WORKING PAPERS

N° 1332

May 2022

# "Encompassing Tests for Nonparametric Regressions"

Elia Lapenta and Pascal Lavergne

Toulouse School of Economics

# Encompassing Tests for Nonparametric Regressions*

Elia Lapenta† and Pascal Lavergne‡

January 31, 2022

**Abstract**

We set up a formal framework to characterize encompassing of nonparametric models through the $L^2$ distance. We contrast it to previous literature on the comparison of nonparametric regression models. We then develop testing procedures for the encompassing hypothesis that are fully nonparametric. Our test statistics depend on kernel regression, raising the issue of bandwidth's choice. We investigate two alternative approaches to obtain a "small bias property" for our test statistics. We show the validity of a wild bootstrap method, and we illustrate the attractive features of our tests for small and moderate samples.

**Keywords**: Encompassing, Nonparametric Regression, Bootstrap, Bias Correction, Locally Robust Statistic.

**JEL Classification**: C01, C12, C14

*P. Lavergne acknowledges funding from ANR under grant ANR-17-EURE-0010 (Investissements d'Avenir program).

†*CREST and ENSAE. Email: elia.lapenta@ensae.fr* Address correspondence: CREST, 5 Avenue Le Chatelier, 91120 Palaiseau, FRANCE.

‡*Toulouse School of Economics. Email: lavergnetse@gmail.com* Address correspondence: Toulouse School of Economics, 1 Esplanade de l'Université, 31080 Toulouse Cedex 06, FRANCE.

# 1 Introduction

The encompassing principle was introduced in econometrics by Hendry and Richard (1982), Gourieroux, Monfort, and Trognon (1983), and Mizon and Richard (1986), and further developed in Gourieroux and Monfort (1995), Florens, Hendry, and Richard (1996), and Dhaene, Gourieroux, and Scaillet (1998) among others. It provides a natural principle for choosing between two competing theories. A new theory must be able to accommodate the results obtained by a concurrent older one. An extensive survey is provided in Bontemps and Mizon (2008).

Our goal is to propose encompassing tests for nonparametric models. In that aim, the main steps are (i) to formally define encompassing for nonparametric models, (ii) to develop fully nonparametric encompassing tests, and (iii) to show asymptotic validity of a wild bootstrap method for asymptotic inference. Our first contribution is thus to formally set up a framework to precisely define encompassing for nonparametric regression models. We discuss nonparametric encompassing with respect to previous literature on the comparison of such models, whether nested or non nested, see below for references. We show that encompassing reduces neither to significance of some variables nor to the comparison of the models' theoretical fit. Hence, the null hypothesis of encompassing cannot generally be tested with existing procedures. Our second contribution is to propose fully nonparametric encompassing tests for regression models. Existing encompassing tests rely on parametric functional forms, except for Bontemps, Florens, and Richard (2008) who propose a test aimed at assessing a consequence of encompassing. The new tests we develop directly test the encompassing hypothesis. They are based on an empirical process estimating a continuum of unconditional moments following the ICM principle introduced by Bierens (1982). Our third contribution is to develop a wild bootstrap method and to show that it provides asymptotically correct inference.

Our fourth contribution is to propose and investigate two approaches to obtain a "small bias property." Our test statistic depends on an empirical process involving a first-step nonparametric estimator. The small bias property of a semiparametric estimator is that its bias converges to zero faster than the pointwise and integrated bias of the nonparametric estimator on which it is based. A distinguishing feature is that the estimator is $\sqrt{n}$-consistent even when the nonparametric estimator on which it is based converges at the optimal nonparametric rate. Without this property, using a first-step nonparametric estimator necessitates some undersmoothing, and this complicates practical implementation. Newey, Hsieh, and Robins (2004) have developed a generic technique based on twicing kernels to obtain semiparametric estimators with small bias. We develop here two alternative methods that yield a small bias for the empirical process of interest. The first one uses a bias-corrected kernel estimator based on the boosting principle (Di Marzio and Taylor, 2008; Park, Lee, and Ha, 2009). The

second approach is to make the empirical process of interest *locally robust* with respect to the nonparametric regression. This has previously been used successfully in semiparametric estimation, see Newey (1990) for an early example and Chernozhukov, Escanciano, Ichimura, Newey, and Robins (2022) for a general approach. We here adapt the approach to apply it to an empirical process. We show that these two methods yield a small bias property in the asymptotic expansion of our test statistic. This allows for a larger set of smoothing parameters, so the test is expected to be less sensitive to the bandwidth choice.

