Inference in Non-Parametric/Semi-Parametric Moment Equality Models with Shape Restrictions

Yu Zhu*

July 30, 2019

Abstract

This paper studies the inference problem of an infinite-dimensional parameter with a shape restriction. This parameter is identified by arbitrarily many unconditional moment equalities. The shape restriction leads to a convex restriction set. I propose a test of the shape restriction, which controls size uniformly and applies to both point- and partially identified models. The test can be inverted to construct confidence sets after imposing the shape restriction. Monte Carlo experiments show the finite-sample properties of this method. In an empirical illustration, I apply the method to ascending auctions held by the US Forest Service and show that imposing shape restrictions can significantly improve inference.

JEL codes: C12, C14

Key words: Non-Parametric/Semi-Parametric Models, Partial Identification, Shape Restrictions, Unconditional Moments

*Bank of Canada, 234 Wellington Street, Ottawa, Ontario, Canada. Email: zhuyuzl57@gmail.com. I am very grateful to Jack Porter, Amit Gandhi and Xiaoxia Shi for their advice and guidance. I am also very grateful to the co-editor Andres Santos and the two anonymous referees for their insightful comments and suggestions that greatly improve the paper. I would like to thank Jason Allen, Andrés Aradillas-López, Geoffrey Dunbar, Serafin Grundl, Bruce Hansen, and Jonathan Hill for their very helpful comments. The views expressed in this paper are solely those of the author and may differ from official Bank of Canada views. No responsibility for them should be attributed to the Bank.
1 Introduction

Economic theories often imply shape restrictions. For example, utility functions are weakly concave in consumption, and demand functions are weakly decreasing in price. These shape restrictions can be useful in empirical studies in two ways. First, they can be used to test competing theories. Second, they provide information that can improve inference on underlying unknown parameters. The latter is especially relevant if the unknown parameters contain functions and a researcher, concerned with misspecification, is unwilling to impose parametric assumptions. In such a case, imposing shape restrictions can lead to significantly smaller confidence sets.

This paper studies an inference problem in which an infinite-dimensional unknown parameter is point-identified or partially identified by unconditional moment equalities. A researcher would like to test whether the unknown parameter satisfies a shape restriction and/or to construct confidence sets after imposing the shape restriction. The unconditional moments may be nonlinear and their sample analog may be discontinuous. The number of the unconditional moments can be finite, countably infinite, or a continuum. The shape restriction may involve a large number of inequalities and the parameters that satisfy the shape restriction form a convex set. Many commonly used shape restrictions such as non-negativity, monotonicity, concavity, first-order and second-order stochastic dominance satisfy these requirements.

This framework has many applications, and it accommodates conditional and/or unconditional moment equality and/or inequality models and, in particular, indirect inference in structural estimation. For example, in labor economics, a researcher may want to test whether an individual’s reservation wage is decreasing in spells of unemployment (Gutknecht (2016)). In the inference problem of the shape-invariant Engel curve studied in Blundell, Chen, and Kristensen (2007), one may want to test, for example, whether the Engel curve of food is weakly decreasing. Or one may want to impose that the Engel curve is weakly decreasing and construct its confidence set. In first-price auctions, Zhu and Grundl (2014) show that under a weak exclusion condition, inference of bidders’ risk aversion is equivalent to testing a first-order stochastic dominance relation. In ascending auctions, one may want to infer
valuation distributions with a first-order stochastic dominance restriction that arises from asymmetry in bidders. In these examples, the unknown parameters are identified either by conditional moments (the first two examples), or by a continuum of unconditional moments obtained from indirect inference methods (the last two examples).

In this paper, I propose a test for shape restrictions. It can be inverted to construct confidence sets. The test statistic is obtained by minimizing the integrated squared sample moments subject to the shape restriction. The challenge is that the null distribution of the test statistic depends on a minimization problem taken over an unknown local parameter space. A naive estimator of the local parameter space may over-estimate its size and lead to underestimation of the critical value. To obtain the critical value, I propose a bootstrap procedure that estimates the local parameter space via rescaled sample moments. The main insight is that this amounts to estimating the true local parameter space by shrinking it towards 0. If the restriction set is convex, the true local parameter is convex and contains 0. Therefore, it contains the estimated local parameter space. This ensures that the bootstrap statistic does not over-estimate the local parameter space and validates the test. This method is related to the moment inequality literature where the parameter of interest is finite-dimensional such as Andrews and Soares (2010); Pakes, Porter, Ho, and Ishii (2011); Andrews and Shi (2013); and, in particular, Gandhi, Lu, and Shi (2012), but the insight is different.

My test applies to both point-identified models and partially identified models, which is a desirable feature because, in many cases, moment conditions only partially identify the unknown parameters.\footnote{For example, Canay, Santos, and Shaikh (2013) show that in the non-parametric IV regression, the completeness assumption, which is sufficient and necessary for point identification, is not testable. In second-price auction models with incomplete data, the underlying parameter of interest is only partially identified. Finally, in moment inequality models, partial identification arises naturally.} One limitation of the paper is that the theory does not apply to problems with weak identification. To allow for weak identification is important but more challenging.

In an empirical illustration, I apply my method to infer the marginal valuation distributions of bidders in ascending auctions of timber held by the US Forest Service (USFS). In these auctions, bidders are asymmetric: big firms and firms with manufacturing capacity tend to value timber more than small firms and firms without manufacturing capacity. The
valuations of all bidders in an auction are correlated. Because of the auction format, bids other than the transaction price (the second highest bid) may not correctly reflect bidders’ valuations. Partial identification prevails in this setup. Therefore, it fits my method perfectly. Adapting the indirect inference method developed in Bierens and Song (2012) to ascending auctions, I match the model-predicted joint distribution of the transaction price and the winner’s identity with the data. This leads to a continuum of unconditional moments that identify the unknown parameter. The key insight is that the asymmetry in bidders implies that the valuation distribution of big firms first-order stochastically dominates (FOSD) that of the small firms. Results show that imposing the FOSD restriction greatly tightens the confidence set and helps to exclude the valuation distribution obtained by incorrectly assuming that bidders are symmetric and have independent valuations. To the best of my knowledge, the inference procedure is new in ascending auctions.

This paper contributes to the growing literature on inference with shape restrictions. It is closely related to Chernozhukov, Newey, and Santos (2015) (CNS hereafter), which provides a general treatment of inference problems with shape restrictions and partial identification. Different from their method, I use a bootstrap statistic that relies on the convexity of the restriction set. As a result, in partially identified models, my test can be faster to compute. Also, unlike CNS, my test can be applied to indirect inference problems with a continuum of moments.

Other recent work in this area includes Chetverikov (2012); Chetverikov and Wilhelm (2017); Freyberger and Horowitz (2015); Freyberger and Reeves (2017); Gutknecht (2016); and Horowitz and Lee (2017). All of these papers focus on non-parametric regression or non-parametric IV regression models, except Freyberger and Reeves (2017), which considers a general setup but requires point identification. This paper also generalizes Hong (2017) and Santos (2012) by allowing for shape restrictions defined by inequalities.

This paper also contributes to the growing literature on semi-parametric/non-parametric conditional moment models. Recent work includes Newey, Powell, and Vella (1999); Newey and Powell (2003); Ai and Chen (2003); Hall and Horowitz (2005); Blundell, Chen, and Kristensen (2007); Darolles, Fan, Florens, and Renault (2011); Chen, Chernozhukov, Lee, and Newey (2014); Chen and Pouzo (2009, 2012, 2015) and Tao (2014).
The paper is organized as follows. Section 2 introduces the econometric framework. Section 3 provides a detailed guide on how to conduct the test and how to obtain confidence sets. Section 4 provides a heuristic illustration of the method. Section 5 lays out the theory. Readers who are less interested in the technical details can skip this section. Section 6 presents Monte Carlo experiments to evaluate the finite-sample performance of the method. Section 7 applies the method to ascending auctions using timber auctions. Section 8 concludes. Technical proofs are collected in the Appendix.

2 Setup and Examples

The parameter of interest $\theta_F$ is infinite-dimensional and lives in a known parameter space $\Theta$. Denote $W$ as a $D_W$-dimensional random vector, which takes values in $W \subseteq \mathbb{R}^{D_W}$ and has a distribution $F$. The unknown parameter $\theta_F$ satisfies moment conditions:

$$\mathbb{E}_F \rho_t (W, \theta_F) = 0 \quad \forall t \in T,$$

where $\rho_t (\cdot, \cdot) : \mathbb{R}^{D_W} \times \Theta \to \mathbb{R}$ is a real function indexed by $t$, and $T$ is an index set. $\mathbb{E}_F$ is the expectation taken under the distribution $F$. The cardinality of $T$ can be finite, countably infinite, or a continuum. The moment function $\rho_t$ may depend on $\theta$ only through its value at data points as in Ai and Chen (2003) and Chen and Pouzo (2012), i.e., $\rho_t (W, \theta) = \rho_t (W, \theta (W))$. Or it may depend on $\theta$ as a whole. The latter often arises in indirect inference of structural models. For an example in first-price auctions, see Bierens and Song (2012).

Even in simple models nested in this framework, partial identification may arise and it is not possible to test partial identification uniformly.\footnote{See Santos (2012) for an example of partial identification in the non-parametric IV case and Canay, Santos, and Shaikh (2013) for an impossibility result on the testability of identification.} It is therefore desirable to have an approach robust to partial identification. To this end, define the identified set $\Theta_F$ as the collection of all $\theta$s which satisfy the moment conditions under $F$, i.e.,

$$\Theta_F = \{ \theta \in \Theta : \mathbb{E}_F \rho_t (W, \theta) = 0, \forall t \in T \}.$$
Notice that $\Theta_F$ is allowed to be a singleton, in which case $\theta_F$ is point-identified.

A researcher may want to test whether $\theta_F$ satisfies some shape restriction or to impose the shape restriction and draw an inference on $\phi (\theta_F)$, where $\phi : \Theta \to \mathbb{R}^{D_\phi}$ is a linear functional and $D_\phi$ is a positive integer. This paper focuses on the case where the shape restriction is defined by a large number of inequalities. In other words, the shape restriction defines a restriction set

$$ R = \{ \theta : l(\theta) = a_l \ \forall l \in L, \ \psi(\theta) \geq 0, \ \forall \psi \in \Psi \} , \quad (1) $$

where $L$ is a collection of linear functionals, $a_l \in \mathbb{R}$ is a number associated with the linear functional $l$, and $\Psi$ is a collection of functionals. Both $L$ and $\Psi$ can have finitely or infinitely many elements and $R$ is a convex set. To accommodate partial identification, the testing problem is

$$ H_0 : \Theta_F \cap R \neq \emptyset $$
$$ H_1 : \Theta_F \cap R = \emptyset . \quad (2) $$

This problem is a generalization of Hong (2017) and Santos (2012), who focus on $R$ defined by equalities of linear operators. A key difference is that if $R$ is defined by inequalities, it has boundary points. Therefore, the inference procedure must account for the possibility that $\theta_F$ is close to or at the boundary of $R$. The formulation of $R$ covers two inference problems: testing the shape restriction and constructing confidence sets of linear functionals with the shape restriction.

**Example 2.1** (Non-parametric IV Regression). Consider the model

$$ Y = \theta_F (X) + \epsilon , $$

where $\mathbb{E}_F [\epsilon | X] \neq 0$. Let $Z$ be a set of instruments and $\mathbb{E}_F [\epsilon | Z] = 0$. Here $X$ and $Z$ are $D_X$- and $D_Z$-dimensional real random vectors. The goal is to test whether $\theta_F (x)$ is weakly increasing in the first element of $x$, where $x$ is a $D_X$-dimensional vector and its $k$-th element is denoted by $x_k$. Many economic applications fit into this class. One example is to test whether an individual’s reservation wage is decreasing in spells of unemployment, e.g.,
Gutknecht (2016). The identified set is defined by the conditional moment equality

\[ \Theta_F = \left\{ \theta \in \Theta : \mathbb{E}_F [Y - \theta(X) | Z] = 0 \right\}. \]

Because of the well-known ill-posed inverse problem, \( \Theta_F \) might not be a singleton. The restriction set is \( R = \{ \theta : \frac{\partial \theta(x)}{\partial x_1} \geq 0, \forall x \} \). It is convex and defined by a continuum of linear functionals: \( \psi_x \theta = \frac{\partial \theta(x)}{\partial x_1}, \forall x \). Following Bierens (1990) and Stinchcombe and White (1998), the conditional moment is equivalent to a continuum of unconditional moments with a set of weight functions \( w(\cdot, t) \):

\[ \Theta_F = \left\{ \theta \in \Theta : \mathbb{E}_F [Y - \theta(X)] w(Z, t) = 0, \forall t \in T \right\}, \]

where \( T \subset \mathbb{R}^{D_Z} \) has a positive Lebesgue measure. Therefore, \( W = (X, Y, Z) \) and \( \rho_t (W, \theta) = [Y - \theta(X)] w(Z, t) \). One may also want to impose monotonicity to infer \( \theta_F(x) \) for some \( x \). Then the linear functional \( \phi(\theta) = \theta(x) \). Shape restrictions can greatly help inference in this model. Chetverikov and Wilhelm (2017) show that monotonicity can improve rates of convergence. If the model is partially identified, shape restrictions can also dramatically reduce the identified set. In extreme cases, moment conditions alone may not provide any information. But with shape restrictions, point identification can be achieved. Section B in the Appendix provides such an example. Also notice that the same testing problem but with conditional quantile restrictions also fits into my framework.

**Example 2.2** (Shape-Invariant Engel Curve). Blundell, Chen, and Kristensen (2007) study semi-nonparametric estimation of the shape-invariant Engel curve. In this estimation problem, a researcher observes \( W = (Y_1, Y_2, X, Z) \) from a household, where \( Y_1 \) is the fraction of expenditure on a particular good, say food; \( Y_2 \) is the total expenditure; \( X \) is a vector of household characteristics; and \( Z \) is the total income. He is interested in estimating

\[ Y_1 = h(Y_2 - X\beta) + \epsilon, \]

where \( h \) is the shape-invariant Engel curve and is specified non-parametrically. The total expenditure \( Y_2 \) may be endogenous while the total income \( Z \) serves as an instrument, i.e.,
$\mathbb{E}_F (\epsilon | Y_2, X) \neq 0$ while $\mathbb{E}_F (\epsilon | Z, X) = 0$. In this case, $\theta = (\beta, h)$ and one can transform the conditional moments to a continuum of unconditional moments as in the previous example. The researcher may want to test whether the fraction of expenditure on a particular good is decreasing in total expenditure, or may want to conduct an inference on $h$ imposing this decreasing assumption. For example, in advanced economy like the UK, it is natural to think that the fraction of expenditure on food decreases as total expenditure increases. In this case, the restriction set is $R = \{ \theta : h' (x) \geq 0, \forall x \}$, which is a convex set.

**Example 2.3** (Asymmetric Ascending Auctions). In an ascending auction, bidders cry out their bids in ascending order until only one bidder remains. The remaining bidder wins the good and pays his last bid. The auction outcome is equivalent to a sealed bid second-price auction, where the bidder with the highest valuation wins the good and the transaction price is the second highest bid. Consider the model where bidders are asymmetric and have correlated private valuations. Bidders are divided into strong bidders and weak bidders. Let $F_s (v)$ be the marginal valuation distribution of strong bidders and $F_w (v)$ be the marginal valuation distribution of weak bidders. Strong bidders are more likely to have higher valuations for the auctioned object than weak bidders, i.e., $F_s (v) \leq F_w (v)$ for every $v \in \mathbb{R}$. Correlation among bidders’ valuations is modeled by a Gaussian Copula function. For simplicity, I focus only on auctions with one strong and one weak bidder. Let $\Phi_2 (\cdot, \cdot, r)$ be a 2-dimensional normal distribution function with mean 0, variance 1, and correlation $r$. And let $\Phi$ be a standard normal distribution function. Then in an auction, the joint distribution of valuations is $C (F_s (\cdot), F_w (\cdot), r)$, where $C (x_1, x_2, r) = \Phi_2 (\Phi^{-1} (x_1), \Phi^{-1} (x_2), r)$. The primitive of this model is $\theta = (F_s, F_w, r)$.

For each auction, a researcher observes the transaction price $Y$, which is equal to the second highest valuation. He also observes a variable $X$, which takes value 1 if the winner is a strong bidder and 0 otherwise. He is interested in constructing a conference band for $F_s$ and wants to impose the shape restriction $F_s (v) \leq F_w (v)$ for every $v \in \mathbb{R}$ to improve inference.

One can fit this inference problem into my framework using the indirect inference method. The idea is that, at the true parameter, the joint distribution of the transaction price and the winner’s type predicted by the model should be close to its empirical counterpart. Fol-
following Bierens and Song (2012), one can use characteristic functions to match these two
distributions. Then, model restrictions reduce to $\forall t \in T$

\[
0 = E_F \sin (tY) \mathbf{1}(X = 1) - E \left[ \sin \left( t \tilde{Y} \right) \mathbf{1}(\tilde{X} = 1) | \theta \right], \\
0 = E_F \sin (tY) \mathbf{1}(X = 0) - E \left[ \sin \left( t \tilde{Y} \right) \mathbf{1}(\tilde{X} = 0) | \theta \right], \\
0 = E_F \cos (tY) \mathbf{1}(X = 1) - E \left[ \cos \left( t \tilde{Y} \right) \mathbf{1}(\tilde{X} = 1) | \theta \right], \\
0 = E_F \cos (tY) \mathbf{1}(X = 0) - E \left[ \cos \left( t \tilde{Y} \right) \mathbf{1}(\tilde{X} = 0) | \theta \right],
\]

where $T$ is an interval on the real line that contains 0; $\tilde{X}$ and $\tilde{Y}$ are the winner’s type and
the transaction price generated by the model under parameter $\theta$. The restriction set is $R = \{ \theta : F_s (v) \leq F_w (v) \text{ for all } v \in \mathbb{R} \}$, which is a convex set. It is well-known that in this
setup, $F_s$ is only partially identified. See Aradillas-López, Gandhi, and Quint (2013); Coey, Larsen, Sweeney, and Waisman (2017); Komarova (2013). In Section 7, I apply this method
to USFS timber auctions.

There are many other applications that fit into my framework. They include the inference
of risk aversion in first-price auctions considered in Zhu and Grundl (2014) and Examples
2.2-2.4 in CNS. In particular, my framework also allows for conditional moment inequalities
following the strategy used in Examples 2.3 and 2.4 in CNS.

3 Inference Procedure

The researcher observes an independently identically distributed (i.i.d.) random sample
$\{W_i\}_{i=1}^n$ from $F$. Denote the probability measure that induces $F$ by $P_F$.\(^4\)

\(^3\)Unlike in Aradillas-López, Gandhi, and Quint (2013) and Coey, Larsen, Sweeney, and Waisman (2017),
variation in the number of bidders is not used in this example, and the focus is on the marginal valuation
distribution.

\(^4\)Formally, $\{W_i\}_{i=1}^n$ is the first $n$ coordinate of $\{W_i\}_{i=1}^\infty$, which is the identity mapping on the infinite
product probability space $(\mathcal{W}^\infty, \mathcal{S}^\infty, P_F^\infty)$. Here $\mathcal{S}^\infty$ is the product $\sigma$-algebra and $P_F^\infty$ is the product measure.
This paper relies heavily on convergence of empirical processes. It is understood that if an empirical process
is not measurable, convergence is under the outer probability on $(\mathcal{W}^\infty, \mathcal{S}^\infty, P_F^\infty)$.
The Test Statistic  Let $\mu (t)$ be a probability measure on $T$ such that $\mathbb{E}_F \rho_t (W_i, \theta) = 0$ for all $t \in T$ if and only if

$$Q_F (\theta) = \int_T [\mathbb{E}_F \rho_t (W_i, \theta)]^2 d\mu (t) = 0.$$  

Throughout the paper, I assume that such $\mu$ exists. If $T$ contains finitely many or countably infinitely many elements, it is easy to find such a $\mu$. If $T$ is a closed rectangular in an Euclidean space and $\mathbb{E}_F \rho_t (W_i, \theta)$ is continuous in $t$, $\mu$ can be the uniform probability measure on $T$.

Ideally, one can compute the infimum (inf) of the criterion $Q_F (\theta)$ on the restriction set $R$ and reject the null hypothesis if it is not 0. Because $Q_F (\theta)$ is unknown, replace it with its sample analog

$$Q_n (\theta) = \int_T \left[ \frac{1}{n} \sum_{i=1}^n \rho_t (W_i, \theta) \right]^2 d\mu (t). \quad (3)$$

The test statistic for the restriction set $R$ is

$$T_n (R) = \inf_{\theta \in \Theta_n \cap R} nQ_n (\theta), \quad (4)$$

where $\{\Theta_n\}_{n=1}^\infty$ is a sequence of increasing sieve spaces that approximate $\Theta$ well enough. Here inf is used to allow for potentially discontinuous $Q_n (\theta)$, which arises in, for example, IV quantile regression models. Also define

$$\hat{\Theta}_n (R) = \arg \inf_{\theta \in \Theta_n \cap R} nQ_n (\theta). \quad (5)$$

Reject the null hypothesis if $T_n (R)$ exceeds some critical value. Although $T_n (R)$ is based on a CVM-type criterion function, other criterion functions may also be used. One possibility is the Kolmogorov-Smirnov (KS) criterion, which is used in Hong (2017) and Santos (2012). Theories developed in this paper carry naturally over to the case with the KS criterion. This paper focuses on CVM because it is easier to compute.\textsuperscript{5}

\textsuperscript{5}If $\rho_t$ is linear in $\theta$, one only needs to solve a quadratic programming problem for many shape restrictions if the criterion function is of the CVM-type. But a KS-type criterion function requires solving a complicated nonlinear optimization problem. In general, the CVM-type criterion tends to have fewer local optimums. This computational difference matters especially because I use a bootstrap method to obtain critical values.
Bootstrap Critical Value  This paper develops a bootstrap method to obtain the critical value. Define
\[ G_{n,F}(\theta,t) = n^{-1/2} \sum_{i=1}^{n} \left[ \rho_t(W_i,\theta) - \mathbb{E}_F \rho_t(W_i,\theta) \right]. \]

Let \( \{W_i^*\}_{i=1}^{n} \) be a sequence of i.i.d. draws from the empirical distribution and
\[ G_n^*(\theta,t) = n^{-1/2} \sum_{i=1}^{n} \left[ \rho_t(W_i^*,\theta) - \frac{1}{n} \sum_{i=1}^{n} \rho_t(W_i,\theta) \right], \]
The bootstrap statistic is defined as
\[ T_n^*(R) = \inf_{(\gamma_n,\lambda_n) \in I_n} nQ_n^*(\theta_n^*(\gamma_n,\lambda_n,R),\gamma_n,0), \tag{6} \]
where
\[
nQ_n^*(\theta,\gamma_n,\lambda_n) = \int_T \left[ G_n^*(\theta,t) + n^{-1/2} \gamma_n \sum_{i=1}^{n} \rho_t(W_i,\theta) \right]^2 d\mu(t) + \lambda_n Q_n(\theta), \tag{7} \]
\[ \theta_n^*(\gamma_n,\lambda_n,R) \in \hat{\Theta}_n^*(\gamma_n,\lambda_n,R) = \arg\inf_{\theta \in \Theta_n \cap R} nQ_n^*(\theta,\gamma_n,\lambda_n), \tag{8} \]
\[ I_n \subseteq \bar{I}_n = \{ (\gamma,\lambda) \in \mathbb{R}^2 | n\gamma^2 + \lambda = \kappa_n, \lambda \geq 0, \gamma \geq 0 \}. \tag{9} \]

If \( \hat{\Theta}_n^*(\gamma_n,\lambda_n,R) \) contains more than one point, one can assign any of them to \( \theta_n^*(\gamma_n,\lambda_n,R) \).\(^6\)
Here \( I_n \) can be any non-empty subset of \( \bar{I}_n \), and \( \kappa_n \) is a tuning parameter which diverges at a rate slower than \( n \).\(^7\) One may set \( \kappa_n = n/\ln n \). Reject the null hypothesis if \( T_n(R) > C_n^*(1-\alpha + \eta,R) + \eta \), where \( \eta \) is an arbitrarily small positive number and \( C_n^*(1-\alpha + \eta,R) \) is the \( 1-\alpha+\eta \)-th quantile of the distribution of \( T_n^*(R) \) evaluated at the empirical distribution.

One can set \( \eta \) to a small positive number such as \( 10^{-8} \). Under certain conditions discussed later, it can be set to 0.

Under a proper \( \kappa_n \), any non-empty \( I_n \subseteq \bar{I}_n \) leads to a valid critical value. A natural question is how to pick \( I_n \). This paper recommends using \( I_n = \bar{I}_n \) for computationally simple problems and a two-point set \( I_n = \{ (0,\kappa_n), (\sqrt{\kappa_n/n},0) \} \) for computationally intensive problems. This is because different values of \( \gamma_n \) and \( \lambda_n \) have different strengths and weaknesses. On the

\(^6\)In practice, it suffices to obtain an approximate minimum. I thank the referee for pointing this out.

