) = \frac{2}{N^{3/2}h^2} \sum_{n=1}^{N} \sum_{i<n} \sum_{j<i} H_2(w_n, w_i, w_j|\theta)
\]

In order to establish the asymptotic behavior of \( C_N(\theta) \) we will need to consider two cases: \( \theta \in \Theta_0 \) and \( \theta \notin \Theta_0 \).

Case 1: \( \theta \in \Theta_0 \)

If \( \theta \in \Theta_0 \), then \( E[v_n(\theta) | Z = z_n] = 0 \). This implies \( E[H_2(\theta)|w_n] = 0 \) and hence \( H_2(w_n, w_i, w_j|\theta) \) is degenerate. We derive the asymptotic distribution of \( C_N(\theta) \) by using its Hoeffding decomposition \( C_N(\theta) = C_N^I(\theta) + C_N^H(\theta) \), where:

\[
C_N^I(\theta) = \frac{2}{N^{3/2}h^2} \sum_{n=1}^{N} \sum_{i<n} \sum_{j<i} E[H_2(\theta)|w_n, w_i] + E[H_2(\theta)|w_i, w_j] + E[H_2(\theta)|w_n, w_j] \\
C_N^H(\theta) = \frac{2}{N^{3/2}h^2} \sum_{n=1}^{N} \sum_{i<n} \sum_{j<i} H_2(w_n, w_i, w_j|\theta) - E[H_2(\theta)|w_n, w_i] - E[H_2(\theta)|w_i, w_j] - E[H_2(\theta)|w_n, w_j]
\]

We first analyze the asymptotic behavior of \( C_N^I(\theta) \). By the i.i.d. assumption, and since \( \theta \in \Theta_0 \), we have that \( E[v_n(\theta)|w_i, w_j, z_n] = 0 \). Thus, a straightforward calculation and rearranging terms shows that:

\[
C_N^I(\theta) = \frac{2(N-1)}{N^{3/2}h^2} \sum_{n=1}^{N} \sum_{i<n} \left( \int K \left( \frac{z_i - z}{h} \right) K \left( \frac{z_n - z}{h} \right) f_Z(z)dz \right) v_n(\theta)v_i(\theta) \tag{35}
\]

The asymptotic distribution of the expression in (35) was analyzed in Hall (1984) under slightly different assumptions. For the sake of completeness I include his derivations in the present paper, following his notation as closely as possible.

Let \( W_{ni} = \int K \left( \frac{z_i - z}{h} \right) K \left( \frac{z_n - z}{h} \right) f_Z(z)dz \), \( Y_n = v_n(\theta) \sum_{i=1}^{n-1} W_{ni} v_i(\theta) \) and use (35) to establish:

\[
N^{3/2}h^2 C_N^I(\theta) = \sum_{n=1}^{N} \sum_{i<n} W_{ni} v_n(\theta) v_i(\theta) = \sum_{n=2}^{N} Y_n \tag{36}
\]

Let \( \mathcal{F}_n \) denote the \( \sigma \)-field generated by \( \{Z_1, \ldots, Z_n, Z_{n+1}\} \) and \( \{V_1(\theta), \ldots, V_n(\theta)\} \) for \( 0 \leq n \leq N \). Then \( E[Y_n|\mathcal{F}_{n-1}] = 0 \) for all \( n \), which implies the sequence \( \{S_N = \sum_{n=2}^{N} Y_n, \mathcal{F}_N\} \) is a martingale. Following Hall (1984), I aim to prove a central limit theorem for \( S_N \). Let \( \sigma^2(z_n, \theta) = E[v_n^2(\theta)|Z = z_n] \) and note that by the i.i.d. assumption, the conditional variance of \( S_N \) is given by:

\[
V_N = \sum_{n=2}^{N} E[Y_n^2|\mathcal{F}_{n-1}] = \sum_{n=2}^{N} \sigma^2(z_n, \theta) \left[ \sum_{i=1}^{n-1} W_{ni} v_i(\theta) \right]^2 = \sum_{n=2}^{N} \sigma^2(z_n, \theta) \sum_{i=1}^{n-1} W_{ni}^2 v_i^2(\theta) + 2 \sum_{n=2}^{N} \sigma^2(z_n, \theta) \sum_{i=1}^{n-1} \sum_{k<i} W_{ni} W_{nk} v_i(\theta) v_k(\theta) = V_{N1} + V_{N2} \tag{37}
\]

From Lemma (.1), it follows that \( N^{-2}h^{-3/2}V_N \xrightarrow{p} \sigma^2_C(\theta)/4 \). In addition, Lemma (.2) verifies a Lindenberg-Feller type condition for the martingale \( S_N \). Thus, by Corollary 3.1 of Hall and Heyde (1980) and the continuous mapping theorem we conclude that \( 2N^{-1}h^{-3/2} \sigma_C(\theta)^{-1} \sum_{i=1}^{N} Y_i \xrightarrow{L} N(0, 1) \). Hence, using (36) we obtain:

\[
C_N^I(\theta) = \frac{\sigma_C(\theta)}{Nh^2} \Phi_{C_N} \tag{38}
\]

where \( \Phi_{C_N} \) is a random variable satisfying \( \Phi_{C_N} \xrightarrow{L} N(0, 1) \). I now show that \( C_N^I(\theta) \) is asymptotically negligible. First we need to compute \( E[H_2^2(\theta)] \). In (39) expand the square, notice that since \( \theta \in \Theta_0 \) cross terms disappear, and do the change of variables \( u = (z_i - z_n)/h \) and \( v = (z_j - z_n)/h \) to obtain the equality:

\[
E[H_2^2(\theta)] = 3h^2d \int K^2(u)K^2(v)v_i^2(\theta)v_j^2(\theta)f_{ZV}(z_n + hu, v_i)f_{ZV}(z_n + hv, v_j)f_Z(z_n)du dv_i dv_j \tag{39}
\]
Since \( \sigma^2(z, \theta) \), \( f_Z(z) \) and \( K(u) \) are bounded it follows that \( E[H_2^2(\theta)] = O(h^{2d}) \). We can now show that \( C_N^{II}(\theta) \) is asymptotically negligible. The first inequality in (40) follows by expanding the square, noticing that the cross terms disappear and using that projections reduce second moments. The second inequality follows by \( E[H_2^2(\theta)] = O(h^{2d}) \).

\[
E[(C_N^{II}(\theta))^2] \leq \frac{4}{N^6 h^{4d}} \sum_{n=1}^{N} \sum_{i<n} \sum_{j<i} E[H_2^2(\theta)] \lesssim \frac{1}{N^3 h^{2d}} \tag{40}
\]

Markov’s inequality and (40) imply that \( C_N^{II}(\theta) = O_p(N^{-\frac{1}{2}} h^{-d}) \). Since \( Nh^d \to \infty \), it follows that \( Nh^d C_N^{II}(\theta) \to 0 \).

We combine this result with (38) and \( C_N(\theta) = C_N^{I}(\theta) + C_N^{II}(\theta) \) to conclude that if \( \theta \in \Theta_0 \), then:

\[
C_N(\theta) = Q_N(\theta) - \hat{B}_N(\theta) = \frac{\sigma_{\xi}(\theta)}{Nh^d} \Phi_C \tag{41}
\]

This concludes the proof of the theorem for the case \( \theta \in \Theta_0 \). We now look at the case \( \theta \notin \Theta_0 \).

Case II: \( \theta \notin \Theta_0 \)

I begin by showing that if \( \theta \notin \Theta_0 \), then for \( h \) small enough \( E[H_2(\theta)|w_n] \neq 0 \). In (42) do the change of variables \( u = (z - z_n)/h \) to obtain the first equality. Since \( E[v(\theta)|Z] \) and \( f_Z(z) \) are bounded and continuous, the dominated convergence theorem implies the second result in (42). The final result in (42) is definitional.

\[
h^{-2d} E[H_2(\theta)] = 3 \int \left( \int K(u) v(\theta)f_{ZV}(z_n + hu, v)du dv \right)^2 f_Z(z_n)dz_n \to 3E \left[ (E[v(\theta)|Z])^2 f_Z^2(Z) \right] \equiv \mu^2(\theta) \tag{42}
\]

Since \( \mu^2(\theta) > 0 \) for \( h \) small enough, \( E[H_2(\theta)] \neq 0 \) and hence \( E[H_2(\theta)|w_n] \neq 0 \) either, which together with (34) implies \( E[H_2(\theta)|w_n] \neq E[H_2(\theta)] \). Thus, the kernel \( H_2(w_n, w, w, \theta) \) is not degenerate for \( h \) small enough. We proceed by decomposing \( C_N(\theta) = C_N^{I}(\theta) + C_N^{II}(\theta) + D_N(\theta) \), where:

\[
C_N^{I}(\theta) = \frac{2}{N^{3\alpha_d}} \sum_{n=1}^{N} \sum_{i<n} \sum_{j<i} E[H_2(\theta)|w_n] + E[H_2(\theta)|w_i] + E[H_2(\theta)|w_j] - 3E[H_2(\theta)]
\]
\[
C_N^{II}(\theta) = \frac{2}{N^{3\alpha_d}} \sum_{n=1}^{N} \sum_{i<n} \sum_{j<i} H_2(w_n, w, w, \theta) - E[H_2(\theta)|w_n] - E[H_2(\theta)|w_i] - E[H_2(\theta)|w_j] + 2E[H_2(\theta)]
\]
\[
D_N(\theta) = \frac{2}{N^{3\alpha_d}} \sum_{n=1}^{N} \sum_{i<n} \sum_{j<i} E[H_2(\theta)]
\]

We will establish the asymptotic behavior of \( C_N(\theta) \) by analyzing the terms \( C_N^{I}(\theta) \), \( C_N^{II}(\theta) \) and \( D_N(\theta) \). We regroup the terms in \( C_N^{I}(\theta) \) to obtain a triangular array to which we can apply a central limit theorem:

\[
C_N^{I}(\theta) = \frac{(N-1)(N-2)}{N^{3\alpha_d h^{2d}}} \sum_{n=1}^{N} E[H_2(\theta)|w_n] - E[H_2(\theta)] \tag{43}
\]

In order to show the asymptotic normality of \( C_N^{I}(\theta) \), we begin by calculating \( E[(E[H_2(\theta)|w_n])^2] \). In (44) use the i.i.d. assumption and the change of variables \( u = (z_i - z_n)/h \) and \( s = (z_j - z_n)/h \) to obtain the first equality. Since \( E[v(\theta)|Z] \) and \( f_Z(Z) \) are bounded and continuous the dominance convergence theorem applies, which implies the second result in (44). The last equality is definitional.

\[
h^{-4d} E \left[ (E[H_2(\theta)|w_n])^2 \right] = \int \left( \int K(u)v(\theta)f_{ZV}(hu + z_n, v)du dv \right)^2 f_Z(z_n)dz_n dv_n + 2 \int K(u)K(s - u)v(\theta)f_{ZV}(hs + z_n, v)vduds \int dz_n dv_n \tag{44}
\]
With results (42) and (44) we can now proceed to check the conditions for Liapounov’s CLT for \( C_N^I(\theta) \). First note that Jensen’s inequality and the assumptions that \( E[Y^{1+\delta}] < \infty \) for some \( \delta > 0 \) and \( \theta \) being bounded imply that \( E[(E[H_2(\theta)|w_n])^{1+\delta}] < \infty \). Let \( \sigma_D^2(\theta) = \mu_D^2(\theta) - (\mu_D^2(\theta))^2 \) and note that by (42) and (44) we have \( \text{Var} \left[ \sum_{n=1}^N E[H_2(\theta)|w_n] - E[H_2(\theta)] \right] = Nh^{4d}(\sigma_D^2(\theta) + o(1)) \). Now use the i.i.d. assumption and the inequality \( E[(a - E[a])^{2+\delta}] \leq 2^{1+\delta} E[a^{2+\delta}] \) to get the first inequality in (45). Arguments similar to those in (44) can be used to show that \( E \left[ (E[H_2(\theta)|w_n])^{2+\delta} \right] = O(h^{2d(2+\delta)}) \), which implies the third inequality in (45).

\[
\lim_{N \to \infty} \sum_{n=1}^N E \left[ \left( \frac{E[H_2(\theta)|w_n] - E[H_2(\theta)]}{\sqrt{Nh^{2d}(\sigma_D^2(\theta) + o(1))}} \right)^{2+\delta} \right] \leq \lim_{N \to \infty} \frac{N}{Nh^{2d}} \frac{N}{2^{1+\delta}} E \left[ (E[H_2(\theta)|w_n])^{2+\delta} \right] \leq \lim_{N \to \infty} \frac{1}{N^2} = 0 \tag{45}
\]

Equation (45) concludes verifying the conditions of Liapounov’s CLT. Thus, the continuous mapping theorem and (43) imply that for a random variable \( \Phi_{D,N} \xrightarrow{L} N(0,1) \) we have:

\[
C_N^I(\theta) = \frac{\sigma_D(\theta)}{\sqrt{N}} \Phi_{D,N} \tag{46}
\]

I now show that \( C_N^{II}(\theta) \) is asymptotically negligible. In (47), expand the square, and notice that cross terms with just one or no variables in common have zero expectation. Terms with two elements in common have expectation equal to \( E[(E[H_2(\theta)|w_n, w_i])^2] - 2E[(E[H_2(\theta)|w_n])^2] + (E[H_2(\theta)])^2 \). Since projections decrease second moments, the expectation of terms with all elements in common is less than or equal to \( E[H_2(\theta)] \). We use these results and \( E[(E[H_2(\theta)|w_n])^2] \geq (E[H_2(\theta)])^2 \) by Jensen’s inequality to derive the first inequality in (47). Calculations similar to those in (42) can be used to show that \( E[H_2^2(\theta)] = O(h^{2d}) \). In addition, Jensen’s inequality for conditional expectations and (44) imply that \( E[(E[H_2(\theta)|w_n, w_i])^2] = O(h^{4d}) \). Combining these result with the i.i.d. assumption we obtain the last equality in (47).

\[
E \left[ (C_N^{II}(\theta))^2 \right] \leq \frac{1}{N^{3}h^{4d}} E \left[ H_2^2(\theta) \right] + \frac{1}{N^{2}h^{2d}} E \left[ (E[H_2(\theta)|w_n, w_i])^2 \right] = O(N^{3}h^{2d} + N^{2}) \tag{47}
\]

Markov’s inequality and (47) imply that \( C_N^{II}(\theta) = O_p(N^{\frac{2}{3}}h^{d} + N) \). Therefore, since \( Nh^{d} \to \infty \), it follows that \( \sqrt{N}C_N^{II}(\theta) = o_p(1) \). Combining this result with (46) and the continuous mapping theorem implies:

\[
C_N^I(\theta) + C_N^{II}(\theta) = \frac{\sigma_D(\theta)}{\sqrt{N}} \Phi_{D,N} \tag{48}
\]

To conclude the analysis of the asymptotic behavior of \( C_N(\theta) \) we need to evaluate \( D_N(\theta) \). The i.i.d. assumption and result (42) imply the first two derivations in (49). The final equality in (49) follows from the definition of \( \mu_D^2(\theta) \).

\[
D_N(\theta) = \frac{(N-1)(N-2)}{3N^{2}h^{2d}} E[H_2(\theta)] \to \lim E[(E[v(\theta)|Z = z)^2] f_Z(z) = \frac{\mu_D^2(\theta)}{3} \tag{49}
\]

Finally, recall that \( C_N(\theta) = C_N^I(\theta) + C_N^{II}(\theta) + D_N(\theta) \), and hence by (48) we conclude that if \( \theta \notin \Theta_0 \), then:

\[
C_N(\theta) = Q_N(\theta) - \hat{B}_N(\theta) = \frac{\sigma_D(\theta)}{\sqrt{N}} \Phi_{D,N} + D_N(\theta) \tag{50}
\]

which finishes the proof of the theorem for the case \( \theta \notin \Theta_0 \).