Our work is related to the extensive literature on consistent specification testing based on empirical processes, see Bierens and Ploberger (1997), Stinchcombe and White (1998), Xia, Li, Tong, and Zhang (2004), Escanciano (2006), Delgado and Stute (2008), Lavergne and Patilea (2008) to mention just a few. The main features of our tests compared to previous work are that (i) our empirical process contains a nonparametric kernel estimator, and (ii) due to the form of the null hypothesis we cannot use a density-weighted process, and we thus need to control for a random denominator. Our nonparametric encompassing tests are also connected to the comparison of nonparametric regressions in nested and non nested cases, e.g., Fan and Li (1996), Lavergne and Vuong (1996), Delgado and Manteiga (2001), and Lavergne, Maistre, and Patilea (2015). We show however that the encompassing hypothesis cannot be tested through existing procedures. Our work is also related to the literature on estimation and testing with nonparametric nuisance components, e.g., Escanciano, Jacho-Chávez, and Lewbel (2014) and Mammen, Rothe, and Schienle (2016). The former authors obtain uniform-in-bandwidth expansions for an empirical process similar to the one we consider. We focus here on obtaining a small bias property, but we do not formally establish uniformity in bandwidth.

Our paper is organized as follows. Section 2 formalizes the encompassing notion for nonparametric models and compares our framework to the literature on comparison of nonparametric regressions. Section 3 details the construction of the test statistics and the two approaches used to obtain a small bias property. Section 4 is devoted to the analysis of the asymptotic behavior of our statistics. Since the asymptotic distribution under the null depends on unknown features of the DGP, we establish in Section 5 the validity of a wild bootstrap procedure. Section 6 provides evidence about the small sample performances of our procedures to check that bootstrapping allows to correctly control size and that our tests have power. We investigate thoroughly the influence of the bandwidth as well as of the trimming parameter, which is theoretically necessary, and we evaluate the benefits of each bias-reducing approach. Section 8 contains the proofs of our main results.

## 2 Encompassing for Nonparametric Regressions

The definition of encompassing starts with the definition of the *binding function*, see Gourieroux and Monfort (1995). In a parametric context, we typically start with two competing parameterized families of densities for $Y$, $\mathcal{M}_1 = \{g_1(\cdot, \alpha_1) : \alpha_1 \in A_1\}$ and $\mathcal{M}_2 := \{g_2(\cdot, \alpha_2) : \alpha_2 \in A_2\}$. The pseudo true value of $\alpha_i$, $i = 1$ or 2, is defined as

$$\alpha_i^* = \arg \min_{\alpha_i \in A_i} d(f(\cdot), g_i(\cdot, \alpha_i)),$$

where $f(\cdot)$ is the true density of $Y$ and $d$ is some divergence between the two distributions, for instance the Kullback-Leibler divergence $\mathbb{E}_f [\log (f(\cdot)/g(\cdot, \alpha))]$, with expectation taken with respect to $f(\cdot)$. The biding function $b(\alpha_1)$ is a correspondence between an element of model $\mathcal{M}_1$ and the element of model $\mathcal{M}_2$ that is closest to it. Specifically,

$$b(\alpha_1) = \arg \min_{\alpha_2 \in A_2} d(g_1(\cdot, \alpha_1), g_2(\cdot, \alpha_2)).$$

We then say that $\mathcal{M}_1$ encompasses model $\mathcal{M}_2$ if $\alpha_2^* = b(\alpha_1^*)$. That is, $\mathcal{M}_1$ encompasses $\mathcal{M}_2$ if the pseudo-true value of the latter can be obtained from the pseudo-true value of the former.

Here we focus on two nonparametric competing models to explain $Y$, where $\mathcal{M}_W$ uses covariates $W$ and Model $\mathcal{M}_X$ uses $X$. A nonparametric model with a specific set of covariates is a function of these variables. To define the pseudo-true value that corresponds to the best explanation of $Y$, we consider the $\mathcal{L}_2$ distance. That is, $\mathcal{M}_W$ and $\mathcal{M}_X$ are respectively defined as $\mathcal{L}^2(W)$, the space of the square integrable functions of $W$, and $\mathcal{L}^2(X)$. The "pseudo-true functions" are then the regression functions

$$\begin{aligned} m(W) &= \arg \min_{g \in \mathcal{L}^2(W)} \mathbb{E}\,(Y - g(W))^2 = \mathbb{E}\,(Y|W), \\ \text{and } m_X(X) &= \arg \min_{h \in \mathcal{L}^2(X)} \mathbb{E}\,(Y - h(X))^2 = \mathbb{E}\,(Y|X). \end{aligned} \tag{1}$$

The binding function is similarly defined as

$$b(g(W)) = \arg \min_{h \in \mathcal{L}^2(X)} \mathbb{E}\,(g(W) - h(X))^2 = \mathbb{E}\,(g(W)|X)\,.$$

We thus say that $\mathcal{M}_W$ encompasses model $\mathcal{M}_X$ if

$$b(m(W)) = m_X(X) \text{ a.s.} \Leftrightarrow \mathbb{E}\,(\mathbb{E}\,(Y|W)|X) = \mathbb{E}\,(Y|X) \text{ a.s.} \tag{2}$$

That is, the regression function of $Y$ on $X$ can be obtained from the regression function of $Y$ on $W$. Similarly, $\mathcal{M}_X$ encompasses model $\mathcal{M}_W$ if $\mathbb{E}\,(\mathbb{E}\,(Y|X)|W) = \mathbb{E}\,(Y|W)$ almost

surely. In what follows, we explore the implications of the definition of encompassing for nonparametric regressions.