\(^7\)See Section 4 for a heuristic illustration of the method with \( \lambda_n = 0 \).
one hand, \( \sqrt{n\gamma_n \mathbb{E}_t \rho_t(W_i, \theta)} \) leads to a better approximation of the local parameter space if \( \gamma_n \) is larger. For example, if \((\gamma_n, \lambda_n) = (0, \kappa_n)\), the method uses 0 as a conservative estimate of the local parameter space. The resulting critical value controls size but is conservative asymptotically. If \((\gamma_n, \lambda_n) = (\kappa_n, 0)\), the test has the correct size asymptotically in many cases.\(^8\) On the other hand, estimation error
\[
\left| n^{-1/2} \sum_{i=1}^{n} \rho_t(W_i, \theta) - \sqrt{n \gamma_n \mathbb{E}_t \rho_t(W_i, \theta)} \right|
\]
is increasing in \( \gamma_n \), which can worsen the finite-sample performance. Because of this trade-off, it is desirable to have multiple elements in \( I_n \) to combine the advantages of different values of \((\gamma_n, \lambda_n)\). Naturally, \( I_n = \bar{T}_n \) yields the smallest critical value. But simulations suggest that it is important to include \((0, \kappa_n)\) and \(\left(\sqrt{\kappa_n/n}, 0\right)\). Beyond that, adding more elements does not significantly improve performance. If \((0, \kappa_n) \in I_n\), the resulting critical value remains bounded under any fixed alternative.\(^9\) It is also worth noting that computing the infimum in (6) is a one-dimensional optimization problem if \( I_n = \bar{T}_n \), while computing \(\theta_n^* (\gamma_n, \lambda_n, R)\) involves an optimization problem which has the same dimension as \(\Theta_n \cap R\). In total, computing one bootstrap statistic involves an optimization problem that, at most, has a dimension of 1 plus the dimension of \(\Theta_n \cap R\). It reduces further if \( I_n = \{ (0, \kappa_n), \left(\sqrt{\kappa_n/n}, 0\right) \}\). This is very convenient for computationally intensive problems. In contrast, the bootstrap statistic proposed in CNS needs to compute an optimization problem which has a dimension twice that of \(\Theta_n \cap R\) if partial identification is a concern. The computational gain of my method is non-negligible if the dimension of \(\Theta_n \cap R\) is high and the moments are difficult to evaluate.

Confidence Sets To construct a confidence set for \( \phi(\theta_F) \) after imposing some restriction \( S \), let \( R_{a,\phi,S} = \{ \theta \in S : \phi(\theta) = a \} \) and \( \mathbb{K} \) be the parameter space of \( \phi(\theta_F) \). A confidence set with \( 1 - \alpha \) confidence level is
\[
C_n (1 - \alpha) = \{ a \in \mathbb{K} : T_n (R_{a,\phi,S}) \leq C_n^* (1 - \alpha + \eta, R_{a,\phi,S}) + \eta \}.
\]

A Step-by-Step Guide To construct a test of significant level \( \alpha \):

\(^8\)My test is asymptotically non-similar, i.e., the test does not have the same size uniformly. This is, however, not deficient, as pointed out by Andrews (2012), who shows that asymptotically similar tests have very poor power if there are inequality constraints. If there are no inequality constraints, the test can achieve the exact size asymptotically as shown in Section 5.2.

\(^9\)In contrast, the bootstrap methods considered in Santos (2012); Hong (2017); and Gandhi, Lu, and Shi (2012) all lead to diverging critical values under fixed alternatives.
(a) Compute the test statistic $T_n(R)$ according to equation (4).

(b) Draw an i.i.d. sample $\{W_i^n\}_{i=1}^n$ from the empirical distribution $F_n$ and form $G_n^*(\theta, t)$.

(c) Compute the bootstrap statistic using equation (6). One may set $\kappa_n = n/\ln n$.\(^\text{10}\) I recommend $I_n = \bar{I}_n$ if computation is not a concern and $I_n = \{(0, \kappa_n), (\sqrt{\kappa_n/n}, 0)\}$ in computationally intensive problems.

(d) Repeat (b)-(c) $B$ times and collect all the bootstrap statistics from each repetition.

(e) Use $\eta$ plus the $(1 - \alpha + \eta)$-th quantile of all the bootstrap statistics as the critical value. Reject the null hypothesis if $T_n(R)$ exceeds this critical value. One may set $\eta$ to be a very small number, say $10^{-8}$.

4 A Heuristic Illustration

This section heuristically illustrates how the method works and why convexity is crucial using the non-parametric IV setup as in Example 2.1. I focus on the bootstrap statistic with $\lambda_n = 0$, i.e., $I_n = \{(\sqrt{\kappa_n/n}, 0)\}$. To abstract from the complications of sieve approximation and partial identification, I further assume that $\Theta_n = \Theta$ and $\theta_F$ is point-identified.

By the definition of the test statistic,

$$T_n(R) = \min_{\theta \in \Theta \cap R} \int_{t \in T} \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \rho_t(W_i, \theta) \right]^2 d\mu(t)$$

$$= \min_{\theta \in \Theta \cap R} \int_{t \in T} \left\{ G_{n,F}(\theta, t) + \sqrt{n} \left[ E_F \rho_t(W_i, \theta) - E_F \rho_t(W_i, \theta_F) \right] \right\}^2 d\mu(t).$$

Under the null, the minimizer obtained from calculating $T_n(R)$ is close to $\theta_F$ for sufficiently large $n$ under appropriate assumptions. If $G_{n,F}(\theta, t)$ is continuous in $\theta$, it is close to $G_{n,F}(\theta_F, t)$. Then use the definition of $\rho_t$ in Example 2.1 and the change of variable

\(^{10}\)Although very desirable, it is difficult to develop a data-driven method to choose $\kappa_n$. I instead use simulations to provide guidance on $\kappa_n$. Results suggest that $n/\ln n$ is a good choice.
\[ \Delta = \sqrt{n} (\theta - \theta_F) \] to obtain

\[ T_n (R) = \min_{\Delta \in V_n(\theta_F)} \int_{t \in T} \{ G_n,F (\theta_F, t) - E_F w (Z_i, t) \Delta (X_i) \}^2 d\mu (t) + o_{\mathbb{P}} (1), \tag{11} \]

where \( o_{\mathbb{P}} (1) \) is a term that converges to 0 in probability and

\[ V_n(\theta_F) = \{ \Delta : \theta_F + \Delta / \sqrt{n} \in \Theta \cap R \}. \]

Here \( -E_Fw (Z, t) \Delta (X) \) is essentially the directional derivative of \( E_F \rho_t (W_i, \theta) \) at \( \theta_F \) in the direction \( \Delta \). All the directions that need to be evaluated are determined by \( V_n(\theta_F) \), which is the true local parameter space (LPS).

To estimate the null distribution of \( T_n (R) \), the main challenge lies in estimating \( V_n(\theta_F) \) because it is, roughly speaking, discontinuous in \( \theta_F \). To see this, consider the case where \( R \) is the set of non-negative functions. If \( \theta_F = 0 \), \( V_n(\theta_F) \) converges to a space that contains only non-negative functions. If \( \theta_F \) is strictly positive but arbitrarily close to 0, \( V_n(\theta_F) \) converges to a space that contains all functions. Due to this discontinuity, one cannot naively plug an estimated \( \theta_F \) into \( V_n(\theta_F) \) to obtain an estimate of \( V_n(\theta_F) \).

This paper observes that directional derivatives can be approximated by differences between moment conditions evaluated at \( \theta_F + \Delta \) and those evaluated at \( \theta_F \). Although the true value of \( \theta_F \) is unknown, \( E_F \rho_t (W_i, \theta_F) \) is known to be 0. This implies these differences do not depend on \( \theta_F \). If \( \gamma_n \to 0 \), \( \sqrt{n} \gamma_n \sum_i \rho_t (W_i, \theta) \) converges to \( \sqrt{n} \gamma_n E_F \rho_t (W_i, \theta) \) and

\[ T_n^* (R) = \min_{\theta \in \Theta \cap R} \int_{t \in T} \left\{ G_n^* (\theta, t) + \sqrt{n} \gamma_n \sum_i \rho_t (W_i, \theta) \right\}^2 d\mu (t) \]

\[ = \min_{\theta \in \Theta \cap R} \int_{t \in T} \left\{ G_n^* (\theta, t) + \sqrt{n} \gamma_n E_F \rho_t (W_i, \theta) \right\}^2 d\mu (t) + o_{\mathbb{P}} (1). \]

If \( \sqrt{n} \gamma_n \to \infty \) sufficiently fast, the minimizer obtained from calculating \( T_n^* (R) \) is close to \( \theta_F \). Then, similar as (11) but with \( \Delta = \sqrt{n} \gamma_n (\theta - \theta_F) \),

\[ T_n^* (R) = \min_{\Delta \in V_n^* (\theta_F)} \int_{t \in T} \{ G_n^* (\theta_F, t) - E_F w (Z_i, t) \Delta (X_i) \}^2 d\mu (t) + o_{\mathbb{P}} (1), \tag{12} \]
where my estimated LPS is

\[ V_n^\gamma (\theta_F) = \left\{ \Delta : \theta_F + \frac{\Delta}{\sqrt{n} \gamma_n} \in \Theta \cap R \right\}. \]

Because \( G_n^* (\theta_F, t) \) approximates the distribution of \( G_{n,F} (\theta_F, t), T_n^* (R) \) leads to a valid critical value if \( V_n^\gamma (\theta_F) \subseteq V_n (\theta_F) \). This is guaranteed if both \( \Theta \) and \( R \) are convex because \( \theta_F \in \Theta \cap R \) and

\[ \theta_F + \frac{\Delta}{\sqrt{n}} = (1 - \gamma_n) \theta_F + \gamma_n \left[ \theta_F + \frac{\Delta}{\sqrt{n} \gamma_n} \right]. \]

If there is partial identification and \( \rho_t \) is not linear, a similar argument based on the first-order expansion of the moment conditions goes through.

Figure I illustrates how the rescaled moments approximate the true LPS and why convexity is important. In all graphs, \( V_n (\theta_F) / \sqrt{n} \) and \( V_n^\gamma (\theta_F) / \sqrt{n} \) are defined as the collection of all elements in \( V_n (\theta_F) \) and \( V_n^\gamma (\theta_F) \) divided by \( \sqrt{n} \). Loosely, I will also refer to \( V_n (\theta_F) / \sqrt{n} \) and \( V_n^\gamma (\theta_F) / \sqrt{n} \) as the true LPS and the estimated LPS in the rest of this section. Figure Ia is constructed under \( R = \{ \theta : \theta (x) \geq 0 \text{ for all } x \} \), which is convex. And \( \theta_F \) is the blue solid line that lies in \( R \). The true LPS and the estimated LPS are the areas above their boundaries, respectively. Both spaces are convex because if two functions lie above one of the two boundaries, their convex combinations also lie above that boundary. The estimated LPS is obtained by shrinking the true LPS toward 0 by a factor of \( \gamma_n \). The yellow area is shared by the true LPS and the estimated LPS, while the green area is in the true LPS but
not in the estimated LPS. Because of the convexity, the estimated LPS is contained in the true LPS. This implies that $T_n(R)$ is obtained by a minimization problem over a larger set compared to $T^*_n(R)$. Therefore, the distribution of $T_n(R)$ is first-order stochastically dominated by that of $T^*_n(R)$ asymptotically, which validates the bootstrap critical value. Notice that the green area is smaller if $\gamma_n$ is larger, which means that a large $\gamma_n$ yields a better approximation of the true LPS for any given $n$. It is also worth noting that even though $\gamma_n$ converges to 0, the limit of $V^\gamma_n(\theta_F)$ can be the same as that of $V_n(\theta_F)$ as $n \to \infty$.

Figure Ib is constructed under $R = \{ \theta : \theta(x) \geq 0 \text{ for all } x \text{ or } \theta(x) \leq -0.5 \text{ for all } x \}$, which is not convex. Now the true LPS has two boundary lines: one passes through the origin and the other lies below the x-axis. It consists of functions that lie above the higher boundary and functions that lie below the lower boundary. Obviously, it is not convex. Again, the estimated LPS is obtained by shrinking the true LPS towards 0. As in Figure Ia, the yellow region is shared by both the true LPS and the estimated LPS, and the green region lies in the true LPS but not in the estimated LPS. But there is now a magenta area, which lies in the estimated LPS but not in the true LPS. Therefore, $T_n(R)$ is no longer obtained by a minimization problem over a larger region compared to $T^*_n(R)$. Hence, the size control may not hold. Such differences can matter asymptotically and one can construct examples where uniform size control fails due to lack of convexity.

Lastly, it is useful to compare my method with CNS. The yellow region in Figure Ic corresponds to the estimator of LPS in CNS under $R = \{ \theta : \theta(x) \geq 0 \text{ for all } x \}$. It is obtained by shifting the boundary of $V_n\left(\hat{\theta}_F\right) / \sqrt{n}$ towards 0, where $\hat{\theta}_F$ is a constrained estimator of $\theta_F$. The size of the shift at a point $x$ is the minimum of a tuning parameter and the distance from $\hat{\theta}_F(x)$ to the boundary of the restriction set. I refer to this tuning parameter as the shifting parameter. This method differs from mine in two ways. First, because the rescaled moments exploit the fact that the moments are 0 at $\theta_F$, my estimated LPS is not explicitly based on an estimator of $\theta_F$. Second, my method shrinks the true LPS toward 0 by a factor of $\gamma_n$. Compared to Figure Ia, the green region in Figure Ic, which is the set of function values that are in the true LPS but not in the LPS of CNS, is wider around the origin but gets narrower away from the origin. This highlights the difference between the two methods.
5 Asymptotic Theory

5.1 Uniform Size Control

In the rest of the paper, I assume that $\Theta$ is a convex space. The theoretic results of this paper focus on the following class of distribution functions.

**Definition 5.1.** Let $\mathcal{F}$ be a collection of distribution functions $F$ such that

1. The moment function $\rho_t(\cdot, \theta)$ is measurable for all $(\theta, t) \in \Theta \times T$.

2. There exists a measurable function $F$ such that $|\rho_t(\cdot, \theta)| \leq F(\cdot)$ for all $(\theta, t) \in \Theta \times T$, and $\sup_{F \in \mathcal{F}} \mathbb{E}_F (W_i)^2 < \infty$.

3. The set of functions $\varrho = \{\rho_t(\cdot, \theta) : (\theta, t) \in \Theta \times T\}$ is Donsker and pre-Gaussian uniformly in $F \in \mathcal{F}$.

Part (1) in Definition 5.1 is a mild condition and is automatically satisfied if $\rho_t(W_i, \theta)$ is continuous in $W_i$ for all $(\theta, t) \in \Theta \times T$. Part (2) requires that there is an envelope function for $\rho_t(\cdot, \theta)$. And the envelope function has a uniformly bounded second moment. Part (3) is the key to the uniform convergence rates and the validity of the test. Under the null hypothesis, the test statistic can be expressed as a continuous function applied to $G_{n,F}$, and the bootstrap statistic can be expressed as the same function applied to $G_{n}^*$. Part (3) then allows me to invoke a result in Linton, Song, and Whang (2010) to establish uniform size control. This condition can be satisfied after regularizing the parameter space, following Newey and Powell (2003) and Santos (2012). Section D in the Appendix provides low-level conditions that guarantee Part (3).

**Definition 5.2.** Let $\mathcal{J} = \{(F, R) : F \in \mathcal{F}, R \in \mathcal{R}, \text{ and } \Theta_F \cap R \neq \emptyset\}$, where $\mathcal{R}$ is a set of non-empty, convex sets.

The set $\mathcal{J}$ consists of pairs of distribution function and restriction set that satisfy the null hypothesis. I now introduce the main result of this paper, which shows that the proposed test controls size uniformly on $\mathcal{J}$.

**Theorem 5.1.** Under Assumptions C.1-C.5, for any non-empty $I_n \subseteq \bar{I}_n$ and $\eta > 0$.
\( (1). \ \limsup_{n \to \infty} \sup_{(F,R) \in \mathcal{J}} \mathbb{P}_F \left( T_n (R) > C_n^* (1 - \alpha + \eta, R) + \eta \right) \leq \alpha. \) If in addition Assumption C.7 holds, this inequality holds with \( \eta = 0. \)

\( (2). \ \text{If} \ F \in \mathcal{F}, \ \Theta_F \cap R = \emptyset, \ \text{then} \ \mathbb{P}_F \ \text{almost surely}, \ T_n (R) > C_n^* (1 - \alpha + \eta, R) + \eta \ \text{as} \ n \to \infty. \)

Theorem 5.1 establishes the validity of the test. In general, Theorem 5.1(1) holds with weak inequality. But if the shape restriction does not involve inequality constraints, the test can achieve the exact size, as will be shown in the next section. Here \( \eta \) appears for technical reasons. Under Assumption C.7, which consists of an anti-concentration condition and a condition that handles the degenerate case, \( \eta \) can then be set to 0.

Now I briefly discuss the assumptions. More discussions can be found in Section C in the Appendix. Assumption C.1 requires that \( \Theta \cap R \) is compact, and that \( \mathbb{E}_F \rho_t (W, \theta) \) is continuous in \( \theta \). It guarantees that the test statistic diverges to infinity under a fixed alternative, which ensures the second claim of Theorem 5.1. Assumption C.2 requires that the sieve spaces approximate \( \Theta \) well such that the approximation bias can be ignored. Assumption C.3 is necessary to obtain the convergence rates of \( \hat{\Theta}_n (R) \) and \( \hat{\Theta}_n^* (\gamma_n, \lambda_n, R) \). These rates are important for size control. Assumption C.4 serves two roles. First, combined with Definition 5.1(3), it implies that \( G_{n,F} \) and \( G_{n}^* \) are asymptotically equi-continuous in \( \theta \). This allows us to replace \( \theta \) by \( \theta_F \) in (11) and (12). Second, it guarantees that the moment conditions can be approximated by their linear expansion. Therefore, the heuristic argument in Section 4 based on linearity can be extended to nonlinear moments. Lastly, Assumption C.5 puts restrictions on \( \kappa_n \). It implicitly rules out some slow convergence rate including the severely ill-posed problems.

Remark 5.1. Assumptions C.4(i), (ii) and the convexity of \( \Theta \) and \( R \) are needed to use a general \( I_n \). If \( I_n = \{(0, \kappa_n)\} \), my test remains valid without these assumptions. In this case, Assumption C.5 can also be relaxed. However, this comes at the cost of being conservative. As illustrated in Section 4, a lower \( \gamma_n \) yields a potentially more conservative approximation. If \( I_n = \{(0, \kappa_n)\} \), \( \gamma_n \) is set to be 0, which is the smallest value.

Remark 5.2. An immediate consequence of Theorem 5.1 is the validity of \( C_n (1 - \alpha) \). To see
this, just notice that one can set $\mathcal{R} = \{ R_{a,\phi,S} : a \in \mathbb{K} \}$ and

$$
\liminf_{n \to \infty} \inf_{F \in \mathcal{F}, \theta_F \in S} \mathbb{P}_F \left( \phi(\theta_F) \in \mathcal{C}_n \left( 1 - \alpha \right) \right) \\
\geq \liminf_{n \to \infty} \inf_{F \in \mathcal{F}, \theta_F \in S} \mathbb{P}_F \left( T_n \left( R_{\phi(\theta_F),\phi,S} \right) \leq C_n^* \left( 1 - \alpha + \eta, R_{\phi(\theta_F),\phi,S} \right) + \eta \right) \\
\geq \liminf_{n \to \infty} \inf_{(F,R) \in \mathcal{J}} \mathbb{P}_F \left( T_n \left( R \right) \leq C_n^* \left( 1 - \alpha + \eta, R \right) + \eta \right) \geq 1 - \alpha.
$$

**Corollary 5.1.** Under Assumptions C.1-C.6,

$$
\liminf_{n \to \infty} \inf_{F \in \mathcal{F}, \theta_F \in S} \mathbb{P}_F \left( \phi(\theta_F) \in \mathcal{C}_n \left( 1 - \alpha \right) \right) \geq 1 - \alpha.
$$

**Remark 5.3.** To make the bootstrap procedure even faster, one might also consider bootstrapping

$$
\hat{T}_n^*(R) = \inf_{\theta \in \Theta_n \cap \hat{\Theta}_n(R)} \int T \left[ \mathcal{G}_n^* (\theta, t) + \frac{\gamma_n}{\sqrt{n}} \sum_{i=1}^{n} \rho_t \left( W_i, \theta \right) \right]^2 d\mu (t),
$$

where $\epsilon > 0$ is a constant and $\hat{\Theta}_n^\epsilon (R)$ is the $\epsilon$-neighborhood of $\hat{\Theta}_n (R)$. One can prove that this bootstrap is valid following the same method used in this paper.

### 5.2 Non-Conservativeness without Inequality Restrictions

If there are no inequality constraints, one can show that my test delivers the correct size with slightly stronger assumptions. I first put more structure on the sieve space by focusing on linear sieves. Let $\{ p_k \}_{k=1}^{\infty}$ be the set of basis functions of $\Theta$. Define $P_{kn} = (p_1, p_2, \cdots, p_{k_n})$ and $\beta$ to be a column vector. The sieve spaces are $\Theta_n = \{ \theta \in \Theta : \theta = P_{kn} \beta, \beta \in \mathbb{R}^{k_n} \}$, where $k_n$ is a positive integer which diverges to infinity as $n \to \infty$.

**Definition 5.3.** The restriction $R$ is an affine set if $R = \{ \theta : l(\theta) = a_l \forall l \in L \}$, where $L$ is a set of linear functionals.

The next result shows that the test has the exact size under the point-wise asymptotics.

**Theorem 5.2.** Suppose that Assumptions C.1-C.5 hold and $R$ is an affine set. Let $(F,R) \in \mathcal{J}$ satisfies Assumption G.1 and $\Theta_F$ lies in the interior of $\Theta$. If $\left( \sqrt{k_n/n}, 0 \right) \in I_n$ and the
limiting distribution of $T_n(R)$ is continuous and strictly increasing at its $1 - \alpha$-th quantile, then
$$\lim_{n \to \infty} \mathbb{P}_F(T_n(R) > C_n^\alpha (1 - \alpha, R)) = \alpha.$$ 

One crucial condition is that \((\sqrt{k_n/n}, 0) \in I_n\). It guarantees that the bootstrap statistic properly estimates the local parameter space. The test proposed by Hong (2017) can be interpreted as a version of our test where $I_n$ contains only $(0, \kappa_n)$ and therefore does not necessarily have the exact asymptotic size.

**Corollary 5.2.** Suppose that $\mathcal{R}$ consists of only affine sets and \((\sqrt{k_n/n}, 0) \in I_n\), and that Assumptions C.1-C.5, C.7 hold. If there exists some \((F,R) \in \mathcal{J}\) such that (i) it satisfies Assumption G.1; (ii) $\Theta_F$ lies in the interior of $\Theta$; and (iii) the limiting distribution of $T_n(R)$ is continuous and strictly increasing at its $1 - \alpha$-th quantile, then
$$\limsup_{n \to \infty} \sup_{(F,R) \in \mathcal{J}} \mathbb{P}_F(T_n(R) > C_n^\alpha (1 - \alpha, R)) = \alpha.$$ 

This corollary is a direct consequence of Theorem 5.1 and Theorem 5.2. The former shows that the asymptotic size is uniformly no more than $\alpha$. The latter guarantees that $\alpha$ is achieved under a particular \((F,R)\).

I end this section with a discussion on power comparison between my method and CNS. I focus only on the difference caused by the different ways to estimate the local parameter space because the two methods apply to different classes of models. In general, neither of these two methods dominates the other. This can be seen by comparing Figures Ia and Ic. There are functions that lie in the yellow region in Figure Ia but not in the yellow region in Figure Ic and vice versa. Indeed, one can construct examples where for each sequence of rescaling parameters in my test, there exists a sequence of shifting parameters and a sequence of data generating processes along which the test based on CNS has better size control or power. Similarly, for each sequence of shifting parameters, one can find a sequence of rescaling parameters and a sequence of data generating processes along which my test has better size control or power.