**Lemma 1.** Let \( \sigma^2(z_n, \theta) = E[v_n^2|\theta](Z = z_n) \) and \( W_{ni} = f K (\frac{z_n - z}{h}) K (\frac{z_i - z}{h}) f_Z(z) dz \). If \( \theta \in \Theta_0 \) and Assumptions 1-5 hold, \( V_{N1} = \sum_{i=2}^{N} \sigma^2(z_i, \theta) \sum_{n=1}^{i-1} W_{ni} v_n^2(\theta) \) and \( \text{Var} V_{N2} = 2 \sum_{i=2}^{N} \sigma^2(z_i, \theta) \sum_{n=1}^{i-1} \sum_{k < n} W_{ni} W_{ki} v_n(\theta) v_k(\theta) \), then:

\[
N^{-2}h^{-3d} V_{N1} \overset{p}{\to} \sigma_C^2(\theta)/4 \quad N^{-2}h^{-3d} V_{N2} \overset{p}{\to} 0
\]
Proof of Lemma 1.1: This proof is a simple adaptation from Hall (1984) and has been included in this paper for the sake of completeness. First I show that \( N^{-2} h^{-3d} V_{N1} \overset{p}{\rightarrow} \sigma^2(\theta)/4 \). Let \( Z_N \) be the \( \sigma \)-field generated by \( \{ Z_1, \ldots, Z_n \} \). Use \( E[v_i^2(\theta)|Z_N] = \sigma^2(z_i, \theta) \) by the i.i.d. assumption and let \( \mu_4(z_i) = E[(v_i^2(\theta) - \sigma^2(z_i, \theta))^2|Z = z_i] \) to obtain the first two equalities in (51). The third equality in (51) follows from \( \sigma^2(z_i, \theta) \) being bounded.

\[
E \left[ (V_{N1} - E[V_{N1}|Z_N])^2 | Z_N \right] = \sum_{j=1}^{N-1} E \left[ (v_j^2(\theta) - \sigma^2(z_j, \theta))^2 | Z_N \right] \left( \sum_{i=j+1}^{N} \sigma^2(z_i, \theta) W_{ij}^2 \right)^2
\]

\[
= \sum_{j=1}^{N-1} \mu_4(z_j) \left( \sum_{i=j+1}^{N} \sigma^2(z_i, \theta) W_{ij}^2 \right)^2 \lesssim \sum_{j=1}^{N-1} \mu_4(z_j) \left( \sum_{i=j+1}^{N} W_{ij}^2 \right)^2
\]

Since \( K(u) \) is bounded, it follows that \( W_{ij} \lesssim \int K \left( \frac{z_j - z_i}{h} \right) f_Z(z) dz \). Use \( f_Z(z) \) being bounded and let \( u = (z - z_i)/h \) to obtain \( \int K \left( \frac{z_j - z_i}{h} \right) f_Z(z) dz = h^d \int K(u)f_Z(z_i + hu)du = O(h^d) \). We combine this result with (51) to show:

\[
E \left[ (V_{N1} - E[V_{N1}|Z_N])^2 | Z_N \right] \lesssim Nh^{3d} \sum_{j=1}^{N-1} \mu_4(z_j) \sum_{i=j+1}^{N} W_{ij}
\]

I now examine the expectation of the right hand side of (52). Using the i.i.d. assumption and doing the change of variables \( u = (z - z_j)/h \) and \( v = (z_i - z_j)/h \) we derive the first two equalities in (52). The last equality follows from \( f_Z(z) \) being bounded and \( E[\mu_4(z)] < \infty \).

\[
E \left[ \sum_{j=1}^{N-1} \mu_4(z_j) \sum_{i=j+1}^{N} W_{ij} \right] \leq N^2 \int \mu_4(z_j) K \left( \frac{z_j - z_i}{h} \right) K \left( \frac{z_i - z_j}{h} \right) f_Z(z)f_Z(z_i)f_Z(z_j)dzdz_j
\]

\[
= N^2 h^{2d} \int \mu_4(z_j)K(u)(u - v)f_Z(z_j + hu)f_Z(z_j + hv)dzdu = O(N^2 h^{2d})
\]

For positive random variables \( R_n \), \( E[R_n] = O(\delta_n) \) implies that \( Z_n = O_p(\delta_n) \). Therefore, combining (52) and (53) we conclude that \( E \left[ (V_{N1} - E[V_{N1}|Z_N])^2 | Z_N \right] = O_p(N^3 h^{5d}) \). Using Markov’s inequality we thus derive:

\[
V_{N1} = E[V_{N1}|Z_N] + O_p(N^2 h^{2d})
\]

To conclude the analysis of \( V_{N1} \) we need to examine \( E[V_{N1}|Z_N] \), starting by looking at \( E[V_{N1}] \). The law of iterated expectations, the definition of \( W_{ij} \) and the change of variables \( u = (z - z_j)/h \) and \( v = (z_i - z_j)/h \) imply the first two equalities in (55). Since \( \sigma^2(z, \theta) \) and \( f_Z(z) \) are both bounded and continuous, the dominated convergence theorem implies the final result.

\[
\frac{1}{N^2 h^{2d}} E[V_{N1}] = \frac{N(N - 1)}{2N^2 h^{4d}} E[\sigma^2(z_i, \theta)\sigma^2(z_j, \theta) W_{ij}^2]
\]

\[
= \frac{N - 1}{2N} \int \sigma^2(z_j, \theta)\sigma^2(z_i + hv, \theta) \left[ \int K(u)f_Z(z_j + hu)du \right]^2 f_Z(z_j + hv)dzdu \rightarrow \sigma^2(\theta)/4
\]

Now write \( E[V_{N1}|Z_N] - E[V_{N1}] \) as a U-Statistic \( U_N \), where \( U_N = \sum_{n=1}^{N} \sum_{i<n} \sigma^2(z_i, \theta)\sigma^2(z_n, \theta) W_{ni}^2 - E[V_{N1}] \). Its U-Statistic projection, \( S_N \) is \( S_N = (N - 1) \sum_{n=1}^{N} E[\sigma^2(z_i, \theta)\sigma^2(z_n, \theta) W_{ni}^2 | Z_n] - E[V_{N1}] \), which gives the first equality in (56). The second inequality follows by evaluating \( E[S_N^2] \) and noting that \( E[(S_N - U_N)^2] \) is of order \( N^2 \). For the third inequality, use Jensen’s inequality for conditional expectations and \( \sigma^2(z, \theta) \) being bounded.

\[
E[(E[V_{N1}|Z_N] - E[V_{N1}])^2] = E[S_N^2] + E[(U_N - S_N)^2] \lesssim N^3 E \left[ (E[\sigma^2(z_i, \theta)\sigma^2(z_n, \theta) W_{ni}^2 | Z_n]) \right] \lesssim N^3 E[W_{ni}^2]
\]

We now analyze \( E[W_{ni}^2] \). Use the definition of \( W_{ni} \) and do the change of variables \( u = (z - z_n)/h \) and \( v = (z_i - z_n)/h \) to derive the first two equalities in (57). Finally, for the third equality, use \( f_Z(z) \) and \( K(u) \) being bounded, Jensen’s
inequality, and \( K(u)K(u-v) \) being integrable.

\[
E[W_{n1}^4] = \int \left( \int K \left( \frac{z_i - z}{h} \right) K \left( \frac{z_n - z}{h} \right) f_Z(z) dz \right)^4 f_Z(z_i)f_Z(z_n)dz_idz_n = h^{4d} \int \left( \int K(u)K(u-v)f_Z(z_j+hu)du \right)^4 f_Z(z_n+hu)f_Z(z_n)dvdz_n = O(h^{5d})
\]  

Combining (56) and (57) we get that \( E[(E[V_{N1}|Z_N] - E[V_{N1})^2] = O(N^3h^{5d}) \). Therefore, Markov’s inequality implies \( E[V_{N1}|Z_N] = E[V_{N1}] + O(P(N^2h^{5d})) \). From (54), it follows that \( V_{N1} = E[V_{N1}] + O(P(N^2h^{5d})) \). Finally, using \( Nh^d \to \infty \) and (55), it follows that \( N^{-2}h^{-3d}V_{N1} \xrightarrow{p} \sigma^2_z(\theta)/4 \), which establishes the first claim of the Lemma.

I now show that \( N^{-2}h^{-3d}V_{N2} \xrightarrow{p} 0 \). In (58) expand the square and use that since \( \theta \in \Theta_0, E[v_{n}(\theta)|Z_N] = 0 \) to obtain the first equality. The second inequality follows from \( \sigma^2_z(\theta) \) being bounded.

\[
E[V_{N2}^2|Z_N] = 4 \sum_{n=1}^{N-1} \sum_{j<n} \sigma^2(z_n, \theta)\sigma^2(z_j, \theta) \left[ \sum_{i=j+1}^{N} \sigma^2(z_i, \theta)W_{ni}W_{ij} \right]^2 \leq \sum_{n=1}^{N-1} \sum_{j<n} \left[ \sum_{i=j+1}^{N} W_{ni}W_{ij} \right]^2 \]

Now we evaluate the expectation of (58). Expanding the square, using the i.i.d. assumption and the fact that \( 1 \leq n < j < N - 1 \), we get the equality in (59).

\[
E \left[ \left( \sum_{i=j+1}^{N} W_{in}W_{ij} \right)^2 \right] = (N-J)E[W_{12}^2W_{13}^2] + (N-J)(N-J-1)E[W_{13}W_{14}W_{23}W_{24}]
\]

We proceed to derive a bound for \( E[W_{13}W_{14}W_{23}W_{24}] \). A number of steps need to be followed to derive the expression in (60). First use the definition of \( W_{ij} \) and do a change of variables of the form \( u = (z_i - z)/h \) in all the inside integrals, this yields a term of order \( h^{4d} \). Then, do the change of variables \( u_2 = (z_2 - z_1)/h \), \( u_3 = (z_3 - z_1)/h \) and \( u_4 = (z_4 - z_1)/h \), which yields an additional term \( h^{3d} \). Finally, we drop some of the \( f_Z(z) \) terms, since they are bounded and \( K(u)K(u+v) \) is integrable.

\[
E[W_{13}W_{14}W_{23}W_{24}] = h^{7d} \int \left[ \int K(u)K(u+u_3)du \right] \left[ \int K(u)K(u+u_4)du \right] \left[ \int K(u)K(u+u_2)du \right] \left[ \int K(u)K(u+u_4-u_2)du \right] f_Z(z_1)dz_1du_2du_3du_4
\]

Result (60) establishes that \( E[W_{13}W_{14}W_{23}W_{24}] = O(h^{7d}) \). In addition, since we have shown \( W_{ij} = O(h^{d}) \) and from (55) \( E[W_{ij}^2] = O(h^{3d}) \), it follows that \( E[W_{12}^2W_{13}^2] = O(h^{5d}) \). Combining these results with (59) we conclude that

\[
E \left[ \left( \sum_{i=j+1}^{N} W_{in}W_{ij} \right)^2 \right] = (Nh^{5d} + N^2h^{7d})
\]

Hence, the law of iterated expectation and (58) imply (61).

\[
E \left[ (N^{-2}h^{-3d}V_{N2})^2 \right] = N^{-4}h^{-6d} \sum_{n=1}^{N} \sum_{j<n} O(Nh^{5d} + N^2h^{7d}) = O(N^{-1}h^{-d} + h^d)
\]

Since \( Nh^d \to \infty \), result (61) establishes that \( E \left[ (N^{-2}h^{-3d}V_{N2})^2 \right] = o(1) \). Markov’s inequality then implies \( N^{-2}h^{-3d}V_{N2} = o_p(1) \), thus concluding the proof of the Lemma. ■

**Lemma 2.** Let \( W_{ni} = \int K \left( \frac{z_i - z}{h} \right) f_Z(z)dz, Y_i = v_i(\theta) \sum_{n=1}^{i-1} W_{ni}v_n(\theta) \) and \( \mathcal{F}_i \) be the sigma-field generated by \( \{Z_1, \ldots, Z_i, Z_{i+1} \} \) and \( \{V_1(\theta), \ldots, V_i(\theta) \} \). If \( \theta \in \Theta_0 \) and Assumptions 1-5 hold, then for each \( \epsilon > 0 \):

\[
N^{-2}h^{-3d} \sum_{i=2}^{N} E \left[ Y_i^2 I[|Y_i| > \epsilon Nh^{3/2}] |\mathcal{F}_{i-1} \right] \xrightarrow{p} 0
\]
Proof of Lemma 2: This proof is a simple adaptation from Hall (1984) and has been included in this paper for the sake of completeness. Let $S_i = \sum_{n=1}^{i-1} W_{ni} \nu_n(\theta)$, and note that $Y_i = v_i(\theta)S_i$. Also note that for any $k > 0$, the event $|Y_i| > \epsilon N h^{d/2}$ entails either $|v_i(\theta)| > k$ or $kS_i > \epsilon N h^{d/2}$. Combine this fact with $E[S_i|\mathcal{F}_{i-1}] = S_i$ to deduce that:

$$E \left[ Y_i^2 \mathbb{1}\{|Y_i| > \epsilon N h^{d/2}\}|\mathcal{F}_{i-1}\right] \leq \mathbb{S}_i^2 E[v_i^2(\theta) \mathbb{1}\{|v_i(\theta)| > k\}|\mathcal{F}_{i-1}] + \sigma^2(z_i, \theta)\mathbb{S}_i^2 \mathbb{1}\{|S_i| > \epsilon N h^{d/2}\}$$

(62)

where $\sigma^2(z_i, \theta) = E[v_i^2(\theta)|Z = z_i]$. Let $Z_i$ be the $\sigma$-field generated by $\{Z_1, \ldots, Z_{i+1}\}$. By the i.i.d. assumption and $\theta \in \Theta_0$, it follows that $E[S_i^2|\mathcal{Z}_{i-1}] = \sum_{n=1}^{i-1} \sigma^2(z_i, \theta)W_{ni}^2$ and that $E[v_i^2(\theta) \mathbb{1}\{|v_i(\theta)| > k\}|\mathcal{F}_{i-1}] = E[v_i^2(\theta) \mathbb{1}\{|v_i(\theta) > k\}|\mathcal{Z}_{i-1}]$. Combine (62) together with these results, $Z_i \subset \mathcal{F}_i$, the law of iterated expectations and $\sigma^2(z_i, \theta)$ being bounded to conclude that:

$$\frac{1}{N^2 h^{3d}} \sum_{i=2}^{N} E \left[ Y_i^2 \mathbb{1}\{|Y_i| > \epsilon N h^{d/2}\}|\mathcal{Z}_{i-1}\right] \leq \frac{1}{N^2 h^{3d}} \sum_{i=2}^{N} \left( \mathbb{S}_i^2 E[v_i^2(\theta) \mathbb{1}\{|v_i(\theta)| > k\}|\mathcal{Z}_{i-1}] + E \left[ \sigma^2(z_i, \theta)\mathbb{S}_i^2 \mathbb{1}\{|S_i| > \epsilon N h^{d/2}\}|\mathcal{Z}_{i-1}\right]\right)$$

(63)

Since $Z_i \subset \mathcal{F}_i$, and $Y_i^2 \mathbb{1}\{|Y_i| > \epsilon N h^{d/2}\} \geq 0$, showing $N^{-2} h^{-3d} \sum_{i=2}^{N} E \left[ Y_i^2 \mathbb{1}\{|Y_i| > \epsilon N h^{d/2}\}|\mathcal{Z}_{i-1}\right] \xrightarrow{p} 0$ will imply that $N^{-2} h^{-3d} \sum_{i=2}^{N} E \left[ Y_i^2 \mathbb{1}\{|Y_i| > \epsilon N h^{d/2}\}|\mathcal{F}_{i-1}\right] \xrightarrow{p} 0$ and hence establish the Lemma. With this purpose we examine the two terms in the right hand side of (63). For the first term, I will show that its expectation converges to zero and hence, since it is a positive random variable, that it converges to zero in probability. Apply the Cauchy-Schwarz inequality for conditional expectations to the term $E[v_i^2(\theta) \mathbb{1}\{|v_i(\theta)| > k\}|\mathcal{Z}_{i-1}]$, which implies the second inequality in (64). The same calculations as in (53) show that $\sum_{i=1}^{N} E \left[ \mu_4(z_i) \right] \sum_{n=1}^{i-1} W_{ni}^4 = O(N^2 h^{2d})$, which in turn gives the third inequality in (64).

$$\frac{1}{N^2 h^{3d}} \sum_{i=2}^{N} \left( \mathbb{S}_i^2 E[v_i^2(\theta) \mathbb{1}\{|v_i(\theta)| > k\}|\mathcal{Z}_{i-1}] + E \left[ \sigma^2(z_i, \theta)\mathbb{S}_i^2 \mathbb{1}\{|S_i| > \epsilon N h^{d/2}\}|\mathcal{Z}_{i-1}\right]\right)$$

Thus, as $k \to \infty$, $N^{-2} h^{-3d} \sum_{i=2}^{N} \left( \mathbb{S}_i^2 E[v_i^2(\theta) \mathbb{1}\{|v_i(\theta)| > k\}|\mathcal{Z}_{i-1}] + E \left[ \sigma^2(z_i, \theta)\mathbb{S}_i^2 \mathbb{1}\{|S_i| > \epsilon N h^{d/2}\}|\mathcal{Z}_{i-1}\right]\right) \xrightarrow{p} 0$. I now proceed to examine the second term in the right hand side of (63). For this purpose I begin by examining $E[S_i^4|\mathcal{Z}_{i-1}]$. In (65) expand the fourth power and use that $\theta \in \Theta_0$ to obtain the first equality. The second inequality follows by letting $\mu_4(z_i) = E[v_i^4(\theta)Z = z_i]$, $\sigma^2(z_i, \theta)$ being bounded using $W_{ni} \leq Ch^d$.

$$E[S_i^4|\mathcal{Z}_{i-1}] = \sum_{n=1}^{i-1} E[v_i^4(\theta)|\mathcal{Z}_{i-1}] W_{ni}^4 + \sum_{n=1}^{i-1} \sum_{j<n} \sigma^2(z_n, \theta)\sigma^2(z_j, \theta) W_{ni}^2 W_{nj}^2 \leq h^{3d} \sum_{n=1}^{i-1} \mu_4(z_n) W_{ni} + N h^{3d} \sum_{n=1}^{i-1} W_{ni}$$

(65)

The same calculations as in (53) show that $E\left[ \sum_{n=1}^{i-1} \mu_4(z_n) W_{ni} + \sum_{n=1}^{i-1} W_{ni}\right] = O(N h^{3d})$. Therefore, (65) and the law of iterated expectations imply that $E[S_i^4] = O(N^2 h^{2d})$. The first inequality in (66) follows from this result, the Cauchy-Schwarz inequality and Chebychev’s inequality. In addition, since $E[S_i^2|\mathcal{Z}_{i-1}] = \sum_{n=1}^{i-1} \sigma^2(z_n, \theta) W_{nj}^2$, the law of iterated expectations and the same calculations as in (55) show $E[S_i^2] \lesssim N h^{3d}$, which in turn implies the second
inequality in (66).
\[
\frac{1}{N^2 h^{3d}} \sum_{i=1}^{N} E \left[ S_{\epsilon}^2 \mathbf{1}\{k|S|_i > \epsilon Nh^{\frac{d}{3}}\} \right] \lesssim \frac{Nh^{\frac{d}{3}}}{N^2 h^{3d}} \sum_{i=1}^{N} \left[ \frac{1}{N^2 h^{3d}} E[S_{\epsilon}^2] \right]^{\frac{1}{2}} \lesssim \frac{Nh^{\frac{d}{3}}}{N^2 h^{3d}} \sum_{i=1}^{N} \frac{1}{N^2} = \frac{1}{N^2 h^{\frac{d}{3}}} \tag{66}
\]
Markov’s inequality, \(Nh^d \to \infty\) and (66) imply \(N^{-2} h^{-3d} \sum_{i=1}^{N} \frac{1}{N^2} \lesssim N h^{\frac{d}{3}}\). Together with (63) and (64) this completes the proof of the Lemma. ■