## 2.1 Encompassing and Model Fit

Lavergne and Vuong (1996) proposed comparing nonparametric regression models on the basis of their theoretical fit and considered the hypotheses

$$\begin{aligned} H_0 &: \mathbb{E}\left[Y - m(W)\right]^2 - \mathbb{E}\left[Y - m_X(X)\right]^2 = 0\,, \\ H_W &: \mathbb{E}\left[Y - m(W)\right]^2 - \mathbb{E}\left[Y - m_X(X)\right]^2 < 0\,, \\ H_X &: \mathbb{E}\left[Y - m(W)\right]^2 - \mathbb{E}\left[Y - m_X(X)\right]^2 > 0\,. \end{aligned}$$

Non rejection of the null hypothesis $H_0$ means that both models have the same theoretical fit. Rejection of $H_0$ in favor of either $H_W$ or $H_X$ indicates which model dominates the other. This framework is quite general since it does not make a distinction between nested and non nested situations and treats the two competing models symmetrically. Lavergne and Vuong (1996) built a test of $H_0$ against $H_W$ and $H_X$ based on the comparison of the empirical analogs of the models' fit. They found that the asymptotic distribution of their statistic is degenerate under $H_0$ when the two models are "generalized nested regressions," that is when either $\mathbb{E}\left(\mathbb{E}\left(Y|W\right)|X\right) = \mathbb{E}\left(Y|X\right)$ a.s. or $\mathbb{E}\left(\mathbb{E}\left(Y|X\right)|W\right) = \mathbb{E}\left(Y|W\right)$ a.s. Hence their test cannot be used when there is encompassing, and our tests developed below are complementary to theirs. It could also be used as a preliminary test to check whether their test can be entertained.[1]

Comparing nonparametric regressions through their fit seems natural. One model can have a better fit than another without encompassing it. Conversely, does encompassing imply a better fit for the encompassing model? As we detail below, the answer is yes: if $\mathcal{M}_W$ encompasses $\mathcal{M}_X$, the theoretical fit of $\mathbb{E}\left(Y|W\right)$ is at least as good as the one of $\mathbb{E}\left(Y|X\right)$, and is strictly better except when $\mathbb{E}\left(Y|W\right) = \mathbb{E}\left(Y|X\right)$ a.s.

**Proposition 2.1.** *If $\mathcal{M}_W$ encompasses $\mathcal{M}_X$,*

*(a) $H_X$ cannot hold,*

*(b) $H_0$ holds iff $\mathbb{E}\left(Y|W\right) = \mathbb{E}\left(Y|X\right)$ a.s. iff $\mathcal{M}_X$ encompasses $\mathcal{M}_W$.*

*Proof.* Consider Statement (a). Since $\mathbb{E}\left(\mathbb{E}\left(Y|W\right)|X\right) = \mathbb{E}\left(Y|X\right)$ a.s.,

$$E\left[Y - \mathbb{E}\left(Y|X\right)\right]^2 = \mathbb{E}\left[Y - \mathbb{E}\left(Y|W\right)\right]^2 + \mathbb{E}\left[\mathbb{E}\left(Y|W\right) - \mathbb{E}\left(Y|X\right)\right]^2\,,$$

[1] We leave the study of the formal properties of such a two-step procedure for future work.

as the cross-term cancels, and $E\left[Y - \mathbb{E}\left(Y|X\right)\right]^2 \geq \mathbb{E}\left[Y - \mathbb{E}\left(Y|W\right)\right]^2$.

Consider Statement (b). It is obvious that if $\mathbb{E}\left(Y|W\right) = \mathbb{E}\left(Y|X\right)$ a.s. then $H_0$ holds and both models encompass each other. By Lavergne and Vuong (1996, Lemma 1), if one model encompasses the other and $H_0$ holds, it should be that $\mathbb{E}\left(Y|W\right) = \mathbb{E}\left(Y|X\right)$. Finally, if both models encompass each other, then $H_0$ holds by Statement (a). ■

Hence, except in the case where the two regressions are equal, encompassing implies a strictly better fit for the encompassing model. If the two regressions are equal, and thus encompass each other, then each regression depends solely on the set of variables that are common to $W$ and $X$ (or more generally on the intersection of the sigma-algebra generated by the two sets).

## 2.2 Encompassing and Significance

When the two sets of regressors are nested, specifically when $W = (X, Z)$, then it is clear that (2) holds. A more interesting question is whether $\mathcal{M}_X$ encompasses $\mathcal{M}_W$, that is whether

$$\mathbb{E}\left(\mathbb{E}\left(Y|X\right)|X, Z\right) = \mathbb{E}\left(Y|X\right) = \mathbb{E}\left(Y|X, Z\right) \text{ a.s.}$$

In this setup, encompassing is equivalent to whether $Z$ is significant in the regression function of $Y$ on $X$ and $Z$. Significance testing in nonparametric regressions has been a focus of extensive work (Ait-Sahalia, Bickel, and Stoker, 2001; Delgado and Manteiga, 2001; Fan and Li, 1996; Lavergne, 2001; Lavergne and Vuong, 2000).