**Example 5.1** (Asymptotic Size). Consider the case where $W_i = (X_i, Y_i)$ is bivariate normal with the identity covariance matrix and has mean $\theta_{F_n} = (\theta_{1,n}, \theta_{2,n})$ as $n \to \infty$. One would like
to test $H_0 : \theta_F \geq 0$ against $H_1 : \theta_F \neq 0$. Then the moment conditions are $E_F(W_i - \theta) = 0$, where $\theta = (\theta_1, \theta_2)$. The test statistic that gives equal weights to both moment conditions is

$$T_n(R) = \min_{(\theta_1, \theta_2) \geq 0} \frac{1}{\sqrt{n}} \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \theta_1) - \sqrt{n} (\theta_1 - \theta_1) \right]^2 + \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Y_i - \theta_2) - \sqrt{n} (\theta_2 - \theta_2) \right]^2$$

$$= \min_{(h_1, h_2) \in V_n(\theta_{F_n})} \frac{1}{\sqrt{n}} \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \theta_1) - h_1 \right]^2 + \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Y_i - \theta_2) - h_2 \right]^2,$$

where the true local parameter space is

$$V_n(\theta_{F_n}) = \{ (h_1, h_2) : h_1/\sqrt{n} \geq -\theta_1, h_2/\sqrt{n} \geq -\theta_2 \}.$$

Similarly, the bootstrap statistic with $I_n = \{ (\sqrt{n}/\kappa_n) \}$ is

$$T_n^* (R) = \min_{(h_1, h_2) \in V_n^*(\theta_{F_n})} \frac{1}{\sqrt{n}} \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \theta_1) - h_1 \right]^2 + \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Y_i - \theta_2) - h_2 \right]^2 + o_p_F (1),$$

where $V_n^*(\theta_{F_n}) = \{ (h_1, h_2) : h_1/\sqrt{n} \geq -\theta_1, h_2/\sqrt{n} \geq -\theta_2 \}.$

If one follows the method proposed in CNS to estimate the local parameter space, one can obtain a bootstrap statistic

$$T_n^{*, CNS} (R) = \min_{(h_1, h_2) \in V_n^{*, CNS} (\hat{\theta}_n)} \frac{1}{\sqrt{n}} \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \theta_1) - h_1 \right]^2 + \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Y_i - \theta_2) - h_2 \right]^2,$$

where $\hat{\theta}_n = (\hat{\theta}_1, \hat{\theta}_2)$ is the minimizer obtained from the minimization problem for $T_n(R)$, $r_n$ is a sequence of positive numbers that converges to 0, and

$$V_n^{r_n} (\hat{\theta}_n) = \{(h_1, h_2) : h_1/\sqrt{n} \geq -\hat{\theta}_1 + \min (h_1, r_n), h_2/\sqrt{n} \geq -\hat{\theta}_2 + \min (h_2, r_n) \}.$$

Given $\gamma_n = O (\sqrt{1/\ln n})$, let $r_n = 1/\gamma_n \sqrt{n}$, $\hat{\theta}_1, \hat{\theta}_2 = 0$ and $\theta_2 = c/\gamma_n \sqrt{n}$ where $c$ is a constant. And let $Z_1$ and $Z_2$ be two independent standard normal random variables. Along $F_n$, $T_n(R)$ and $T_n^*(R)$ converge to $\min (Z_1, 0)^2/2$ and $[\min (Z_1, 0)^2 + \min (Z_2 + c, 0)^2]/2$ in

21
distribution, respectively.

Suppose \( c = 1.01 \). Because \( \hat{\theta}_{2,n} = \theta_{2,n} + O_{F_n} \big(n^{-1/2}\big) \),

\[
P_{F_n} \left( \hat{\theta}_{2,n} > r_n \right) = P_{F_n} \left( \hat{\theta}_{2,n} - \theta_{2,n} > r_n - \theta_{2,n} \right) = P_{F_n} \left( O_{F_n} \big(n^{-1/2}\big) > \frac{1 - c}{\gamma_n \sqrt{n}} \right)
\]

converges to 1. This suggests that along \( F_n \), \( T_{n,CNS}^* (R) \) converges to \( \min (Z_1, 0)^2 / 2 \). Therefore, \( T_{n,CNS}^* (R) \) leads to a test that has the exact asymptotic size for all \( \alpha > 0.5 \) along \( F_n \). On the other hand, a 10% test using \( T_n^* (R) \) has a rejection probability that converges to 8.75%, while a 5% test has a rejection probability that converges to 4.4%. Therefore, \( T_{n,CNS}^* (R) \) has better size control.

Now suppose \( c = 0.99 \). Given any \( r_n \) such that \( \sqrt{\ln n/n} = O \left( r_n \right) \), let \( \gamma_n = 1/r_n \sqrt{n} \). Let \( \theta_{1,n} \) and \( \theta_{2,n} \) be the same as the above. Along \( F_n \), \( P_{F_n} \left( \hat{\theta}_{2,n} > r_n \right) \rightarrow 0 \) and \( T_{n,CNS}^* (R) \) converges to \( \left[ \min (Z_1, 0)^2 + \min (Z_2, 0)^2 \right] / 2 \). Then a 10% test and a 5% test based on \( T_{n,CNS}^* (R) \) have asymptotic sizes 4.29% and 1.99%, respectively. The corresponding values based on \( T_n^* (R) \) are 8.69% and 4.37%, respectively. Along this \( F_n \), my method has better size control.

**Example 5.2** (Local Power). Consider the same testing problem as in the previous example. Construct \( F_n \) in the same way as above except that \( \theta_{1,n} = -1/\sqrt{n} \). Then \( F_n \) is a sequence of local alternatives. Along \( F_n \), the limiting distribution of \( T_n^* (R) \) is \( \min (Z_1 - 1, 0)^2 / 2 \). The limiting distributions of the bootstrap statistics are the same as the ones calculated in the previous example. If \( c = 1.01 \), a 5% test using \( T_n^* (R) \) rejects with a probability converging to 36.07%, while a 5% test using \( T_{n,CNS}^* (R) \) rejects with a probability converging to 38.91%, which is about 3% higher. If \( c = 0.99 \), a 5% test based on \( T_n^* (R) \) has a rejection probability that converges to 35.94%, while a test based on \( T_{n,CNS}^* (R) \) has a rejection probability that converges to 23.63%. Now \( T_n^* (R) \) has better power. A similar result holds for a 10% test.

### 6 Monte Carlo Experiments

This section considers three data generating processes (DGPs). The first two DGPs follow Example 2.1, where critical values are relatively easy to calculate. This allows me to investigate extensively how the performance of the test depends on the parameter space, \( I_n \),
κ_n, and the approximation of the integration. The last DGP follows Example 2.3, where critical values are more difficult to calculate. I use it to illustrate how the test works in an environment that is close to my empirical application and to provide some guidance on the choice of κ_n.

6.1 Monotonicity in Non-Parametric IV Regression

Now consider Example 2.1 with the following two DGPs.

**DGP I** This DGP is considered in Santos (2012). Generate a latent i.i.d. sample by

\[
\begin{pmatrix}
X^* \\
Z^* \\
\epsilon^*
\end{pmatrix}
\sim N
\begin{pmatrix}
1 & 0.5 & 0.3 \\
0.5 & 1 & 0 \\
0.3 & 0 & 1
\end{pmatrix}.
\]  

(14)

Construct Y, X, Z, ε by setting

\[
X = 2 (\Phi (X^*/3) - 0.5), \; Z = 2 (\Phi (Z^*/3) - 0.5), \; \epsilon = \epsilon^*,
\]  

(15)

where \(\Phi\) is the distribution function of a standard normal random variable.

**DGP II** \(Z \sim \text{Uniform}[0, 1]\) and \(X = Z + U\) where \(U \sim \text{Uniform}[-1, 1]\) and is independent of \(Z\). In addition, \(\epsilon = \eta + 0.1\epsilon^*\) where \(\epsilon^* \sim N(0, 1)\) and is independent of \((U, X, Z)\). Newey, Powell, and Vella (1999) show that \(\theta_F\) is point-identified up to a constant with the additional moment condition \(E_F(U|Z) = 0\). I use only the moment condition \(E_F(\epsilon|Z) = 0\). Then the model is in general partially identified but can be point-identified with shape restrictions in some cases. For more details, see Section B in the Appendix.

For both DGPs, I consider \(\theta_F = f_i, \; i = 1, 2, 3, 4\) where \(f_1(X) = 0, \; f_2(X) = X - 4\phi(X - 1), \; f_3(X) = X (1 - X), \; \text{and} \; f_4(X) = X - 8 \times \phi(X - 1.15)\). Here \(\phi\) is the standard normal density function. These functions are plotted in Figure II. \(f_1\) is a constant function. \(f_2\) is strictly increasing, while \(f_3\) and \(f_4\) are decreasing in some regions. However, because this model may be partially identified, this does not mean that data generated by \(f_3\) and \(f_4\)
violate the null hypothesis. In fact, in DGP I with $\theta_F = f_3$, the null hypothesis is satisfied. In DGP II, only $f_1$ and $f_2$ satisfy the null hypothesis.

I generate i.i.d. samples with 1000 observations. The sieve space is chosen to be spanned by B-splines. In DGP I, I use knots $\{-1, -1, -\frac{1}{3}, \frac{1}{3}, 1, 1\}$ to approximate $f_1$, $f_2$, and $f_3$, and knots $\{-1, -1, -\frac{3}{4}, -\frac{1}{2}, \frac{1}{2}, \frac{3}{4}, 1, 1\}$ for $f_4$. In model II, I use knots $\{-1, -1, 0, 1, 2, 2\}$ for $f_1$, $f_2$, $f_3$, and

$$\{-1, -1, -0.625, -0.25, 0.125, 0.5, 0.875, 1.25, 1.625, 2, 2\}$$

for $f_4$. Different knots are used because $X$ is supported on $[-1, 1]$ in DGP I and on $[-1, 2]$ in DGP II. I also use more knots for $f_4$ because it is harder to approximate. The weight function is defined as $w(z, t) = \phi \left( \frac{z - t}{t_2} \right)$ and $t = (t_1, t_2) \in T$ with $T = [-0.8, 0.8] \times [0.05, 0.2]$, where $\phi$ is a standard normal density. This weight function is also used in Santos (2012). I approximate the numerical integration by averaging over points in $T$. To evaluate the sensitivity of the test to the numerical integration, I consider two cases. In the first case, I average over 14 points, which is constructed by taking the product of 7 equally spaced points on $[-0.8, 0.8]$ and 2 equally spaced points on $[0.05, 0.2]$. In the second case, I average over 350 points obtained by taking the product of 35 equally spaced points on $[-0.8, 0.8]$ and 10 equally spaced points on $[0.05, 0.2]$. This effectively uses 350 moments to approximate the test. I restrict the parameter space such that the B-spline coefficients are no more than $B$ in
absolute values. The value of $B$ ranges from 5 to 100 to evaluate its influence.

I consider the following choices of $I_n$: (1) $I_n = \left\{ \left( \sqrt{\kappa_n/n}, 0 \right) \right\}$. This leads to a test that is in spirit similar to Gandhi, Lu, and Shi (2012) and is denoted by GLS. (2) $I_n = \{(0, \kappa_n)\}$, which is otherwise the same as in Hong (2017) except it is based on a CvM criterion function and does not include the penalty term. Since this choice uses the most conservative estimate for the local parameter space, I denote it as LF (least favorable). (3) I also consider bootstrap statistics with richer $I_n$, i.e.,

$$I_n = \left\{ (\gamma, \lambda) \mid n\gamma^2 + \lambda = \kappa_n, \lambda \geq 0, \gamma \geq 0, \gamma = \frac{i}{N} \sqrt{\frac{\kappa_n}{n}}, i = 0, 1, 2, \ldots, N \right\}.$$  

This bootstrap statistic is the minimum of several bootstrap statistics with different $(\gamma_n, \lambda_n)$ and hence is called the Minimum Bootstrap statistic (MB). To assess the sensitivity of the method to $I_n$, I report tests with $N = 1$ (MB1) and $N = 30$ (MB2). The performance of the test can be sensitive to $\kappa_n$. Therefore, I report results for $\kappa_n = n^{2/3}$, $n/\ln n$ to evaluate its impact. Critical values are computed with 500 bootstrap statistics and $\eta$ is set to 0. Setting $\eta = 10^{-8}$ yields identical results. I simulate 1000 random samples to compute the rejection probability. The significance level is set to 10%.

Table 1 summarizes the rejection probability on the simulated data. The first column shows the true $\theta_F$. The other columns report rejection probabilities of corresponding tests.
Table 1: Rejection Probability with a 10% Significance Level under \( n = 1000 \). GLS: \( I_n = \left\{ \left( \sqrt{\kappa_n/n}, 0 \right) \right\} \); LF: \( I_n = \{0, \kappa_n\} \); MB1: \( I_n = \{0, \kappa_n, \sqrt{\kappa_n/n}, 0\} \); and MB2: \( I_n \) contains 31 equally spaced points between \((0, \kappa_n)\) and \(\left( \sqrt{\kappa_n/n}, 0 \right)\). The number of moments corresponds to how many points are used to approximate the integral. And \( B \) is the bound imposed on the parameter space.

<table>
<thead>
<tr>
<th>( \theta_F )</th>
<th>GLS</th>
<th>LF</th>
<th>MB1</th>
<th>MB2</th>
<th>GLS</th>
<th>LF</th>
<th>MB1</th>
<th>MB2</th>
<th>GLS</th>
<th>LF</th>
<th>MB1</th>
<th>MB2</th>
<th>GLS</th>
<th>LF</th>
<th>MB1</th>
<th>MB2</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \kappa_n = n/\ln n )</td>
<td>( \kappa_n = n^{2/3} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>DGP I</strong></td>
<td><strong>DGP II</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( B = 5, 14 ) moments</td>
<td>( B = 5, 350 ) moments</td>
<td>( B = 5, 14 ) moments</td>
<td>( B = 5, 350 ) moments</td>
<td>( B = 10, 14 ) moments</td>
<td>( B = 10, 350 ) moments</td>
<td>( B = 10, 14 ) moments</td>
<td>( B = 10, 350 ) moments</td>
<td>( B = 100, 14 ) moments</td>
<td>( B = 100, 350 ) moments</td>
<td>( B = 100, 14 ) moments</td>
<td>( B = 100, 350 ) moments</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( f_1 )</td>
<td>0.017</td>
<td>0.002</td>
<td>0.056</td>
<td>0.058</td>
<td>0.028</td>
<td>0.006</td>
<td>0.068</td>
<td>0.069</td>
<td>0.042</td>
<td>0.003</td>
<td>0.072</td>
<td>0.076</td>
<td>0.05</td>
<td>0.006</td>
<td>0.079</td>
<td>0.081</td>
</tr>
<tr>
<td>( f_2 )</td>
<td>0.025</td>
<td>0.002</td>
<td>0.049</td>
<td>0.05</td>
<td>0.016</td>
<td>0.004</td>
<td>0.04</td>
<td>0.043</td>
<td>0.026</td>
<td>0</td>
<td>0.039</td>
<td>0.039</td>
<td>0.027</td>
<td>0.003</td>
<td>0.06</td>
<td>0.061</td>
</tr>
<tr>
<td>( f_3 )</td>
<td>0.011</td>
<td>0</td>
<td>0.018</td>
<td>0.02</td>
<td>0.006</td>
<td>0</td>
<td>0.02</td>
<td>0.02</td>
<td>0.014</td>
<td>0.001</td>
<td>0.02</td>
<td>0.02</td>
<td>0.015</td>
<td>0</td>
<td>0.023</td>
<td>0.023</td>
</tr>
<tr>
<td>( f_4 )</td>
<td>0.812</td>
<td>0.64</td>
<td>0.909</td>
<td>0.914</td>
<td>0.785</td>
<td>0.67</td>
<td>0.912</td>
<td>0.914</td>
<td>0.847</td>
<td>0.674</td>
<td>0.918</td>
<td>0.919</td>
<td>0.863</td>
<td>0.706</td>
<td>0.937</td>
<td>0.937</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
\kappa_n &= n/\ln n \\
\kappa_n &= n^{2/3}
\end{align*}
\]
First, let us focus on the size control. Under DGP I, the rejection probabilities should be no larger than 10% for $f_1$, $f_2$, and $f_3$, while under DGP II, the rejection probabilities should be no larger than 10% for $f_1$ and $f_2$. There are several interesting findings. First, the MBs perform very well. If $\kappa_n = n/\ln n$, they have the right size control in both DGPs. In particular, with $f_1$ under DGP II, the rejection probabilities are very close to the nominal size 10%. While under DGP I with $f_1$, the rejection probability is also not far from 10%. This is mostly true with $\kappa_n = n^{2/3}$ except that in some cases, there is slight over-rejection. Second, MB1 performs significantly better than LF and GLS. The latter two can be very conservative. This highlights the importance of including multiple points in $I_n$. Third, MB1 performs similarly to MB2 although it has much fewer points in $I_n$. This suggests that the method is not sensitive to the choice of $I_n$ as long as both $(0, \kappa_n)$ and $\left(\sqrt{\kappa_n/n}, 0\right)$ are included. Fourth, the test does not seem to be sensitive to the choices of $B$ and the number of moments in terms of size control.

Now let us turn to power. First, notice that GLS and LF have much lower power compared to MB1 and MB2. The power difference can be more than 20%. This again highlights the power gain by an $I_n$ with multiple elements. Second, MB1 and MB2 have almost identical performance. This is good news for computationally intensive problems because one can set $I_n = \{(0, \kappa_n), (\sqrt{\kappa_n/n}, 0)\}$ without losing much. Third, $B$ has a limited impact on the results. Fourth, using more moments may decrease (against $f_3$ in DGP II) or increase (against $f_4$ in DGP II) power in finite samples. This is not surprising. Intuitively, if one includes informative moments, power increases. Otherwise, power can decrease. Lastly, a smaller $\kappa_n$ increases the power of the test in certain cases. This is because a smaller $\kappa_n$ increases the size of the imposed identified set and hence can lead to smaller critical values. But it is worth noting that this can also lead to over-rejection under the null.

### 6.2 Asymmetric Ascending Auctions

**DGP III** Now consider Example 2.3 where $F_s$ and $F_w$ are beta distributions with parameters $(2, 3)$ and $(2, 3.5)$ respectively. The true value of $r$ in the Gaussian Copula function is 0.8. Each auction has two bidders. One of them is strong and the other is weak. This DGP closely follows the empirical application. The goal here is to construct a confidence interval.
for $F_s$ evaluated at 0.5, whose true value is 0.6875.

I generate 500 auctions and use B-splines with knots $\{0, 0, \frac{1}{6}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{5}{6}, 1, 1\}$ to approximate $F_s$ and $F_w$. Because $F_s$ and $F_w$ are distribution functions, I impose that they are weakly increasing and $F_s(0) = F_w(0) = 0$, $F_s(1) = F_w(1) = 1$. I impose the FOSD restriction, i.e., $F_s(v) \leq F_w(v)$ for all $v$. I also impose that bidders’ valuations are positively correlated and that the correlation is not too low, i.e., $r \geq 0.5$.\(^{11}\)

Given $\theta$, theoretic moments predicted by the model involve integrals that do not have closed forms. I approximate the theoretic moments by Monte Carlo integration with 30,000 simulated auctions. The measure $\mu$ is chosen to be the uniform probability measure on $T = [-1, 1]$. I approximate the integrals under $\mu$ by averaging over 100 equally spaced points in $[-1, 1]$. For the bootstrap statistic, I consider GLS, LF, and MB1 as in the previous subsection. I report MB3 with $I_n = \{(0, n/\ln n), (\sqrt{1/\ln n}, 0), (\sqrt{1/2\ln n}, n/2\ln n)\}$ instead of MB2 because of computational difficulties. As in Section 6.1, I restrict the coefficients on the B-splines to be less than $B$ in absolute values. I only consider $B = 5$ and $B = 100$.

Again, $\eta = 0$ and $\eta = 10^{-8}$ lead to identical rejection probabilities. To assess the sensitivity of my method to $\kappa_n$, I consider $\kappa_n = n^{2/3}$ and $n/\ln n$. Critical values are computed based on 200 bootstrap statistics and results are obtained from 500 independent samples. The nominal size of these tests is 10%.

Figure III reports rejection probabilities against different values of $F_s(0.5)$. The first row is constructed under $B = 5$. If $\kappa_n = n/\ln n$ as shown in Figure IIIa, all four tests have correct size control. Similar to the results from DGP I and DGP II, MB1 and MB3 are less conservative than GLS and LF under the null and have much better power. Including both $(0, \kappa_n)$ and $\left(\sqrt{\kappa_n/n}, 0\right)$ in $I_n$ significantly improves power. Interestingly, LF seems to be very conservative and does not have any power against alternatives larger than the true value. Also, notice that MB1 and MB3 have identical rejection probabilities, i.e., adding $\left(\sqrt{1/2\ln n}, n/2\ln n\right)$ does not seem to affect the performance of the test. Figure IIIb is constructed under $\kappa_n = n^{2/3}$. It is very similar to Figure IIIa. Figure IIIc compares the rejection probabilities of MB1 under $\kappa_n = n/\ln n$ and $\kappa_n = n^{2/3}$. The performance of the

\(^{11}\)Fan, He, and Li (2015) also use such parametric restrictions on copula functions to partially identify primitives in auctions with correlated values.
test is not sensitive to $\kappa_n$ under this DGP. Overall, the test seems to perform slightly better with $\kappa_n = n/\ln n$. This finding is suggestive on the choice of $\kappa_n$ in Section 7 because the application is very close to DGP III.

The second row of Figure III is constructed with $B = 100$. It is similar to the first row. In particular, the rejection probabilities are very to those under $B = 5$, and the performance of the tests are not sensitive to the choice of $\kappa_n$.

It is worth noting that accounting for partial identification is desirable here because the model is only partially identified. At the same time, the dimension of the optimization problem is relatively high (17 parameters) and the moment conditions are nonlinear and difficult to evaluate. Solving an optimization problem with a dimension twice that of the sieve space can be very time consuming. For such applications, the method proposed in this paper can be very attractive.

To sum up, all the simulation results suggest that MB1 works very well. Adding more points to $I_n$ only slightly improves the performance of the test. In addition, $\kappa_n = n/\ln n$ seems to be a good choice and results are not sensitive to the value of $B$. 


Figure III: Rejection Probability with a 10% Significance Level Test. GLS: $I_n = \{(\sqrt{\kappa_n/n}, 0)\}$; LF: $I_n = \{(0, \kappa_n)\}$; MB1: $I_n = \{(0, \kappa_n), (\sqrt{\kappa_n/n}, 0)\}$; and MB3: $I_n = \{(0, n/\ln n), (\sqrt{1/\ln n}, 0), (\sqrt{1/2 \ln n}, n/2 \ln n)\}$. And B is the bound imposed on the parameter space. True value is 0.6875.
7 Empirical Application: USFS Ascending Auctions

This section applies my method to infer marginal valuation distributions in ascending auctions held by the USFS. In these auctions, bidders are timber firms and the auctioned goods are timber tracts. Bidders in these auctions are asymmetric: big firms and firms with manufacturing capacity are more likely to have higher values compared to small firms and firms without manufacturing capacity. The former are thus strong bidders and the latter are weak bidders. For a discussion on the asymmetry, see Athey, Levin, and Seira (2011). In addition, these bidders have correlated valuations, and ignoring this correlation can lead to inconsistent inference. For example, see Aradillas-López, Gandhi, and Quint (2013) and Aradillas-López, Gandhi, and Quint (2016). Therefore, the econometric framework described in Example 2.3 is appropriate for this analysis.

I focus on auctions from region 5 of USFS that are not set-aside auctions. From each auction, I observe every bidder’s identity, all the bids, and a set of auction characteristics. I classify bidders as strong or weak according to the SBA classification. To simplify the analysis, I use auctions with 1 strong bidder and 1 weak bidder. To eliminate outliers, I exclude auctions with the highest 5% of transaction prices. This gives me a sample with 218 observations. The average transaction price is $51.13 with a standard deviation of $38.82.

Following Aradillas-López, Gandhi, and Quint (2013), I assume that the transaction price equals the second highest valuation among all bidders. To keep the illustration simple, I do not include any auction-level characteristics.

The goal is to construct a point-wise confidence band for $F_s$. I implement my method as described in Section 6.2. The only difference is that now the support of the marginal valuation distributions is the interval between the lowest and the highest transaction prices observed in the data. In light of the findings in Section 6, I set $\kappa_n = n / \ln n$, $I_n = \{(0, n / \ln n), (\sqrt{1 / \ln n}, 0), (\sqrt{1 / 2 \ln n}, n / 2 \ln n)\}$, and $\eta = 0$. I also set $B = 5$ because in all simulations in Section 6, the value of $B$ does not seem to change results much, and a larger $B$ typically leads to longer computing time. As a comparison, I also consider a version of my test without imposing FOSD.

The results are shown in Figure IV. The dashed lines with circle markers are the upper
and lower bounds of the point-wise 90% confidence set with FOSD, while the ones with cross markers are bounds of the confidence set without FOSD. The confidence set with FOSD is much narrower compared to the one without FOSD. I also compute the estimated value distribution assuming symmetry and independence, which is shown by the red curve. Imposing FOSD allows me to exclude the red curve, at least at certain points, from the conference set. Without FOSD, I cannot exclude any part of the red curve. This shows that impose the shape restriction (FOSD) can greatly improve inference.

8 Conclusion

This paper proposes a test of a shape restriction on an infinite-dimensional parameter that is identified by unconditional moment equalities. This test can be inverted to construct confidence sets. It is very general and has many applications. One attractive feature of the test is that it exploits convexity to strike a balance between computing time and the need to approximate the local parameter space. Specifically, if partial identification is a concern, my method involves optimization problems that have a much lower dimension compared with the method proposed in CNS. This can be crucial if the dimension of the sieve space is high and/or in structural estimation where moments are difficult to evaluate. I illustrate my method in an empirical application to ascending auctions. The results show that imposing
shape restrictions can greatly improve inference.