Proof of Theorem 3.1: Throughout this proof I will use the same notation as in Theorem (.1). The proof follows directly from the representation derived in Theorem (.1) and the consistency of \(\hat{\sigma}_C(\theta)\) established in Lemma (.3). If \(\theta \in \Theta_0\), then Theorem (.1) implies:
\[
\frac{Nh^{\frac{d}{3}}}{\sigma_C(\theta)} (Q_N(\theta) - \hat{B}_N(\theta)) = \frac{\sigma_C(\theta)}{\hat{\sigma}_C(\theta)} \Phi_{C_N}
\tag{67}
\]
Lemma (.3) establishes that \(\sigma_C(\theta) \xrightarrow{p} \sigma_C(\theta)\). Therefore, (67) and the continuous mapping theorem imply that
\[
\frac{Nh^{\frac{d}{3}}}{\hat{\sigma}_C(\theta)} (Q_N(\theta) - \hat{B}_N(\theta)) \xrightarrow{L} N(0,1).
\]
Now suppose that \(\theta \notin \Theta_0\), then Theorem (.1) implies:
\[
\frac{Nh^{\frac{d}{3}}}{\sigma_C(\theta)} (Q_N(\theta) - \hat{B}_N(\theta)) = (Nh^d) \frac{\sigma_C(\theta)}{\hat{\sigma}_C(\theta)} \left( \Phi_{D_N} + \sqrt{N} \Phi_{D_N} \right)
\tag{68}
\]
Since \(D_N(\theta) \longrightarrow \mu_d(\theta) > 0\), it follows that \(\sqrt{N} \Phi_{D_N} \longrightarrow +\infty\), and since \(\Phi_{D_N} \xrightarrow{L} N(0,1)\), a tight random variable, (68) implies that if \(\theta \notin \Theta_0\), then
\[
\frac{Nh^{\frac{d}{3}}}{\hat{\sigma}_C(\theta)} (Q_N(\theta) - \hat{B}_N(\theta)) \xrightarrow{p} +\infty,
\]
which conclude the proof of the Theorem. ■

APPENDIX C - Proof of Theorem 3.2, Corollary 3.1 and Auxiliary Theorems .2, .3,.4 and Lemmas .3 and .4

Theorem 2. Let \(\mathcal{F}\) be a set of \(P\) canonical symmetric functions in \(L^2(X^m)\). Define \(I_m^N\) to be the set of distinct \(m\)-tuples from a sample of size \(N\) and define the U-Statistic \(U_n^m(f) = N^{-m} \sum_{I_m^N} f(x_{i_1}, \ldots, x_{i_m})\). Let \(D_N\) be the diameter of \(\mathcal{F}\) under \(\rho_{mN}\). Then there exists a constant \(K\) depending only on \(m\) such that:
\[
E \left[ \sup_{f,g \in \mathcal{F}} N^{-\frac{d}{2}} \left| U_n^m(f) - U_n^m(g) \right| \right] \leq K \int_0^{D_N} \left[ \log \mathcal{N} \left( \mathcal{F}, \rho_{mN}, \epsilon \right) \right]^{\frac{d}{2}} d\epsilon
\]

Proof of Theorem 2: This result is derived as an intermediate step in several proofs in Arcones & Gine (1993) and follows arguments similar to the classical empirical process case. I replicate their arguments and present them as a single theorem for easy reference. First we use a result from de la Pena, Proposition 2.1 in Arcones & Gine (1993), to derive that for \(\mathcal{F}\) a class of \(P\)-canonical, symmetric and square integrable functions:
\[
E \left[ \sup_{f,g \in \mathcal{F}} N^{-\frac{d}{2}} \left| U_n^m(f) - U_n^m(g) \right| \right] = E \left[ \sup_{f,g \in \mathcal{F}} N^{-\frac{d}{2}} \left| \sum_{I_m^N} f(x_{i_1}, \ldots, x_{i_m}) - g(x_{i_1}, \ldots, x_{i_m}) \right| \right]
\]
\[
\lesssim E \left[ \sup_{f,g \in \mathcal{F}} N^{-\frac{d}{2}} \left| \sum_{I_m^N} \epsilon_{i_1} \ldots \epsilon_{i_m} \left( f(x_{i_1}, \ldots, x_{i_m}) - g(x_{i_1}, \ldots, x_{i_m}) \right) \right| \right]
\tag{69}
\]
where \(\{\epsilon_i\}_{i=1}^{N}\) are i.i.d. Rademacher random variables independent of \(\{X_i\}_{i=1}^{N}\). Condition on \(\{X_i\}_{i=1}^{N}\), and let \(E_c[\cdot]\) denote the expectation over \(\{\epsilon_i\}_{i=1}^{N}\). Let \(\hat{U}_n^m(f) = N^{-\frac{d}{2}} \sum_{I_m^N} \epsilon_{i_1} \ldots \epsilon_{i_m} f(x_{i_1}, \ldots, x_{i_m})\). In (70) we aim to find a bound for \(E_c[\exp\{|\hat{U} - g|^{\frac{d}{2}}\}]\). For the first equality in (70), use the series representation of \(e^x\) and the monotone convergence theorem. Note that terms in the sum with \(i \leq m\) are less than \(1/i!\) by Jensen’s inequality. To bound
terms with \( i > m \), we will use Proposition 2.2 in Arcones & Gine (1993), which implies that 
\[
\left( E_x[|\bar{U}_n^m(f-g)|^q] \right)^{1/q} \leq 
\left( \frac{q-1}{q} \right)^{\frac{q}{2}} \left( E_x[|\bar{U}_n^m(f-g)|^2] \right)^{\frac{1}{2}}
\]
for all \( 1 < q < p \leq \infty \) and all \( f, g \in \mathcal{F} \).

\[
E_x \left[ \exp \left\{ \left( \frac{|\bar{U}_n^m(f-g)|}{E_x[|\bar{U}_n^m(f-g)|^2]} \right)^{\frac{1}{2}} \right\} \right] \leq \sum_{i=1}^{m} \frac{1}{i!} \left( E_x[|\bar{U}_n^m(f-g)|^2] \right) \leq \sum_{i=1}^{m} \frac{1}{i!} \left( \frac{2i-m}{m} \right)^{i} \tag{70}
\]

Equation (70) implies that \( E_x \left[ \exp \left\{ \left( \frac{|\bar{U}_n^m(f-g)|}{E_x[|\bar{U}_n^m(f-g)|^2]} \right)^{\frac{1}{2}} \right\} \right] \leq c \), where \( c \) is a constant depending only on \( m \). Standard chaining arguments, such as Proposition 2.6 in Arcones & Gine (1993), imply the first inequality in (71) with \( D \) the diameter of \( \mathcal{F} \) under the semimetric \( E_x[|\bar{U}_n^m(f-g)|^2]^{\frac{1}{2}} \). Furthermore, it follows by the independence of the \( \{\epsilon_i\}_{i=1}^{N} \) that upon expanding the square in \( E_x[|\bar{U}_n^m(f-g)|^2] \) all cross terms disappear. Therefore, direct calculation shows that \( E_x[|\bar{U}_n^m(f-g)|^2]^{\frac{1}{2}} = \rho_{mn}(f,g) \), which gives us the second equality in (71).

\[
E_x \left[ \sup_{f,g \in \mathcal{F}} \left| \bar{U}_n^m(f-g) \right| \right] \leq K \int_{0}^{D} \left[ \log N(\mathcal{F},E_x[|\bar{U}_n^m(f-g)|^2]^{\frac{1}{2}},\epsilon) \right]^{\frac{1}{2}} d\epsilon \leq K \int_{0}^{D} \left[ \log N(\mathcal{F},\rho_{mn}(f,g),\epsilon) \right]^{\frac{1}{2}} d\epsilon \tag{71}
\]

To conclude the proof, take expectation of \( \{X_t\}_{t=1}^{N} \) on (71) and combine the resulting inequality with (69).

**Theorem 3.** Let \( \mathcal{F} = \{f_t: t \in T\} \) be a class of functions such that \( |f_s(x) - f_t(x)| \leq d(s,t)F(x) \) for every \( s \) and \( t \) and some fixed function \( F \). Then for any norm \( \| \cdot \| \),
\[
N(\mathcal{F},\| \cdot \|,2\epsilon\|F\|) \leq N(T,d,\epsilon)
\]

**Proof of Theorem 3:** Refer to Theorem 2.7.11 in van der Vaart & Wellner.

**Theorem 4.** For a constant \( K \) depending only on \( k \) and \( m_0 \), \( \log N(\Theta,\| \cdot \|_\infty,\epsilon) \leq K \left( \frac{1}{\epsilon} \right)^{\frac{m_0+k+1}{m_0}} \)

**Proof of Theorem 4:** Fix \( \epsilon > 0 \). By Lemma A.2 from Gallant and Nychka (1987), there is a constant \( C \) depending only on \( m \) and \( \delta_0 \) such that \( |D^\lambda(1 + z')^{\delta_0}| \leq C(1 + z')^{\delta_0} \) for all \( \lambda \leq m \) and \( z \in \mathbb{R}^d \). Therefore, as Gallant and Nychka (1987) note:

\[
\max_{|\lambda| \leq m} \sup_{z \in \mathbb{R}^d} |D^\lambda \theta(z)|(1 + z')^{\delta_0} \leq \left( \sum_{|\lambda| \leq m + m_0} \int [D^\lambda \theta(z)(1 + z')^{\delta_0}]^2 dz \right)^{\frac{1}{2}} \leq B \tag{72}
\]

The first inequality follows by the Sobolev Imbedding Theorem. Let \( Z_\epsilon^k = \left\{ z \in \mathbb{R}^d : z \leq \left( \frac{B^*}{\epsilon^2} \right)^{\frac{1}{2}} - 1 \right\} \), and note that for any \( \theta_1(z), \theta_2(z) \in \Theta \), equation (72) implies that \( \sup_{z \in Z_\epsilon^k} |\theta_1(z) - \theta_2(z)| \leq \sup_{z \in Z_\epsilon^k} |\theta_1(z) + \sup_{z \in Z_\epsilon^k} |\theta_2(z)| < \epsilon \) for some \( B^* \). This implies that \( \sup_{z \in \mathbb{R}^d} |\theta_1(z) - \theta_2(z)| < \epsilon \) if and only if \( \sup_{z \in Z_\epsilon^k} |\theta_1(z) - \theta_2(z)| < \epsilon \). Therefore, without loss of generality, when calculating \( N(\Theta,\epsilon,\| \cdot \|_\infty) \) we may assume that \( \theta(z) = 0 \) for all \( z \in Z_\epsilon^k \) and \( \theta \in \Theta \). Then, by Theorem 2.7.1 from van der Vaart and Wellner (2000), there exists a constant \( K_0 \) depending only on \( m \) and \( d \) such that:

\[
\log N(\epsilon,\Theta,\| \cdot \|_\infty) \leq K_0 \lambda(Z_\epsilon^1) \left( \frac{1}{\epsilon} \right)^{\frac{1}{2}} \tag{73}
\]
where \( \lambda(\cdot) \) is the Lebesgue measure, and \( Z_{1/2} = \{ z \in \mathbb{R}^d : ||z - Z_{1/2}|| < 1 \} \). If \( z \in Z_{1/2} \) then for some \( \tilde{z} \in Z_{1/2} \), \( ||z|| \leq ||z - \tilde{z}|| + ||\tilde{z}|| < 1 + \left( \frac{B^*}{c^*} \right)^{1/n} \), which implies that \( Z_{1/2} \subseteq \{ z \in \mathbb{R}^d : ||z|| \leq 1 + \left( \frac{B^*}{c^*} \right)^{1/n} \} \). Thus, \( \lambda(Z_{1/2}) \lesssim \left[ 1 + \left( \frac{B^*}{c^*} \right)^{1/n} \right]^d \). Combining this result with (73), it follows that for \( \epsilon \) small enough:

\[
\log N(\epsilon, \Theta, ||\cdot||_\infty) \lesssim \left[ 1 + \left( \frac{B^*}{c^*} \right)^{1/n} \right]^d \left( \frac{1}{\epsilon} \right)^{\frac{d}{n}} \lesssim \left( \frac{1}{\epsilon} \right)^{\frac{m+\delta_0+\mu}{n\alpha_0}} \tag{74}
\]

which establishes the theorem. 

**Lemma 3.** Under Assumptions 1-8, \( \sup_{\theta} |\hat{\sigma}_{CI}^2(\theta) - \sigma_{CI}^2(\theta)| \overset{p}{\to} 0 \).

**Proof of Lemma 3:** The proof will proceed by showing that \( E[\hat{\sigma}_{CI}^2(\theta)] \) converges uniformly in \( \theta \) to \( \sigma_{CI}^2(\theta) \) and by establishing a uniform law of large numbers for \( \hat{\sigma}_{CI}^2(\theta) \). To simplify notation, let:

\[
H_3(w_n, w_i, w_j, w_k) = \int K \left( \frac{z_n - z_i}{h} \right) K \left( \frac{z_i - z_j}{h} \right) K \left( \frac{z_j - z_k}{h} \right) K \left( \frac{z_k - z_j}{h} \right) v_{i}^2(\theta) v_{j}^2(\theta) v_{k}^2(\theta) dz
\]

and note that \( \hat{\sigma}_{CI}^2(\theta) = \frac{24}{N(N-1)^2} \sum_{n=1}^{N} \sum_{i<n}^{N} \sum_{j<i}^{N} H_3(w_n, w_i, w_j, w_k) \). When the context is clear, the dependence on \((w_n, w_i, w_j, w_k)\) will be suppressed. I begin by showing that \( \sup_{\theta} |\hat{\sigma}_{CI}^2(\theta) - E[\hat{\sigma}_{CI}^2(\theta)]| \overset{p}{\to} 0 \). Let \( P_1^N(H_3(\theta)) \) be the \( i \)th term in the Hoeffding decomposition of the U-Statistic \( \hat{\sigma}_{CI}^2(\theta) - E[\hat{\sigma}_{CI}^2(\theta)] \):

\[
P_1^N(H_3(\theta)) = \frac{4}{N^2} \sum_{n=1}^{N} E[H_3(\theta)|w_n] - E[H_3(\theta)]
\]

\[
P_2^N(H_3(\theta)) = \frac{24(N-1)}{N^2} \sum_{n=1}^{N} \sum_{1<i<n} E[H_3(\theta)|w_i, w_n] - E[H_3(\theta)|w_n] - E[H_3(\theta)|w_i] + E[H_3(\theta)]
\]

\[
P_3^N(H_3(\theta)) = \frac{24(N-1)^2}{N^2} \sum_{n=1}^{N} \sum_{1<i<n} \sum_{j<i} E[H_3(\theta)|w_i, w_j, w_n] - E[H_3(\theta)|w_n, w_i, w_j] + E[H_3(\theta)|w_n, w_j, w_i] - E[H_3(\theta)|w_i, w_j, w_n] + E[H_3(\theta)|w_i, w_n, w_j] + E[H_3(\theta)|w_n, w_i, w_j] - E[H_3(\theta)|w_i, w_j, w_n] + E[H_3(\theta)|w_n, w_i, w_j] - E[H_3(\theta)|w_n, w_j, w_i]
\]

\[
P_4^N(H_3(\theta)) = \hat{\sigma}_{CI}^2(\theta) - E[\hat{\sigma}_{CI}^2(\theta)] = P_1^N(H_3(\theta)) - P_2^N(H_3(\theta)) + P_3^N(H_3(\theta)) + P_4^N(H_3(\theta))
\]

The terms \( P_i^N(H_3(\theta)) \) are P-canonical U-Statistics, which will allow us to use the maximal inequality from Theorem (2). By construction, \( \hat{\sigma}_{CI}^2(\theta) - E[\hat{\sigma}_{CI}^2(\theta)] = P_1^N(H_3(\theta)) + P_2^N(H_3(\theta)) + P_3^N(H_3(\theta)) + P_4^N(H_3(\theta)) \), and therefore:

\[
\sup_{\theta} |\hat{\sigma}_{CI}^2(\theta) - E[\hat{\sigma}_{CI}^2(\theta)]| \leq \sum_{i=1}^{4} \sup_{\theta} |P_i^N(H_3(\theta))| \tag{76}
\]

I now proceed to show that \( \sup_{\theta} |P_i^N(H_3(\theta))| \overset{p}{\to} 0 \) for \( 1 \leq i \leq 4 \), beginning with \( \sup_{\theta} |P_1^N(H_3(\theta))| \). Let \( \tilde{\Theta}_{1N} = \{ h^{-d}(E[H_3(\theta)|w_n] - E[H_3(\theta)] : \theta \in \Theta \} \), where we note the dependence on \( h \) and therefore on \( N \). The U-Statistic \( P_1^N(H_3(\theta)) \) is linear in the indexing set \( \tilde{\Theta}_{1N} \), and thus Theorem (2) and \( N(\tilde{\Theta}_{1N}, \rho, \epsilon) \leq N(||(\tilde{\Theta}_{1N}, \rho, 2\epsilon) \) imply the first inequality in (77). From Lemma (4), there is a function \( F_N(w_i, w_j, w_k) \) such that \( H_3(w_n, w_i, w_j, w_k) \lesssim F_N(w_n, w_i, w_j, w_k) \). Therefore, for \( \theta_1, \theta_2 \in \tilde{\Theta}_{1N}, |\theta_1(w_n) - \theta_2(w_n)| \lesssim G_1(w_n)||\theta_1 - \theta_2||_\infty \), where \( G_1(w_n) = h^{-d}(E[F_N|w_n] + E[F_N]) \), which implies that the class \( \tilde{\Theta}_{1N} \) is Lipschitz in \( \Theta \). Thus, apply Theorem (3) and do the change of variables \( u = \epsilon/2|G_1(N)|_{1N} \) to get the second inequality in (77). The inequality \( D_N \leq M||G_1N||_{1N} \) for some constant \( M \) for all \( N \), Jensen’s inequality and Theorem (4) imply the third inequality in (77). The final result follows from \( \frac{(m_0+\delta_0)k}{\alpha n \rho_0} < \frac{1}{4} \) and that, from Lemma (4),
Therefore, the class $\Theta$ is Lipschitz in $\Theta$, and thus, following the same arguments as in (77), we derive the first inequality in (78). The second inequality in (78) follows from Theorem (4), $D_N \leq M ||G_{2N}||_{2N}$ for some $M$ for all $N$ and Jensen’s inequality. From Lemma (4) we know that $E[(E[F_N|w_n])^2] = O(h^{d_2})$, $E[(E[F_N|w_n])^2] = O(h^{d_2})$, and $E[F_N] = O(h^{d_2})$, and hence the Cauchy Schwarz inequality implies that $E[G_{2N}^2]$ is uniformly bounded in $N$. We combine this result with $\frac{(m_0 + b_0)k}{\delta_0 m_0} < \frac{1}{2}$ to obtain the last inequality in (78).