Consider now two nonnested sets of regressors, and assume the regressors $X$ are not significant once we control for $W$, that is

$$\mathbb{E}\left(Y|W, X\right) = \mathbb{E}\left(Y|W\right) \text{ a.s.} \tag{3}$$

By conditioning again with respect to $X$, one obtains that

$$\mathbb{E}\left(Y|X\right) = \mathbb{E}\left[\mathbb{E}\left(Y|W\right)|X\right] \text{ a.s.}$$

so that $\mathcal{M}_W$ encompasses $\mathcal{M}_X$. Bontemps et al. (2008) consider (3) as their hypothesis of interest, since it implies encompassing. The other direction of the implication, however, does not hold: if $\mathcal{M}_W$ encompasses $\mathcal{M}_X$, then it is not necessarily true that the covariates $X$ are not significant in the nonparametric regression of $Y$ onto $(W, X)$, as shown below.

**Proposition 2.2.** *$\mathcal{M}_W$ encompasses $\mathcal{M}_X$ if and only if*

$$\mathbb{E}\left(Y|W, X\right) = \mathbb{E}\left(Y|W\right) + g(W, X) \qquad \textit{with } \mathbb{E}\left[g(W, X)|W\right] = \mathbb{E}\left[g(W, X)|X\right] = 0 \textit{ a.s.}$$

*Proof.* By definition, $Y = \mathbb{E}(Y|W) + \varepsilon$ with $\mathbb{E}(\varepsilon|W) = 0$. Hence,

$$\mathbb{E}(Y|W,X) = \mathbb{E}(Y|W) + \mathbb{E}(\varepsilon|W,X) = \mathbb{E}(Y|W) + g(W,X)\,.$$

Conditioning on $W$ yields $\mathbb{E}[g(W,X)|W] = 0$. Conditioning on $X$ and using the fact that $\mathcal{M}_W$ encompasses $\mathcal{M}_X$ yields $\mathbb{E}[g(W,X)|X] = 0$. This shows necessity. Sufficiency similarly follows by conditioning $\mathbb{E}(Y|W,X) = \mathbb{E}(Y|W) + g(W,X)$ on $X$ and using $\mathbb{E}[g(W,X)|X] = 0$. ■

The previous result highlights that for non nested sets of regressors, the encompassing property does not reduce to the significance of some regressors in the complete regression. As long as $g(W,X)$ is not a.s. equal to zero, $\mathcal{M}_W$ encompasses $\mathcal{M}_X$, but $X$ is a significant covariate in $\mathbb{E}(Y|W,X)$. The following provides a concrete example.

**Example.** Let $W$ be continuous and symmetrically distributed around 0 with density $f_W(\cdot)$ and

$$f_{X|W}(x|w) = \mathbb{I}(w \geq 0)\varphi(x) + \mathbb{I}(w < 0)\psi(x)\,,$$

where $\varphi$ and $\psi$ are two densities with mean 0 and $\mathbb{I}$ is the indicator function. Then

$$f_X(x) = 0.5\varphi(x) + 0.5\psi(x) \quad \text{and} \quad f_{W|X}(w|x) = 2f_W(w)\frac{\mathbb{I}(w \geq 0)\varphi(x) + \mathbb{I}(w < 0)\psi(x)}{\varphi(x) + \psi(x)}\,.$$

Let

$$Y = m(W) + h(W)X + \eta\,, \qquad \mathbb{E}(\eta|W,X) = 0\,,$$

where $\mathbb{E}[h(W)\mathbb{I}(W \geq 0)] = \mathbb{E}[h(W)\mathbb{I}(W < 0)] = 0$. The function $h(W)$ could for instance be chosen as $s(W^2) - \mathbb{E}\,s(W^2)$, with $s(\cdot)$ any non-constant integrable function. Then

$$\mathbb{E}(X|W) = \mathbb{I}(W \geq 0)\int x\varphi(x)\,dx + \mathbb{I}(W < 0)\int x\psi(x)\,dx = 0\,.$$

Moreover $\mathbb{E}[h(W)|X] = 0$. It follows that $\mathbb{E}[h(W)X|W] = \mathbb{E}[h(W)X|X] = 0$, and $\mathcal{M}_W$ encompasses $\mathcal{M}_X$, but $X$ is significant in $\mathbb{E}(Y|W,X)$. Furthermore, $\mathcal{M}_X$ does not encompass $\mathcal{M}_W$ in general. If this holds, then $m(W) = \mathbb{E}(Y|X)$ a.s. from Proposition 2.1. Integrating both sides with respect to the marginal density $f_X(\cdot)$ implies that $m(W)$ and thus $\mathbb{E}(Y|X)$ are both constant a.s.