References


A Notation List

The parameter space $\Theta$ is equipped with a norm $\| \cdot \|$ and $d_s$ is the metric induced by $\| \cdot \|$. Let $\| \cdot \|_s$ be a chosen norm on the parameter space $\Theta$ under which the convergence property of $\hat{\Theta}_n(R)$ is important. Denote the metric induced by $\| \cdot \|_s$ as $d_s$. Notice $\| \cdot \|_{s}$ can depend on $F$. For example, if $\theta = (\theta^P, \theta^N)$ where $\theta^P$ is a finite-dimensional real vector and $\theta^N$ is a vector-valued function on $W$, then a common choice is

$$\| \theta \|^2_F = \| \theta^P \|^2_E + \| \theta^N \|^2_{2,F},$$

where $\| \cdot \|_E$ is the Euclidean norm and $\| \theta^N \|^2_{2,F} = \int \| \theta_N(W) \|^2_E dF$ is the usual $L^2$-norm under the distribution $F$. Also define

$$\Theta_{n,F}(R) = \arg \min_{\theta \in \Theta_n \cap R} Q_F(\theta)$$

to be the set of minimizers of the population criterion function on the sieve space. In addition, define the projection of $\theta \in \Theta$ onto $\Theta_{n,R}$ under $d_s$ as

$$\Pi_{n,R} \theta = \arg \min_{\theta \in \Theta_n \cap R} d_s\left( \tilde{\theta}, \theta \right).$$

Table 2: Some Important Notations

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B_F^\epsilon(\theta)$</td>
<td>${ \theta \in \Theta : d_F(\theta, \Theta) \leq \epsilon }$</td>
</tr>
<tr>
<td>$\tilde{d}_F(A, B)$</td>
<td>$\sup_{a \in A} \inf_{b \in B} d_F(a, b)$, directed Hausdorff metric between $A, B \subset \Theta$</td>
</tr>
<tr>
<td>$d_F(A, B)$</td>
<td>$d_F({a}, B)$ for $a \in \Theta$ and $B \subset \Theta$</td>
</tr>
<tr>
<td>$d_{w,F}(a,b)$</td>
<td>$\sqrt{[\mathbb{E}_F \rho_t(W_i, a) - \mathbb{E}_F \rho_t(W_i, b)]^2} d\mu(t)$ for $a, b \in \Theta$</td>
</tr>
<tr>
<td>$\tilde{d}_{w,F}(A, B)$</td>
<td>$\sup_{a \in A} \inf_{b \in B} d_{w,F}(a, b)$ for $A, B \subset \Theta$</td>
</tr>
<tr>
<td>$d_{w,F}(a,b)$</td>
<td>$\tilde{d}_{w,F}({a}, B)$ for $a \in \Theta$ and $B \subset \Theta$</td>
</tr>
<tr>
<td>$\mathbb{E}^*$</td>
<td>Expectation under the empirical distribution.</td>
</tr>
<tr>
<td>$\Theta_F^\epsilon$</td>
<td>${ \theta \in \Theta : d_F(\theta, \Theta_F) \leq \epsilon }$</td>
</tr>
<tr>
<td>$\Theta_{n,F}^\epsilon(R)$</td>
<td>${ \theta \in \Theta_n \cap R : d_F(\theta, \Theta_{n,F}(R)) \leq \epsilon }$</td>
</tr>
<tr>
<td>$R_{n,F}^\epsilon(\theta)$</td>
<td>$\Theta_n \cap R \cap B_{F}^\epsilon(\theta)$</td>
</tr>
<tr>
<td>$\mathbb{1}(\cdot)$</td>
<td>Indicator function</td>
</tr>
<tr>
<td>$a \vee b$</td>
<td>Maximum of $a$ and $b$</td>
</tr>
<tr>
<td>$\downarrow$</td>
<td>Converge from above</td>
</tr>
<tr>
<td>$\uparrow$</td>
<td>Converge from below</td>
</tr>
<tr>
<td>$\rightarrow$</td>
<td>Converge to</td>
</tr>
<tr>
<td>$\rightarrow^d$</td>
<td>Converge in distribution</td>
</tr>
<tr>
<td>$\rightarrow^{d^*}$</td>
<td>Converge in distribution under the empirical distribution</td>
</tr>
</tbody>
</table>

Let $\{\delta_n\}_{n=1}^\infty$ be a sequence of non-negative numbers and $X_{n,F}$ be a sequence of random variables indexed
by $n$ and $F$. Say $X_{n,F} = O_{p_F}(\delta_n)$ uniformly in $F \in \mathcal{F}$ if and only if (iff) for every $\epsilon > 0$, there exists $M < \infty$ such that $\limsup_{n \to \infty} \sup_{F \in \mathcal{F}} \mathbb{P}_{F}(|X_{n,F}| > M\delta_n) < \epsilon$ and $X_{n,F} = o_{p_F}^\alpha(\delta_n)$ uniformly in $F \in \mathcal{F}$ if for every $\epsilon > 0$, there exists $M < \infty$ such that $\limsup_{n \to \infty} \sup_{F \in \mathcal{F}} \mathbb{P}_{F}(|X_{n,F}| > M\delta_n) = 0$. Similarly, $X_{n,F,R} = O_{p_F}(\delta_n)$ uniformly in $(F,R) \in \mathcal{J}$ if for every $\epsilon > 0$, there exists $M < \infty$ such that $\limsup_{n \to \infty} \sup_{(F,R) \in \mathcal{J}} \mathbb{P}_{F}(X_{n,F,R} > M\delta_n) < \epsilon$ and $X_{n,F,R} = o_{p_F}^\alpha(\delta_n)$ uniformly in $(F,R) \in \mathcal{J}$ if for every $\epsilon > 0$, there exists $M < \infty$ such that $\limsup_{n \to \infty} \sup_{(F,R) \in \mathcal{J}} \mathbb{P}_{F}(X_{n,F,R} > M\delta_n) = 0$. Because $(F,R) \in \mathcal{J}$ implies $F \in \mathcal{F}$, $X_{n,F} = O_{p_F}(\delta_n) (X_{n,F} = o_{p_F}^\alpha(\delta_n))$ uniformly in $(F,R) \in \mathcal{J}$ if $X_{n,F} = O_{p_F}(\delta_n) (X_{n,F} = o_{p_F}^\alpha(\delta_n))$ uniformly in $F \in \mathcal{F}$. Define the first-order directional derivative of $E_{\mathcal{F}} \rho_t(W_i, \cdot)$ at $\theta$ in the direction $\Delta$ as

$$\frac{dE_{\mathcal{F}} \rho_t(W_i, \theta)}{d\theta}[\Delta] = \lim_{\tau \downarrow 0} \frac{E_{\mathcal{F}} \rho_t(W_i, \theta + \tau \Delta) - E_{\mathcal{F}} \rho_t(W_i, \theta)}{\tau}.$$ 

Let $\ell^\infty(\Theta \times \mathcal{T})$ be the space of bounded functions on $\Theta \times \mathcal{T}$ equipped with the standard sup norm $\|\cdot\|_\infty$. Let $d_{\infty}^m$ be the metric induced by $\|\cdot\|_\infty$. $\Gamma_{\kappa,F,m,R} : \ell^\infty(\Theta \times \mathcal{T}) \to \mathbb{R}$ are functions indexed by $\kappa \geq 0, F \in \mathcal{F}, m \in \mathbb{Z}^+, R \in \mathcal{R}$.

$$\Gamma_{\kappa,F,m,R}(\omega) = \inf_{\hat{\theta} \in \Theta_{m,F}(R)} \inf_{\hat{\theta} \in B_F(\hat{\theta}) \cap \Theta_m \cap \mathcal{R}} \int_{\mathcal{T}} \omega(\hat{\theta}, t) + \sqrt{m} \frac{dE_{\mathcal{F}} \rho_t(W_i, \hat{\theta})}{d\theta} \left[\theta - \hat{\theta}\right]^2 d\mu(t).$$

Notice that if $\kappa_1 > \kappa_2$, $\Gamma_{\kappa_1,F,m,R}(\omega) \leq \Gamma_{\kappa_2,F,m,R}(\omega)$ because inf in the curly bracket is taken over a larger set under $\kappa_1$. Later on, I show that the test statistic and the bootstrap statistic can be approximated asymptotically by $\Gamma_{\kappa,F,m,R}(\cdot)$ applied to the empirical process and the bootstrap empirical process, respectively.

Lastly, denote the weak limit of $G_{n,F}(\theta, t)$ by $G_F$. And for any function $g : \ell^\infty(\Theta \times \mathcal{T}) \to \mathbb{R}$, let $q(1 - \alpha, g, F)$ be the $1 - \alpha$-th quantile of $g(G_F)$ and $q^*_n(1 - \alpha, g)$ be the $1 - \alpha$-th quantile of $g(G^*_n)$.

### B Identification Power of Shape Restrictions

It is commonly known that shape restrictions can play an important role for identification in structural models, for example in BLP estimation and auctions, see Gandhi, Lu, and Shi (2012); Zhu and Grundl (2014); Komarova (2013); and Fan, He, and Li (2015). But even in the simple non-parametric IV regression model, shape restrictions can greatly help identification. In extreme cases, without shape restrictions, the parameter of interest is not identified at all with moment conditions, i.e., one cannot rule out any point from the parameter space. But once a shape restriction is imposed, the parameter of interest is point-identified. This section provides such an example.

Consider the setup in Example 2.1 where the joint distribution of $(X, Y, Z)$ is such that

$$Z \sim \text{Uniform}[0, 1], \ X | Z \sim \text{Uniform}[Z, Z + 1].$$

(16)
And $\epsilon$ can be arbitrarily correlated with $X$ but satisfies $\mathbb{E}_F (\epsilon | Z) = 0$. For example $\epsilon = Z (X - Z + \epsilon_1 - 1/2)$, where $\epsilon_1$ is a random variable with mean 0 and independent of both $X$ and $Z$. From this point on, I restrict $\theta_F$ to be a bounded continuous function and the parameter of interest is $\theta_F (x_0)$ for some $x_0 \in [0, 2]$.

Suppose that one uses the conditional moment restriction

$$
\mathbb{E}_F [Y - \theta (X) | Z] = 0.
$$

(17)

Then, the identified set of $\theta_F (x_0)$ is $(\infty, \infty)$. To see this, notice that $\theta = \hat{\theta} = \hat{\theta} + \hat{\theta}$ satisfies (17) iff $\mathbb{E}_F [\hat{\theta} (X) | Z] = 0$, which holds iff $\hat{\theta}$ is a periodic function with period 1 and $\int_0^1 \hat{\theta} (x) \, dx = 0$. For any real number $a$, there exists $\hat{\theta}$ with $\hat{\theta} (x_0) = a$. Hence, $\theta_F (x_0)$ can take any value in $(\infty, \infty)$.

Now suppose $\theta_F$ is weakly increasing and is constant on $[b, c] \subseteq [0, 2]$ with $c - b \geq 1$. After imposing the weakly increasing restriction, $\theta_F (x_0)$ is point-identified. To see this, first notice $\hat{\theta}$ has to be a constant function. Otherwise, it must decrease in some region of $[b, c]$ because it has period 1, which means $\theta_F + \hat{\theta}$ violates the increasing restriction. In addition, because $\hat{\theta}$ satisfies $\int_0^1 \hat{\theta} (x) \, dx = 0$, it must be that $\hat{\theta} = 0$. This implies that $\theta_F$ is point-identified.

C Assumptions for Uniform Size Control

**Assumption C.1.** (i) For every $R \in \mathcal{R}$, $\Theta \cap R$ is compact under $d_s$; (ii) there exist $C_1, C_2 \in (0, \infty)$ such that $C_1 d_{w,F} \leq d_F \leq C_2 d_s$ for every $F \in \mathcal{F}$.

**Assumption C.2.** The sieve space $\Theta_n$ satisfies the following conditions. (i) For every $n$, $\Theta_n \subseteq \Theta$ is finite-dimensional, convex, and closed under $d_s$. (ii) As $n \to \infty$, $\sup_{R \in \mathcal{R}} \sup_{\theta \in \Theta \cap R} d_s (\theta, \Theta_n \cap R) = o (1)$. (iii) As $n \to \infty$, $\sup_{R \in \mathcal{R}} \sup_{\theta \in \Theta \cap R} d_s (\theta, \Theta_n \cap R) = o (n^{-1/2})$.

**Assumption C.3.** (i) For any $\epsilon > 0$, the following quantity

$$
S_n (\epsilon) = \sup_{(F,R) \in \mathcal{J}} \left[ \inf_{\theta \in \Theta_n \cap R} \inf_{d_F (\theta, \Theta_n, F (R)) > \epsilon} \mathbb{Q}_F (\theta) - \inf_{\theta \in \Theta_n \cap R} \mathbb{Q}_F (\theta) \right]
$$

satisfies that $\sqrt{n} S_n (\epsilon) \to \infty$ if $n \to \infty$. (ii) There exists an $\epsilon > 0$ and $\nu \geq 1$ such that

$$
\zeta_n = \sup_{(F,R) \in \mathcal{J}} \sup_{\theta \in \Theta_n, F (R)} d_F (\theta, \Theta_n, F (R))^{\nu}
$$

satisfies that $\frac{d \zeta_n}{d \epsilon} [\theta_2 - \theta_1]$ exists for every $\theta_1, \theta_2 \in \Theta_F$, for every $n > 0$.

**Assumption C.4.** (i) There exists $\epsilon > 0$ such that $\frac{d \mathbb{E}_F W (\theta_1)}{d \theta} [\theta_2 - \theta_1]$ exists for every $\theta_1, \theta_2 \in \Theta_F$.
\(t \in \mathcal{T}\), and \(F \in \mathcal{F}\). (ii) There exist \(\epsilon > 0\) and \(\xi_n > 0\) such that \(\xi_n \to 0\), \(\sqrt{n}\xi_n \to \infty\) as \(n \to \infty\) and

\[
\sup_{(F,R) \in \mathcal{J}} \sup_{\theta_1, \theta_2 \in \Theta_{\alpha}(R)} \sup_{(W_i, \theta_1, \theta_2) \in \xi_n, t \in \mathcal{T}} \left| \mathbb{E}_F \rho_t(W_i, \theta_2) - \mathbb{E}_F \rho_t(W_i, \theta_1) - \frac{d \mathbb{E}_F \rho_t(W_i, \theta_1)}{d \theta} [\theta_2 - \theta_1] \right| = o \left( n^{-1/2} \right).
\]

(iii) There exists \(C_D < \infty\) such that for all \(F \in \mathcal{F}\)

\[
\sup_{t \in \mathcal{T}} \mathbb{E}_F [\rho_t(W_i, \theta_1) - \rho_t(W_i, \theta_2)]^2 \leq C_D d_F(\theta_1, \theta_2)^2.
\]

**Assumption C.5.** The tuning parameter \(\kappa_n\) satisfies \(\kappa_n \to \infty\), \(\kappa_n \ln \ln n \to 0\), and \(\zeta_n / \sqrt{n} = o(\xi_n^\epsilon)\).

**Assumption C.6.** (i) \(\Theta \cap S\) is convex and compact under \(d_s\). (ii) \(\phi\) is linear and \(\phi(\theta_F)\) lives in a known compact set \(K \subset \mathbb{R}^D\), for all \(F \in \mathcal{F}\). (iii) \(d_s(\theta, \Theta \cap R_{n,\phi,S}) = o(n^{-1/2})\).

**Assumption C.7.** Let \(\Gamma_{m,J} = \Gamma_{\xi_m,F,m,R}\) where \(J = (F,R)\) and let \(C_F(1-\alpha, \Gamma_{m,J})\) be the \(1-\alpha\)-th quantile of \(\Gamma_{m,J}(G_F)\) and

\[
\mathcal{J}_n^\epsilon(\alpha) = \{ J \in \mathcal{J} : C_F(1-\alpha, \Gamma_{m,J}) > \epsilon \text{ if } m \geq n \}.
\]

(i) For every \(\epsilon > 0\), if \(\eta \downarrow 0\), then

\[
\limsup_{n \to \infty} \sup_{J \in \mathcal{J}_n^\epsilon(\alpha)} \mathbb{P}_F \left( C_F(1-\alpha - \eta, \Gamma_{n,J}) - \eta < \Gamma_{n,J}(G_F) \leq C_F(1-\alpha + \eta, \Gamma_{n,J}) + \eta \right) \to 0.
\]

(ii) For every \(\epsilon_n > 0\) and \(\epsilon_n \to 0\), \(\limsup_{n \to \infty} \sup_{(F,R) \in \mathcal{J} \setminus \mathcal{J}_n^{\epsilon_n}(\alpha)} \mathbb{P}_F(T_n(R) > 0) < \alpha\).

Assumption C.1(i) holds after regularizing the parameter space following Santos (2012). As in Santos (2012), the choice of \(\Theta\) depends on \(R\). For example, if \(R\) involves derivatives of \(\theta\), \(\Theta\) needs to be chosen accordingly. For more details, see Section D. Assumption C.1(ii) is satisfied with many commonly used norms. For example, if \(d_s\) is the sup metric and \(d_F\) is induced by the \(L^2\) norm under \(F\), then standard smoothness assumptions on \(\rho_t(W_i, \theta)\) imply Assumption C.1(ii). Assumption C.1 ensures that \(Q_F(\cdot)\) is continuous under \(d_s\) and that \(\inf_{\theta \in \Theta \cap R} Q_F(\theta) = \min_{\theta \in \Theta \cap R} Q_F(\theta)\). It implies that \(\min_{\theta \in \Theta \cap R} Q_F(\theta) > 0\) under any fixed alternative, which is crucial for the power of the test. Assumption C.2(i) is a mild requirement. Together with Assumption C.1(i), it implies the compactness of \(\Theta_n \cap R\) under \(d_s\) and hence \(\Theta_n(R)\) is well-defined.\(^\text{12}\) Convexity of \(\Theta_n \cap R\) can be relaxed if \(\frac{d \mathbb{E}_F \rho_t(W_i, \theta)}{d \theta} [\Delta]\) is linear in \(\Delta\) and uniformly bounded. Assumptions C.2(ii) and (iii) require the sieve spaces to approximate \(\Theta \cap R\) well enough under \(d_s\) uniformly in \(R \in \mathcal{R}\). Under Assumption C.1(ii), this means that the sequence of sieve spaces approximates \(\Theta \cap R\) well under \(d_F\) uniformly in \((F,R) \in \mathcal{J}\). If \(d_s\) is the sup metric and one regularizes \(\Theta\) following Newey and Powell (2003) and Santos (2012), then the sieve spaces can be spanned by orthogonal polynomials or splines with sufficiently many orders or knots. See

\(^\text{12}\)If one is only concerned with size control, one can directly assume that \(\Theta_n \cap R\) is compact for every \(n\) and \(R \in \mathcal{R}\). Then Assumption C.1(i) is not needed.
Newey (1997) for more discussion. Notice, unlike in Hong (2017) and Santos (2012), there is no restriction on how the boundary points of $\Theta$ can be approximated by interior points.

Assumption C.3(i) is similar to Assumption 4.1(ii) in CNS. It allows that as $n$ increases, $\Theta_{n,F} (R)$ becomes less well separable, i.e., $S_n(\epsilon)$ can converge to 0 for a given $\epsilon$. However, the convergence rate must be slower than $\sqrt{n}$. This assumption rules out weak identification. It, however, does not require $\tilde{d}_F (\Theta_{n,F} (R), \Theta_F \cap R) \to 0$ as $n \to \infty$, i.e., the minimizers of $Q_F (\theta)$ on the sieve space are allowed to be far away from the true identified set under $d_F$.

Assumption C.3(ii) is otherwise a standard assumption in the sieve literature except that I also take sup across all $F \in \mathcal{F}$. Notice that $d_{w,F}$ is related to the weak norm considered in Ai and Chen (2003). There are several differences. First, since I am focusing on models with a large number of unconditional moments, $d_{w,F}$ is based on the integration of these unconditional moments under $\mu$. Second, $d_{w,F}$ does not allow for a general weight matrix, but it can be extended in this direction. Third, $d_{w,F}$ does not involve the directional derivatives at the true value of the unknown parameter. This is more convenient for models with partial identification. It is also worth pointing out that in non-parametric IV models, one can choose weight functions and $\mu$ such that $d_{w,F}$ is arbitrarily close to the weak norm defined in Ai and Chen (2003). Also, $\zeta_n$ is closely related to the measure considered in CNS, who transform conditional moments into countably infinitely many moments using basis functions. In fact, given the set of basis functions, one can construct $w$ and $\mu$ such that my method essentially uses the same set of unconditional moments as in CNS. Then $\zeta_n$ with $\nu = 1$ is different from its counterpart in CNS under $r = 2$ in two ways. First, it uses all the moments for any finite $n$, while CNS use a finite subset of these moments. Second, within the finite subset of moments used in CNS, $\zeta_n$ typically give each moment a different weight compared to CNS.

If $\mathcal{F}$ is known, $\zeta_n$ can be directly calculated. Alternatively, if $\mathcal{F} = \{F_0\}$ and $\nu$ is known, one may estimate $\zeta_n$ by replacing $d_F$ and $d_{w,F}$ by their sample analog and replacing $\Theta_{n,F} (R)$ by a consistent estimator. As a special case, if $\nu = 1$, $d_F$ is the $L^2$ norm, $\rho_k$ is linear in $\theta$ for all $t$, and $\Theta_n$ is a linear space with a finite number of basis functions, then one can estimate $\zeta_n$ following the method proposed in Chen and Christensen (2015) if $\theta_F$ is point-identified.

Assumptions C.4(i) and (ii) imply that the moment conditions can be approximated by their linear expansion in a neighborhood of $\Theta_{n,F} (R)$. Combined with other assumptions, it rules out cases where all directional derivatives are 0 at the boundary of $\Theta_{n,F} (R)$, i.e., local identification failure as considered in Dovonon and Renault (2013) and Lee and Liao (2018). Assumption C.4(iii) accompanied by Definition 5.1(3) implies that the empirical process $G_{n,F} (\theta, t)$ is asymptotically equi-continuous in $\theta$.

Assumptions C.1-C.4 are sufficient to guarantee that the test statistic is approximated asymptotically by $\Gamma_{\zeta_n,F,n,R} (G_{n,F})$ under the null hypothesis. Assumptions C.5 is needed to validate the bootstrap statistic. It guarantees that the rescaled moments can be approximated by their first-order expansion around the identified set. If the model is linear in $\theta$, $\xi_n$ can be any $o(1)$ sequence and $\kappa_n$ can be any sequence that does not diverge too fast. If the model is nonlinear and $\nu = 1$, usually $\xi_n$ can be any $o(n^{-1/4})$ sequence.
Then \( \kappa_n \) needs to satisfy \( \zeta_n/\sqrt{n} = o(n^{-1/4}) \) and \( \kappa_n \ln \ln n/n \to 0 \). In some cases, such \( \kappa_n \) may not exist. In particular, these requirements rule out severely ill-posed nonlinear models. Similar assumptions can also be found in the semi-parametric conditional moment literature if endogenous variables enter the unknown functions. For an example, see Chen and Pouzo (2009). It is worth noting that if \( I_n = \{(0, \kappa_n)\} \), this assumption can be relaxed to only require \( \zeta_n/\sqrt{n} = o(1) \). Assumption C.6 guarantees that the confidence set for \( \phi(\theta_F) \) is valid. It is a version of Assumption C.2 with \( \mathcal{R} = \{R_{a,0,S} : a \in \mathbb{K}\} \).

Assumption C.7 is required only if one would like to set \( \eta = 0 \). Assumption C.7(i) holds if there exists \( \delta > 0 \) such that uniformly in \( J \in \mathcal{J} \) and sufficiently large \( m \), \( \Gamma_{m,J}(G_F) \) has a density bounded away from 0 and from above on \( ((C_F (1 - \alpha, \Gamma_{m,J}) - \delta) \lor 0, C_F (1 - \alpha, \Gamma_{m,J}) + \delta) \). Assumption C.7(ii) handles the case where \( C_F (1 - \alpha, \Gamma_{m,J}) \) is equal to or arbitrarily close to 0. It is related to Assumption A.2 in Bugni, Canay, and Shi (2017).

### D Choices of \( \Theta \) and Low-Level Sufficient Conditions for Definition 5.1(3)

Define \( \Lambda = (\Lambda_1, \Lambda_2, \cdots, \Lambda_{D_W}) \) to be a \( D_W \)-dimensional vector of non-negative integers and \( w = (w_1, w_2, \cdots, w_{D_W}) \in \mathcal{W} \) to be a \( D_W \)-dimensional vector of real numbers. Let \( |\Lambda| = \sum_{i=1}^{D_W} \Lambda_i \) and \( d^A f = \partial^{|\Lambda|} f(w)/\partial^{\Lambda_1}w_1 \cdots \partial^{\Lambda_{D_W}}w_{D_W} \) where \( f: \mathcal{X} \to \mathbb{R} \) is a function. Introduce the norms

\[
\|f\|_c = \max_{|\Lambda| \leq m} \sup_{w \in \mathcal{W}} \left| d^A f(w) \right| (1 + w_1^r)^{\delta/2},
\]

\[
\|f\|_{sob}^2 = \int \max_{|\Lambda| \leq m+m_0} \left| d^A f(w) \right|^2 (1 + w_1^r)^{\delta_0} dw,
\]

where \( m \) and \( m_0 \) are two integers. Let \( \mathcal{E}^{sob}_B(\mathcal{W}) = \{ f : \mathcal{W} \to \mathbb{R} \text{ s.t. } \|f\|_{sob} \leq B \} \) where \( 0 < B < \infty \) is a known constant. Define \( \Theta = \Theta^P \times \mathcal{E}^{sob}_B(\mathcal{W})^{D_N} \) where \( \Theta^P \) is a compact subset of \( \mathbb{R}^{D_P} \) and \( D_N \) is a positive integer. In addition, let \( \theta = (\theta^P, \theta^N) \) where \( \theta^P \in \Theta^P \) and \( \theta^N \in \mathcal{E}^{sob}_B(\mathcal{W})^{D_N} \). Define

\[
\|\theta\|_s = \|\theta\|_{\infty} = \max \left\{ \|\theta^P\|_{\infty,E}, \sup_{w \in \mathcal{W}} \|\theta^N(w)\|_{\infty,E} \right\}
\]

\[
= \max \left\{ \|\theta^P\|_{\infty,E}, \sup_{w \in \mathcal{W}} \|\theta^1_N(w)\|, \sup_{w \in \mathcal{W}} |\theta^2_N(w)|, \cdots, \sup_{w \in \mathcal{W}} |\theta^N_D(w)| \right\},
\]

where \( \|\cdot\|_{\infty,E} \) is the sup norm on the Euclidean space and \( \theta^N_j(w) \) is the \( j \)-th element of \( \theta^N(w) \). The following is Assumption 2.1 in Santos (2012).