\[ N^{-1}h^{-\frac{3}{2}}E \left[ N h^{-\frac{3}{2}} \sup_{\Theta} |P_N^3(\Theta)| \right] \leq N^{-1}h^{-\frac{3}{2}}E \left[ \int_0^{D_N} \int_0^{m+2\theta_0} \|G_{2N}||_{2N} \log N(\Theta, \| \cdot \|_\infty, u) du \right] \\
\leq N^{-1}h^{-\frac{3}{2}} \int_0^{D_N} \left( \frac{1}{u} \right) \left( \frac{(m_0 + b_0)k}{\delta_0 m_0} \right) du (E[G_{2N}^2])^{\frac{1}{2}} = O(N^{-1}h^{-\frac{3}{2}}) \quad (78) \]

Since $N^d \rightarrow \infty$, Markov’s inequality and (78) imply that $\sup_{\Theta} |P_N^3(\Theta)| \rightarrow 0$. I now examine $\sup_{\Theta} \left| P_N^3(\Theta) \right|$. Let $\Theta_{3N} = \{ h^{-3d}(E[H_3(\Theta)|w_n, w_j]) - E[H_3(\Theta)|w_n, w_j] - E[H_3(\Theta)|w_n, w_j] - E[H_3(\Theta)|w_n, w_j] + E[H_3(\Theta)|w_n, w_j] + E[H_3(\Theta)|w_n, w_j] - E[H_3(\Theta)] \}$. Note that for $\hat{\Theta}_{31}, \hat{\Theta}_{32} \in \Theta_{3N}$, we have $|\hat{\Theta}_{31}(w_n, w_j) - \hat{\Theta}_{32}(w_n, w_j)| \leq G_{3N}(w_n, w_j) ||\theta_1 - \theta_2||_\infty$, where $G_{3N}(w_n, w_j) = h^{-3d}(E[F_N|w_n, w_j] + E[F_N|w_n, w_j] + E[F_N|w_n, w_j] + E[F_N|w_n, w_j] + E[F_N|w_n, w_j] + E[F_N|w_n, w_j] + E[F_N|w_n, w_j] + E[F_N|w_n, w_j])$. Therefore, the class $\Theta_{3N}$ is Lipschitz in $\Theta$. The same arguments as in (77) give us the first and second inequalities in (79). Since by Lemma (4), $E[(E[F_N|w_n, w_j])^2] = O(h^{d_2})$, $E[(E[F_N|w_n])^2] = O(h^{d_2})$, and $E[F_N] = O(h^{d_2})$, the Cauchy Schwarz inequality implies $E[G_{3N}^2]$ is uniformly bounded in $N$, which together with $\frac{(m_0 + b_0)k}{\delta_0 m_0} < \frac{1}{2}$ imply the final result in (79):

\[ N^{-\frac{3}{2}}h^{-d}E \left[ N^{\frac{3}{2}}h^{-d} \sup_{\Theta} |P_N^3(\Theta)| \right] \leq N^{-\frac{3}{2}}h^{-d}E \left[ \int_0^{D_N} \int_0^{m+2\theta_0} \|G_{3N}||_{3N} \log N(\Theta, \| \cdot \|_\infty, u) du \right] \\
\leq N^{-\frac{3}{2}}h^{-d} \int_0^{D_N} \left( \frac{1}{u} \right) \left( \frac{3(m_0 + b_0)k}{\delta_0 m_0} \right) du (E[G_{3N}^2])^{\frac{1}{2}} = O(N^{-\frac{3}{2}}h^{-d}) \quad (79) \]

Since $N^d \rightarrow \infty$, (79) and Markov’s inequality imply that $\sup_{\Theta} |P_N^3(\Theta)| \rightarrow 0$. I now proceed to examine $\sup_{\Theta} |P_N^3(\Theta)|$. Define $\Theta_{4N}$ to be the kernel of $P_N^3(\Theta)$ scaled by a factor of $h^{-2d}$. For $\hat{\Theta}_{41}, \hat{\Theta}_{42} \in \Theta_{4N}$, $|\hat{\Theta}_{41}(w_n, w_j, w_k) - \hat{\Theta}_{42}(w_n, w_j, w_k)| \leq G_{4N}(w_n, w_j, w_k) ||\theta_1 - \theta_2||_\infty$, where we define $G_{4N}(w_n, w_j, w_k) = h^{-2d}(F_N(w_n, w_j, w_k) + h^2 G_{4N}(w_n, w_j, w_k) + h^2 G_{4N}(w_n, w_j, w_k) + h^2 G_{4N}(w_n, w_j, w_k))$. This result and the same arguments as in (77) give us the first and second inequalities in (80). By Lemma (4), $E[F_N^2] = O(h^{d_2})$ and as already shown $E[G_{2N}^2]$ is uniformly bounded in $N$. Therefore, the Cauchy Schwarz inequality implies that $E[G_{2N}^2]$ is also
uniformly bounded in $N$. We combine this result with \( \frac{m_u + \delta_k}{m_u m_0} < \frac{1}{2} \) to obtain the last inequality in (80).

\[
N^{-2}h^{-\frac{2d}{2}} E \left[ N^2 h^{-\frac{2d}{2}} \sup_{\Theta} |P_N^d(H_4(\theta))| \right] \lesssim N^{-2}h^{-\frac{2d}{2}} E \left[ \left| G_{4N} \right| \int_0^N \log N(\Theta, \| \cdot \|, u)^2 du \right]
\]

\[
\lesssim N^{-2}h^{-\frac{2d}{2}} \int_0^M \left( \frac{1}{u} \right)^{2n+i+j+k} \log u \left( \frac{h}{2n+i+j+k} \right) du \] (80)

Since $Nh^d \to \infty$, we can use (80) and Markov's inequality to establish that $\sup_{\Theta} |P_N^d(H_3(\theta))| \to 0$. Therefore, combining (81) and (82) we derive:

\[
\sup_{\Theta} |\hat{\sigma}_{CI}^2(\theta) - E[\hat{\sigma}_{CI}^2(\theta)]| \leq \sup_{\Theta} \sup_{i=1}^4 |P_N^d(H_3(\theta))| \to 0 \quad (81)
\]

To complete the proof it is necessary to show $\sup_{\Theta} |E[\hat{\sigma}_{CI}^2(\theta)] - \sigma_{CI}^2(\theta)| \to 0$. The definition of $\hat{\sigma}_{CI}^2(\theta)$ and the change of variables $u_l = (z_l - z)/h$ for $l = n, i, j, k$ imply the first and second equalities in (82). For the third inequality, do a Taylor expansion and notice that Assumption 7 allows for differentiation under the integral sign. Assumption 7 also implies the resulting integral is uniformly bounded in $\theta$, which gives us the final equality in (82):

\[
|E[\hat{\sigma}_{CI}^2(\theta)] - \sigma_{CI}^2(\theta)| = |h^{-d}E[H_3(\theta)] - \sigma_{CI}^2(\theta)| = \left| \int \prod_{l=n+1}^\theta K(u_l) v_l^4(\theta) f_Z(hu_l + z, v_l) du_l dv_l dz - \sigma_{CI}^2(\theta) \right|
\]

\[
= \tilde{h} \int \frac{\partial}{\partial h} \left( \prod_{l=n+1}^\theta K(u_l) v_l^4(\theta) f_Z(hu_l + z, v_l) dv_l \right) dz = O(h) \quad (81)
\]

Since $h \to 0$, (82) establishes that $\sup_{\Theta} |E[\hat{\sigma}_{CI}^2(\theta)] - \sigma_{CI}^2(\theta)| \to 0$. Therefore, combining (81) and (82) we derive:

\[
\sup_{\Theta} |\hat{\sigma}_{CI}^2(\theta) - \sigma^2(\theta)| \leq \sup_{\Theta} \sup_{i=1}^4 |P_N^d(H_3(\theta))| + \sup_{\Theta} |E[\hat{\sigma}_{CI}^2(\theta)] - \sigma_{CI}^2(\theta)| \to 0 \quad (82)
\]

which concludes the proof of the Lemma.

**Corollary 1.** Under Assumptions 1-8, $\sup_{\Theta} \sigma_{CI}^2(\theta) < \infty$ and $\inf_{\Theta} \sigma_{CI}^2(\theta) > 0$. In addition, $\sup_{\Theta} \sigma_{CI}^2(\theta) = O_p(1)$ and $\sup_{\Theta} \sigma_{CI}^2(\theta) = O_p(1)$.

**Proof of Corollary 1:** Since $\sigma_{CI}^2$ does not depend on $\theta$ it is sufficient to examine $\sigma_{CI}^2(\theta)$ and $\hat{\sigma}_{CI}^2(\theta)$. The assumptions that $E[v^2(\theta)|Z]$ and $f_Z(z)$ are uniformly bounded imply $\sup_{\Theta} \sigma_{CI}^2(\theta) = \sup_{\Theta} \sigma_{CI}^2(\theta) = \sup_{\Theta} E \left[ E[v^2(\theta)|Z = z_n] \right] f_Z^2(z_n) < \infty$. Now suppose there exists a sequence $\theta_n$ such that $\sigma_{CI}^2(\theta_n) \to 0$. From Gallant and Nychka (1987) the set $\Theta$ is compact under $\| \cdot \|_{CI}$, and therefore there exists a converging subsequence $\theta_{n_k}$. In (84) we use Fatou’s Lemma to obtain the first equality, and Fatou’s Lemma for conditional expectations to obtain the second inequality.

\[
0 = \lim_{n_k \to \infty} \inf \int (E[v^2(\theta_{n_k})|Z])^4 f_Z^2(z) dz \geq \lim_{n_k \to \infty} \inf \int (E[v^2(\theta_{n_k})|Z])^4 f_Z^2(z) dz 
\]

\[
\geq \int (E[\lim \inf_{n_k \to \infty} v^2(\theta_{n_k})|Z])^4 f_Z^2(z) dz \geq 0 \quad (84)
\]

Therefore, (84) implies that $v^2(\theta_{n_k})$ converges to zero, which is not possible because $\epsilon$ is not a deterministic function of $X$. This contradiction implies $\inf_{\Theta} \sigma_{CI}^2(\theta) > 0$. In combination with Theorem (3), $\sup_{\Theta} \sigma_{CI}^2 < \infty$ implies that $\sup_{\Theta} \sigma_{CI}^2 = O_p(1)$, and $\inf_{\Theta} \sigma_{CI}^2 > 0$ implies that $\sup_{\Theta} \sigma_{CI}^2 = O_p(1)$, which concludes the proof. ■

**Proof of Theorem 3.2:** The strategy of the proof will be to show that $Nh^{\frac{d}{2}} \left( Q_N(\theta) - B_N(\theta) \right)$ is well behaved in a shrinking neighborhood of $\Theta_0$ and diverges to infinity uniformly outside this neighborhood. This result will imply
that \( \inf_{\theta \in R} N h^{\frac{2}{\text{dim}}} \left( Q_N(\theta) - \bar{B}_N(\theta) \right) = \inf_{\theta \in R} N h^{\frac{2}{\text{dim}}} \left( Q_N(\theta) - \bar{B}_N(\theta) \right) + o_p(1) \). I will then complete the proof by showing that \( N h^{\frac{2}{\text{dim}}} \left( Q_N(\theta) - \bar{B}_N(\theta) \right) \overset{p}{\to} G(\theta) \) on \( L^\infty(\Theta_0) \). With this purpose we thus define:

\[
\Theta_0^N = \left\{ \theta \in \Theta : E \left[ \left( f | Y - \theta(X) | Z \right)^2 \right] \right\}
\]

I will set \( \epsilon_N \to 0 \), with \( \epsilon_N \) proportional to \( h' \) for \( l \leq 1, \sqrt{N}h' \to +\infty \) and \( N h^{\frac{d+1}{\text{dim}}} \left( \frac{1}{(\ln N)^{\frac{1}{d}}} \right) \to 0 \), which exists by Assumption 7. Note that for the kernel \( H_2(w_n, w_i, w_j) \) defined in (34), and the U-Statistic \( C_N(\theta) = \sum_{i=1}^{N} \sum_{j<n} H_2(w_n, w_i, w_j) \):

\[
\sup_{\theta \in \Theta} \left( Q_N(\theta) - \bar{B}_N(\theta) \right) = N h^{\frac{2}{\text{dim}}} C_N(\theta)
\]

In addition, let \( P_i^{\text{CN}}(C_N(\theta)) \) be the \( i \)-th term in the Hoeffding decomposition of \( C_N(\theta) \):

\[
P_i^{\text{CN}}(C_N(\theta)) = \frac{(N-1)(N-2)}{N^{2\text{dim}}} \sum_{n=1}^{N} E[H_2(\theta)|w_n] - E[H_2(\theta)]
\]

I will examine the behavior of \( N h^{\frac{2}{\text{dim}}} \left( Q_N(\theta) - \bar{B}_N(\theta) \right) \), or equivalently \( N h^{\frac{2}{\text{dim}}} C_N(\theta) \), on \( (\Theta_0^N)^c \cap R \) and on \( \Theta_0^N \cap R \).

Part (i): Show \( \inf_{(\Theta_0^N)^c \cap R} \left( \Theta_0^N \cap R \right) \to +\infty \).

I will show that \( \inf_{(\Theta_0^N)^c \cap R} \left( \Theta_0^N \cap R \right) \to +\infty \) by analyzing the terms in the Hoeffding decomposition of \( C_N(\theta) \) and exploiting the equality \( C_N(\theta) = E[C_N(\theta)] + \sum_{i=1}^{3} P_i^{\text{CN}}(C_N(\theta)) \). Hence, observe that:

\[
\inf_{(\Theta_0^N)^c \cap R} N h^{\frac{2}{\text{dim}}} C_N(\theta) \geq \inf_{(\Theta_0^N)^c \cap R} N h^{\frac{2}{\text{dim}}} E[C_N(\theta)] + \inf_{(\Theta_0^N)^c \cap R} \sum_{i=1}^{3} P_i^{\text{CN}}(C_N(\theta))
\]

First I show that \( \inf_{(\Theta_0^N)^c \cap R} N h^{\frac{2}{\text{dim}}} E[C_N(\theta)] + N h^{\frac{2}{\text{dim}}} \inf_{(\Theta_0^N)^c \cap R} P_i^{\text{CN}}(C_N(\theta)) \) diverges to infinity. As a first step I show \( \sup_{\Theta_1} \left| E[C_N(\theta)] - E \left[ \left( f \cdot \theta(X) \right)^2 \right] \right| \to 0 \). In (88) use the definition of \( H_2(w_n, w_i, w_j) \) and do the change of variables \( \alpha = (z_1 - z_)/h \) to derive the first equality. To obtain the third equality in (88) a Taylor expansion around \( h = 0 \) of the whole integral and note that Assumption 7 allows us to differentiate under the integral and since \( f \) is Lipschitz in \( \alpha \).
\[ E \left[ \sup_{\Theta} \sqrt{N} |P_{N}^{1}(C(\Theta))| \right] \leq E \left[ \int_{0}^{D_{N}} \log \left[ N(\Theta_{1}, \rho_{1N}, t) \right]^2 dt \right] \leq E \left[ \|G_{1N}\|_{1N} \int_{0}^{D_{N}} |\log N(\Theta, || \cdot ||_{\infty}, u)| du \right] \leq \left[ \int_{0}^{M} \left( \frac{1}{u} \right)^{\frac{(m+\delta)/k}{2m\delta}} \right] (E[G_{1N}^2])^{\frac{1}{2}} = O(1) \] (89)

Markov’s inequality and (89) imply \( \sup_{\Theta} \sqrt{N} |P_{N}^{1}(C(\Theta))| = O_{p}(1) \). We now have the results necessary to show \( \inf_{(\Theta_{0}^{\ast})' \cap R} N \int_{\Theta} E[C(\Theta)] + N \int_{\Theta} E[C(\Theta)] - E[H_{2}(\Theta)|w_{n}, w_{1}] - E[H_{2}(\Theta)|w_{n}] | \leq G_{2N}(w_{n}, w_{1}) |||_{\infty}, where \( G_{2N}(w_{n}, w_{1}) = h^{-3d}(E[J_{N}|w_{n}, w_{1}] + E[J_{N}|w_{n}] + E[J_{N}|w_{n}] + E[J_{N}|w_{n}] + E[J_{N}|w_{n}]) \). Therefore, the class \( \Theta_{2N} \) is Lipschitz in \( \Theta \), and following the same arguments as in (89) we obtain the first two inequalities in (90). From Lemma (4), \( h^{-3d}(E[J_{N}]) \), \( h^{-3d}(E[J_{N}|w_{n}, w_{1}] + E[J_{N}|w_{n}]) \) and \( h^{-3d}(E[J_{N}|w_{n}, w_{1}) \) are all uniformly bounded in \( N \), which implies \( E[G_{2N}^2] \) is uniformly bounded. Together with \( \frac{(m+\delta)/k}{2m\delta} < \frac{1}{2} \), this implies the final equality.

\[ E \left[ \sup_{\Theta} \sqrt{N} |P_{N}^{2}(C(\Theta))| \right] \leq E \left[ \|G_{2N}\|_{2N} \int_{0}^{D_{N}} |\log N(\Theta, || \cdot ||_{\infty}, u)| du \right] \leq \left[ \int_{0}^{M} \left( \frac{1}{u} \right)^{\frac{(m+\delta)/k}{2m\delta}} \right] (E[G_{2N}^2])^{\frac{1}{2}} = O(1) \] (91)

Result (91) and Markov’s inequality imply \( \sup_{\Theta} \sqrt{N} |P_{N}^{2}(C(\Theta))| \) is asymptotically tight. I now proceed to show that \( \sup_{\Theta} \sqrt{N} |P_{N}^{2}(C(\Theta))| \) \( \rightarrow 0 \). Let \( \Theta_{3N} = \{ 2h^{-3d}(H_{2}(\Theta)|w_{n}, w_{1}) - E[H_{2}(\Theta)|w_{n}] - E[H_{2}(\Theta)|w_{n}] \} \). By Lemma (4), if \( \tilde{\theta}_{21}, \tilde{\theta}_{22} \in \Theta_{2N} \), then \( |\tilde{\theta}_{21}(w_{n}, w_{1}) - \tilde{\theta}_{22}(w_{n}, w_{1})| \leq G_{2N}(w_{n}, w_{1}) |||_{\infty} \), where \( G_{2N}(w_{n}, w_{1}) = h^{-3d}(E[J_{N}|w_{n}, w_{1}] + E[J_{N}|w_{n}] + E[J_{N}|w_{n}] + E[J_{N}|w_{n}] + E[J_{N}|w_{n}]) \). Therefore, the class \( \Theta_{3N} \) is Lipschitz in \( \Theta \), and we apply the same arguments as in (89) to obtain the first two inequalities in (92). From Lemma (4), we know that \( h^{-3d}(E[J_{N}]) \), \( h^{-3d}(E[J_{N}]) \), \( h^{-3d}(E[J_{N}|w_{n}, w_{1}] + E[J_{N}|w_{n}, w_{1}) \) and \( h^{-3d}(E[J_{N}|w_{n}, w_{1}] \) are all uniformly bounded in \( N \), which implies \( E[G_{3N}^2] \) is uniformly bounded.
uniformly bounded in $N$, which implies that $E[G_{\tilde{A}_N}^2\theta]$ is uniformly bounded in $N$.