# 3 Tests Statistics

## 3.1 ICM Statistic

We want to test

$$H_0 \,:\, \mathbb{E}\,(E(Y|W)|X) = \mathbb{E}\,(Y|X) \text{ a.s.} \Leftrightarrow \mathbb{E}\,(Y - E(Y|W)|X) = 0 \text{ a.s.}$$

against its logical complement $H_1 = H_0^c$. While $H_0$ is a conditional moment restriction, we can consider instead an equivalent continuum of unconditional moments. Assume $X \in \mathbb{R}^d$ has bounded support, which is without loss of generality, as we can always transform $X$ by a one-to-one function that maps it to a compact set. Then the null hypothesis is equivalent to

$$H_0 \,:\, \mathbb{E}\,[(Y - \mathbb{E}\,(Y|W))\,\varphi(s'X)] = 0 \quad \text{a.s.} \quad \forall s \in \mathcal{S}\,, \tag{4}$$

where $\mathcal{S}$ is a (arbitrary) neighborhood of the origin in $\mathbb{R}^d$ and $\varphi(\cdot)$ is a well-chosen function, see Assumption A below for precise conditions. Some convenient choices for $\varphi(\cdot)$ are as follows. Bierens (1982) shows the previous equivalence for the complex exponential $\varphi(u) = \exp(iu)$, Bierens (1990) considers the exponential $\varphi(u) = \exp(u)$, Bierens and Ploberger (1997) the logistic c.d.f. $\varphi(u) = 1/(1 + \exp(c - u))$, see also Stinchcombe and White (1998). Other types of functions could be used such as indicator functions (Delgado and Manteiga, 2001; Escanciano, 2006).

If we observe a random sample $(Y_i, W_i, X_i)$, $i = 1, \ldots n$, from $(Y, W, X)$, and if we know the precise form of $\mathbb{E}\,(Y|W)$, then Bierens' Integrated Conditional Moment (ICM) statistic for testing $H_0$ is

$$\int_{\mathbb{R}^d} \left| n^{-1/2} \sum_{j=1}^{n} (Y_j - \mathbb{E}\,(Y|W_j))\,\varphi(s'X_j) \right|^2 d\mu(s)\,, \tag{5}$$

where $\mu$ is some probability measure on $\mathcal{S}$, such as the uniform distribution on $\mathcal{S}$. Alternatively, a Kolmogorov-Smirnov type statistic could be considered, but the Cramer-von-Mises form appears to be easier to deal with in practice, see below.

In practice, we use a kernel nonparametric estimator of the conditional expectation $\mathbb{E}\,(Y|W)$. Let $K(\cdot)$ be a kernel on $\mathbb{R}^p$, $h$ a bandwidth, and $K_h(u) = K\,(u_1/h, \ldots u_p/h)$. Define

$$\overline{Y}(w) = (nh^p)^{-1} \sum_{i=1}^{n} Y_i K_h\,(W_i - w)\,.$$

With $e = (1, \ldots, 1)'$, let $\widehat{f}(w) = \overline{e}(w)$, and $\widehat{m}(w) = \overline{Y}(w)/\widehat{f}(w)$.[2]

[2]For notational simplicity we are assuming the same bandwidth across regressors, but our proofs carry over when each regressor has a specific bandwidth.

To control for the random denominator in the kernel estimator, we introduce a trimming factor $\widehat{t}(w) = \mathbb{I}\left(\widehat{f}(w) \geq \tau_n\right)$, where $\tau_n$ converges to zero. Let $\widehat{\varepsilon}_i = Y_i - \widehat{m}(W_i)$, $\phi_s(\cdot) = \varphi(s'\cdot)$, and $\mathbb{P}_n(g) = n^{-1}\sum_{i=1}^n g(Z_i)$ denotes the empirical mean process based on $g(\cdot)$. An ICM test statistic can be built upon the process $\mathbb{P}_n\left(\widehat{\varepsilon}\phi_s\widehat{t}\right)$ as

$$S_n = n\int_{\mathcal{S}} |\mathbb{P}_n\left(\widehat{\varepsilon}\phi_s\widehat{t}\right)|^2 \, d\mu(s)\,.$$

In our practical implementation, we chose $\varphi(\cdot)$ as the complex exponential and $\mu$ to be symmetric around the origin. Let

$$a(z) = \int_{\mathcal{S}} \exp(is'z)\, d\mu(s) = \int_{\mathcal{T}} \cos(s'z)\, d\mu(s)\,,$$

due to the symmetry of $\mu$, $\widehat{\varepsilon} = (\widehat{\varepsilon}_j, j = 1, \ldots n)$, and $\widehat{t}_i = \widehat{t}(W_i)$. Then the statistic becomes

$$\begin{aligned}
&\int_{\mathcal{S}} n^{-1}\sum_{j=1}^n\sum_{m=1}^n \widehat{\varepsilon}_j\widehat{\varepsilon}_m\widehat{t}_j\widehat{t}_m \exp(is'(X_j - X_m))\, d\mu(s) \\
&= n^{-1}\sum_{j=1}^n\sum_{m=1}^n \widehat{\varepsilon}_j\widehat{\varepsilon}_m\widehat{t}_j\widehat{t}_m \int_{\mathcal{S}} \exp(is'(X_j - X_m))\, d\mu(s) = \widehat{\varepsilon}' A \widehat{\varepsilon}\,,
\end{aligned}$$

where $A$ is a matrix with generic element $n^{-1}a\left(X_j - X_m\right)\widehat{t}_j\widehat{t}_m$. In practice the function $a(\cdot)$ is the Fourier transform of $\mu$, so we can choose the latter so that the former has an analytic expression, and computation of the matrix $A$ is fast. To achieve scale invariance, we recommend, as in Bierens (1982), to scale each component of $X$ by a measure of dispersion, such as the empirical standard deviation.