**Assumption D.1.** (i) For \( \mathcal{W} \) bounded, \( \delta_0 = \delta = 0 \) and \( \min \{m_0, m \} > D_W/2 \), while for \( \mathcal{W} \) unbounded, \( m_0 > D_W/2 \) and \( D_W (m + \delta) / (m \delta) < 2 \) and \( \delta > \delta > D_W/2 \). (ii) \( \mathcal{X} \) satisfies a uniform cone condition.

Santos (2012) shows that under Assumption D.1, \( \mathcal{E}^{sob}_B(\mathcal{W}) \) is compact under \( \|\cdot\|_c \). Since \( \|\cdot\|_c \) is stronger than the sup norm, \( \mathcal{E}^{sob}_B(\mathcal{W}) \) is compact under the sup norm. Therefore, \( \Theta \) is compact under \( \|\cdot\|_s \) as it is a product.

42
of compact spaces. Moreover, suppose that the shape restriction is

\[ R = \{ \theta : \partial \theta^N(w) / \partial w_1 \geq 0, \forall w \in W \} . \]

Because partial differentiation operators are continuous with respect to \( \| \cdot \|_c \), \( R \) is closed under \( \| \cdot \|_c \). This implies that \( \Theta \cap R \) is compact under \( \| \cdot \|_c \), and therefore compact under \( \| \cdot \|_s \). If \( R \) involves higher-order derivatives, one needs to adjust \( \Theta \) accordingly to satisfy Assumptions C.1(i). Suppose \( t \in T \) is a \( D_t \)-dimensional real vector. Define \( N(\epsilon, \Theta \times T, \| \cdot \|_s + \| \cdot \|_E) \) to be the smallest \( \epsilon \)-covering number of \( \Theta \times T \) under \( \| \cdot \|_s + \| \cdot \|_E \).

**Lemma D.1.** Under Assumption D.1, if \( T \) is a bounded subset of \( \mathbb{R}^{D_t} \), then

\[
\int_0^\infty \sqrt{\ln N(\epsilon, \Theta \times T, \| \cdot \|_s + \| \cdot \|_E)} d\epsilon < \infty. 
\]

**Proof.** Lemma A.3 in Santos (2012) shows that

\[
\ln N(\epsilon, E_{B}^{sob}(W), \| \cdot \|_\infty) \leq K_1 \left( \frac{1}{\epsilon} \right)^{D_W/(m+\delta)} 
\]

if \( W \) unbounded,

\[
\ln N(\epsilon, E_{B}^{sob}(W), \| \cdot \|_\infty) \leq K_2 \left( \frac{1}{\epsilon} \right)^{m/D_W} 
\]

if \( W \) bounded,

where \( N(\epsilon, E_{B}^{sob}(W), \| \cdot \|_\infty) \) is the smallest \( \epsilon \)-covering number of \( E_{B}^{sob}(W) \) under \( \| \cdot \|_\infty \). Notice that

\[
N(\epsilon, \Theta \times T, \| \cdot \|_s + \| \cdot \|_E) < N \left( \frac{\epsilon}{2}, \Theta, \| \cdot \|_s \right) \times N \left( \frac{\epsilon}{2}, T, \| \cdot \|_E \right) 
\]

\[
\leq N \left( \frac{\epsilon}{2}, \Theta^P, \| \cdot \|_{\infty,E} \right) N \left( \frac{\epsilon}{2}, T, \| \cdot \|_E \right) N \left( \frac{\epsilon}{2}, E_{B}^{sob}(W), \| \cdot \|_\infty \right)^{D_N} . 
\]

Therefore, there exists a constant \( K \) such that

\[
\ln N(\epsilon, \Theta \times T, \| \cdot \|_s + \| \cdot \|_E) \leq K \left( \frac{1}{\epsilon} \right)^{D_W/(m+\delta)} 
\]

if \( X \) unbounded,

\[
\ln N(\epsilon, \Theta \times T, \| \cdot \|_s + \| \cdot \|_E) \leq K \left( \frac{1}{\epsilon} \right)^{D_W/m} 
\]

if \( X \) bounded.

Let \( M = \max_{(\theta, t) \in \Theta \times T} (\| \theta \|_s + \| t \|_E) \). Then under Assumption D.1,

\[
\int_0^\infty \sqrt{\ln N(\epsilon, \Theta \times T, \| \cdot \|_s + \| \cdot \|_E)} d\epsilon \leq K \int_0^M \left( \frac{1}{\epsilon} \right)^{\eta} d\epsilon, 
\]

where \( \eta = D_W (m + \delta) / (2\delta m) \) if \( W \) is unbounded and \( \eta = D_W / 2m \) if \( W \) is bounded. Because \( \eta < 1 \) by Assumption D.1, the right-hand side equals \( KM^{1-\eta}/(1-\eta) < \infty \).
Assumption D.2. \( \rho_t \) and \( F \) satisfy (i) \( \rho_t (\cdot, \theta) \) is measurable for all \((\theta, t) \in \Theta \times T\). (ii) \( \sup_{(\theta, t) \in \Theta \times T} |\rho_t (\cdot, \theta)| \leq F (\cdot) \) for some function \( F \). (iii) There exists \( L (\cdot) \) such that \( \sup_{F \in \mathcal{F}} E_F L (W_i)^2 < \infty \) and

\[
|\rho_{t_1} (\cdot, \theta_1) - \rho_{t_2} (\cdot, \theta_2)| \leq L (\cdot) \|	heta_1 - \theta_2\|_s + \|	heta_1 - \theta_2\|_E.
\]

(iii) \( \limsup_{M \to \infty} \sup_{F \in \mathcal{F}} E_F F (W_i)^2 \mathbb{1} \{ F (W_i) > M \} = 0. \)

Assumption D.2 is a uniform version of the standard smoothness assumption on the moment functions. Assumption D.2(iii) is a uniform integrability condition, which holds if \( \sup_{F \in \mathcal{F}} E_F F (W_i)^2 + \delta < \infty \) for some \( \delta > 0. \)

Lemma D.2. Under Assumptions D.1 and D.2, if \( T \) is a bounded subset of \( \mathbb{R}^D_t \), \( \varrho \) is Donsker and pre-Gaussian uniformly in \( F \in \mathcal{F}. \)

Proof. By Assumption D.2(iii),

\[
\|\rho_{t_1} (\cdot, \theta_1) - \rho_{t_2} (\cdot, \theta_2)\|_{2, F}^2 \leq E_F L (W_i)^2 \|	heta_1 - \theta_2\|_s + \|	heta_1 - \theta_2\|_E^2.
\]

Let \( N \) be the bracketing number. By Theorem 2.7.11 in Van Der Vaart and Wellner (1996),

\[
N \left( 2e \left( \|L\|_{2, F} + 1 \right), \varrho, \|\cdot\|_{2, F} \right) \leq N \left( 2e \|L\|_{2, F}, \varrho, \|\cdot\|_{2, F} \right) \leq N (\epsilon, \Theta \times T, \|\cdot\|_s + \|\cdot\|_E).
\]

Without loss of generality, assume that \( F (\cdot) \geq 1. \) This is because one can always redefine the envelope function to be max \{\( F (\cdot), 1 \). Then

\[
\int_0^\infty \sup_{F \in \mathcal{F}} \sqrt{\ln N \left( e \|F\|_{2, F}, \varrho, \|\cdot\|_{2, F} \right)} \, d\epsilon \\
\leq \int_0^\infty \sup_{F \in \mathcal{F}} \frac{\ln N \left( e \|F\|_{2, F}, \Theta \times T, \|\cdot\|_s + \|\cdot\|_E \right)}{2 (\|L\|_{2, F} + 1)} \, d\epsilon \\
\leq \sup_{F \in \mathcal{F}} \int_0^\infty \frac{2 (\|L\|_{2, F} + 1)}{\|F\|_{2, F}} \sqrt{\ln N (e, \Theta \times T, \|\cdot\|_s + \|\cdot\|_E)} \, d\epsilon,
\]

which is bounded by Lemma D.1. Also by Assumption D.2, \( \lim_{M \to \infty} \sup_{F \in \mathcal{F}} E_F F^2 \mathbb{1} (F > M) = 0. \) Then Theorem 2.8.4 in Van Der Vaart and Wellner (1996) implies that \( \varrho \) is Donsker and pre-Gaussian uniformly in \( F \in \mathcal{F}. \)

\( \square \)

E Asymptotic Approximation of the Test Statistic

Theorem E.1. If Assumptions C.1 to C.4 hold, then
(1) Uniformly in \((F, R) \in \mathcal{J}\), \(T_n(R) \leq \Gamma_{\xi_n,F,n,R}(G_{n,F}) + o_{P_F}(1)\), where the inequality holds as an equality if \(\tilde{d}_F\left(\hat{\Theta}_n(R), \Theta_{n,F}(R)\right) = o_{P_F}(\xi_n)\) uniformly in \((F, R) \in \mathcal{J}\).

(2) If \(F \in \mathcal{F}\) and \(\Theta_F \cap R = \emptyset\), \(T_n/n \to \min_{\theta \in \Theta \cap R} Q_F(\theta), \mathbb{P}_F\) almost surely.

**Proof.** For any \((F, R) \in \mathcal{J}\),

\[
T_n(R) = \inf_{\theta \in \Theta \cap R} \int \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left| \rho_i(W_i, \theta) - \mathbb{E}_F \rho_i(W_i, \theta) - \sqrt{n} \mathbb{E}_F \rho_i(W_i, \theta) \right|^2 d\mu(t)
\]

\[
\leq \inf_{\tilde{d} \in \Theta_{n,F}(R)} \inf_{\hat{\Theta} \in \mathbb{R}_n^m(\tilde{d})} \int \left| \mathbb{E}_F (\theta, t) + \sqrt{n} \mathbb{E}_F \rho_t(W_i, \theta) \right|^2 d\mu(t)
\]

\[
= \inf_{\tilde{d} \in \Theta_{n,F}(R)} \inf_{\hat{\Theta} \in \mathbb{R}_n^m(\tilde{d})} \int \left| \mathbb{E}_F (\hat{\theta}, t) + \sqrt{n} \mathbb{E}_F \rho_t(W_i, \hat{\theta}) \right|^2 d\mu(t)
\]

\[
\equiv T_n(R),
\]

where \(\Delta_n(t) = \Delta_n,1(t) + \Delta_n,2(t) + \Delta_n,3(t)\) with

\[
\Delta_n,1(t) = \mathbb{E}_{n,F}(\theta, t) - \mathbb{E}_{n,F}(\hat{\theta}, t),
\]

\[
\Delta_n,2(t) = \sqrt{n} \left( \mathbb{E}_F \rho_t(W_i, \theta) - \mathbb{E}_F \rho_t(W_i, \hat{\theta}) \right) - \frac{d\mathbb{E}_F \rho_t(W_i, \hat{\theta})}{d\theta} [\theta - \hat{\theta}],
\]

\[
\Delta_n,3(t) = \sqrt{n} \mathbb{E}_F \rho_t(W_i, \hat{\theta}).
\]

I suppress the dependence of the \(\Delta\)s on \(F, R, \theta, \) and \(\hat{\theta}\) to simplify notations. The inequality in \((18)\) holds because I shrink the region over which the inf is taken. By the triangular inequality,

\[
\sqrt{T_n(R)} \leq \sqrt{\Gamma_{\xi_n,F,n,R}(G_{n,F}) + \sum_{j=1}^{3} \sup_{\tilde{d} \in \Theta_{n,F}(R)} \sup_{\hat{\Theta} \in \mathbb{R}_n^m(\tilde{d})} \left\| \Delta_{n,j}(\cdot) \right\|_{2,\mu}^2,
\]

where \(\left\| \Delta_{n,j}(\cdot) \right\|_{2,\mu}^2 = \int \Delta_{n,j}(t)^2 d\mu(t)\). Definition 5.1(3), Assumptions C.4(ii), (iii), and Theorem 2.8.2 in Van Der Vaart and Wellner (1996) imply that

\[
\sup_{\tilde{d} \in \Theta_{n,F}(R)} \sup_{\hat{\Theta} \in \mathbb{R}_n^m(\tilde{d})} \left\| \Delta_{n,1}(\cdot) \right\|_{2,\mu} \leq \sup_{d_F(\theta_1, \theta_2) \leq \xi_n, \theta \in \mathcal{T}} \sup \left| G_{n,F}(\theta_1, t) - G_{n,F}(\theta_2, t) \right| = o_{P_F}(1),
\]

uniformly in \(F \in \mathcal{F}\). In addition, under Assumptions C.4(i) and (ii), uniformly in \((F, R) \in \mathcal{J}\)

\[
\sup_{\tilde{d} \in \Theta_{n,F}(R)} \sup_{\hat{\Theta} \in \mathbb{R}_n^m(\tilde{d})} \left\| \Delta_{n,2}(\cdot) \right\|_{2,\mu} \leq \sup_{d_F(\theta_1, \theta_2) \leq \xi_n, \theta \in \mathcal{T}} \sup \left| \Delta_{n,2}(t) \right| = o(1).
\]
Lastly, because \( \hat{\theta} \in \Theta_{n,F}(R) \), Assumption C.1(ii) implies that for any \( \hat{\theta} \in \Theta \cap R \),

\[
\| \Delta_{n,3} \|_{2,\mu}^2 = n \min_{\theta \in \Theta_{n,R}} Q_F(\theta) \leq nQ_F(\Pi_{n,R}\hat{\theta}) = nd_{\varphi,F}(\Pi_{n,R}\hat{\theta}, \hat{\theta})^2 \leq nC_2d_s(\Pi_{n,R}\hat{\theta}, \hat{\theta})^2.
\]

Then Assumption C.2(iii) and the definition of \( \Pi_{n,R}\hat{\theta} \) imply that uniformly in \( (F, R) \in J \),

\[
\sup_{\hat{\theta} \in \Theta_{n,F}(R)} \sup_{\theta \in \Theta_{n,R}^n(\hat{\theta})} \| \Delta_{n,3} \|_{2,\mu} = o(1).
\]

Therefore, \( \sqrt{T_n(R)} \leq \sqrt{\Gamma_{\xi_n,F,n,R}(\mathbb{G}_{n,F}) + o_p(1)} \) uniformly in \( (F, R) \in J \). Similarly, one can show that \( \sqrt{T_n(R)} \geq \sqrt{\Gamma_{\xi_n,F,n,R}(\mathbb{G}_{n,F}) - o_p(1)} \). Therefore, \( \sqrt{T_n(R)} = \sqrt{\Gamma_{\xi_n,F,n,R}(\mathbb{G}_{n,F}) + o_p(1)} \) uniformly in \( (F, R) \in J \). Because \( \hat{\theta} \in \Theta_{n,R}^n(\hat{\theta}) \), \( \sqrt{\Gamma_{\xi_n,F,n,R}(\mathbb{G}_{n,F})} = o_p(1) \) uniformly in \( (F, R) \in J \). Apply Lemma I.2 with \( A = J \) to get \( T_n(R) = \Gamma_{\xi_n,F,n,R}(\mathbb{G}_{n,F}) + o_p(1) \) uniformly in \( (F, R) \in J \), which implies that \( T_n(R) \leq \Gamma_{\xi_n,F,n,R}(\mathbb{G}_{n,F}) + o_p(1) \) uniformly in \( (F, R) \in J \). If, in addition, \( \tilde{d}_F\left(\hat{\Theta}_n(R), \Theta_{n,F}(R)\right) = o_p(\xi_n) \) uniformly in \( (F, R) \in J \), \( T_n(R) = T_n(R) \) with probability approaching 1 uniformly in \( (F, R) \in J \).

Then \( T_n(R) = \Gamma_{\xi_n,F,n,R}(\mathbb{G}_{n,F}) + o_p(1) \) uniformly in \( (F, R) \in J \).

For the second claim, notice that by definition, \( \varrho \) is Glivenko-Cantelli under any \( F \in F \),

\[
\sup_{(\theta, t) \in \Theta \times T} \| \tilde{\rho}_t(\theta) - E_F \rho_t(W_i, \theta) \| = o(1) \text{ a.s.,} \tag{19}
\]

where \( \tilde{\rho}_t(\theta) = n^{-1} \sum_{i=1}^n \rho_t(W_i, \theta) \). Because \( E_F \rho_t(W_i, \theta) \) is continuous in \( \theta \) and \( \Theta \cap R \) is compact under \( d_s \),

\[
\sqrt{T_n(R)} n \geq \inf_{\theta \in \Theta \cap R} \sqrt{\int_T |E_F \rho_t(W_i, \theta)|^2 d\mu(t) - o(1) \geq \sqrt{\min_{\theta \in \Theta \cap R} \int_T |E_F \rho_t(W_i, \theta)|^2 d\mu(t) - o(1) \text{ a.s.}} \tag{20}
\]

Similarly, by the compactness of \( \Theta \cap R \) and continuity,

\[
\sqrt{T_n(R)} n \leq \inf_{\theta \in \Theta \cap R} \sqrt{\int_T |E_F \rho_t(W_i, \theta)|^2 d\mu(t) + o(1) \rightarrow \sqrt{\min_{\theta \in \Theta \cap R} \int_T |E_F \rho_t(W_i, \theta)|^2 d\mu(t), \text{ a.s.}}
\]

By the continuous mapping theorem, \( T_n(R) / n \rightarrow \min_{\theta \in \Theta \cap R} \int_T |E_F \rho_t(W_i, \theta)|^2 d\mu(t) \) a.s.

Theorem E.1 shows that under the null hypothesis, \( \Gamma_{\xi_n,F,n,R}(\mathbb{G}_{n,F}) \) is, at least, a conservative approximation to \( T_n(R) \) regardless how fast \( \tilde{d}_F\left(\hat{\Theta}_n(R), \Theta_{n,F}(R)\right) \) vanishes. But if it vanishes fast enough, this approximation is asymptotically exact. For example, if \( |d\rho_t(W_i, \theta_1 + \tau(\theta_2 - \theta_1)) / d\tau|_{\tau=0} \leq C \rho_t(\theta_1, \theta_2)^2 \), the approximation is asymptotically exact if \( \tilde{d}_F\left(\hat{\Theta}_n(R), \Theta_{n,F}(R)\right) = o_p(n^{-1/4}) \). If \( \rho_t \) is linear in \( \theta \), the approximation is asymptotically exact if \( \tilde{d}_F\left(\hat{\Theta}_n(R), \Theta_{n,F}(R)\right) = o_p(1) \). Under a fixed alternative, \( T_n(R) \) diverges to infinity at the rate \( n \).
It is worth pointing out that the validity of the approximation does not depend on the shape of the restriction set $R$. In particular, $R$ does not need to be convex. In addition, even if Assumptions C.4(i) and (ii) are not satisfied, one can still obtain a valid uniform asymptotic approximation if one replaces the term
\[ \frac{\sqrt{n}dF_n\rho_k(W_i, \theta)}{d\theta} \] in $\Gamma_{\xi_n,F,n,R}(G_{n,F})$ by $\sqrt{n}\left[ E_F\rho_k(W_i, \theta) - E_F\rho_k(W_i, \hat{\theta}) \right]$.

**F Proof of Theorem 5.1**

To simplify notations, I suppress the dependence of $\hat{\Theta}_n^*$ and $Q_n^*$ on $\gamma_n$, $\lambda_n$, and $R$ when there is no ambiguity. Let $c > 0$ and $\eta > 0$ be two constants. Define

\[ h_{c, \eta}(x) = \begin{cases} 
1 & \text{if } x > c + \eta \\
(x - c) / \eta & \text{if } c + \eta \geq x > c \\
0 & \text{if } x \leq c 
\end{cases} \quad (21) \]

**Lemma F.1.** Under the assumptions of Theorem 5.1, uniformly in $(F, R) \in \mathcal{J}$, $T_n^*(R) \geq \Gamma_{\xi_n,F,n,R}(G_{n,F}^*) + o_{\mathbb{F}}(1)$.

**Proof.** I establish this lemma by proving the following chain of inequalities:

\[ T_n^*(R) \geq \inf_{(\gamma_n, \lambda_n) \in I_n} \inf_{\tilde{\Theta}_n \in \Theta_{n,F}(R)} nQ_n^*(\theta, \gamma_n, 0) + o_{\mathbb{F}}(1) \geq \inf_{(\gamma_n, \lambda_n) \in I_n} \bar{T}_n^*(R, \gamma_n) + o_{\mathbb{F}}(1) \]

\[ = \inf_{(\gamma_n, \lambda_n) \in I_n} \bar{T}_n^*(R, \gamma_n) + o_{\mathbb{F}}(1) \geq \Gamma_{\xi_n,F,n,R}(G_{n,F}^*) + o_{\mathbb{F}}(1) \quad (22) \]

uniformly in $(F, R) \in \mathcal{J}$, where

\[ \bar{T}_n^*(R, \gamma_n) = \inf_{\tilde{\Theta}_n \in \Theta_{n,F}(R)} \inf_{\tilde{\Theta}_n \in \Theta_{n,F}(R)} \int_{\mathbf{T}} \left[ G_n^*(\tilde{\Theta}, \tilde{t}) + \frac{\sqrt{n}dF_n\rho_k(W_i, \tilde{\Theta})}{d\tilde{\Theta}[\theta(\gamma_n, \tilde{\Theta}) - \tilde{\Theta}] + \Delta_n(t)} \right]^2 d\mu(t), \]

\[ \hat{T}_n^*(R, \gamma_n) = \inf_{\tilde{\Theta}_n \in \Theta_{n,F}(R)} \inf_{\tilde{\Theta}_n \in \Theta_{n,F}(R)} \int_{\mathbf{T}} \left[ G_n^*(\tilde{\Theta}, \tilde{t}) + \frac{\sqrt{n}dF_n\rho_k(W_i, \tilde{\Theta})}{d\tilde{\Theta}[\theta(\gamma_n, \tilde{\Theta}) - \tilde{\Theta}] + \Delta_n(t)} \right]^2 d\mu(t), \]

\[ \theta(\gamma_n, \tilde{\Theta}) = \gamma_n(\theta - \tilde{\Theta}) + \tilde{\Theta}, \text{ and} \]

\[ \Delta_n(t) = \gamma_n\sqrt{n}d\rho_k(W_i, \tilde{\Theta}) - \sqrt{n} \frac{dE_F\rho_k(W_i, \tilde{\Theta})}{d\tilde{\Theta}[\theta(\gamma_n, \tilde{\Theta}) - \tilde{\Theta}] + G_n^*(\tilde{\Theta}, t) - G_n^*(\tilde{\Theta}, t)}. \]

Again, to simplify notations, I suppress the dependence of $\Delta_n$ on $F, R, \theta$, and $\hat{\Theta}$. Notice that $\hat{T}_n^*(R, \gamma_n) = \inf_{\tilde{\Theta}_n \in \Theta_{n,F}(R)} \inf_{\tilde{\Theta}_n \in \Theta_{n,F}(R)} nQ_n^*(\theta, \gamma_n, 0)$.
First, by Theorem H.2 and Assumption C.5, if \( \theta^*_n \in \hat{\Theta}^*_n \), then \( \theta^*_n \in \Theta^*_{n,F,R} \) with probability approaching 1 uniformly in \( (\gamma_n, \lambda_n) \in \tilde{T}_n \) and \( (F, R) \in J \). Therefore, uniformly in \( (F, R) \in J \)

\[
T^*_n(R) = \inf_{(\gamma_n, \lambda_n) \in \tilde{T}_n} nQ^*_n(\theta^*_n, \gamma_n, 0) \geq \inf_{(\gamma_n, \lambda_n) \in \tilde{T}_n} \inf_{\hat{\Theta}^*_n \in \Theta^*_{n,F,R}} nQ^*_n(\theta, \gamma_n, 0) + o_{\mathbb{P}}(1).
\]

This establishes the first inequality in (22). Next, notice that if \( \theta \in \Theta^*_{n,F,R} \), then \( \theta \in R^*_{n,F,R}(\hat{\theta}) \) for some \( \hat{\theta} \in \Theta^*_{n,F,R} \). Therefore, the second inequality in (22) follows from

\[
\inf_{\hat{\Theta}^*_n \in \Theta^*_{n,F,R}} nQ^*_n(\theta^*_n, \gamma_n, 0) = \tilde{T}^*_n(R, \gamma_n).
\]