\[
(Nh^d)^{-\frac{1}{2}} E \left[ \sup_{\Theta} \sup_{\epsilon N} \left| P_N^1(C_N(\theta)) \right| \right] \leq (Nh^d)^{-\frac{1}{2}} E \left[ \left| G_{\tilde{A}_N} \right|_{\sup} \int_0^{\frac{D_N}{N}} [\log N(\Theta, \|\cdot\|_\infty, u)] \frac{1}{u} du \right]
\]
\[
\leq (Nh^d)^{-\frac{1}{2}} \left[ \int_0^M \left( \frac{1}{u} \right) \frac{3^{m_0+4k}}{2^{m_0+6k}} \right] (E[G_{\tilde{A}_N}^2(\theta)])^\frac{1}{2} = O(N^{-\frac{1}{2}}h^{-\frac{1}{2}}) \tag{92}
\]
Since $Nh^d \to \infty$, result (92) and Markov’s inequality imply $\sup_{\Theta} N h^d |P_N^1(C_N(\theta))| \to 0$. In (93), use (87) to obtain the first inequality. From (90), \(-\sup_{\Theta} \sqrt{N} |P_N^1(C_N(\theta))| + \sqrt{N} \inf_{(\Theta, \epsilon N)^r} E[C_N(\theta)]\) is positive with probability tending to one. Hence, the first term diverges to positive infinity in probability. Furthermore, from (91), $\sup_{\Theta} N h^d |P_N^2(C_N(\theta))| = O_p(1)$ and from (92), $\sup_{\Theta} N h^d |P_N^3(C_N(\theta))| = O_p(1)$, which implies the final result in (93).

\[
\inf_{(\Theta, \epsilon N)^r} \sup_{\Theta} \left| N h^d C_N(\theta) \right| \geq (Nh^d)^{\frac{1}{2}} \left( -\sup_{\Theta} \sqrt{N} |P_N^1(\theta)| + \sqrt{N} \inf_{(\Theta, \epsilon N)^r} E[C_N(\theta)] \right)
\]
\[
+ \left( -\sup_{\Theta} N h^d |P_N^2(\theta)| - \sup_{\Theta} N h^d |P_N^3(\theta)| \right) \to +\infty \tag{93}
\]
I now derive an inequality that will be necessary to examine the behavior of $N h^d C_N(\theta)$ within $\Theta_{0, N}^r$.

Part (ii): $\inf_{\Theta_{0, N}^r} \sup_{\Theta} N h^d |P_N^1(C_N(\theta))| + O_p(1) \leq \inf_{\Theta_{0, N}^r} \sup_{\Theta} N h^d C_N(\theta) \leq \inf_{\Theta_{0, N}^r} \sup_{\Theta} N h^d |P_N^2(C_N(\theta))| + O_p(1)$

To establish this result we first derive a useful inequality in (94). For the left hand side, we note that $C_N(\theta) = \sum_{i=1}^3 P_N^i(C_N(\theta)) + E[C_N(\theta)] \geq \sum_{i=1}^3 P_N^i(C_N(\theta))$ since $E[C_N(\theta)] \geq 0$. The right hand side of the inequality follows from $\Theta_0 \cap \Theta \subseteq \Theta_{0, N}^r \cap \Theta$ for all $N$ and $E[C_N(\theta)] = 0$ for all $N$ if $\theta \in \Theta_0$.

\[
\sum_{i=1}^3 \inf_{\Theta_{0, N}^r} \sup_{\Theta} N h^d |P_N^i(C_N(\theta))| \leq \inf_{\Theta_{0, N}^r} \sup_{\Theta} N h^d C_N(\theta) \leq \inf_{\Theta_{0, N}^r} \sup_{\Theta} N h^d |P_N^2(C_N(\theta))| + \sum_{i \in \{1, 3\}} \sup_{\Theta_0} N h^d |P_N^i(C_N(\theta))| \tag{94}
\]
We proceed to analyze the terms in the left hand side of (94), beginning with $\inf_{\Theta_{0, N}^r} \sup_{\Theta} N h^d |P_N^1(C_N(\theta))|$. The first inequality in (95) was already derived in (89). For the second inequality in (95) apply Theorem (4). As already shown, $E[G_{\tilde{A}_N}^2\theta]$ is uniformly bounded in $N$. Therefore, Holder’s inequality, Jensen’s inequality and the definition of $D_N$ and $\|G_{\tilde{A}_N}\|_1$ imply the third inequality in (95).

\[
(Nh^d)^{\frac{1}{2}} E \left[ \sqrt{N} \sup_{\Theta_{0, N}^r} \left| P_N^1(C_N(\theta)) \right| \right] \leq (Nh^d)^{\frac{1}{2}} E \left[ \left| G_{\tilde{A}_N} \right|_{\sup} \int_0^{\frac{D_N}{N}} [\log N(\Theta, \|\cdot\|_\infty, u)] \frac{1}{u} du \right]
\]
\[
\leq (Nh^d)^{\frac{1}{2}} \left[ \int_0^M \left( \frac{1}{u} \right) \frac{3^{m_0+4k}}{2^{m_0+6k}} \right] (E[G_{\tilde{A}_N}^2\theta])^{\frac{1}{2}} = O(N^{-\frac{1}{2}}h^{-\frac{1}{2}}) \tag{95}
\]
To control the right hand side of (95), we begin by examining $E[(E[H_2(\theta)|w_n]]^2]$. In (96), use the Cauchy-Schwarz inequality and the change of variables $u = (z_i - z_n)/h$ and $v = (z_j - z_n)/h$ to obtain the first inequality.

\[
h^{-4d} E[(E[H_2(\theta)|w_n])]^{\frac{1}{2}} \leq \int \left[ \int K(u) v_i(\theta) f_{ZV}(hu + z_n, v_i) du dv_i \right]^4 f_Z(z_n) dz_n
\]
\[
+ 4 \int \left[ \int K(v) K(u - v) v_i(\theta) f_{ZV}(hu + z_n, v_i) f_Z(hv + z_n) dv_i du \right]^2 f_Z(z_n, v_n) dz_n dv_n \tag{96}
\]
Since $E[|v(\theta)|\|Z\|$ and $f_Z(z)$ are bounded it follows that $\int K(u) v_i(\theta) f_{ZV}(hu + z_n, v_i) du dv_i$ is bounded in $\theta$. This implies $\int \left[ \int K(u) v_i(\theta) f_{ZV}(hu + z_n, v_i) du dv_i \right]^4 f_Z(z_n) dz_n \leq \int \left[ \int K(u) v_i(\theta) f_{ZV}(hu + z_n, v_i) du dv_i \right]^2 f_Z(z_n) dz_n$.
(88) shows sup_{o} \int [f(K(u,v_1(\theta))f_{ZV}(hu + zn, v_1)du | f_{Z}(zn)dz_n - E[(|v(\theta)|Z)]^2 f_{Z}(Z)] = O(h). Hence, it follows that sup_{\epsilon_{N}} \int [f(K(u,v_1(\theta))f_{ZV}(hu + zn, v_1)du | f_{Z}(zn)dz_n = O(\epsilon_{N}). We now address the second term in (96). In (97), do a Taylor expansion of the integral around h = 0, and note that Assumption 7 allows us to differentiate under the integral sign and also implies that the second term that results from the Taylor expansion is bounded in \theta. Both E[v^2(\theta)|Z] and the density being bounded bounded imply the third inequality in (97).

\[
4 \int \left( \int K(vK(v-u)v(v(\theta)v(v_1(\theta))f_{ZV}(zn, v_1)f_{Z}(zn)du | f_{Z}(zn, v_n)dz_n dv_n + 8h \int K(vK(v-u)v(v_1(\theta)v(\theta)(f_{Z}(\hat{h} + zn, v_1) + f_{Z}(\hat{h} + zn, v_1)v) f_{Z}(zn, v_n) \right. \\
\left. \lesssim E[(|v(\theta)|Z)]^2 f_{Z}(Z)] + O(h) \right)
\]

(97)

Combining (96), (88) and (97) implies sup_{\epsilon_{N}} h^{-4d}E[(|H_{2}(\theta)|w_n)^2] = O(\epsilon_{N}). To finish controlling the right hand side of (95) define \tilde{\Theta}_{1N} = \{h^{-4d}((|H_{2}(\theta)|w_n) - E[|H_{2}(\theta)|])^2 - E[(|H_{2}(\theta)|w_n) - E[|H_{2}(\theta)|]^2] : \theta \in \Theta \}. From Lemma (4), we know that for \tilde{\theta}_{41}, \tilde{\theta}_{42} \in \tilde{\Theta}_{1N}, |\tilde{\theta}_{41}(w_n) - \tilde{\theta}_{42}(w_n)| \lesssim G_{4N}(w_n)||\theta_1 - \theta_2||_{\infty}, where G_{4N}(w_n) = h^{-4d}((|J_{N}|w_n)^2 + 2E[|J_{N}|]^2 + 2E[J_{N}]E[|J_{N}|] + E[|E[J_{N}|]^2]). Hence, the classes \tilde{\Theta}_{1N} are Lipschitz in \Theta and therefore the same arguments as in (89) imply the first two inequalities in (98). By Lemma (4), h^{-2d}E[|J_{N}|] and h^{-4d}E[|E[J_{N}|]^2] are bounded in N, which implies E[|G_{4N}|] is uniformly bounded. Together with \frac{(\frac{\frac{N}{p} + \frac{\frac{N}{p}}{2}}{\frac{N}{p}} + 1}{\frac{N}{p}} < \frac{1}{2},

this implies the last inequality in (98).

\[
E \left[ \frac{\sup_{\Theta_{1N}} \sqrt{N} \left( \sum_{n=1}^{N} \tilde{\theta}(w_n) \right) )}{\sqrt{N}} \right] \lesssim E \left[ \left| G_{4N} \right| \frac{1}{N} \int_{0}^{\frac{N}{p}} \left[ \log \left| \Theta_{1N} \right| \left( \frac{1}{2} \right) du \right] \right] \\
\lesssim \left[ \int_{0}^{N} \left( \frac{1}{u} \right) \left( \frac{\frac{N}{p} + \frac{\frac{N}{p}}{2}}{\frac{N}{p}} + 1 \right) \right] \left( E \left| G_{4N} \right| \right)^{2} = O(1)
\]

(98)

I now show sup_{\epsilon_{N} \cap R} N h^{\frac{1}{2}} \log^{p_{1}}(C_{N}(\theta)) \xrightarrow{P} 0. Result (95) and \Theta_{0}^{\epsilon_{N}} \subseteq \Theta imply the first inequality in (99). From (98), \frac{1}{\sqrt{N}} E \left[ \sup_{\epsilon_{N} \cap R} \frac{1}{\sqrt{N}} \left( \sum_{n=1}^{N} \tilde{\theta}(w_n) \right) \right] = O(\sqrt{N}), and as shown sup_{\epsilon_{N} \cap R} h^{-4d}E[(|H_{2}(\theta)|w_n)^2] = O(\epsilon_{N}). Since \epsilon_{N} has the same order as h for l \leq 1 and \sqrt{Nh} \xrightarrow{l} \infty, we obtain the second equality in (99)

\[
(Nh^{d}) \frac{1}{N} \log^{p_{1}}(C_{N}(\theta)) \xrightarrow{P} 0.
\]

Since \frac{1}{\sqrt{N}} \xrightarrow{P} 0, it follows from Markov’s inequality and (99) that \left| N h^{\frac{1}{2}} \log^{p_{1}}(C_{N}(\theta)) \right| \leq \frac{1}{\sqrt{N}} \log^{p_{1}}(C_{N}(\theta)) \xrightarrow{P} 0. Furthermore, from result (92) \frac{1}{\sqrt{N}} \log^{p_{1}}(C_{N}(\theta)) \xrightarrow{P} 0, which implies \left| N h^{\frac{1}{2}} \log^{p_{1}}(C_{N}(\theta)) \right| \xrightarrow{P} 0. In addition, note that if \theta \in \Theta_{0}, then E[|H_{2}(\theta)|w_n] = 0, and hence \frac{p_{1}}{N}(C_{N}(\theta)) = 0. Combine these results together with inequality (94) to conclude:

\[
\inf_{\epsilon_{N} \cap R} N h^{\frac{1}{2}} P_{N}^{2}(C_{N}(\theta)) + op(1) \leq \inf_{\epsilon_{N} \cap R} N h^{\frac{1}{2}} C_{N}(\theta) \leq \inf_{\epsilon_{N} \cap R} N h^{\frac{1}{2}} P_{N}^{2}(C_{N}(\theta)) + op(1)
\]

(100)

We now proceed to exploit inequality (100) to analyze \inf_{\epsilon_{N} \cap R} N h^{\frac{1}{2}} C_{N}(\theta).

Part (iii): Show \inf_{\epsilon_{N} \cap R} N h^{\frac{1}{2}} C_{N}(\theta) \xrightarrow{P} \inf_{\Theta_{0} \cap R} N h^{\frac{1}{2}} P_{N}^{2}(C_{N}(\theta))

Due to inequality (100), we only need to show \inf_{\epsilon_{N} \cap R} N h^{\frac{1}{2}} P_{N}^{2}(C_{N}(\theta)) \xrightarrow{P} \inf_{\Theta_{0} \cap R} N h^{\frac{1}{2}} P_{N}^{2}(C_{N}(\theta)). For this purpose we will need to show that N h^{\frac{1}{2}} P_{N}^{2}(C_{N}(\theta)) is asymptotically uniformly equicontinuous in probability on \Theta
with respect to the norm \( \| \cdot \|_{L^2(X)} \). That is, we wish to show that for every \( \epsilon, \eta > 0 \), there exists a \( \delta > 0 \) such that:

\[
\lim_{N \to \infty} \sup P \left( \frac{NH}{2} \left| P_N^2(C_N(\theta_1)) - P_N^2(C_N(\theta_2)) \right| > \epsilon \right) < \eta
\]  

(101)

In order to prove (101) we need to define the classes \( \hat{\Theta}_{2N} = \{ \hat{\theta}_{51}, \hat{\theta}_{52} : \hat{\theta}_{51}, \hat{\theta}_{52} \in \hat{\Theta}_{2N}, ||\theta_1 - \theta_2||_{L^2(X)} < \delta \} \), and \( \hat{\Theta}^\epsilon_{2N} = \{ \hat{\theta}_{51}, \hat{\theta}_{52} : \hat{\theta}_{51}, \hat{\theta}_{52} \in \hat{\Theta}_{2N}, ||\theta_1 - \theta_2||_{L^2(X)} < \delta \} \). Since \( NH \frac{1}{2} P_N^2(C_N(\theta)) \) is a P-Canonical U-Statistic, we use Theorem (2) to derive the first inequality in (102). Since \( \hat{\Theta}_{2N} \subseteq \hat{\Theta}^\infty_{2N} \), it follows that \( N(\hat{\Theta}_{2N}, \rho_{2N}, \epsilon) \leq N(\hat{\Theta}^\infty_{2N}, \rho_{2N}, \epsilon) \). By the Cauchy-Schwarz inequality we can form a cover for \( \hat{\Theta}^\infty_{2N} \) of size \( \epsilon \) under \( \rho_{2N} \) by taking the product of two covers of \( \hat{\Theta}_{2N} \) of size \( \epsilon/2 \) and hence \( N(\hat{\Theta}_{2N}, \rho_{2N}, \epsilon) \leq N^2(\hat{\Theta}_{2N}, \rho_{2N}, \epsilon/2) \).

By Lemma (4), for \( \theta_{51}, \theta_{52} \in \hat{\Theta}_{2N} \), \( ||\theta_{51}(w, n) - \theta_{52}(w, n)|| \leq G_{2N}(w, n, \theta_1 - \theta_2) \), where \( G_{2N}(w, n, \theta_1 - \theta_2) \) is the class \( \hat{\Theta}_{2N} \) is Lipschitz in \( \Theta \) and we can use Theorem (3) and the previous results to derive the second inequality in (102). Applying Theorem (4) gives us the third inequality in (102).