The behavior of $\sqrt{n}\mathbb{P}_n\left(\widehat{\varepsilon}\phi_s\widehat{t}\right)$ is studied in detail by Escanciano et al. (2014), who derived a uniform expansion. Hence, the properties of the ICM test based on $S_n$ can be derived from their results. However, these impose undersmoothing in kernel estimation, which ensures that the bias disappears fast enough, but makes the practical bandwidth choice tricky. In what follows, we develop two approaches to avoid undersmoothing, and in particular we allow for an optimal nonparametric bandwidth. This is convenient because there are well-known methods for constructing such bandwidths.

## 3.2 Bias Corrected Estimation

Newey et al. (2004) obtained a small bias property for density weighted average semiparametric estimators. Here we instead rely on an idea developed in the boosting literature for kernel regressions (Di Marzio and Taylor, 2008; Park et al., 2009). Xia et al. (2004) used a similar bias correction in a specification test for a single-index model.

Let us briefly describe the $L^2$ boosting principle. Denote by $\widehat{m}(w;Y,h)$ the kernel estimator $\overline{Y}(w)/\widehat{f}(w)$ with a bandwidth $h$. The method is iterative and works as follows. Initialize $\widehat{m}_0(\cdot)$ as $\widehat{m}(\cdot;Y,h)$. For $l=1,\ldots,L$, (i) compute the residuals $u_{i,l-1}=Y_i-\widehat{m}_{l-1}(W_i;Y,h)$, (ii) apply our estimator to the vector $u_{l-1}$ of residuals to obtain $\widehat{m}(\cdot;u_{l-1},h)$, and (iii) update the estimator as $\widehat{m}_l(\cdot;Y,h)=\widehat{m}_{l-1}(\cdot;Y,h)+\widehat{m}(\cdot;u_{l-1},h)$. At each step, the method extracts remaining information from the residuals to which the same smoother is applied.

In our case, we will restrict to a one-step boosting, that is $L=1$, which is sufficient for our purpose and thus reduces to a bias correction. For our kernel smoother $\widehat{m}(\cdot)$, consider

$$\overline{\widehat{m}}(w)=(nh^p)^{-1}\sum_{i=1}^{n}\widehat{m}(W_i)\widehat{t}_iK_h(W_i-w)\,, \tag{6}$$

where we introduce the trimming $\widehat{t}$ in the above estimator to control for the random denominator of $\widehat{m}$. Then

$$\widehat{B}(w)=\frac{\overline{\widehat{m}}(w)}{\widehat{f}(w)}-\widehat{m}(w)$$

is an estimator of the bias of $\widehat{m}(w)$ and the bias-corrected estimator writes

$$\widetilde{m}(w)=\widehat{m}(w)-\widehat{B}(w)=2\widehat{m}(w)-\frac{\overline{\widehat{m}}(w)}{\widehat{f}(w)}\,. \tag{7}$$

We thus consider $\widetilde{\varepsilon}_i=Y_i-\widetilde{m}(W_i)$ and the bias-corrected ICM statistic

$$S_n^{BC}=n\int_{\mathcal{S}}|\mathbb{P}_n\left(\widetilde{\varepsilon}\phi_s\widehat{t}\right)|^2\,d\mu(s)\,.$$

The form (7) of our bias corrected estimator is similar to the one discussed in Newey et al. (2004, Section 3) who considered density-weighted nonparametric estimators. To apply their technique in our setup would necessitate to apply their correction to both the numerator and the denominator of the regression estimator. We feel more natural and practically more convenient to correct for the bias of the regression estimator as described above.

### 3.3 Locally Robust Process

Locally robust semiparametric estimation has been considered by several authors, see Newey (1990) for an early example. Chernozhukov et al. (2022) consider a general GMM estimation problem, where the moments depend on a first-step nonparametric estimator. Locally robust semiparametric GMM estimators are built from moment conditions that have zero derivatives with respect to the first-step estimator so that the latter does not affect the asymptotic variance of the parameters of interest. They have smaller bias and improved small sample

properties than standard GMM estimators.