Now I prove the equality in (22). By the triangular inequality,

\[
\sqrt{\inf_{(\gamma_n, \lambda_n) \in \tilde{T}_n} \tilde{T}^*_n(R, \gamma_n) + \delta_n} \geq \sqrt{\inf_{(\gamma_n, \lambda_n) \in \tilde{T}_n} \tilde{T}^*_n(R, \gamma_n)} \geq \sqrt{\inf_{(\gamma_n, \lambda_n) \in \tilde{T}_n} \tilde{T}^*_n(R, \gamma_n)} - \delta_n,
\]

where \( \delta_n = \sup_{(\gamma_n, \lambda_n) \in \tilde{T}_n} \sup_{\theta \in \Theta^*_{n,F,R}} \sup_{\hat{\theta} \in R^*_{n,F,R}(\hat{\theta})} \| \Delta_n \|_{2, \mu} \). If \( \delta_n = o_{\mathbb{P}}(1) \) uniformly in \( (F, R) \in J \),

\[
\sqrt{\inf_{(\gamma_n, \lambda_n) \in \tilde{T}_n} \tilde{T}^*_n(R, \gamma_n) + o_{\mathbb{P}}(1)} = \sqrt{\inf_{(\gamma_n, \lambda_n) \in \tilde{T}_n} \tilde{T}^*_n(R, \gamma_n)}
\]

uniformly in \( (F, R) \in J \). Then, because \( \inf_{(\gamma_n, \lambda_n) \in \tilde{T}_n} \tilde{T}^*_n(R, \gamma_n) = O_{\mathbb{P}}(1) \) uniformly in \( (F, R) \in J \), Lemma I.2 implies the equality in (22). To see \( \delta_n = o_{\mathbb{P}}(1) \) uniformly in \( (F, R) \in J \), first notice that by definition,

\[
\frac{\sqrt{n}dE_F\rho_t(W_i, \hat{\theta})}{d\theta} [\theta - \hat{\theta}] = \frac{\sqrt{n}dE_F\rho_t(W_i, \hat{\theta})}{d\theta} [\theta - \hat{\theta}] .
\]

Then one can write \( \Delta_n = \Delta_{n,1} + \Delta_{n,2} + \Delta_{n,3} \) where

\[
\Delta_{n,1}(t) = \gamma_n \sqrt{n} (\hat{\rho}_t(\theta) - E_F\rho_t(W_i, \theta)) ,
\]

\[
\Delta_{n,2}(t) = \gamma_n \sqrt{n} \left( E_F\rho_t(W_i, \theta) - \frac{dE_F\rho_t(W_i, \hat{\theta})}{d\theta} [\theta - \hat{\theta}] \right) ,
\]

\[
\Delta_{n,3}(t) = G^*_n(\theta, t) - G^*_n(\hat{\theta}, t) .
\]

By the triangular inequality, \( \| \Delta_n \|_{2, \mu} \leq \| \Delta_{n,1} \|_{2, \mu} + \| \Delta_{n,2} \|_{2, \mu} + \| \Delta_{n,3} \|_{2, \mu} \). Because \( \gamma_n \sqrt{n} \leq \sqrt{\kappa_n} \) and \( \kappa_n \ln n/n \to 0 \), Definition 5.1(3) implies that uniformly in \( (F, R) \in J \)

\[
\sup_{(\gamma_n, \lambda_n) \in \tilde{T}_n} \sup_{\hat{\Theta}^*_n \in \Theta^*_{n,F,R}} \sup_{\hat{\theta} \in R^*_{n,F,R}(\hat{\theta})} \| \Delta_{n,1} \|_{2, \mu} \leq \sup_{(\theta, t) \in \Theta \times T} \sqrt{\kappa_n} |\hat{\rho}_t(\theta) - E_F\rho_t(W_i, \theta)| = o_{\mathbb{P}}(1) .
\]

48
Next, by the triangular inequality and the fact that \( \gamma_n \sqrt{n} \leq \sqrt{k_n} \),
\[
\|\Delta_{n,2}\|_{2,\mu} \leq \sqrt{k_n} \left[ \int_T \left| \mathbb{E}_F \rho_t (W_i, \theta) - \mathbb{E}_F \rho_t \left( W_i, \hat{\theta} \right) \right| \frac{d\mathbb{E}_F \rho_t (W_i, \hat{\theta})}{d\theta} \left[ \theta - \hat{\theta} \right] \right] \, d\mu (t) \\
+ \sqrt{k_n} \left[ \int_T \left| \mathbb{E}_F \rho_t \left( W_i, \hat{\theta} \right) \right|^2 \, d\mu (t) \right].
\]

Therefore, uniformly in \( (F, R) \in \mathcal{J} \), \( \sup_{(\gamma_n, \lambda_n) \in \xi_n} \sup_{\theta \in \Theta_{n,F} (R)} \sup_{\theta \in R_n^{\xi_n} (\hat{\theta})} \|\Delta_{n,2}\|_{2,\mu} = o (1) \) by Assumptions C.1(ii), C.2(iii) and C.4(ii).

Lastly, if \( \omega \in \ell^\infty (\Theta \times \mathbb{T}) \), define \( g_{F,\xi} (\omega) = \sup_{d_F (\theta, \hat{\theta}) \leq \xi} \sup_{t \in \mathbb{T}} |\omega (\theta, t) - \omega (\hat{\theta}, t)| \). I first show that for any \( \epsilon > 0 \)
\[
\lim_{n \to \infty} \sup_{F \in \mathcal{F}} \mathbb{P} \left( g_{F,\xi_n} (G_n^*) > \epsilon \right) = \lim_{n \to \infty} \sup_{F \in \mathcal{F}} \mathbb{E}_F \mathbb{1} \left( g_{F,\xi_n} (G_n^*) > \epsilon \right) = 0.
\]

Notice \( \mathbb{1} (g_{F,\xi_n} (G_n^*) > \epsilon) \leq h_{\epsilon/2,\epsilon/2} \circ g_{F,\xi_n} (G_n^*) \) where \( h_{\epsilon/2,\epsilon/2} \) is defined by (21). Then \( \mathcal{H} = \{ h_{\epsilon/2,\epsilon/2} \} \) and \( \mathcal{G} = \{ g_{F,\xi} : \xi > 0, \ F \in \mathcal{F} \} \) satisfy the assumptions of Lemma H.2. Then Lemma H.2 implies that
\[
\lim_{n \to \infty} \sup_{F \in \mathcal{F}} \sup_{\xi > 0} \left| \mathbb{E}_F h_{\epsilon/2,\epsilon/2} \circ g_{F,\xi} (G_n^*) - \mathbb{E} h_{\epsilon/2,\epsilon/2} \circ g_{F,\xi} (G_F) \right| = 0.
\]

The last inequality holds by the Jensen’s inequality. This means that
\[
\lim_{n \to \infty} \sup_{F \in \mathcal{F}} \sup_{\xi > 0} \left| \mathbb{E}_F h_{\epsilon/2,\epsilon/2} \circ g_{F,\xi_n} (G_n^*) - \mathbb{E} h_{\epsilon/2,\epsilon/2} \circ g_{F,\xi_n} (G_F) \right| = 0. \tag{23}
\]

By Definition 5.1(3), \( \vartheta \) is uniformly pre-Gaussian. Then (23) and Assumption C.4(iii) suggest
\[
\lim_{n \to \infty} \sup_{F \in \mathcal{F}} \mathbb{P} \left( \mathbb{1} (g_{F,\xi_n} (G_n^*) > \epsilon) \right) \leq \lim_{n \to \infty} \sup_{F \in \mathcal{F}} \mathbb{E}_F h_{\epsilon/2,\epsilon/2} \circ g_{F,\xi_n} (G_n^*) \right) \right| \right) = 0.
\]

Uniformly in \( (F, R) \in \mathcal{J} \), \( \sup_{(\gamma_n, \lambda_n) \in \xi_n} \sup_{\theta \in \Theta_{n,F} (R)} \sup_{\theta \in R_n^{\xi_n} (\hat{\theta})} \|\Delta_{n,2}\|_{2,\mu} = o_{\mathbb{P}} (1) \). The above bounds imply that \( \delta_n = o_{\mathbb{P}} (1) \) uniformly in \( (F, R) \in \mathcal{J} \) and the equality in (22) follows.

To prove the last inequality in (22), I only need to show that if \( \theta \in R_n^{\xi_n} (\hat{\theta}) \), then \( \theta (\gamma_n, \theta, \hat{\theta}) \in R_n^{\xi_n} (\hat{\theta}) \) for sufficiently large \( n \) and all \( (\gamma_n, \lambda_n) \in \tilde{I}_n \). Notice that by definition, \( \gamma_n \leq \sqrt{k_n/n} < 1 \) for every
\((\gamma_n, \lambda_n) \in \tilde{I}_n\). Therefore, \(\theta \left(\gamma_n, \theta, \tilde{\theta}\right) \in \Theta_n \cap R\) by Assumption C.2(i) and the fact that \(\theta, \tilde{\theta} \in \Theta_n \cap R\). Then, 
\[\theta \left(\gamma_n, \theta, \tilde{\theta}\right) \in R_n^{\epsilon}(\tilde{\theta})\]
for all \((\gamma_n, \lambda_n) \in \tilde{I}_n\) because \(d_F \left(\theta \left(\gamma_n, \theta, \tilde{\theta}\right), \tilde{\theta}\right) \leq \xi_n\).

\[\square\]

**Proof of Theorem 5.1.** **Proof of the first claim** Because \(T_n^*(R)\) is weakly decreasing in \(I_n\), I only need to prove the first claim under \(I_n = \tilde{I}_n\). Notice that by Lemma F.1, \(C_n^*(1 - \alpha, R) \geq q_n^* \left(1 - \alpha, \Gamma_{\xi_n, F,n,R} + o_{\mathcal{F}}(1)\right)\) uniformly in \((F, R) \in \mathcal{J}\). Define \(\tilde{G} = \{\Gamma_{\kappa, F,m,R} : \kappa \geq 0, (F, R) \in \mathcal{J}, m \in \mathbb{Z}^+\}\). Then by Theorem E.1,

\[
\limsup_{n \to \infty} \sup_{(F,R) \in \mathcal{J}} \mathbb{P}_F \left(T_n(R) > C_n^*(1 - \alpha + \eta, R) + \eta\right)
\leq \limsup_{n \to \infty} \sup_{(F,R) \in \mathcal{J}} \mathbb{P}_F \left(\Gamma_{\xi_n, F,n,R}(G_{n,F}) > q_n^* \left(1 - \alpha + \frac{\eta}{2}, \Gamma_{\xi_n, F,n,R}\right) + \frac{\eta}{2}\right)
\leq \limsup_{n \to \infty} \sup_{(F,R) \in \mathcal{J}} \sup_{g \in \tilde{G}} \mathbb{P}_F \left(g(G_{n,F}) > q(1 - \alpha, g, F) + \frac{\eta}{4}\right)
+ \limsup_{n \to \infty} \sup_{(F,R) \in \mathcal{J}} \sup_{g \in \tilde{G}} \mathbb{P}_F \left(q(1 - \alpha, g, F) > q_n^* \left(1 - \alpha + \frac{\eta}{2}, g\right) + \frac{\eta}{4}\right).
\]

The second inequality follows because \(\Gamma_{\xi_n, F,n,R} \in \tilde{G}\) for all \(n\) and \((F, R) \in \mathcal{J}\). Lemma H.4 ensures that \(\tilde{G}\) satisfies the assumptions in Proposition H.1. Therefore, (27) and (28) imply that the first term after the second inequality is no larger than \(\alpha\) and the second term is 0. This implies that the test is valid if \(\eta > 0\).

Now suppose that in addition, Assumption C.7 holds. For any \(\epsilon > 0\), Proposition H.1 and Lemmas H.3, H.4 imply that for any sufficiently small \(\eta > 0\),

\[
\limsup_{n \to \infty} \sup_{(F,R) \in \mathcal{J}^*(\alpha)} \mathbb{P}_F \left(T_n(R) > C_n^*(1 - \alpha, R)\right)
\leq \limsup_{n \to \infty} \sup_{J \in \mathcal{J}^*(\alpha)} \mathbb{P}_F \left(\Gamma_{\xi_n, J}(G_{n,F}) > q \left(1 - \alpha - \eta, \Gamma_{\xi_n, J,R}(F) - \eta\right)\right)
\leq \limsup_{n \to \infty} \sup_{J \in \mathcal{J}^*(\alpha)} \mathbb{P}_F \left(\Gamma_{\xi_n, J}(G_{n,F}) > q \left(1 - \alpha + \eta, \Gamma_{\xi_n, J,R}(F) + \eta\right)\right)
+ \limsup_{n \to \infty} \sup_{J \in \mathcal{J}^*(\alpha)} \mathbb{P}_F \left(q \left(1 - \alpha + \eta, \Gamma_{\xi_n, J,R}(F) + \eta\right) \geq \Gamma_{\xi_n, J}(G_{n,F}) > q \left(1 - \alpha - \eta, \Gamma_{\xi_n, J,R}(F) - \eta\right)\right)
\leq \alpha + \limsup_{n \to \infty} \sup_{J \in \mathcal{J}^*(\alpha)} \mathbb{P}_F \left(C_F \left(1 - \alpha + 2\eta, \Gamma_{\xi_n, J,R}\right) + 2\eta \geq \Gamma_{\xi_n, J}(G_{n,F}) > C_F \left(1 - \alpha - 2\eta, \Gamma_{\xi_n, J,R} - 2\eta\right)\right),
\]

where \(\Gamma_{\xi_n, J} = \Gamma_{\xi_n, F,n,R}\) and \(J = (F, R)\). The second term in the last line converges to 0 as \(\eta \to 0\) under Assumption C.7(i). This suggests that \(\limsup_{n \to \infty} \sup_{(F,R) \in \mathcal{J}^*(\alpha)} \mathbb{P}_F \left(T_n(R) > C_n^*(1 - \alpha, R)\right) \leq \alpha\) for all \(\epsilon > 0\). Therefore, there exists \(\epsilon_n \downarrow 0\) such that \(\limsup_{n \to \infty} \sup_{(F,R) \in \mathcal{J}^*(\alpha)} \mathbb{P}_F \left(T_n(R) > C_n^*(1 - \alpha, R)\right) \leq \alpha\).

50
This implies that

$$\limsup_{n \to \infty} \sup_{(F,R) \in J} \mathbb{P}_F \left( T_n(R) > C_n^*(1 - \alpha, R) \right)$$

$$\leq \limsup_{n \to \infty} \left[ \sup_{(F,R) \in J \setminus J_n^{\alpha}(\alpha)} \mathbb{P}_F \left( T_n(R) > C_n^*(1 - \alpha, R) \right) \vee \sup_{(F,R) \in J_n^{\alpha}(\alpha)} \mathbb{P}_F \left( T_n(R) > C_n^*(1 - \alpha, R) \right) \right]$$

$$\leq \left[ \limsup_{n \to \infty} \sup_{(F,R) \in J \setminus J_n^{\alpha}(\alpha)} \mathbb{P}_F \left( T_n(R) > C_n^*(1 - \alpha, R) \right) \right] \vee \left[ \limsup_{n \to \infty} \sup_{(F,R) \in J_n^{\alpha}(\alpha)} \mathbb{P}_F \left( T_n(R) > C_n^*(1 - \alpha, R) \right) \right].$$

Assumption C.7(ii) implies that the first term is at most $\alpha$. And the second term is at most $\alpha$ by the above argument. Therefore, the test is valid with $\eta = 0$.

**Proof of the second claim** For any $(\gamma_n, \lambda_n) \in I_n$

$$\frac{T_n^*(R)}{\kappa_n} \leq \frac{nQ_n^*(\theta_n^*, \gamma_n, 0)}{\kappa_n} \leq \inf_{\theta \in \Theta_n \cap R} nQ_n^*(\theta, \gamma_n, \lambda_n).$$

By Theorem 3.6.2 in Van Der Vaart and Wellner (1996), $G_n^*(\theta, t)$ converges in distribution to $G_F(\theta, t)$ almost surely. This implies $\sup_{(\theta,t) \in \Theta \times T} \left| G_n^*(\theta, t) \right| = O_{\mathbb{P}_F}(1)$ almost surely. In addition, $\sup_{(\theta,t) \in \Theta \times T} \left| \tilde{\rho}_t(\theta) - \mathbb{E}_F \rho_t(W, \theta) \right| = o(1)$ almost surely. Then $\gamma_n \leq \sqrt{n}/n \to 0$ and $\kappa_n \to \infty$ imply that almost surely

$$\frac{T_n^*(R)}{\kappa_n} \leq \inf_{\theta \in \Theta_n \cap R} \left\{ \int_T \left| \mathbb{E}_F \rho_t(W, \theta) \right|^2 d\mu(t) + \int_T \frac{G_n^*(\theta, t)^2}{\kappa_n} d\mu(t) \right\} \leq \frac{\min_{\theta \in \Theta_n \cap R} Q_F(\theta) + o_{\mathbb{P}_F}(1)}{\kappa_n}.$$

Therefore, $T_n^*(R) = O_{\mathbb{P}_F}(\kappa_n)$ almost surely. Because $\Theta \cap R$ is compact and $\Theta_F \cap R = \emptyset$, $\min_{\theta \in \Theta_n \cap R} Q_F(\theta) > 0$ by the continuity of $Q_F(\cdot)$. By Theorem E.1, $T_n(R)$ diverges to infinity at rate $n$ which is faster than $\kappa_n$. Therefore, the second claim holds. If $(0, \kappa_n) \in I_n$, $T_n^*(R) \leq \int_T \left| G_n^*(\theta_n^*, t) \right|^2 d\mu(t) = O_{\mathbb{P}_F}(1)$ almost surely. Hence, under a fixed alternative, the bootstrap critical value does not diverge.

\[\square\]

**G No Inequality Constraints**

Throughout this section, let

$$g(\mathbb{G}_F) = \inf_{\tilde{\theta} \in \Theta_F \cap R} \inf_{v \in V_{\infty}(\tilde{\theta})} \int_T \left[ \mathbb{G}_F \left( \tilde{\theta}, t \right) + v(t) \right]^2 d\mu(t), \quad (24)$$

51
where $V_\infty (\hat{\theta}) = \cup V_n (\hat{\theta})$ with the closure taken under sup norm and

$$V_n (\hat{\theta}) = \left\{ v : v (t) = \frac{d\mathbb{E}_F \rho_k (W_i, \hat{\theta})}{d\theta} [P^{k_n}] \beta, \ l (P^{k_n} \beta) = 0, \ \forall l \in L, \ \|P^{k_n} \beta\|_F < \infty \right\},$$

$$\frac{d\mathbb{E}_F \rho_k (W_i, \hat{\theta})}{d\theta} [P^{k_n}] = \left( \frac{d\mathbb{E}_F \rho_k (W_i, \hat{\theta})}{d\theta} [p_1], \ldots, \frac{d\mathbb{E}_F \rho_k (W_i, \hat{\theta})}{d\theta} [p_{k_n}] \right).$$

In addition, define

$$V_n^c (\hat{\theta}) = \left\{ v : v (t) = \frac{d\mathbb{E}_F \rho_k (W_i, \hat{\theta})}{d\theta} [P^{k_n}] \beta, \ l (P^{k_n} \beta) = 0, \ \forall l \in L, \ \|P^{k_n} \beta\|_F \leq c \right\}.$$

Notice if $n_1 < n_2$ and $c_1 < c_2$, then $V_{n_1} (\hat{\theta}) \subset V_{n_2} (\hat{\theta})$ and $V_{n_1}^c (\hat{\theta}) \subset V_{n_2}^c (\hat{\theta}).$

**Assumption G.1.** (i) $\hat{d}_F (\Theta_{n,F} (R), \Theta_F \cap R) = o (\xi_n);$ (ii) There exist $\epsilon > 0$ such that

$$\sup_{\theta_1, \theta_2 \in \Theta_F \cap R : d_F (\theta_1, \theta_2) \leq \xi_n, \ t \in \mathcal{T}} \left| \mathbb{E}_F \rho_k (W_i, \theta_2) - \mathbb{E}_F \rho_k (W_i, \theta_1) - \frac{d\mathbb{E}_F \rho_k (W_i, \theta_1)}{d\theta} [\theta_2 - \theta_1] \right| = o \left( n^{-1/2} \right).$$

Assumption G.1(i) requires that $\Theta_{n,F} (R)$ converges to $\Theta_F \cap R$ fast enough. Assumption G.1(ii) is the same as Assumption C.4(ii) except that it is only concerned with a single $(F,R)$ and focuses on a neighborhood of $\Theta_F \cap R$ instead of $\Theta_{n,F} (R)$. It allows me to use the first-order expansion as in (18) but around $\Theta_F \cap R$. This is convenient for deriving the asymptotic distribution. Assumption C.4(ii) holds if $\mathbb{E}_F \rho_k (W_i, \theta)$ is sufficiently smooth around $\Theta \cap R$, e.g., Assumption 4.5 in Hong (2017), and $\xi_n$ converges to 0 fast enough.

**Proof of Theorem 5.2.** I follow Hong (2017) to first show that $T_n (R) \rightarrow^d g (\mathcal{G}_F)$. And then I show that $T_n (R)$ converges to the same distribution. By Theorem H.1 and Assumptions G.1(i), $\hat{d}_F \left( \hat{\Theta}_n (R), \Theta_F \cap R \right) = O_F (\xi_n / \sqrt{n}) = o (\xi_n)$. Therefore, one can replace $\Theta_{n,F} (R)$ in (18) by $\Theta_F \cap R$ and follow similar steps there to show that

$$T_n (R) = \inf_{\hat{\theta} \in \Theta_F \cap R} \inf_{\theta \in \Theta_{n,F} (R) \cap \Pi_{n,R} (\hat{\theta})} \int_{\mathcal{T}} \left[ \mathcal{G}_{n,F} \left( \hat{\theta}, t \right) + \sqrt{n} \frac{d\mathbb{E}_F \rho_k (W_i, \hat{\theta})}{d\theta} [\theta - \Pi_{n,R} \hat{\theta}] \right]^2 d\mu (t) + o_F (1).$$

Define $\Delta = \sqrt{n} (\theta - \Pi_{n,R} \hat{\theta})$, then

$$T_n (R) = \inf_{\hat{\theta} \in \Theta_F \cap R} \inf_{\Delta \in \Theta_n : \ l (\Delta) = 0, \ \|\Delta\|_F = \sqrt{n} \xi_n} \int_{\mathcal{T}} \left[ \mathcal{G}_{n,F} \left( \hat{\theta}, t \right) + \frac{d\mathbb{E}_F \rho_k (W_i, \hat{\theta})}{d\theta} [\Delta] \right]^2 d\mu (t) + o_F (1)\ (25)$$

$$= \inf_{\hat{\theta} \in \Theta_F \cap R} \inf_{\Delta \in \Theta_n : \ l (\Delta) = 0, \ \|\Delta\|_F = \sqrt{n} \xi_n} \int_{\mathcal{T}} \left[ \mathcal{G}_{n,F} \left( \hat{\theta}, t \right) + \frac{d\mathbb{E}_F \rho_k (W_i, \hat{\theta})}{d\theta} [\Delta] \right]^2 d\mu (t) + o_F (1).$$

52
Since $\sqrt{n} \xi_n \to \infty$, $T_n (R) \to^d g (G_F)$ by Lemma H.5 and the extended continuous mapping theorem (Theorem 1.11.1 in Van Der Vaart and Wellner (1996)).