We now proceed to control the right hand side of (102) by looking at its Hoeffding decomposition, beginning with \( \sup_{\hat{\Theta}_{2N}} E[\hat{\theta}^2] \). First, note that Jensen’s and the Cauchy-Schwarz inequalities imply that \( E[\hat{\theta}^2] \leq E[(E[H_2(\theta_1) - H_2(\theta_2)]^2) \). Furthermore, since \( v_n(\theta_1)

\[
E \left[ \sup_{||\theta_1 - \theta_2||_{L^2(X)} < \delta} NH \frac{1}{2} \left| P_N^2(C_N(\theta_1)) - P_N^2(C_N(\theta_2)) \right| \right] \leq E \left[ \int_{D_N} \log N(\hat{\Theta}_{2N}^\epsilon, \rho_{2N}, \epsilon) \right] 
\]  

(102)

We now proceed to control the right hand side of (102) by looking at its Hoeffding decomposition, beginning with \( \sup_{\hat{\Theta}_{2N}} E[\hat{\theta}^2] \). First, note that Jensen’s and the Cauchy-Schwarz inequalities imply that \( E[\hat{\theta}^2] \leq E[\hat{\theta}^2] \). Furthermore, since \( v_n(\theta_1)

\[
E[\hat{\theta}^2] \leq 2 \int K^2(\theta_1 - \theta_2) f_{ZX}(z, x) dudvdz + \int K^2(u) K^2(v) dudvdz 
\]  

(103)

Define the class \( \hat{\Theta}_{2N}^\epsilon = \{ \hat{\theta}_{52} : \hat{\theta}_{52} \in \hat{\Theta}_{2N} \} \). We aim to show \( E \left[ \sup_{\hat{\Theta}_{2N}^\epsilon} N^{-1} \sum_{n=1}^{N} \hat{\theta}(w, n) \right] \rightarrow 0 \), but before it is necessary to understand the covering numbers of \( \hat{\Theta}_{2N}^\infty \). In (104), let \( \hat{\theta}_{51}, \hat{\theta}_{52} \in \hat{\Theta}_{2N} \) and use the Cauchy-Schwarz inequality and concavity to derive the first inequality, for \( \hat{\theta}_{51}, \hat{\theta}_{52} \in \hat{\Theta}_{2N}^\infty \). Since \( \theta \in \Theta \) are uniformly bounded and the class \( \hat{\Theta}_{2N} \) is Lipschitz in \( \Theta \), it follows that \( \hat{\theta} \in \hat{\Theta}_{2N}^\infty \) have envelope proportional to \( G_{2N}(w, n) \). Jensen’s inequality then gives us the second result in (104). We notice that the resuling expression is a semimetric on \( \hat{\Theta}_{2N}^\infty \), and we define it to be \( \hat{\rho}_{2N}(\hat{\theta}_{51}, \hat{\theta}_{52}) \), and its associated norm \( \| \cdot \|_{1/2} \).
In (105), apply Theorem (2), use $\hat{\Theta}^\delta_{6N} \subseteq \hat{\Theta}^\infty_{6N}$ and (104) to obtain the first inequality. Since $N(\hat{\Theta}^\infty_{5N}, \bar{\rho}_{1N}, \epsilon) \leq N^2(\hat{\Theta}^\infty_{5N}, \bar{\rho}_{1N}, \epsilon/2)$, and the class $\hat{\Theta}^\infty_{5N}$ is Lipschitz in $\Theta$, we obtain the third inequality in (105). In addition, from (104) and $\hat{\theta} \in \hat{\Theta}^\infty_{5N}$ having envelope $G_{5N}(w_n, w_i)$ it follows that $D_N \leq 2||G_{5N}||_{1N}$. Therefore, Theorem (4) and Jensen’s inequality imply the fourth inequality in (105). Finally, straightforward calculations and Lemma (4) can be used to show that $E \left[ (E[G^2_{5N}(w_n)]^2) \right]$ is bounded in $N$, which implies the last inequality in (105).

$$N^{-\frac{1}{2}}E \left[ \frac{1}{N} \sum_{n=1}^{N} \bar{\theta}(w_n) \right] \leq N^{-\frac{1}{2}}E \left[ \int_{0}^{D_N} \log N(\hat{\Theta}^\infty_{5N}, \bar{\rho}_{1N}, \epsilon) \right] \leq N^{-\frac{1}{2}} \left( E \left[ (E[G^2_{5N}(w_n)]^2) \right] \right)^{\frac{1}{2}} = O(N^{-\frac{1}{2}}) \quad (105)$$

Define the class of functions $\hat{\Theta}^\delta_{2N} = \{ \hat{\theta}(w_n, w_i) - E[\hat{\theta}^2 | w_n] - E[\hat{\theta}^2 | w_i] + E[\hat{\theta}^2] : \hat{\theta} \in \hat{\Theta}^\delta_{2N} \}$. To control the right hand side of (102), we aim to show $E \left[ \sup_{\Theta^\delta_{2N}} N^{-2} \sum_{n=1}^{N} \sum_{i<n} \bar{\theta}(w_n, w_i) \right] \to 0$. First we need to control the covering numbers of $\Theta^\delta_{2N}$ under $\rho_{2N}$. Similarly as in (104), use the Cauchy Schwarz and Jensen’s inequalities and the $\hat{\theta} \in \hat{\Theta}^\delta_{2N}$ having envelope $G_{5N}$ to derive the first inequality in (106) for $\hat{\theta}_{1N}, \hat{\theta}_{2N}$ in $\hat{\Theta}^\infty_{5N}$ and $\hat{\theta}_{51}, \hat{\theta}_{52} \in \hat{\Theta}^\infty_{5N}$. Define $\tilde{\rho}_{2N}(\hat{\theta}_{51}, \hat{\theta}_{52})$ to be result semimetric on $\hat{\Theta}^\infty_{5N}$, and let $|| \cdot ||_{2N}$ be the associated norm.

$$\rho_{2N}(\hat{\theta}_{1N}, \hat{\theta}_{2N}) \geq \left[ \frac{1}{N^2} \sum_{n=1}^{N} \sum_{i<n} \left( \hat{\theta}_{51}(w_n, w_i) - \hat{\theta}_{52}(w_n, w_i) \right)^2 G^2_{5N}(w_n, w_i) \right]^{\frac{1}{2}}$$

$$+ \left[ \frac{1}{N} \sum_{n=1}^{N} E[\hat{\theta}_{51} - \hat{\theta}_{52}]^2 G^2_{5N}(w_n) \right]^{\frac{1}{2}} + \left[ E \left[ (\hat{\theta}_{51} - \hat{\theta}_{52}) G^2_{5N} \right] \right]^{\frac{1}{2}} \equiv \tilde{\rho}_{2N}(\hat{\theta}_{51}, \hat{\theta}_{52}) \quad (106)$$

Similarly, in (105), in (107) apply Theorem (2), use $\hat{\Theta}^\delta_{7N} \subseteq \hat{\Theta}^\infty_{7N}$ and (106) to derive the first inequality. Again we exploit $N(\hat{\Theta}^\infty_{5N}, \tilde{\rho}_{2N}, \epsilon) \leq N^2(\hat{\Theta}^\infty_{5N}, \bar{\rho}_{2N}, \epsilon)$ and $\hat{\Theta}^\infty_{5N}$ being Lipschitz in $\Theta$, together with Theorem (3), to obtain the second inequality in (107). It follows from (106) that $D_N \leq 2||G_{5N}||_{2N}$, and therefore Theorem (4) and Jensen’s inequality imply the third inequality in (107). Straightforward manipulations and Lemma (4) can be used to show that $E \left[ G^2_{5N} \right] \leq h^{-d}$, which gives us the last equality in (107).

$$N^{-1}E \left[ \sup_{\Theta^\delta_{2N}} \frac{1}{N} \sum_{n=1}^{N} \sum_{i<n} \bar{\theta}(w_n, w_i) \right] \leq N^{-1}E \left[ \int_{0}^{D_N} \log N(\hat{\Theta}^\infty_{5N}, \tilde{\rho}_{2N}, \epsilon) \right] \leq N^{-1} \left( E \left[ G^2_{5N} \right] \right)^{\frac{1}{2}} = O(N^{-1}h^{-\frac{d}{2}}) \quad (107)$$

We can now finish showing that $Nh^{\frac{d}{2}} P_{N}^\delta(C_N(\theta))$ is asymptotically uniformly equicontinuous in probability on $\Theta$ with respect to $|| \cdot ||_{L^2(X)}$. The first inequality in (108) follows from Markov’s inequality and (102). From (103), we know $\sup_{\Theta^\delta_{2N}} E \left[ \hat{\theta}_{51}^2 | w_n, w_i \right] \leq \delta^2$. Hence, (103), (105), (107) and $Nh^d \to \infty$ imply the final inequality in (108).

$$P \left( \sup_{||\theta_1 - \theta_2||_{L^2(X)} < \delta} Nh^{\frac{d}{2}} | P_{N}^\delta(C_N(\theta_1)) - P_{N}^\delta(C_N(\theta_2)) | > \epsilon \right) \leq \sup_{\Theta^\delta_{2N}} E \left[ \hat{\theta}_{51}^2 | w_n, w_i \right] + \sup_{\Theta^\delta_{2N}} \left[ \frac{1}{N} \sum_{n=1}^{N} \hat{\theta}_{51}(w_n) \right] + \sup_{\Theta^\delta_{2N}} \left[ \frac{1}{N} \sum_{n=1}^{N} \hat{\theta}_{52}(w_n, w_i) \right] \leq \delta^2 + O(N^{-\frac{1}{2}}) \quad (108)$$

Using (108), we can show that $\inf_{\Theta^\delta_{2N}} Nh^{\frac{d}{2}} C_N(\theta) \to P_{N}^\delta(C_N(\theta))$. Note that $R$, $\Theta_0$ and $\Theta_0^\delta$ are closed under $|| \cdot ||_{C_5}$. Hence, since $\Theta$ is compact under $|| \cdot ||_{C_5}$ by Gallant & Nychka (1987), we conclude that $\Theta_0 \cap R$ and $\Theta_0^\delta \cap R$ are both compact under $|| \cdot ||_{C_5}$. Since $Nh^{\frac{d}{2}} P_{N}^\delta(C_N(\theta))$ is continuous under $|| \cdot ||_{C_5}$, the infimums in (109)
are attained. Let \( \theta^* \) denote the optimum on \( \Theta_0^N \cap R \) and \( \theta^*_P = \arg \min_{\theta \in R} \| \theta^* - \theta \|_{C^5} \) be the projection of \( \theta^* \) to \( \Theta_0 \cap R \) under \( \| \cdot \|_{C^5} \) to obtain the first inequality in (109). Next we note that \( \| \theta^* - \theta^*_P \|_{C^5} \to 0 \). To see this, fix any \( \delta > 0 \), and define \( A^\delta = \{ \tilde{\theta} \in \Theta : \inf_{\theta \in R} \| \tilde{\theta} - \theta \|_{C^5} \geq \delta \} \), which is compact under \( \| \cdot \|_{C^5} \). Since \( E[(E[\epsilon(\theta)]Z)^2] = \overline{f}(Z) \) is continuous under \( \| \cdot \|_{C^5} \) the minimum \( \pi^* = \inf_{\theta \in R} E[(E[\epsilon(\theta)]Z)^2] \) is attained with \( \pi^* > 0 \) as \( A^\delta \cap \Theta_0 \cap R = \emptyset \).

Thus, \( A^\delta \cap \Theta_0^N \cap R = \emptyset \) for all \( \varepsilon < \pi^* \), which implies \( \sup_{\theta_1 \in \Theta_0^N \cap R} \inf_{\theta_2 \in \Theta_0^N \cap R} |\theta_1 - \theta_2|_{C^5} < \delta \) for \( \varepsilon < \pi^* \). This implies \( \| \theta^* - \theta^*_P \|_{C^5} \to 0 \), and hence \( \| \theta^* - \theta^*_P \|_{L^2(X)} \to 0 \). This result implies the second inequality in (109) for some \( \delta_n \to 0 \). Finally, from (108) we know \( \sup_{\theta_1 \in \Theta_0^N \cap R} \inf_{\theta_2 \in \Theta_0^N \cap R} |\theta_1 - \theta_2|_{C^5} < \delta_n \) with \( \inf_{\theta \in R} \| \theta^* - \theta \|_{C^5} \to 0 \).

### Part (v)

Show \( \inf_{\Theta \cap R} N \hat{h}^2 C_\Theta(\theta) \xrightarrow{\mathcal{L}} \inf_{\Theta \cap R} G(\theta) \) when \( \Theta_0 \cap R \neq \emptyset \).

We begin by establishing the asymptotic distribution of \( \inf_{\Theta_0^N \cap R} N \hat{h}^2 C(\theta) \). By Theorem (1), for every \( \theta \in \Theta_0 \),

\[
N \hat{h}^2 P_N^2(C(\theta)) \xrightarrow{\mathcal{L}} N(0, \hat{C}^2(\theta)).
\]

Using the Cramer-Wold device and calculations identical to those in Lemma (1) and Lemma (2), it is possible to show that the statistics \( \left( N \hat{h}^2 P_N^2(C(\theta_1)), \ldots, N \hat{h}^2 P_N^2(C(\theta_K)) \right) \) converge jointly to a multivariate Normal distribution for any finite vector \( (\theta_1, \ldots, \theta_K) \). From (108), the process \( N \hat{h}^2 P_N^2(C(\theta)) \) is asymptotically equicontinuous in probability on \( \Theta \) and therefore also on \( \Theta_0 \) with respect to \( \| \cdot \|_{C^5(X)} \). Since convergence of marginals and asymptotic equicontinuity in probability imply convergence in Law on \( \mathcal{L}^\infty(\Theta_0) \), see for example Theorem 1.5.4 and Theorem 1.5.7 in van der Vaart & Wellner (1996), we conclude that \( N \hat{h}^2 P_N^2(C(\theta)) \) converges in Law to a tight Gaussian process on \( \mathcal{L}^\infty(\Theta_0) \). In (110) we examine the asymptotic distribution of \( N \hat{h}^2 C(\theta) \) on \( \mathcal{L}^\infty(\Theta_0) \). The first equality in (110) follows from the Hoeffding decomposition of \( C(\theta) \) and \( P_N(C(\theta)) \to 0 \) for \( \theta \in \Theta_0 \). By (92), \( N \hat{h}^2 P_N^2(C(\theta)) \xrightarrow{p} 0 \) in \( \mathcal{L}^\infty(\Theta_0) \), and hence Slutsky’s Lemma implies the final result in (110).

\[
N \hat{h}^2 C(\theta) = N \hat{h}^2 P_N^2(C(\theta)) + N \hat{h}^2 P_N^2(C(\theta)) \xrightarrow{\mathcal{L}} G(\theta)
\]

where \( G(\theta) \) is a Gaussian process defined on \( \mathcal{L}^\infty(\Theta_0) \). We now use this result to conclude the proof. In (111) note that the second term is \( o_p(1) \) since \( P \left( \inf_{(\theta_0^N) \cap R} \inf_{(\epsilon_0^N) \cap R} N \hat{h}^2 C(\theta) \right) \geq \inf_{(\theta_0^N) \cap R} N \hat{h}^2 C(\theta) \| \right) > \epsilon \) \( P \left( \inf_{(\theta_0^N) \cap R} N \hat{h}^2 C(\theta) \geq \inf_{(\theta_0^N) \cap R} N \hat{h}^2 C(\theta) \right) \xrightarrow{p} 0 \), because as shown in Part (iii) \( \inf_{\Theta_0^N \cap R} N \hat{h}^2 C(\theta) \) is asymptotically tight, while \( \inf_{(\theta_0^N) \cap R} N \hat{h}^2 C(\theta) \xrightarrow{p} +\infty \) by (93). These results imply the second equality in (111). The final result in (111) follows from (110) and the continuous mapping theorem.

\[
\inf_{\theta \in \Theta} N \hat{h}^2 C(\theta) = \inf_{(\theta_0^N) \cap R} \left( \inf_{\theta \in \Theta} N \hat{h}^2 C(\theta) \right) \leq \inf_{(\theta_0^N) \cap R} N \hat{h}^2 C(\theta) + \inf_{(\epsilon_0^N) \cap R} N \hat{h}^2 C(\theta) \leq \inf_{(\theta_0^N) \cap R} N \hat{h}^2 C(\theta) + o_p(1) \xrightarrow{\mathcal{L}} \inf_{\Theta \cap R} G(\theta)
\]

which concludes the proof of the first claim of the Theorem. We now proceed to establish the second claim.

### Part (v)

Show \( \inf_{\Theta \cap R} N \hat{h}^2 C(\theta) \xrightarrow{p} +\infty \) when \( \Theta_0 \cap R = \emptyset \).
Let $\pi^* = \inf_{\theta \in \Theta} E \left[ \left| E[Y - \theta(X)] Z \right|^2 f_Z^2(Z) \right]$. Since $\Theta \cap R$ is compact under $||.|||_{C^4}$ and $E \left[ \left| E[Y - \theta(X)] Z \right|^2 f_Z^2(Z) \right]$ is continuous under $||.|||_{C^4}$ the infimum is attained at some function $\theta^*$. Since $\Theta_0 \cap R = \emptyset$ and $\Theta_0 \subseteq \Theta$, it follows that $\theta^* \notin \Theta_0$ and hence $\pi^* > 0$. Thus, $\Theta \cap R \subseteq (\Theta_0^c)^c$ for $\epsilon_N$ small enough, which together with (93) implies:

$$\inf_{\theta \in \Theta \cap R} Nh^\frac{4}{5} C_N(\theta) \geq \inf_{(\Theta_0)^c} Nh^\frac{4}{5} C_N(\theta) \xrightarrow{P} +\infty \quad (112)$$

which concludes the proof of the Theorem.

**Lemma 4.** Define the kernels $H_3(w_n, w_i, w_j | \theta)$ and $H_4(w_n, w_i, w_j, w_k | \theta)$ as in (34) and (75) respectively. If Assumptions 1- 8 hold, then:

1. There are $J_N(w_n, w_i, w_j)$ such that $|H_2(w_n, w_i, w_j | \theta_1) - H_2(w_n, w_i, w_j | \theta_2)| \leq J_N(w_n, w_i, w_j) ||\theta_1 - \theta_2||_{\infty}$. In addition, the quantities $h^{-2d} E[J_N], h^{-4d} E[(E[J_N|w_n])^2], h^{-3d} E[(E[J_N|w_n, w_i])^2]$ and $h^{-2d} E[F_N^2]$ are all uniformly bounded in $N$.

2. There are $F_N(w_n, w_i, w_j, w_k)$ such that $|H_3(w_n, w_i, w_j, w_k | \theta_1) - H_3(w_n, w_i, w_j, w_k | \theta_2)| \leq F_N(w_n, w_i, w_j, w_k) ||\theta_1 - \theta_2||_{\infty}$. In addition, the quantities $h^{-3d} E[(E[F_N|w_n])^2], h^{-7d} E[(E[F_N|w_n, w_i])^2], h^{-6d} E[(E[F_N|w_n, w_i, w_j])^2], h^{-4d} E[F_N]$ and $h^{-5d} E[F_N^2]$ are all uniformly bounded in $N$.