In our case, we have a continuum of moment conditions. Our goal is to modify these moment conditions so that (i) we can still test for our null hypothesis of interest using these modified moments, and (ii) estimation becomes "adaptive" with respect to the nonparametric regression. Hence, we consider the moments

$$\mathbb{E}\left[\left(Y-\mathbb{E}\left(Y|W\right)\right)\left(\varphi(s'X)-\mathbb{E}\left(\varphi(s'X)|W\right)\right)\right]=0 \quad \forall s\in\mathcal{S}\,.$$

Because for any $s$, $\mathbb{E}\left[\left(Y-\mathbb{E}\left(Y|W\right)\right)\mathbb{E}\left(\varphi(s'X)|W\right)\right]=0$, the above moment conditions are equivalent to (4). We show below that for the empirical equivalent of these moments, estimation of nonparametric components has no first order effect.

Let $\iota_s(W)=\mathbb{E}\left(\phi_s(X)|W\right)$ and its estimator $\widehat{\iota}_s(w)=\overline{\phi_s(X)}(w)/\widehat{f}(w)$. Our locally robust approach thus delivers the statistic

$$S_n^{LR}=n\int_{\mathcal{S}}\left|\mathbb{P}_n\left(\widehat{\varepsilon}(\phi_s-\widehat{\iota}_s)\widehat{t}\right)\right|^2 d\mu(s)\,.$$

Newey et al. (2004) note that their bias correction based on twicing kernels is equivalent to a locally robust density weighted average. By contrast, our bias correction detailed in the last section does not yield the same statistic as the one based on the locally robust process.

In practice, there is no need to estimate $\iota_s(W)$ for each $s\in\mathcal{S}$ and to integrate, as

$$\begin{aligned}S_n^{LR}=&S_n-2n^{-1}\sum_{j=1}^n\sum_{m=1}^n\widehat{\varepsilon}_j\widehat{\varepsilon}_m\widehat{t}_j\widehat{t}_m\int_{\mathcal{S}}\overline{\phi}_s(X_j)\widehat{\iota}_s(X_m)\,d\mu(s)\\&+n^{-1}\sum_{j=1}^n\sum_{m=1}^n\widehat{\varepsilon}_j\widehat{\varepsilon}_m\widehat{t}_j\widehat{t}_m\int_{\mathcal{S}}\widehat{\iota}_s(X_j)\widehat{\iota}_s(X_m)\,d\mu(s)\,.\end{aligned}$$

Each term can be easily computed as a quadratic form: if $\mathbb{K}$ is the $n\times n$ matrix of generic element $K_h(W_i-W_j)/[nh^p\widehat{f}(W_j)]$, then $S_n^{LR}=\widehat{\varepsilon}'A\widehat{\varepsilon}-2\widehat{\varepsilon}'A\mathbb{K}\widehat{\varepsilon}+\widehat{\varepsilon}'\mathbb{K}'A\mathbb{K}\widehat{\varepsilon}$, so the locally robust version of the statistic is practically straightforward to compute.

# 4 Asymptotic Analysis

We here focus on the asymptotic expansion of the empirical processes on which our test statistics $S_n^{BC}$ and $S_n^{LR}$ are based. We do not formally consider the empirical process entering $S_n$: its properties would be similar but would necessitate to assume some undersmoothing.

We first introduce some definitions. Define the differential operator

$$\partial^l g(w) = \frac{\partial^{|l|}}{\partial^{l_1} w_1 \dots \partial^{l_p} w_p} g(w) \qquad l = (l_1, .., l_p)', \ |l| = l_1 + .. + l_p \, .$$

**Definition 4.1.** (a) $\mathcal{G}_\lambda(\mathcal{A}) = \left\{ g : \mathcal{A} \mapsto \mathbb{R} : \quad \sup_{a \in \mathcal{A}} |\partial^l g(a)| < M \text{ for all } |l| \leq \lambda \right\}$. (b) $\mathcal{K}^r_\lambda$ is the class of product univariate kernels $k(\cdot)$ such that $k(\cdot)$ is of order $r$, $\lambda$ times continuously differentiable with uniformly bounded derivatives, symmetric about zero, and with bounded support.

**Assumption A.** *(i)* $(Y_i, W_i, X_i), i = 1, \dots n$, *is a random sample from* $(Y, W, X)$. $\mathcal{Y} \subset \mathbb{R}$, $\mathcal{W} \subset \mathbb{R}^p$, *and* $\mathcal{X} \subset \mathbb{R}^d$, *the supports of* $Y$, $W$, *and* $X$, *are bounded. (ii)* $\mathcal{S}$ *is a bounded compact subset of* $\mathbb{R}^d$ *containing a neighborhood of the origin. (iii)* $\varphi(\cdot)$ *is an analytic non polynomial function with* $\partial^l \varphi(0) \neq 0$ *for all* $l \in \mathbb{N}$.

**Assumption B.** *(i)* $f(\cdot) \in \mathcal{G}_r(\mathbb{R}^p)$ *and* $m(\cdot) \in \mathcal{G}_r(\mathcal{W})$, *with* $r \geq \lceil (p+1)/2 \rceil$.[3] *(ii)* $\iota_s(\cdot) \in \mathcal{G}_r(\mathcal{W})$ *uniformly in* $s \in \mathcal{S}$, *with* $r \geq \lceil (p+1)/2 \rceil$. *(iii)* $K(\cdot) \in \mathcal{K}^r_\lambda$ *with* $\lambda \geq \lceil (p+1)/2 \rceil$.