Now I show that $T_n^* (R)$ converges to the same distribution. Theorem H.2 and Assumptions G.1(i) and C.5 imply that $d_F \left( \hat{\Theta}_n (\gamma_n, \lambda_n, R), \Theta_F \cap R \right) = o (\xi_n)$. Therefore, one can use the same argument as in the proof of Lemma F.1 but with $\Theta_{n,F} (R)$ replaced by $\Theta_F \cap R$ to show

$$T_n^* (R) \geq \inf_{\theta \in \Theta_F \cap R} \inf_{v \in V_n^0} \int_T \left[ G_n^* \left( \hat{\theta}, t \right) + v (t) \right]^2 d\mu (t) + o_{P_F} (1).$$

Because $\left( \sqrt{\kappa_n / n}, 0 \right) \in I_n$, almost surely

$$T_n^* (R) \leq \inf_{\theta \in \Theta_n \cap R} nQ_n^* \left( \theta, \sqrt{\frac{\kappa_n}{n}}, 0 \right) = \inf_{\theta \in \Theta_n \cap R} \int_T \left[ G_n^* (\theta, t) + \sqrt{\kappa_n} \rho_k (\theta) \right]^2 d\mu (t)
\leq \inf_{\theta \in \Theta_F \cap R} \inf_{\theta \in R^+_n (\Pi_n, \theta)} \int_T \left[ G_n^* (\theta, t) + \sqrt{\kappa_n} \rho_k (\theta) \right]^2 d\mu (t)
= \inf_{\theta \in \Theta_F \cap R} \inf_{\theta \in R^+_n (\Pi_n, \theta)} \int_T \left[ G_n^* (\hat{\theta}, t) + \sqrt{\kappa_n} \rho_k (W_i, \hat{\theta}) \right]^2 d\mu (t) + o (1)
= \inf_{\theta \in \Theta_F \cap R} \inf_{\theta \in R^+_n (\Pi_n, \theta)} \int_T \left[ G_n^* (\hat{\theta}, t) + \sqrt{\kappa_n} \rho_k (W_i, \hat{\theta}) \right]^2 d\mu (t) + o_{P_F} (1)
= \inf_{\theta \in \Theta_F \cap R} \inf_{\theta \in R^+_n (\Pi_n, \theta)} \int_T \left[ G_n^* (\hat{\theta}, t) + \sqrt{\kappa_n} \rho_k (W_i, \hat{\theta}) \right]^2 d\mu (t) + o_{P_F} (1).$$

Now define $\Delta = \sqrt{\kappa_n} \left( \theta - \Pi_n, \hat{\theta} \right)$ and follow the same steps in (25) to obtain

$$T_n^* (R) \leq \inf_{\theta \in \Theta_F \cap R} \inf_{v \in V_n^0} \frac{\kappa_n}{\sqrt{\kappa_n}} \int_T \left[ G_n^* (\hat{\theta}, t) + v (t) \right]^2 d\mu (t) + o_{P_F} (1).$$

Notice that because $\zeta_n^* / \sqrt{\kappa_n} = o (\xi_n)$, $\sqrt{\kappa_n} \xi_n \to \infty$. By Lemma H.5 and the extended continuous mapping theorem (Theorem 1.11.1 in Van Der Vaart and Wellner (1996)), almost surely,

$$\inf_{\theta \in \Theta_F \cap R} \inf_{v \in V_n^0} \int_T \left[ G_n^* (\hat{\theta}, t) + v (t) \right]^2 d\mu (t) \to^d g (G_F),$$

$$\inf_{\theta \in \Theta_F \cap R} \inf_{v \in V_n^0} \int_T \left[ G_n^* (\hat{\theta}, t) + v (t) \right]^2 d\mu (t) \to^d g (G_F).$$

This suggests that $T_n^* (R) \to^d g (G_F)$ almost surely. The theorem follows because the distribution of $g (G_F)$ is continuous and strictly increasing at its $1 - \alpha$-th quantile.
H Useful Results

Lemma H.1. If Assumptions C.1-C.3 hold, then \( \hat{d}_F(\Theta_n(R), \Theta_{n,F}(R)) = o_{P_F}(1) \) uniformly in \((F, R) \in J\).

Theorem H.1. If Assumptions C.1 to C.3 hold, then uniformly in \((F, R) \in J\),
\[
\hat{d}_{w,F}(\Theta_n(R), \Theta_{n,F}(R)) = O_{P_F}(n^{-1/2}),
\]
and \( \hat{d}_F(\Theta_n(R), \Theta_{n,F}(R)) = O_{P_F}(\zeta_n n^{-1/2}) \).

The first claim in Theorem H.1 says that under \( \hat{d}_{w,F} \), the convergence rate is parametric even though \( \theta \) contains unknown functions. Hong (2017) obtains the same convergence rates in conditional moment equality models with semi/non-parametric unknown parameters. My result is stronger in the sense that the rates are valid uniformly on \( J \). It holds because Definition 5.1(3) implicitly restricts the complexity of \( \Theta \). With a more complex parameter space, the convergence rate can be lower. The convergence rate under \( d_F \) depends on the degree of ill-posedness and is generally slower than the parametric rate because \( \zeta_n \rightarrow \infty \).

Theorem H.2. If Assumptions C.1 to C.3 hold and \( \kappa_n \rightarrow \infty, \kappa_n \ln \ln n/n \rightarrow 0 \), then
\[
\limsup_{n \rightarrow \infty} \sup_{(F,R) \in J} \sup_{(\gamma_n,\lambda_n) \in I_n} \mathbb{P}_F \left( \| \Theta_n(\gamma_n,\lambda_n,R) - \Theta_{n,F}(R) \|_\infty > M \kappa^{-1/2}_n \right) = 0,
\]
\[
\limsup_{n \rightarrow \infty} \sup_{(F,R) \in J} \sup_{(\gamma_n,\lambda_n) \in I_n} \mathbb{P}_F \left( \| \Theta_n(\gamma_n,\lambda_n,R) - \Theta_{n,F}(R) \|_\infty > M \kappa^{-1/2}_n \right) = 0,
\]
\[
\limsup_{n \rightarrow \infty} \sup_{(F,R) \in J} \sup_{(\gamma_n,\lambda_n) \in I_n} \mathbb{P}_F \left( \| \Theta_n(\gamma_n,\lambda_n,R) - \Theta_{n,F}(R) \|_\infty > M \kappa^{-1/2}_n \right) = 0.
\]

Notice that the convergence rate of \( \hat{\Theta}_n^*(\gamma_n,\lambda_n,R) \) has an upper bound that depends only on \( \kappa_n \), which allows me to freely change \( \gamma_n \) without significantly changing the imposed identified set. This is possible because of the additional penalty term \( \lambda_n Q_n(\theta) \).

Proposition H.1. Let \( M(\cdot, \cdot) \) be a function on \( \mathbb{R}^2 \) such that \( \sup_{(x_1,x_2) \in K} M(x_1, x_2) < \infty \) for any bounded set \( K \subset \mathbb{R}^2 \). \( G \) is a collection of non-negative functions on \( \ell^\infty(\Theta \times T) \) such that for any \( g \in G \) and \( \omega_1, \omega_2 \in \ell^\infty(\Theta \times T) \),
\[
|g(\omega_1) - g(\omega_2)| < M(\|\omega_1\|_\infty, \|\omega_2\|_\infty) d^\infty(\omega_1, \omega_2).
\] (26)

Then, for any \( \eta > 0 \) and \( \alpha \in [0,1] \)
\[
\limsup_{n \rightarrow \infty} \sup_{F \in F} \sup_{g \in G} \mathbb{P}_F \left( g(G_{n,F}) > q_n(1 - \alpha, g, F) + \eta \right) \leq \alpha,
\] (27)
\[
\limsup_{n \rightarrow \infty} \sup_{F \in F} \sup_{g \in G} \mathbb{P}_F \left( q(1 - \alpha, g, F) > q_n^*(1 - \alpha + \eta, g) + \eta \right) = 0.
\] (28)
Lemma H.2. Let $\mathcal{H}$ be a class of bounded Lipschitz functions with Lipschitz constant $\gamma$ and bound $B$. If the set of functions $G$ satisfies the assumptions of Proposition H.1, then

$$\limsup_{n \to \infty} \sup_{F \in \mathcal{F}} \sup_{h \in \mathcal{H}} \sup_{g \in G} \left| \mathbb{E} \left[ h \left( G_n^* \right) - \mathbb{E} h \left( G_F \right) \right] \right| = 0, \quad (29)$$

$$\limsup_{n \to \infty} \sup_{F \in \mathcal{F}} \sup_{h \in \mathcal{H}} \sup_{g \in G} \left| \mathbb{E} \left[ h \left( G_n, F \right) - \mathbb{E} h \left( G_F \right) \right] \right| = 0. \quad (30)$$

Lemma H.3. Under the assumptions of Proposition H.1, for any $\eta > 0$

$$\limsup_{n \to \infty} \sup_{F \in \mathcal{F}} \sup_{g \in G} \sup_{a \leq b \in \mathbb{R}} \left[ \mathbb{P} \left( g \left( G_n, F \right) \in [a, b] \right) - \mathbb{P} \left( g \left( G_F \right) \in [a - \eta, b + \eta] \right) \right] \leq 0. \quad (31)$$

Lemma H.4. For all $m \in \mathbb{Z}^+$, $(F, R) \in \mathcal{J}$, $\kappa \geq 0$, and $\omega_1, \omega_2 \in \ell^\infty (\Theta \times \mathcal{T})$,

$$| \Gamma_{\kappa, F, m, R} (\omega_1) - \Gamma_{\kappa, F, m, R} (\omega_2) | \leq 4 \max (\| \omega_1 \|_{\infty}, \| \omega_2 \|_{\infty}) d_{\infty} (\omega_1, \omega_2).$$

Lemma H.5. Let $\{c_n\}_{n=1}^{\infty}$ be a sequence of positive numbers that diverges to infinity. Then for any $\{\omega_n\}_{n=1}^{\infty}$ and $\omega$ in $\ell^\infty (\Theta \times \mathcal{T})$ such that $\| \omega_n - \omega \|_{\infty} \to 0$, $g_n (\omega_n) \to g (\omega)$ where $g$ is defined by (24) and

$$g_n (\omega_n) = \inf_{\theta \in \Theta \cap R} \inf_{v \in V_n (\theta)} \int_{\mathcal{T}} [\omega (\theta, t) + v (t)]^2 d\mu (t).$$

I Proofs for Results in Section H

Proof of Lemma H.1. Notice that Lemma 1.1 implies that $\forall \epsilon > 0$, uniformly in $(F, R) \in \mathcal{J}$,

$$\mathbb{P} \left( d_F \left( \hat{\Theta}_n (R), \Theta_{n,F} (R) \right) > \epsilon \right) \leq \mathbb{P} \left( \inf_{\theta \in \Theta \cap R} \inf_{d_F (\theta, \Theta_{n,F} (R)) > \epsilon} Q_n (\theta) \leq \inf_{\Theta \cap R} Q_n (\theta) \right),$$

$$\leq \mathbb{P} \left( \inf_{\theta \in \Theta \cap R} \inf_{d_F (\theta, \Theta_{n,F} (R)) > \epsilon} Q_n (\theta) \leq \inf_{\Theta \cap R} Q_n (\theta) \right) \leq O_{\mathbb{P}} \left( \frac{1}{\sqrt{n}} \right),$$

$$\leq \mathbb{P} \left( S_n (\epsilon) \leq O_{\mathbb{P}} \left( \frac{1}{\sqrt{n}} \right) \right) \to 0.$$ 

The third inequality holds by Assumption C.3(i).

Proof of Theorem H.1. I begin with the first claim. Let $\theta_{F,R}$ be a point in $\Theta_F \cap R$. This point exists as long as $(F, R) \in \mathcal{J}$. By Lemma 1.1,

$$| Q_n (\Pi_n, R \theta_{F,R}) - Q_F (\Pi_n, R \theta_{F,R}) | \leq 2d_{w,F} (\Pi_n, R \theta_{F,R}, \theta_{F,R}) A_{n,F} + B_{n,F},$$

55
where \( A_{n,F} = O_{P_F} \left( n^{-1/2} \right) \), \( B_{n,F} = O_{P_F} \left( n^{-1} \right) \) uniformly in \( F \in \mathcal{F} \). By Assumptions C.1(ii) and C.2(iii),

\[
2d_{w,F} \left( \Pi_{n,R}\theta_{F,R}, \theta_{F,R} \right) A_{n,F} \leq \frac{2C_1}{C_2} d_{s} \left( \Pi_{n,R}\theta_{F,R}, \theta_{F,R} \right) A_{n,F} = o_{P_F} \left( \frac{1}{n} \right),
\]

which implies uniformly in \((F,R) \in \mathcal{J}\),

\[
|Q_n \left( \Pi_{n,R}\theta_{F,R} \right) - Q_F \left( \Pi_{n,R}\theta_{F,R} \right)| = O_{P_F} \left( \frac{1}{n} \right).
\]

Because \( d_{w,F} (\theta, \Theta_F) = d_{w,F} (\theta, \Theta_F \cap R) \) if \((F,R) \in \mathcal{J}\), Lemma 1.1 implies

\[
Q_n (\theta) \geq Q_F (\theta) - 2d_{w,F} (\theta, \Theta_F) A_{n,F} - B_{n,F}.
\]

Therefore, for sufficiently large \( n \)

\[
\mathbb{P}_F \left( \sqrt{n}d_{w,F} \left( \hat{\Theta}_n (R), \Theta_F \cap R \right) > M \right) \\
\leq \mathbb{P}_F \left( \inf_{\theta \in \Theta_n \cap R : d_{w,F} (\theta, \Theta_F \cap R) > M/\sqrt{n}} Q_n (\theta) \leq Q_n (\Pi_{n,R}\theta_{F,R}) \right) \\
\leq \mathbb{P}_F \left( \inf_{\theta \in \Theta_n \cap R : d_{w,F} (\theta, \Theta_F \cap R) > M/\sqrt{n}} \left[ Q_F (\theta) - 2d_{w,F} (\theta, \Theta_F \cap R) A_{n,F} - B_{n,F} \right] \leq Q_F (\Pi_{n,R}\theta_{F,R}) + O_{P_F} \left( n^{-1} \right) \right) \\
\leq \mathbb{P}_F \left( \inf_{x \geq M} \left( x^2 - 2x\sqrt{n}A_{n,F} \right) \leq O_{P_F} (1) \right).
\]

The first inequality holds because \( d_F (\Pi_{n,R}\theta_{F,R}, \Theta_F \cap R) < M/\sqrt{n} \). The second inequality holds by (32) and (33). Because \( Q_F (\theta) = d_{w,F} (\theta, \Theta_F \cap R)^2 \) and \( B_{n,F} = O_{P_F} \left( n^{-1} \right) \), one can multiply both sides by \( n \) and use the change of variable \( x = \sqrt{n}d_{w,F} (\theta, \Theta_F \cap R) \) to obtain the last inequality. Lastly, notice that \( \sqrt{n}A_{n,F} = O_{P_F} (1) \) uniformly in \((F,R) \in \mathcal{J}\). As \( M \to \infty \), the probability in the last line of (34) converges to 0 uniformly in \((F,R) \in \mathcal{J}\) because \( \inf_{x \geq M} \left( x^2 - 2x\sqrt{n}A_{n,F} \right) \) diverges to infinity. This concludes the first claim.

For the second claim, notice that for any \( \theta_1 \in \Theta, \theta_2 \in \Theta_F \cap R \), and \( \theta_3 \in \Theta_{n,F} (R) \),

\[
d_{w,F} (\theta_1, \theta_2) \leq d_{w,F} (\theta_1, \theta_3) + d_{w,F} (\theta_2, \theta_3).
\]

By the definition of \( \Theta_{n,F} (R) \), \( d_{w,F} (\theta_2, \theta_3) = \sqrt{Q_F (\theta_3)} \leq \sqrt{Q_F (\Pi_{n,R}\theta_2)} = d_{w,F} (\Pi_{n,R}\theta_2, \theta_2) = o \left( n^{-1/2} \right) \).

This shows that \( d_{w,F} (\theta_1, \theta_2) \leq d_{w,F} (\theta_1, \theta_3) + o \left( n^{-1/2} \right) \). Because this holds for all \( \theta_2 \in \Theta_F \cap R \) and \( \theta_3 \in \Theta_{n,F} (R) \), \( d_{w,F} (\theta_1, \Theta_F \cap R) \leq d_{w,F} (\theta_1, \Theta_{n,F} (R)) + o \left( n^{-1/2} \right) \) uniformly in \((F,R) \in \mathcal{J}\). Similarly, \( d_{w,F} (\theta_1, \Theta_{n,F} (R)) \leq d_{w,F} (\theta_1, \Theta_F \cap R) + o \left( n^{-1/2} \right) \). This implies that uniformly in \( \Theta_1 \subset \Theta \) and \((F,R) \in \mathcal{J}\)

\[
\tilde{d}_{w,F} (\Theta_1, \Theta_{n,F} (R)) = \tilde{d}_{w,F} (\Theta_1, \Theta_F \cap R) + o \left( n^{-1/2} \right).
\]
Therefore, \( \tilde{d}_{w,F} \left( \hat{\Theta}_n (R), \Theta_{n,F} (R) \right) = \tilde{d}_{w,F} \left( \hat{\Theta}_n (R), \Theta_F \cap R \right) + o \left( n^{-1/2} \right) = O_{PF} \left( 1 / \sqrt{n} \right) \).

For the last claim, by Lemma H.1, \( \hat{\Theta}_n (R) \subset \Theta_{n,F} (R) \) with probability converging to 1 uniformly on \( \mathcal{J} \). Then by Assumption C.3(ii)

\[
\limsup_{n \to \infty} \sup_{(F,R) \in \mathcal{J}} \mathbb{P}_F \left( \tilde{d}_F \left( \hat{\Theta}_n (R), \Theta_{n,F} (R) \right) > M \left( \frac{\zeta_n}{\sqrt{n}} \right) \right) \\
\leq \limsup_{n \to \infty} \sup_{(F,R) \in \mathcal{J}} \mathbb{P}_F \left( \sup_{\theta \in \hat{\Theta}_n (R)} d_F (\theta, \Theta_{n,F} (R)) > M' \frac{\zeta_n}{\sqrt{n}} \right) \\
\leq \limsup_{n \to \infty} \sup_{(F,R) \in \mathcal{J}} \mathbb{P}_F \left( \sqrt{n} \zeta_n \sup_{\theta \in \hat{\Theta}_n (R)} d_{w,F} (\theta, \Theta_{n,F} (R)) > M' \zeta_n \right).
\]

As \( M \) goes to infinity, the last term goes to zero by the second part of this proposition.

\[
\square
\]

**Proof of Theorem H.2.** To simplify exposition, I suppress the dependence of \( \hat{\Theta}_n^* \) on \( \gamma_n, \lambda_n, R \) and the dependence of \( Q_n^* \) on \( \gamma_n, \lambda_n \). Again, \( \theta_{F,R} \) is an arbitrary point in \( \Theta_F \cap R \). Because \( \sqrt{n} \gamma_n \leq \sqrt{k_n} \),

\[
n Q_n^* (\theta) = \int_T G_n^* (\theta, t)^2 \text{d} \mu (t) + 2 \sqrt{n} \int_T G_n^* (\theta, t) \bar{\rho}_n (\theta) \text{d} \mu (t) + \kappa_n \int_T \bar{\rho}_n (\theta)^2 \text{d} \mu (t) \\
\geq \int_T G_n^* (\theta, t)^2 \text{d} \mu (t) + A_n (\theta) - B_n (\theta) \equiv Q_n^* (\theta)
\]

\[
n Q_n^* (\Pi_{n,R} \theta_{F,R}) \leq \int_T G_n^* (\Pi_{n,R} \theta_{F,R}, t)^2 \text{d} \mu (t) + A_n (\Pi_{n,R} \theta_{F,R}) + B_n (\Pi_{n,R} \theta_{F,R}) \equiv Q_n^* (\Pi_{n,R} \theta_{F,R})
\]

where

\[
A_n (\theta) = \kappa_n \int_T \bar{\rho}_n (\theta)^2 \text{d} \mu (t), \quad B_n (\theta) = 2 \sqrt{n} \int_T G_n^* (\theta, t)^2 \text{d} \mu (t) \int_T \bar{\rho}_n (\theta)^2 \text{d} \mu (t).
\]

Notice \( Q_n^* \) and \( Q_n^* \) are both independent of \( (\gamma_n, \lambda_n) \). Because \( d_{w,F} (\Pi_{n,R} \theta_{F,R}, \Theta_F \cap R) = o \left( n^{-1/2} \right) \leq M / \sqrt{k_n} \) for sufficiently large \( n \),

\[
\mathbb{P}_F \left( \sup_{(\gamma_n, \lambda_n) \in \bar{I}_n} \sqrt{k_n} \tilde{d}_{w,F} \left( \hat{\Theta}_n^*, \Theta_F \cap R \right) > M \right) \\
\leq \mathbb{P}_F \left( \exists (\gamma_n, \lambda_n) \in \bar{I}_n : \inf_{\theta \in \Theta_F \cap R : d_{w,F} (\theta, \Theta_F \cap R) > M / \sqrt{n}} n Q_n^* (\theta) \leq n Q_n^* (\Pi_{n,R} \theta_{F,R}) \right) \\
\leq \mathbb{P}_F \left( \inf_{\theta \in \Theta_F \cap R : d_{w,F} (\theta, \Theta_F \cap R) > M / \sqrt{n}} \left[ Q_n^* (\theta) - Q_n^* (\Pi_{n,R} \theta_{F,R}) \right] \leq 0 \right).
\]

To prove the first claim, I only need to show that for \( M \) sufficiently large, the last line is arbitrarily small uniformly in \( (F,R) \in \mathcal{J} \) as \( n \to \infty \).
To see this, first notice by the triangular inequality,
\[
\sqrt{\kappa_n Q_F (\theta)} + \sqrt{\frac{\kappa_n}{n}} \sup_{(\theta, t) \in \Theta \times T} |G_{n,F} (\theta, t)| \geq \sqrt{A_n (\theta)} \geq \sqrt{\kappa_n Q_F (\theta)} - \sqrt{\frac{\kappa_n}{n}} \sup_{(\theta, t) \in \Theta \times T} |G_{n,F} (\theta, t)|.
\]
Because \( \sup_{(\theta, t) \in \Theta \times T} |G_{n,F} (\theta, t)| = O_{\mathbb{P}_F} (1) \) uniformly in \( F \in \mathcal{F} \) and \( \kappa_n/n = o (1) \), this implies \( \forall \theta \in R \)
\[
\sqrt{A_n (\theta)} = \sqrt{\kappa_n Q_F (\theta)} + o_{\mathbb{P}_F} (1) = \sqrt{\kappa_n d_{w,F} (\theta, \Theta_F \cap R) + o_{\mathbb{P}_F} (1)}
\]
uniformly in \( F \in \mathcal{F} \). Then by Assumption C.2(iii), uniformly in \( (F, R) \in \mathcal{J} \),
\[
\sqrt{A_n (\Pi_{n,R} \theta, F, R)} = o \left( \sqrt{\frac{\kappa_n}{n}} \right) + o_{\mathbb{P}_F} (1) = o_{\mathbb{P}_F} (1).
\]
Because \( g \) is uniformly Donsker, by Lemma A.2. in Linton, Song, and Whang (2010), for any \( \epsilon > 0 \)
\[
\lim_{n \to \infty} \sup_{F \in \mathcal{F}} \mathbb{P}_F \left( \sup_{(\theta, t) \in \Theta \times T} \left| G_n^* (\theta, t) \right| > M \right) \leq \lim_{n \to \infty} \sup_{F \in \mathcal{F}} \mathbb{P}_F \left( \sup_{(\theta, t) \in \Theta \times T} \left| G_F (\theta, t) \right| > M - \epsilon \right),
\]
which converges to 0 as \( M \to \infty \). Hence, \( \sup_{(\theta, t) \in \Theta \times T} \left| G_n^* (\theta, t) \right| = O_{\mathbb{P}_F} (1) \) uniformly in \( F \in \mathcal{F} \). This and (37) imply that
\[
\sqrt{B_n (\Pi_{n,R} \theta, F, R)} = 2 \sqrt{\int_T G_n^* (\Pi_{n,R} \theta, F, R, t)^2 d\mu (t)} \sqrt{A_n (\Pi_{n,R} \theta, F, R)} = o_{\mathbb{P}_F} (1)
\]
uniformly in \( (F, R) \in \mathcal{J} \).

Next, notice by (36), \( \forall \theta \in R \)
\[
A_n (\theta) - B_n (\theta) = [\sqrt{\kappa_n d_{w,F} (\theta, \Theta_F \cap R)} + o_{\mathbb{P}_F} (1)] \left( \sqrt{\kappa_n d_{w,F} (\theta, \Theta_F \cap R)} + o_{\mathbb{P}_F} (1) \right) - 2 \sqrt{\int_T G_n^* (\theta, t)^2 d\mu (t)}.
\]
Define \( x = \sqrt{\kappa_n d_{w,F} (\theta, \Theta_F \cap R)} \). Then \( \forall \theta \in R \), uniformly in \( (F, R) \in \mathcal{J} \)
\[
Q_{n,1}^* (\theta) - Q_{n,2}^* (\Pi_{n,R} \theta, F, R) = \int_T G_n^* (\theta, t)^2 d\mu (t) - \int_T G_n^* (\Pi_{n,R} \theta, F, R, t)^2 d\mu (t) + A_n (\theta) - B_n (\theta)
\]
\[
= - [A_n (\Pi_{n,R} \theta, F, R) + B_n (\Pi_{n,R} \theta, F, R)]
\]
\[
= - |O_{\mathbb{P}_F} (1)| + x^2 - |O_{\mathbb{P}_F} (1)| x + o_{\mathbb{P}_F} (1).
\]
Then, the first claim of the lemma follows because
\[
\lim_{M \uparrow \infty} \limsup_{n \to \infty} \sup_{(F, R) \in \mathcal{J}} \mathbb{P}_F \left( \inf_{d_{w,F} (\theta, \Theta_F \cap R) > M} \frac{1}{M} \left| Q_{n,1}^* (\theta) - Q_{n,2}^* (\Pi_{n,R} \theta, F, R) \right| \leq 0 \right)
\]
\[
= \lim_{M \uparrow \infty} \limsup_{n \to \infty} \sup_{(F, R) \in \mathcal{J}} \mathbb{P}_F \left( \inf_{x > M} (-|O_{\mathbb{P}_F} (1)| + x^2 - |O_{\mathbb{P}_F} (1)| x + o_{\mathbb{P}_F} (1)) \leq 0 \right) = 0.
\]
The second claim holds because of (35). The third claim follows by Assumption C.3(ii).

\[
\text{Proof of Proposition H.1.} \quad I \text{ start with proving (27). Let } h_{c,\eta} \text{ be defined as in (21). Then } \mathcal{H} = \{h_{c,\eta} : c \in \mathbb{R}\} \text{ is a collection of Lipschitz functions with Lipschitz constant } 1/\eta \text{ and upper bound 1. By (30) in Lemma H.2}
\]

\[
\limsup_{n \to \infty} \sup_{F \in \mathcal{F}} \sup_{g \in \mathcal{G}} \sup_{c \in \mathbb{R}} |\mathbb{E}_F h_{c,\eta} \circ g (\mathcal{G}_{n,F}) - \mathbb{E}_F h_{c,\eta} \circ g (\mathcal{G}_F)| = 0. \tag{38}
\]

Because \(\mathbb{E}_F h_{c,\eta} \circ g (\mathcal{G}_{n,F}) \geq \mathbb{P}_F (g (\mathcal{G}_{n,F}) > c + \eta)\) and \(\mathbb{E}_F h_{c,\eta} \circ g (\mathcal{G}_F) \leq \mathbb{P}_F (g (\mathcal{G}_F) > c)\), (38) implies

\[
\limsup_{n \to \infty} \sup_{F \in \mathcal{F}} \sup_{g \in \mathcal{G}} \sup_{c \in \mathbb{R}} [\mathbb{P}_F (g (\mathcal{G}_{n,F}) > c + \eta) - \mathbb{P}_F (g (\mathcal{G}_F) > c)] \leq 0. \tag{39}
\]

Therefore,

\[
\limsup_{n \to \infty} \sup_{F \in \mathcal{F}} \sup_{g \in \mathcal{G}} \sup_{c \in \mathbb{R}} [\mathbb{P}_F (g (\mathcal{G}_{n,F}) > q (1 - \alpha, g, F) + \eta) - \mathbb{P}_F (g (\mathcal{G}_F) > q (1 - \alpha, g, F))] \\
\leq \limsup_{n \to \infty} \sup_{F \in \mathcal{F}} \sup_{g \in \mathcal{G}} \sup_{c \in \mathbb{R}} [\mathbb{P}_F (g (\mathcal{G}_{n,F}) > c + \eta) - \mathbb{P}_F (g (\mathcal{G}_F) > c)] \leq 0.
\]

This implies (27) because \(\mathbb{P}_F (g (\mathcal{G}_F) > q (1 - \alpha, g, F)) \leq \alpha\).