**Proof of Lemma 4:** In (118), note that $|v_n(\theta_1)v_1(\theta_1) - v_n(\theta_2)v_1(\theta_2)| \leq (|y_n| + |\theta_2(x_n)|) ||\theta_1(x_i) - \theta_2(x_i)|| + (|y_i| + |\theta_1(x_i)||\theta_1(x_i) - \theta_2(x_i))$ and use that $\theta \in \Theta$ are uniformly bounded to derive the result holds for $W(y_n, y_i, y_j) = (1 + |y_n| + |y_i| + |y_j|)$. Let $J_N(w_n, w_i, w_j)$ be the right hand side of (113) to establish the first claim of the Lemma.

$$|H_2(w_n, w_i, w_j | \theta_1) - H_2(w_n, w_i, w_j | \theta_2)| \leq \left( K \left( \frac{z_i - z_n}{h} \right) K \left( \frac{z_i - z_n}{h} \right) + K \left( \frac{z_n - z_i}{h} \right) K \left( \frac{z_n - z_i}{h} \right) \right) W(y_n, y_i, y_j) \quad (113)$$

I now calculate the appropriate moments of $J_N(w_n, w_i, w_j)$. In (118) do the change of variables $u = (z_i - z_n)/h$ and $v = (z_j - z_n)/h$ to obtain the first equality. Note that $E[|Y||Z]$ is bounded since $E[|Y - \theta(X)||Z]$ is bounded and $0 \in \Theta$, and hence, since $f_Z(z)$ is bounded, the second inequality follows for some $K$ not depending on $N$.

$$h^{-2d} E[J_N] = 3 \int K(u)K(v)W(y_n, y_i, y_j)f_{Z Y}(hu + z_n, y_i)f_{Z Y}(hv + z_n, y_j)f_{Z Y}(z_n, y_n)dz_n dy_n dudvdy_j \leq K \int K(u)K(v)f_Z(z_n)dudvdz_n = K \quad (114)$$

In (115) the Cauchy Schwarz inequality implies the first inequality. In both expectations, do the change of variables $u = (z_i - z_n)/h$ and $v = (z_j - z_n)/h$, and integrate out $y_i$ and $y_j$ using the assumptions that $E[|Y||Z]$ and $f_Z(z)$ are bounded to obtain the second inequality in (115) for some $K$ not depending on $N$. Exploit the fact that the kernel $K(u)$ is a density to derive the final equality.

$$h^{-4d} E[(E[J_N|w_n])^2] \leq 8h^{-4d} E \left[ \left( E \left[ K \left( \frac{z_i - z_n}{h} \right) K \left( \frac{z_j - z_n}{h} \right) W(y_n, y_i, y_j|w_n) \right] \right)^2 \right]$$

$$+ 8h^{-4d} E \left[ \left( E \left[ K \left( \frac{z_n - z_i}{h} \right) K \left( \frac{z_j - z_i}{h} \right) W(y_n, y_i, y_j|w_n) \right] \right)^2 \right] \leq KE \left( \int K(u)K(v)dudv \right)^2 (1 + |Y|)^2$$

$$+ KE \left( \int K(u)K(v-u)dudv \right)^2 (1 + |Y|)^2 \quad (115)$$

53
In (116) use the Cauchy Schwarz inequality to obtain the first inequality. Let $u = (z_i - z_n)/h$ and $v = (z_j - z_n)/h$, integrate out $y_i$ and $y_j$ using the assumptions that $E[|Y||Z|], E[|Y^2||Z]$ and $f_Z(z)$ are all bounded in $z$ to obtain the second inequality in (116) for some $K$ not depending on $N$.

$$h^{-3d}E[(E[|w_n|w_n]|w_i)^2] \leq 8h^{-3d}E \left[ \left( \sum_{n,j} K \left( \frac{z_i - z_n}{h} \right) \left( \frac{z_j - z_n}{h} \right) W(y_n, y_i, y_j) \right) |w_n, w_i|^2 \right]$$

In (117), do the change of variables $u = (z_i - z_n)/h, v = (z_j - z_n)/h$ and integrate out $y_i$ and $y_j$ using the assumptions that $E[|Y||Z|], E[|Y^2||Z]$ and $f_Z(z)$ are all bounded in $z$ to obtain the result for some $K$ not depending on $N$.

$$h^{-2d}E[J_N^2] \leq K E \left[ \left( \int (K(u)K(v) + K(u)K(v - u) + K(u - v)K(v - u))^2 \right) dudv(1 + |Y|)^2 \right]$$

I now proceed to establish the second claim of the Lemma. In (118), use $a^2 - b^2 = (a - b)(a + b)$ and the $\theta \in \Theta$ being uniformly bounded to obtain the result for $G(y_n, y_i, y_j, y_k)$ a polynomial of order two in $(|y_n|, |y_i|, |y_j|, |y_k|)$.

$$|H_3(w_n, w_i, w_j, w_k\theta_1) - H_3(w_n, w_i, w_j, w_k\theta_2)| \leq \left( \int K \left( \frac{z_n - z}{h} \right) K \left( \frac{z_i - z}{h} \right) K \left( \frac{z_j - z}{h} \right) K \left( \frac{z_k - z}{h} \right) dz \right) G(y_n, y_i, y_j, y_k)||\theta_1 - \theta_2||_\infty$$

Let $F_N(w_n, w_i, w_j, w_k) = \left( \int K \left( \frac{z_n - z}{h} \right) K \left( \frac{z_i - z}{h} \right) K \left( \frac{z_j - z}{h} \right) K \left( \frac{z_k - z}{h} \right) dz \right) G(y_n, y_i, y_j, y_k)$ to establish the second claim of the Lemma. I now calculate the appropriate moments of $F_N(w_n, w_i, w_j, w_k)$. In (119) do the change of variables $u = (z - z_n)/h, v = (z_i - z_n)/h, w = (z_j - z_n)/h$ and $t = (z_k - z_n)/h$, integrate out $(y_n, y_i, y_j, y_k)$ and use that $E[|Y||Z|], E[|Y^2||Z]$ and $f_Z(z)$ are bounded to obtain the first inequality for some $K$ not depending on $N$. The last equality simply follows from the kernel $K(u)$ being a density.

$$h^{-4d}E[F_N] \leq K \int (K(u)K(v - u)K(w - u)K(t - u)f_Z(z_n)dudvdwtdz_n = K$$

In (120) do the change of variables $u = (z - z_n)/h, v = (z_i - z_n)/h, w = (z_j - z_n)/h$ and $t = (z_k - z_n)/h$, integrate out $(y_n, y_i, y_j, y_k)$ and use that $E[|Y||Z|], E[|Y^2||Z]$ and $f_Z(z)$ are bounded to derive the first inequality, where $G_1(y_n)$ is a polynomial of order four in $|y_n|$ and $K$ is a constant not depending on $N$. The second equality in (119) follows from the kernel $K(u)$ being a density.

$$h^{-8d}E[(E[F_N]|w_n)^2] \leq K \int (K(u)K(v - u)K(w - u)K(t - u)dudvdwtdz_n)^2 G_1(y_n)f_Y(y_n)dy_n = KE[\tilde{G}_1(y_n)]$$

In (121) do the change of variables $u = (z - z_n)/h, v = (z_i - z_n)/h, w = (z_j - z_n)/h$ and $t = (z_k - z_n)/h$, integrate out $(y_i, y_k)$ and use the same arguments as in (119) to obtain the first inequality for $\tilde{G}_2(y_n, y_i)$ a polynomial of order four in $(|y_n|, |y_i|)$ and $\tilde{K}$ a scalar not depending on $N$. Integrate out the $y_i$ and use that $E[|Y||Z|], E[|Y^2||Z]|, E[|Y^4||Z]|, E[|Y^4||Z]$ and $f_Z(z)$ are all bounded in $z$ together with $K(u)$ being a density to obtain the last inequality.
for a constant $K_2$ not depending on $N$.

$$h^{-7d}E[(E[F_N|w_n, w_i])^2]$$

$$\leq K_1 \int \left( \int K(u)K(v-u)K(w-u)K(t-u)dudwdt \right)^2 \tilde{G}_2(y_n, y_i)f_{ZY}(hv + z_n, y_i)f_{ZY}(z_n, y_n)dv dy dz dy dz y_n$$

$$\leq K_2 \left[ \int \left( \int K(u)K(v-u)du \right)^2 dv \right] E[\tilde{G}_1(y_n)] \quad (121)$$

In (122) do the change of variables $u = (z - z_n)/h$, $v = (z_i - z_n)/h$, $w = (z_j - z_n)/h$ and $t = (z_k - z_n)/h$, integrate out $y_k$ and use the same arguments as in (119) to obtain the first inequality for $\tilde{G}_3(y_n, y_i, y_j)$ a polynomial of order four in $(|y_n|, |y_i|, |y_j|)$ and $K_1$ a scalar not depending on $N$. Again we integrate out $(y_i, y_j)$, and exploit the assumption that $E[|Y|^2], E[|Y|^4]$, $E[|Y^3|^2], E[|Y^4|]$ and $f_Z(z)$ are bounded together with $K(u)$ being a density to obtain the last inequality in (122) for some constant $K_2$ not depending on $N$.

$$h^{-4d}E[(E[F_N|w_n, w_i, w_j])^2]$$

$$\leq K_1 \int \left( \int K(u)K(v-u)K(w-u)K(t-u)dudwdt \right)^2 \tilde{G}_3(y_n, y_i, y_j)f_{ZY}(hv + z_n, y_i)f_{ZY}(hv + z_n, y_j)f_{ZY}(z_n, y_n)$$

$$\leq K_2 \left[ \int \left( \int K(u)K(v-u)K(w-u)K(t-u)du \right)^2 dv dw \right] E[\tilde{G}_1(y_n)] \quad (122)$$

In (122) do the change of variables $u = (z - z_n)/h$, $v = (z_i - z_n)/h$, $w = (z_j - z_n)/h$ and $t = (z_k - z_n)/h$. Also integrate out $(y_n, y_i, y_j, y_k)$ using that $E[|Y|^2], E[|Y^3|^2], E[|Y^4|]$ and $f_Z(z)$ are bounded to obtain the inequality in (123) for a constant $K$ not depending on $N$.

$$h^{-5d}E[F_N^2] \leq K \int \left( \int K(u)K(v-u)K(w-u)K(t-u)dudw \right)^2 dv dw dt$$

By (119), (120), (121), (122) and (123), the quantities $h^{-bd}E[(E[F_N|w_n])^2]$, $h^{-7d}E[(E[F_N|w_n, w_i])^2]$, $h^{-4d}E[F_N^2]$, $h^{-6d}E[(E[F_N|w_n, w_i, w_j])^2]$ and $h^{-5d}E[F_N^2]$ are uniformly bounded, which concludes the proof of the Lemma.

**Proof of Corollary 3.1:** Since Theorem (3.2) implies $I_N(R) \xrightarrow{p} G(\theta^U)$ when $\Theta_0 \cap R = \{\theta^U(x)\}$, the first claim of the Corollary can be proved by showing that $\hat{\sigma}_C^{-1}(\theta^*)$ is a consistent estimator for the variance of $G(\theta^U)$. I begin by using standard arguments to show $||\theta^*(x) - \theta^U(x)||_{C^4} \xrightarrow{p} 0$. Let $M_N(\theta) = (Nh^2)^{-1}T_N(\theta)$ and $M(\theta) = E[(E[Y - \theta(X)|Z])^2 f_Z^2(Z)]$. For $P_N^2(C_N(\theta))$, $P_N^3(C_N(\theta))$ and $P_N^4(C_N(\theta))$ defined in the proof Theorem (3.2):

$$\sup_{\Theta \cap R} |M_N(\theta) - M(\theta)| \leq \sum_{i=1}^3 \sup_{\Theta} |P_N^i(C_N(\theta))| + \sup_{\Theta} |E[M_N(\theta) - M(\theta)]| \quad (124)$$

From (88) we know $\sup_{\Theta} |E[M_N(\theta) - M(\theta)]| \xrightarrow{p} 0$, while from (89), (90) and (91) we know that $\sup_{\Theta} |P_N^2(C_N(\theta))|$, $\sup_{\Theta} |P_N^3(C_N(\theta))|$ and $\sup_{\Theta} |P_N^4(C_N(\theta))|$ are all $o_p(1)$. Therefore, (124) implies $\sup_{\Theta \cap R} |M_N(\theta) - M(\theta)| \xrightarrow{p} 0$. In addition, since $\Theta \cap R$ is compact under $||\theta||_{C^4}$, $M_N(\theta)$ and $M(\theta)$ are continuous under $||\theta||_{C^4}$, $\hat{\theta}^U(x) = \arg \min_{\Theta \cap R} M(\theta)$ and $\theta^*(x) \in \arg \min M_N(\theta)$, it follows from standard arguments, such as for example Corollary 3.2.3 in van der Vaart and Wellner (1998), that $||\theta^* - \theta^U||_{C^4} \xrightarrow{p} 0$. Also note that $|\hat{\sigma}_C^{-1}(\theta^*) - \sigma_C^{-1}(\theta^U)| \leq \hat{\sigma}_C^{-1}(\theta^*)|\sigma_C^{-1}(\theta^U) - \sigma_C(\theta^*)| + |\sigma_C(\theta^*) - \sigma_C(\theta^*)| \xrightarrow{p} 0$ by Lemma (3.3), Corollary (1) and the continuity of $\sigma_C(\theta)$ under $||\theta||_{C^4}$. Therefore, Theorem (3.2) and the continuous mapping theorem imply:

$$\frac{1}{\hat{\sigma}_C(\theta^*)}I_N(R) \xrightarrow{c} \frac{1}{\sigma_C(\theta^U)} G(\theta^U) = N(0, 1) \quad (125)$$

55
where the last equality follows from Theorem (.1), which shows the asymptotic variance of $G(\theta^U)$ is given by $\sigma_C^2(\theta^U)$. This establishes the first claim of the Corollary. The second claim that $\tilde{\sigma}_C^{-1}(\theta^*)I_N(R) \xrightarrow{p} +\infty$ follows directly from Corollary (.1) and Theorem (3.2).

\[ \Box \]

APPENDIX D - Proof of Theorems 4.1 and 4.2

Proof of Theorem 4.1: Throughout the proof I will use the same notation as in the proof of Theorem (3.2). Suppose $\Theta_0 \cap R = \emptyset$. Since $\Theta_j \cap R \subseteq \Theta \cap R$ for all $j$, Theorem (3.2) implies $\inf_{\Theta_j \cap R} T_N(\theta) \geq \inf_{\Theta \cap R} T_N(\theta) \xrightarrow{p} +\infty$. Similarly, since by Corollary (.1) $\sup_{\Theta} \tilde{\sigma}_C(\theta) = O_p(1)$, it follows that $\tilde{\sigma}_C(\theta^*)T_N(\theta^*) \geq \inf_{\Theta} \tilde{\sigma}_C^{-1}(\theta)\inf_{\Theta \cap R} T_N(\theta) \xrightarrow{p} +\infty$.

Now suppose $\Theta_0 \cap R \neq \emptyset$. Since $\Theta_0 \cap R$ is compact under $|| \cdot ||_{C5}$, $N\tilde{\sigma}_C^2(\theta)$ is continuous under $|| \cdot ||_{C5}$, $\inf_{\Theta \cap R} N\tilde{\sigma}_C^2(\theta)$ is attained at a point we denote $\theta^*$. Furthermore, since $\Theta \cap R$ is compact under $|| \cdot ||_{C5}$ it is also compact under $|| \cdot ||_{L_2(\mathcal{X})}$. Therefore, $\Theta_j \cap R = \Theta_j \cap R \cap \Theta$ is compact under $|| \cdot ||_{L_2(\mathcal{X})}$, since $\Theta_j$ is closed under $|| \cdot ||_{L_2(\mathcal{X})}$. Let $\theta^*_p$ denote the $|| \cdot ||_{L_2(\mathcal{X})}$ projection of $\theta^*$ onto $\Theta_j \cap R$ and note $\theta^*_p \in \Theta_j \cap R$ by compactness. Using $\theta^*_p$, $\theta^*_p$ and $\inf_{\Theta \cap R} T_N(\theta) = \inf_{\Theta \cap R} T_N(\theta) + o_p(1)$ by Theorem (3.2), we get the first inequality in (126). Since $E[|Y - \theta^*(X)|^2] = 0$ because $\theta^* \in \Theta_0$, Jensen’s inequality and $f_\theta(z)$ being bounded imply $E[(|E[Y - \theta^*_p(X)|^2]|f_\theta^2(z))] \leq E[(|E[Y - \theta^*_p(X)|^2]|f_\theta(z))] \leq ||\theta^* - \theta^*_p||_{L_2(\mathcal{X})}^2$. Thus, since $\sup_{\Theta \cap R} \inf_{\Theta_j \cap R} ||\theta - \theta_j||_{L_2(\mathcal{X})}$ is finite under $N$ large enough, $\theta^*_p \in \Theta_0 \cap R$. Together with $\theta^* \in \Theta_0 \cap R$ and the Hoeffding decomposition of $C(\theta)$ this implies the third inequality in (126) for $\delta_n = \sup_{\Theta \cap R} \inf_{\Theta_j \cap R} ||\theta - \theta_j||_{L_2(\mathcal{X})}$. Since $\sup_{\Theta \cap R} N\tilde{\sigma}_C^2(\theta) p_n \propto 0$ from (99), $\sup_{||\theta - \theta_j||_{L_2(\mathcal{X})} < \delta_n} \tilde{\sigma}_C \propto 0$ from (96) and $\sup_{\Theta \cap R} \inf_{\Theta_j \cap R} (\Theta \cap R \cap \Theta)$ is continuous under $|| \cdot ||_{L_2(\mathcal{X})}$.

\begin{align*}
&\inf_{\Theta \cap R} T_N(\theta) = \inf_{\Theta_j \cap R} T_N(\theta) \geq \tilde{\sigma}_C^{-1}(\theta^*)N\tilde{\sigma}_C^2(\theta^*) + o_p(1) \geq \sup_{\Theta \cap R} \tilde{\sigma}_C(\theta)N\tilde{\sigma}_C^2(\theta^*) + o_p(1) \geq \sup_{\Theta \cap R} \tilde{\sigma}_C(\theta)N\tilde{\sigma}_C^2(\theta^*) + o_p(1) \xrightarrow{p} 0
\end{align*}

Since $\Theta_j \cap R \subseteq \Theta \cap R$ it follows that $\inf_{\Theta_j \cap R} T_N(\theta) \xrightarrow{p} 0$, and therefore (126) implies $\inf_{\Theta_j \cap R} T_N(\theta) = \inf_{\Theta \cap R} T_N(\theta) + o_p(1)$. In order to establish that $\tilde{\sigma}_C^{-1}(\theta^*)T_N(\theta^*) \xrightarrow{p} 0$ when $\Theta_0 \cap R \neq \emptyset$, I first show $||\theta^*_p - \theta^*_U||_{L_2(\mathcal{X})} \rightarrow 0$. Let $M(\theta) = E[(|E[Y - \theta(X)|^2]|f_\theta^2(z))]$ and $M(\theta) = (\tilde{\sigma}_C^2(\theta)^{-1}T_N(\theta)$. We begin in (127) by showing that $M(\theta)$ is continuous under $|| \cdot ||_{L_2(\mathcal{X})}$. The second inequality in (127) follows by the Cauchy-Schwarz inequality, Jensen’s inequality for conditional expectations and $f_\theta(z)$ being uniformly bounded. In addition, I have already shown the sets $\Theta \cap R$ and $\Theta_j \cap R$ are compact under $|| \cdot ||_{L_2(\mathcal{X})}$ and in (124) that $\sup_{\Theta \cap R} M(\theta) - M(\theta^*) \xrightarrow{p} 0$. Furthermore, by assumption, $\Theta_0 \cap R$ is a singleton so that $\theta^*_U$ is the unique minimizer of $M(\theta)$ in $\Theta \cap R$ and $\sup_{\Theta \cap R} \inf_{\Theta_j \cap R} ||\theta - \theta_j||_{L_2(\mathcal{X})} \rightarrow 0$. Therefore, by Lemma A.1 in Newey & Powell (2003) it follows that $||\theta^*_p - \theta^*_U||_{L_2(\mathcal{X})} \rightarrow 0$. Lemma A.1 requires continuity of $M(\theta)$ under $|| \cdot ||_{L_2(\mathcal{X})}$, which has not been shown. This is unimportant, however, since continuity is only necessary to ensure that the minimum is attained, which in this case is implied by continuity of $M(\theta)$ and compactness of $\Theta_j \cap R$ under $|| \cdot ||_{C5}$ instead of $|| \cdot ||_{L_2(\mathcal{X})}$. In (128) I show $\sigma_C^2(\theta)$ is continuous under $|| \cdot ||_{L_2(\mathcal{X})}$. For the first inequality use $f_\theta(z)$ and $E[(|Y - \theta(X)|^2)^2]Z$ being uniformly bounded in $\theta \in \Theta$. The second inequality follows from the Cauchy-Schwarz inequality for conditional expectations, Jensen’s inequality and $E[(|2Y + \theta_1(X) + \theta_2(X)|^2)Z]$ being bounded as a result of $\theta \in \Theta$ and $E[Y^2|Z]$.
being bounded (from Assumption 4 and 0 \in \Theta).

\[ |\sigma_C^2(\theta_1) - \sigma_C^2(\theta_2)| \leq E \left[ |E[(Y - \theta_1(X))^2|Z] - E[(Y - \theta_2(X))^2|Z]| \right] \lesssim ||\theta_1 - \theta_2||_{L^2(X)} \]  

(128)

Therefore, by Lemma (.3) and Corollary (1.1) it follows that |\sigma_C^{-1}(\theta_j) - \sigma_C^{-1}(\theta_{U})| \leq \sigma(C(\theta_j)) \sigma(C(\theta_{U})) (\sigma(C(\theta_j)) - \sigma(C(\theta_{U}))) \overset{P}{\rightarrow} 0. Combining this result with the continuous mapping theorem, (126) and Theorem (3.2) this implies \( \hat{\sigma}_C^{-1}(\theta_j)T_N(\theta_j) \overset{L}{\rightarrow} N(0,1) \), which concludes the proof of the Theorem. 

**Proof of Theorem 4.2:** The claim that \( P(\hat{I}_N(R) > \hat{q}_{1-\alpha}) \overset{P}{\rightarrow} \alpha \) when \( \Theta_0 \cap R \neq \emptyset \) follows immediately from Part (i) of Theorem 2.6.1 in Politis, Romano & Wolf (1999), Theorems (4.1) and (3.2), and the assumption that the 1 - \alpha quantile of \( \inf_{\theta \in \Theta \cap R} G(\theta) \) is continuous. In addition, (124) implies \( N\hat{\sigma}_N^{-1}I_N(\theta) \overset{P}{\rightarrow} E \left[ (E[Y - \theta(X)|Z])^2 f_2(Z) \right] \) in \( L^\infty(\Theta) \) and therefore by the continuous mapping theorem \( (N\hat{\sigma}_N^{-1})^{-1}I_N(R) \overset{P}{\rightarrow} \inf_{\theta \in \Theta \cap R} E \left[ (E[Y - \theta(X)|Z])^2 f_2(Z) \right] \). We now use this result to show that \( (N\hat{\sigma}_N^{-1})^{-1}I_N(R) \overset{P}{\rightarrow} \inf_{\theta \in \Theta \cap R} E \left[ (E[Y - \theta(X)|Z])^2 f_2(Z) \right] \). Since \( \Theta \cap R \) is compact and \( C_N(\theta) \) is continuous under \( ||\cdot||_{C5} \) the infimum in \( I_N(R) \) is attained at some point \( \theta^* \). By letting \( \theta_\alpha \) be the projection of \( \theta^* \) onto \( \Theta J \cap R \) under \( ||\cdot||_{\infty} \) we obtain the first inequality in (129). The second inequality in (129) follows from Lemma (.4) and the definition of \( C_N(\theta) \).

\[ (N\hat{\sigma}_N^{-1})^{-1}I_N(R) - (N\hat{\sigma}_N^{-1})^{-1}I_N(R) \geq C_N(\theta^*) - C_N(\theta_\alpha) \]

\[ \geq \frac{2}{N^2 \hat{\sigma}_N^{-2}} \sum_{n=1}^{N} \sum_{i=1}^{n} \sum_{j=1}^{i} J_N(w_n, w_i, w_j) \sup_{\theta \in \Theta \cap R} \inf_{\theta \in \Theta J \cap R} ||\theta - \theta_j||_{\infty} \]  

(129)

Lemma (.4) implies \( E \left[ \frac{2}{N^2 \hat{\sigma}_N^{-2}} \sum_{n=1}^{N} \sum_{i=1}^{n} \sum_{j=1}^{i} J_N(w_n, w_i, w_j) \right] = O(1) \), and therefore Markov’s inequality implies \( \frac{2}{N^2 \hat{\sigma}_N^{-2}} \sum_{n=1}^{N} \sum_{i=1}^{n} \sum_{j=1}^{i} J_N(w_n, w_i, w_j) = O_p(1) \). Furthermore, for every \( \theta \in \Theta J \cap R \), there exists by assumption a sequence \( \theta_j \in \Theta J \cap R \) such that \( ||\theta - \theta_j||_{L^2(X)} \rightarrow 0 \). Since \( \Theta \cap R \) is compact under \( ||\cdot||_{\infty} \) and hence \( \{\theta_j\} \) contains a subsequence \( \{\theta_{j_n}\} \) that is convergent under \( ||\cdot||_{\infty} \). Thus, since \( ||\theta - \theta_j||_{L^2(X)} \rightarrow 0 \), it must be that \( ||\theta - \theta_{j_n}||_{\infty} \rightarrow 0 \). Together with the assumption that \( \Theta J \subseteq \Theta J_{n+1} \), this implies that for every \( \theta \in \Theta \cap R \), \( \inf_{\theta \in \Theta J \cap R} ||\theta - \theta_j||_{\infty} \rightarrow 0 \). Since \( \Theta \cap R \) is compact under \( ||\cdot||_{\infty} \), the convergence is actually uniform and hence \( \sup_{\theta \in \Theta \cap R} \inf_{\theta \in \Theta J \cap R} ||\theta - \theta_j||_{\infty} \rightarrow 0 \), which together with (129) implies that \( (N\hat{\sigma}_N^{-1})^{-1}I_N(R) - (N\hat{\sigma}_N^{-1})^{-1}I_N(R) \geq C_N(\theta^*) - C_N(\theta_\alpha) \overset{P}{\rightarrow} 0 \). On the other hand, since \( \Theta J \cap R \subseteq \Theta \cap R \), \( (N\hat{\sigma}_N^{-1})^{-1}I_N(R) - (N\hat{\sigma}_N^{-1})^{-1}I_N(R) \leq 0 \), and thus we derive the first equality in (130).

\[ (N\hat{\sigma}_N^{-1})^{-1}I_N(R) = (N\hat{\sigma}_N^{-1})^{-1}I_N(R) + o_p(1) \overset{P}{\rightarrow} E \left[ (E[Y - \theta(X)|Z])^2 f_2(Z) \right] \]  

(130)

The derivation in (130) concludes verifying the conditions for part (ii) of Theorem 2.6.1 in Politis, Romano & Wolf (1999), and thus we conclude that if \( \Theta_0 \cap R = \emptyset \) then \( P(\hat{I}_N(R) > \hat{q}_{1-\alpha}) \overset{P}{\rightarrow} 1 \).
## APPENDIX E - Tables

Table 5: Sensitivity Analysis

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>$b_N$</th>
<th>$h$</th>
<th>$h_{bg}$</th>
<th>$J_N$</th>
<th>$I_N(R)$</th>
<th>$F$</th>
<th>[H_N]</th>
<th>[H_{bg}]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi_G(\bar{X}) \geq 11.0%$</td>
<td>70</td>
<td>0.03</td>
<td>0.34</td>
<td>3</td>
<td>-0.0016</td>
<td>0.10</td>
<td>6%</td>
<td>2%</td>
</tr>
<tr>
<td>$\phi_G(\bar{X}) \geq 11.6%$</td>
<td>70</td>
<td>0.03</td>
<td>0.34</td>
<td>3</td>
<td>-0.0014</td>
<td>0.13</td>
<td>2%</td>
<td>0%</td>
</tr>
<tr>
<td>$\phi_G(\bar{X}) \geq 12.2%$</td>
<td>70</td>
<td>0.03</td>
<td>0.34</td>
<td>3</td>
<td>0.0002</td>
<td>0.95</td>
<td>2%</td>
<td>0%</td>
</tr>
<tr>
<td>$\phi_G(\bar{X}) \geq 11.0%$</td>
<td>100</td>
<td>0.03</td>
<td>0.28</td>
<td>3</td>
<td>-0.0016</td>
<td>0.04</td>
<td>2%</td>
<td>0%</td>
</tr>
<tr>
<td>$\phi_G(\bar{X}) \geq 11.6%$</td>
<td>100</td>
<td>0.03</td>
<td>0.28</td>
<td>3</td>
<td>-0.0014</td>
<td>0.09</td>
<td>2%</td>
<td>0%</td>
</tr>
<tr>
<td>$\phi_G(\bar{X}) \geq 12.2%$</td>
<td>100</td>
<td>0.03</td>
<td>0.28</td>
<td>3</td>
<td>0.0002</td>
<td>0.91</td>
<td>2%</td>
<td>0%</td>
</tr>
<tr>
<td>$\phi_G(\bar{X}) \geq 11.0%$</td>
<td>130</td>
<td>0.03</td>
<td>0.24</td>
<td>3</td>
<td>-0.0016</td>
<td>0.03</td>
<td>2%</td>
<td>0%</td>
</tr>
<tr>
<td>$\phi_G(\bar{X}) \geq 11.6%$</td>
<td>130</td>
<td>0.03</td>
<td>0.24</td>
<td>3</td>
<td>-0.0014</td>
<td>0.06</td>
<td>2%</td>
<td>0%</td>
</tr>
<tr>
<td>$\phi_G(\bar{X}) \geq 12.2%$</td>
<td>130</td>
<td>0.03</td>
<td>0.24</td>
<td>3</td>
<td>0.0002</td>
<td>0.90</td>
<td>2%</td>
<td>0%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>$b_N$</th>
<th>$h$</th>
<th>$h_{bg}$</th>
<th>$J_N$</th>
<th>$I_N(R)$</th>
<th>$F$</th>
<th>[H_N]</th>
<th>[H_{bg}]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi_E(\bar{X}) \geq 11.0%$</td>
<td>30</td>
<td>0.1</td>
<td>0.48</td>
<td>2</td>
<td>-0.0011</td>
<td>0.36</td>
<td>2%</td>
<td>0%</td>
</tr>
<tr>
<td>$\phi_E(\bar{X}) \geq 13.0%$</td>
<td>30</td>
<td>0.1</td>
<td>0.48</td>
<td>2</td>
<td>0.0007</td>
<td>0.86</td>
<td>2%</td>
<td>0%</td>
</tr>
<tr>
<td>$\phi_E(\bar{X}) \geq 15.0%$</td>
<td>30</td>
<td>0.1</td>
<td>0.48</td>
<td>2</td>
<td>0.0058</td>
<td>0.97</td>
<td>2%</td>
<td>0%</td>
</tr>
<tr>
<td>$\phi_E(\bar{X}) \geq 11.0%$</td>
<td>70</td>
<td>0.1</td>
<td>0.30</td>
<td>2</td>
<td>-0.0011</td>
<td>0.36</td>
<td>2%</td>
<td>0%</td>
</tr>
<tr>
<td>$\phi_E(\bar{X}) \geq 13.0%$</td>
<td>70</td>
<td>0.1</td>
<td>0.30</td>
<td>2</td>
<td>0.0007</td>
<td>0.72</td>
<td>2%</td>
<td>0%</td>
</tr>
<tr>
<td>$\phi_E(\bar{X}) \geq 15.0%$</td>
<td>70</td>
<td>0.1</td>
<td>0.30</td>
<td>2</td>
<td>0.0058</td>
<td>0.73</td>
<td>2%</td>
<td>0%</td>
</tr>
<tr>
<td>$\phi_E(\bar{X}) \geq 11.0%$</td>
<td>130</td>
<td>0.2</td>
<td>0.96</td>
<td>2</td>
<td>-0.0012</td>
<td>0.28</td>
<td>2%</td>
<td>0%</td>
</tr>
<tr>
<td>$\phi_E(\bar{X}) \geq 13.0%$</td>
<td>130</td>
<td>0.2</td>
<td>0.96</td>
<td>2</td>
<td>0.0008</td>
<td>0.96</td>
<td>2%</td>
<td>0%</td>
</tr>
<tr>
<td>$\phi_E(\bar{X}) \geq 15.0%$</td>
<td>130</td>
<td>0.2</td>
<td>0.96</td>
<td>2</td>
<td>0.0058</td>
<td>1.00</td>
<td>2%</td>
<td>0%</td>
</tr>
<tr>
<td>$\phi_E(\bar{X}) \geq 11.0%$</td>
<td>50</td>
<td>0.2</td>
<td>0.72</td>
<td>2</td>
<td>-0.0012</td>
<td>0.11</td>
<td>2%</td>
<td>0%</td>
</tr>
<tr>
<td>$\phi_E(\bar{X}) \geq 13.0%$</td>
<td>50</td>
<td>0.2</td>
<td>0.72</td>
<td>2</td>
<td>0.0008</td>
<td>0.80</td>
<td>2%</td>
<td>0%</td>
</tr>
<tr>
<td>$\phi_E(\bar{X}) \geq 15.0%$</td>
<td>50</td>
<td>0.2</td>
<td>0.72</td>
<td>2</td>
<td>0.0058</td>
<td>0.89</td>
<td>2%</td>
<td>0%</td>
</tr>
<tr>
<td>$\phi_E(\bar{X}) \geq 11.0%$</td>
<td>70</td>
<td>0.2</td>
<td>0.59</td>
<td>2</td>
<td>-0.0012</td>
<td>0.11</td>
<td>2%</td>
<td>0%</td>
</tr>
<tr>
<td>$\phi_E(\bar{X}) \geq 13.0%$</td>
<td>70</td>
<td>0.2</td>
<td>0.59</td>
<td>2</td>
<td>0.0008</td>
<td>0.77</td>
<td>2%</td>
<td>0%</td>
</tr>
<tr>
<td>$\phi_E(\bar{X}) \geq 15.0%$</td>
<td>70</td>
<td>0.2</td>
<td>0.59</td>
<td>2</td>
<td>0.0058</td>
<td>0.72</td>
<td>2%</td>
<td>0%</td>
</tr>
<tr>
<td>$\phi_E(\bar{X}) \geq 11.0%$</td>
<td>30</td>
<td>0.3</td>
<td>1.44</td>
<td>2</td>
<td>-0.0012</td>
<td>0.76</td>
<td>2%</td>
<td>0%</td>
</tr>
<tr>
<td>$\phi_E(\bar{X}) \geq 13.0%$</td>
<td>30</td>
<td>0.3</td>
<td>1.44</td>
<td>2</td>
<td>0.0008</td>
<td>0.99</td>
<td>2%</td>
<td>0%</td>
</tr>
<tr>
<td>$\phi_E(\bar{X}) \geq 15.0%$</td>
<td>30</td>
<td>0.3</td>
<td>1.44</td>
<td>2</td>
<td>0.0063</td>
<td>1.00</td>
<td>2%</td>
<td>0%</td>
</tr>
<tr>
<td>$\phi_E(\bar{X}) \geq 11.0%$</td>
<td>50</td>
<td>0.3</td>
<td>1.08</td>
<td>2</td>
<td>-0.0012</td>
<td>0.10</td>
<td>2%</td>
<td>0%</td>
</tr>
<tr>
<td>$\phi_E(\bar{X}) \geq 13.0%$</td>
<td>50</td>
<td>0.3</td>
<td>1.08</td>
<td>2</td>
<td>0.0008</td>
<td>0.87</td>
<td>2%</td>
<td>0%</td>
</tr>
<tr>
<td>$\phi_E(\bar{X}) \geq 15.0%$</td>
<td>50</td>
<td>0.3</td>
<td>1.08</td>
<td>2</td>
<td>0.0063</td>
<td>0.97</td>
<td>2%</td>
<td>0%</td>
</tr>
<tr>
<td>$\phi_E(\bar{X}) \geq 11.0%$</td>
<td>70</td>
<td>0.3</td>
<td>0.89</td>
<td>2</td>
<td>-0.0012</td>
<td>0.16</td>
<td>2%</td>
<td>0%</td>
</tr>
<tr>
<td>$\phi_E(\bar{X}) \geq 13.0%$</td>
<td>70</td>
<td>0.3</td>
<td>0.89</td>
<td>2</td>
<td>0.0008</td>
<td>0.80</td>
<td>2%</td>
<td>0%</td>
</tr>
<tr>
<td>$\phi_E(\bar{X}) \geq 15.0%$</td>
<td>70</td>
<td>0.3</td>
<td>0.89</td>
<td>2</td>
<td>0.0063</td>
<td>0.85</td>
<td>2%</td>
<td>0%</td>
</tr>
</tbody>
</table>
References