Now I show (28). Because \(h_{q(1-\alpha,g,F)\eta}(x) < 1 \left( x > q(1-\alpha,g,F) - \eta \right)\) and \(\mathbb{E}_F h_{q(1-\alpha,g,F)\eta} \circ g (\mathcal{G}_F) \geq \alpha, \forall \epsilon > 0,\)

\[
\limsup_{n \to \infty} \sup_{F \in \mathcal{F}} \sup_{g \in \mathcal{G}} \mathbb{P}_F \left( q (1 - \alpha, g, F) > q^*_F (1 - \alpha + \epsilon, g) + \eta \right) \\
\leq \limsup_{n \to \infty} \sup_{F \in \mathcal{F}} \sup_{g \in \mathcal{G}} \mathbb{P}_F \left( \mathbb{E}^* \mathbb{1}(g (\mathbb{G}^*_n) > q (1 - \alpha, g, F) - \eta) < \alpha - \epsilon \right) \\
\leq \limsup_{n \to \infty} \sup_{F \in \mathcal{F}} \sup_{g \in \mathcal{G}} \mathbb{P}_F \left( \mathbb{E}^* h_{q(1-\alpha,g,F)\eta} \circ g (\mathbb{G}^*_n) - \alpha < - \epsilon \right) \\
\leq \limsup_{n \to \infty} \sup_{F \in \mathcal{F}} \left( \mathbb{E}^* h_{q(1-\alpha,g,F)\eta} \circ g (\mathbb{G}^*_n) - \mathbb{E}_F h_{q(1-\alpha,g,F)\eta} \circ g (\mathbb{G}_F) < - \epsilon \right) \\
\leq \frac{1}{\epsilon} \limsup_{n \to \infty} \sup_{F \in \mathcal{F}} \sup_{g \in \mathcal{G}} \left| \mathbb{E}^* h_{q(1-\alpha,g,F)\eta} \circ g (\mathbb{G}^*_n) - \mathbb{E}_F h_{q(1-\alpha,g,F)\eta} \circ g (\mathbb{G}_F) \right|.
\]

Notice that \(\mathcal{G}\) and \(\mathcal{H}\) satisfy the assumptions of Lemma H.2. Then (28) follows because by (29),

\[
\limsup_{n \to \infty} \sup_{F \in \mathcal{F}} \mathbb{E}_F \sup_{g \in \mathcal{G}} |\mathbb{E}^* h_{q(1-\alpha,g,F)\eta} \circ g (\mathbb{G}^*_n) - \mathbb{E}_F h_{q(1-\alpha,g,F)\eta} \circ g (\mathbb{G}_F)| \\
\leq \limsup_{n \to \infty} \sup_{F \in \mathcal{F}} \sup_{c \in \mathbb{R}} \sup_{g \in \mathcal{G}} |\mathbb{E}^* h_{c,\eta} \circ g (\mathbb{G}^*_n) - \mathbb{E}_F h_{c,\eta} \circ g (\mathcal{G}_F)| = 0.
\]

\[
\text{Proof of Lemma H.2.} \quad I \text{ only prove (29). Equation (30) follows in a similar way. First, let } K \text{ be a positive}
\]

59
number. For \( g \in \mathcal{G} \), define \( g_K(\omega) = g(\omega_K) \) where \( \omega_K(\theta, t) = \text{sign}(\omega(\theta, t)) \min(|\omega(\theta, t)|, K) \) and \( \mathcal{G}_K = \{g_K : g \in \mathcal{G}\} \). Notice \( \omega_K(\theta, t) \in \ell^\infty(\Theta \times T) \), and it is bounded by \( K \). Therefore, \( h \circ g_K(\omega) \) is Lipschitz with a constant \( \sup_{|x_1| \leq K, |x_2| \leq K} M(x_1, x_2) \gamma \) and has a bound \( B \) for all \( g \in \mathcal{G} \) and \( h \in \mathcal{H} \). By the triangular inequality,

\[
\mathbb{E}_F \sup_{h \in \mathcal{H}} \sup_{g \in \mathcal{G}} \left| \mathbb{E}^* h \circ g(G_n^*) - \mathbb{E} h \circ g(G_F) \right| \leq D_1 + D_2 + D_3,
\]

\[
D_1 = \mathbb{E}_F \sup_{h \in \mathcal{H}} \sup_{g \in \mathcal{G}} \left| \mathbb{E}^* h \circ g(G_n^*) - \mathbb{E}^* h \circ g_K(G_n^*) \right|, \quad D_2 = \mathbb{E}_F \sup_{h \in \mathcal{H}} \sup_{g \in \mathcal{G}} \left| \mathbb{E}^* h \circ g(G_F) - \mathbb{E} h \circ g_K(G_F) \right|,
\]

\[
D_3 = \mathbb{E}_F \sup_{h \in \mathcal{H}} \sup_{g \in \mathcal{G}} \left| \mathbb{E}^* h \circ g_K(G_n^*) - \mathbb{E} h \circ g_K(G_F) \right|.
\]

Notice \( D_1 \leq 2B \mathbb{P}_F \left( \sup_{(\theta, t) \in \Theta \times T} |G_n^*(\theta, t)| > K \right) \) and \( D_2 \leq 2B \mathbb{P}_F \left( \sup_{(\theta, t) \in \Theta \times T} |G_F(\theta, t)| > K \right) \). Lastly, because \( h \) is bounded by \( B \), \( \forall \varepsilon > 0, \)

\[
D_3 \leq \varepsilon + 2B \mathbb{P}_F \left( \sup_{F \in \mathcal{F}} \sup_{h \in \mathcal{H}} \sup_{g \in \mathcal{G}} \left| \mathbb{E}^* h \circ g_K(G_n^*) - \mathbb{E} h \circ g_K(G_F) \right| > \varepsilon \right).
\]

By Lemma A.2 in Linton, Song, and Whang (2010), the second term on the right-hand side converges to 0 as \( n \to \infty \). Hence, \( \lim_{n \to \infty} \sup_{F \in \mathcal{F}} D_3 \leq \varepsilon \) for every \( \varepsilon > 0 \). Consequently, \( \forall \varepsilon > 0, \ K > 0 \)

\[
\lim_{n \to \infty} \mathbb{E}_F \sup_{h \in \mathcal{H}} \sup_{g \in \mathcal{G}} \left| \mathbb{E}^* h \circ g(G_n^*) - \mathbb{E} h \circ g(G_F) \right| \leq 2B \lim_{n \to \infty} \sup_{F \in \mathcal{F}} \mathbb{P}_F \left( \sup_{(\theta, t) \in \Theta \times T} |G_n^*(\theta, t)| > K \right) + 2B \lim_{n \to \infty} \sup_{F \in \mathcal{F}} \mathbb{P}_F \left( \sup_{(\theta, t) \in \Theta \times T} |G_F(\theta, t)| > K \right) + \varepsilon.
\]

By Lemma A.2. in Linton, Song, and Whang (2010) and Definition 5.1(3), the first two terms are arbitrarily small for sufficiently large \( K \) and \( \varepsilon \) can be arbitrarily small. Therefore, (29) follows.

\[ \square \]

**Proof of Lemma H.3.** Just notice that the set of function \( \mathcal{H} = \{h_{a, \eta} - h_{b, \eta} : a \leq b \in \mathbb{R}\} \) is a collection of bounded Lipschitz functions with Lipschitz constant \( \gamma \leq 2/\eta \). Then by Lemma H.2,

\[
\sup_{F \in \mathcal{F}} \sup_{g \in \mathcal{G}} \sup_{a \leq b \in \mathbb{R}} \left[ \mathbb{P}_F \left( g(G_{n,F}) \in [a, b] \right) - \mathbb{P} \left( g(G_F) \in [a - \eta, b + \eta] \right) \right] \leq \sup_{F \in \mathcal{F}} \sup_{g \in \mathcal{G}} \sup_{a \leq b \in \mathbb{R}} \left[ \mathbb{E}_F \left( h_{a-\eta} - h_{b+\eta} \right) \circ g(G_{n,F}) - \mathbb{P} \left( g(G_F) \in [a - \eta, b + \eta] \right) \right] \leq 0.
\]

\[ \square \]

**Proof of Lemma H.4.** Without loss of generality, assume \( \Gamma_{\kappa,F,m,R}(\omega_1) > \Gamma_{\kappa,F,m,R}(\omega_2) \). By the definition
of $\Gamma_{\kappa,F,m,R}$, $\forall \epsilon > 0$, $\exists \tilde{\theta} \in \Theta_{m,F}$ ($R$) and $\exists \theta \in R_{m,F}^\kappa (\tilde{\theta})$ such that

$$\int_{T} \left| \omega_2 (\bar{\theta}, t) + \sqrt{m} \frac{dE_F \rho_k (W_i, \bar{\theta})}{d\theta} \left[ \theta - \bar{\theta} \right] \right|^2 d\mu (t) < \Gamma_{\kappa,R,F,m} (\omega_2) + \epsilon. \quad (40)$$

Therefore,

$$\Gamma_{\kappa,F,m,R} (\omega_1) - \Gamma_{\kappa,F,m,R} (\omega_2) \leq \int_{T} \left| \omega_1 (\bar{\theta}, t) + \sqrt{m} \frac{dE_F \rho_k (W_i, \bar{\theta})}{d\theta} \left[ \theta - \bar{\theta} \right] \right|^2 d\mu (t) - \int_{T} \left| \omega_2 (\bar{\theta}, t) + \sqrt{m} \frac{dE_F \rho_k (W_i, \bar{\theta})}{d\theta} \left[ \theta - \bar{\theta} \right] \right|^2 d\mu (t) + \epsilon \leq d_{\infty} (\omega_1, \omega_2) \int_{T} \left| \omega_2 (\bar{\theta}, t) + \sqrt{m} \frac{dE_F \rho_k (W_i, \bar{\theta})}{d\theta} \left[ \theta - \bar{\theta} \right] \right|^2 d\mu (t) \leq 2d_{\infty} (\omega_1, \omega_2) \int_{T} \left| \omega_2 (\bar{\theta}, t) + \sqrt{m} \frac{dE_F \rho_k (W_i, \bar{\theta})}{d\theta} \left[ \theta - \bar{\theta} \right] \right|^2 d\mu (t) + \epsilon.$$

Because $\bar{\theta} \in R_{m,F}^\kappa (\tilde{\theta})$, $\Gamma_{\kappa,R,F,m} (\omega_2) \leq \int_{T} \left| \omega_2 (\bar{\theta}, t) \right|^2 d\mu (t) \leq \| \omega_2 \|_{\infty}^2$. Therefore, by (40) and the Jensen’s inequality

$$\int_{T} \left| \omega_2 (\bar{\theta}, t) + \sqrt{m} \frac{dE_F \rho_k (W_i, \bar{\theta})}{d\theta} \left[ \theta - \bar{\theta} \right] \right|^2 d\mu (t) \leq \sqrt{\Gamma_{\kappa,R,F,m} (\omega_2) + \sqrt{\epsilon}} \leq \| \omega_2 \|_{\infty} + \sqrt{\epsilon} \leq \max (\| \omega_1 \|_{\infty}, \| \omega_2 \|_{\infty}) + \sqrt{\epsilon}. \leq \sqrt{\Gamma_{\kappa,R,F,m} (\omega_2) + \sqrt{\epsilon}} \leq \max (\| \omega_1 \|_{\infty}, \| \omega_2 \|_{\infty}) + \sqrt{\epsilon}. \leq \sqrt{\Gamma_{\kappa,R,F,m} (\omega_2) + \sqrt{\epsilon}} \leq \max (\| \omega_1 \|_{\infty}, \| \omega_2 \|_{\infty}) + \sqrt{\epsilon}.$$

In addition, $\int_{T} \left| \omega_1 (\bar{\theta}, t) - \omega_2 (\bar{\theta}, t) \right| d\mu (t) \leq 2 \max (\| \omega_1 \|_{\infty}, \| \omega_2 \|_{\infty})$. Therefore,

$$\Gamma_{\kappa,F,m,R} (\omega_1) - \Gamma_{\kappa,F,m,R} (\omega_2) \leq 4d_{\infty} (\omega_1, \omega_2) \left[ \max (\| \omega_1 \|_{\infty}, \| \omega_2 \|_{\infty}) + \sqrt{\epsilon} \right] + \epsilon.$$

Because $\epsilon$ can be any positive number, the lemma follows.

Proof of Lemma H.5. First notice that the zero function lies in both $V_{\kappa}^{\infty} (\theta)$ and $V_{\infty} (\theta)$ for all $\theta$. By the definition of inf, $\forall \epsilon_1 > 0$, one can find $\bar{\theta} \in \Theta_F \cap R$ and $\bar{v} \in V_{\infty} (\bar{\theta})$ such that

$$g (\omega) + \epsilon_1 > \int_{T} \left[ \omega (\bar{\theta}, t) + \bar{v} (t) \right]^2 d\mu (t), \leq \int_{T} \left[ \omega (\bar{\theta}, t) + \bar{v} (t) \right]^2 d\mu (t) \leq \int_{T} \omega (\bar{\theta}, t)^2 d\mu (t) \leq \| \omega \|_{\infty}^2.$$
Because \( \bar{v} \in V_\infty (\bar{\theta}) \), for any \( \epsilon_2 > 0 \), there exists an \( \bar{v} \in \cup V_n (\theta) \) such that \( \sup_{t \in T} |\bar{v}(t) - \bar{v}(t)| < \epsilon_2 \). Because \( c_n \to \infty \), \( \cup V_n (\theta) = \cup V_{cn} (\theta) \). Consequently, \( \bar{v} \in V_{cn} (\theta) \) for sufficiently large \( n \). Therefore, for sufficiently large \( n \)

\[
g_n (\omega_n) - g (\omega) \leq \int_T [\omega_n (\bar{\theta}, t) + \bar{v}(t)]^2 d\mu(t) - \int_T [\omega (\bar{\theta}, t) + \bar{v}(t)]^2 d\mu(t) + \epsilon_1
\]

\[
= \int_T [\omega_n (\bar{\theta}, t) + \bar{v}(t) - \omega (\bar{\theta}, t) - \bar{v}(t)] \omega_n (\bar{\theta}, t) + \bar{v}(t) + \omega (\bar{\theta}, t) + \bar{v}(t) d\mu(t) + \epsilon_1
\]

\[
\leq (\|\omega_n - \omega\| + \|\bar{v}(t) - \bar{v}(t)\|) \int_T [\omega_n (\bar{\theta}, t) + \bar{v}(t) + \omega (\bar{\theta}, t) + \bar{v}(t)] d\mu(t) + \epsilon_1
\]

\[
\leq (\|\omega_n - \omega\| + \|\bar{v}(t) - \bar{v}(t)\|) \left[ 2 \int_T [\omega (\bar{\theta}, t) + \bar{v}(t)] d\mu(t) + (\|\omega_n - \omega\| + \|\bar{v}(t) - \bar{v}(t)\|) \right] + \epsilon_1
\]

\[
\leq (\|\omega_n - \omega\| + \|\bar{v}(t) - \bar{v}(t)\|) \left[ 2 \int_T [\omega (\bar{\theta}, t) + \bar{v}(t)]^2 d\mu(t) + (\|\omega_n - \omega\| + \|\bar{v}(t) - \bar{v}(t)\|) \right] + \epsilon_1
\]

\[
\leq (\|\omega_n - \omega\| + \|\bar{v}(t) - \bar{v}(t)\|) \left( 2 \|\omega\| + \|\omega_n - \omega\| + \|\bar{v}(t) - \bar{v}(t)\| \right) + \epsilon_1
\]

\[
\leq (\|\omega_n - \omega\| + \epsilon_2) \left( 2 \|\omega\| + \|\omega_n - \omega\| + \epsilon_2 \right) + \epsilon_1,
\]

where the fourth inequality holds by Jensen’s inequality. Therefore,

\[
\lim \sup_{n \to \infty} [g_n (\omega_n) - g (\omega)] \leq \epsilon_2 \left( 2 \|\omega\| + \epsilon_2 \right) + \epsilon_1.
\]

Because \( \epsilon_1 \) and \( \epsilon_2 \) can be arbitrarily small, \( \lim \sup_{n \to \infty} [g_n (\omega_n) - g (\omega)] \leq 0 \). Similarly, for any \( \epsilon_1 > 0 \) and each \( n \), one can find \( \theta_n \in \Theta_F \cap R \) and \( v_n \in V_{cn} (\theta_n) \) such that

\[
g_n (\omega_n) + \epsilon_1 > \int_T [\omega_n (\theta_n, t) + v_n(t)]^2 d\mu(t),
\]

\[
\int_T [\omega_n (\theta_n, t) + v_n(t)]^2 d\mu(t) \leq \int_T [\omega_n (\theta_n, t)]^2 d\mu(t) \leq \|\omega_n\|^2.
\]

Similarly, one can show that

\[
g (\omega) - g_n (\omega_n) \leq \int_T [\omega (\theta_n, t) + v_n(t)]^2 d\mu(t) - \int_T [\omega_n (\theta_n, t) + v_n(t)]^2 d\mu(t) + \epsilon_1
\]

\[
\leq (\|\omega_n - \omega\| \left( 2 \|\omega\| + 3 \|\omega_n - \omega\| \right) + \epsilon_1 \to \epsilon_1,
\]

which implies that \( \lim \inf_{n \to \infty} [g_n (\omega_n) - g (\omega)] \geq 0 \). Therefore, \( |g_n (\omega_n) - g (\omega)| \to 0 \).

\[
\square
\]

**Lemma I.1.** There exist random variables \( A_{n,F} \) and \( B_{n,F} \) such that uniformly in \( F \in F \), \( A_{n,F} = O_P (n^{-1/2}) \), \( B_{n,F} = O_P (n^{-1}) \), and

\[
|Q_F (\theta) - Q_n (\theta)| \leq 2d_{w,F} (\theta, \Theta_F) A_{n,F} + B_{n,F} = O_P \left( n^{-1/2} \right).
\] (41)
Proof. First notice that
\[
|Q_F(\theta) - Q_n(\theta)| = \left| \frac{\int_T \mathcal{G}_{n,F}(\theta, t)^2 \, d\mu(t)}{n} + 2\frac{\int_T \mathcal{G}_{n,F}(\theta, t) \mathbb{E}_F B(W_t, \theta) \, d\mu(t)}{\sqrt{n}} \right|
\]
\[
\leq 2d_{w,F}(\theta, \Theta_F) \frac{\int_T \mathcal{G}_{n,F}(\theta, t)^2 \, d\mu(t)}{\sqrt{n}} + \frac{\int_T \mathcal{G}_{n,F}(\theta, t)^2 \, d\mu(t)}{n}
\]
\[
\leq 2d_{w,F}(\theta, \Theta_F) \sup_{(\theta, t) \in \Theta \times T} \mathbb{E}_F \mathcal{G}_{n,F}(\theta, t) \mid \mathcal{G}_{n,F}(\theta, t) \mid \frac{\max_{(\theta, t) \in \Theta \times T} \mathbb{E}_F |\mathcal{G}_{n,F}(\theta, t)|^2}{n}
\]
\[
= 2d_{w,F}(\theta, \Theta_F) A_{n,F} + B_{n,F},
\]
where \(A_{n,F} = \sup_{(\theta, t) \in \Theta \times T} \mathbb{E}_F \mathcal{G}_{n,F}(\theta, t) \mid \mathcal{G}_{n,F}(\theta, t) \mid \frac{\max_{(\theta, t) \in \Theta \times T} \mathbb{E}_F |\mathcal{G}_{n,F}(\theta, t)|^2}{n} \). By Definition 5.1(3),
\[
\sup_{(\theta, t) \in \Theta \times T} \mathbb{E}_F \mathcal{G}_{n,F}(\theta, t) \mid \mathcal{G}_{n,F}(\theta, t) \mid = O_{\mathbb{P}_F}(1) \text{ uniformly in } F \in \mathcal{F}.
\]
Consequently, \(A_{n,F} = O_{\mathbb{P}_F}(n^{-1/2})\) and \(B_{n,F} = O_{\mathbb{P}_F}(n^{-1})\) uniformly in \(F \in \mathcal{F}\). Lastly, notice that \(d_{w,F}(\theta, \Theta_F)^2 \leq \sup_{F \in \mathcal{F}} \mathbb{E}_F F(W_t)^2 < \infty\). Therefore, \(2d_{w,F}(\theta, \Theta_F) A_{n,F} + B_{n,F} = O_{\mathbb{P}_F}(n^{-1/2})\) uniformly in \(F \in \mathcal{F}\).

\[\square\]

Lemma I.2. \(\{X_{n,a}\}_{n=1}^{\infty}\) and \(\{Y_{n,a}\}_{n=1}^{\infty}\) are sequences of random variables indexed by \(a \in \mathcal{A}\) and defined on the same probability space. For every \(a \in \mathcal{A}\), \(\mathbb{P}_a\) is a probability measure on this probability space. \(g\) is a continuous function on the real line. If \(\lim_{n \to \infty} \sup_{a \in \mathcal{A}} \mathbb{P}_a(|X_{n,a} - Y_{n,a}| > \epsilon) = 0 \forall \epsilon > 0\) and \(X_{n,a}\) is asymptotically tight uniformly in \(a \in \mathcal{A}\), then \(\forall \epsilon > 0\), \(\lim_{n \to \infty} \sup_{a \in \mathcal{A}} \mathbb{P}_a(|g(X_{n,a}) - g(Y_{n,a})| > \epsilon) = 0\).

Proof. By Assumption, \(\forall \eta > 0\), \(\exists M > 0\) such that \(\lim_{n \to \infty} \sup_{a \in \mathcal{A}} \mathbb{P}_a(|X_{n,a}| > M) < \eta\). Because \(g\) is continuous, it is absolutely continuous on \([-2M, 2M]\). For any \(\epsilon > 0\), there exists \(0 < \delta < M\) such that \(|g(x) - g(y)| \leq \epsilon\) if \(|x - y| \leq \delta \forall x, y \in [-2M, 2M]\).

\[
\mathbb{P}_a(|g(X_{n,a}) - g(Y_{n,a})| > \epsilon) \leq \mathbb{P}_a(|X_{n,a} - Y_{n,a}| > \delta, |X_{n,a}| \leq M, |Y_{n,a}| \leq 2M) + \mathbb{P}_a(|X_{n,a}| \leq M, |Y_{n,a}| > 2M) + \mathbb{P}_a(|X_{n,a}| > M)
\]
\[
\leq \mathbb{P}_a(|X_{n,a} - Y_{n,a}| > \delta) + \mathbb{P}_a(|X_{n,a} - Y_{n,a}| > M) + \mathbb{P}_a(|X_{n,a}| > M).
\]

The first two terms converge to 0 uniformly in \(a \in \mathcal{A}\). Therefore,
\[
\lim_{n \to \infty} \sup_{a \in \mathcal{A}} \mathbb{P}_a(|g(X_{n,a}) - g(Y_{n,a})| > \epsilon) \leq \lim_{n \to \infty} \sup_{a \in \mathcal{A}} \mathbb{P}_a(|X_{n,a}| > M) \leq \eta.
\]

This concludes the proof because \(\eta\) can be any positive number.

\[\square\]