Define the uniform convergence rate of the kernel density estimator as

$$d_n = \sqrt{\frac{\log n}{n h_n^p}} + h_n^r \, .$$

**Assumption C.** *(i)* $h_n \tau_n^{-1} = o(1)$. *(ii)* $d_n \tau_n^{-3} = o(n^{-1/4})$. *(iii)* $d_n h^{-|l|} \tau_n^{-(2+|l|)} = o(1)$ *for* $|l| = \lceil (p+1)/2 \rceil$.

**Assumption D.** *(i)* $p_n = \Pr\left(f(W) \leq 3\tau_n/2\right)$ *is such that* $p_n = o(n^{-1/2})$ , $p_n h_n^{-p} \tau_n^{-2} = o(n^{-1/4})$, *and* $p_n h_n^{-(p+|l|)} \tau_n^{-(1+|l|)} = o(1)$ *for* $|l| = \lceil (p+1)/2 \rceil$.

*(ii) There exists* $N$ *such that for all* $n \geq N$, $\mathcal{W}_n = \{w : f(w) \geq \tau_n/2\}$ *is convex.*

Assumption A ensures that the encompassing hypothesis can be written as the continuum of moment conditions (4). In particular, Bierens (2017, Theorem 2.2) builds on previous results by Bierens (1982) and Stinchcombe and White (1998), and shows that A-(iii) is sufficient. This allows for different functions $\varphi(\cdot)$ as previously detailed. Assumption B imposes conditions that are commonly found in the literature on nonparametric estimation. In particular, they impose that the density of $W$ is differentiable over $\mathcal{W}$ and its derivatives go smoothly to zero as we approach to the boundaries of the support.

Assumption C sets the main conditions on the bandwidths. Together with Assumption B, it implies that the kernel estimators asymptotically belong to a class of sufficiently smooth functions with limited entropy/complexity. This is needed for the asymptotic stochastic

[3] $\lceil x \rceil$ denotes the smallest integer above $x$.

equicontinuity of the empirical processes at the basis of our test statistics. Assumptions C and D impose conditions on the bandwidths in connection with the trimming parameter. We need the effect of trimming to vanish quickly enough to avoid large biases. We study the practical influence of trimming in our simulations.

Abstracting from the appearance of the trimming, Condition (ii) in Assumption C ensures that the kernel estimators are $n^{-1/4}$-consistent. It requires in particular that $nh_n^{4r} = o(1)$. Without the bias correction or locally robust approach employed in the construction of our statistics, this condition would become $nh_n^{2r} = o(1)$ to ensure that the bias of the nonparametric estimator is negligible compared to the variability of the empirical process, see Delgado and Manteiga (2001) or Escanciano et al. (2014). This means that our empirical processes have a small bias property (in their stochastic expansion). Our conditions allow the bandwidth to be optimal for nonparametric estimation purposes and avoid the need for undersmoothing. The main practical advantage is that one can use for instance a simple rule-of-thumb.[4] We could extend our theoretical results to stochastic bandwidths to allow for the use of data-driven methods, as was done in other contexts (Andrews, 1995; Lavergne, 2001; Mammen, 1992). The details would be more involved here because of the stochastic trimming, so we do not pursue this issue further.

The following result establishes the influence function representation of the empirical processes used in our statistics $S_n^{BC}$ and $S_n^{LR}$.

**Proposition 4.2.** *Under Assumptions A–D,* $\sqrt{n}\mathbb{P}_n(\widetilde{\varepsilon}\phi_s\widehat{t}) = \sqrt{n}\mathbb{P}_n(\varepsilon(\phi_s - \iota_s)) + o_P(1)$ *and* $\sqrt{n}\mathbb{P}_n(\widehat{\varepsilon}(\phi_s - \widehat{\iota}_s)\widehat{t}) = \sqrt{n}\mathbb{P}_n(\varepsilon(\phi_s - \iota_s)) + o_P(1)$ *uniformly in* $s \in \mathcal{S}$.

From the above result, under $H_0$ the limiting distribution of both empirical processes is the limiting one of $\sqrt{n}\mathbb{P}_n(\varepsilon(\phi_s - \iota_s))$. Let us denote it with $\mathbb{G}_s$. Such a distribution is a zero-mean tight Gaussian process valued in $\ell^\infty(\mathcal{S})$, the space of uniformly bounded functionals over $\mathcal{S}$, and characterized by the collection of covariances $\{\mathbb{E}(Y - m(W))^2(\phi_s(X) - \iota_s(X))(\phi_t(X) - \iota_t(W)) : s, t \in \mathcal{S}\}$. Our asymptotic expansions directly yield

$$S_n^{BC} \quad \text{or} \quad S_n^{LR} \xrightarrow{d} \int |\mathbb{G}_s|^2 \, d\mu(s) \, .$$

However, the result is not useful in practice since the covariance function of $\mathbb{G}_s$ depends on the unknown data generating process. In what follows, we develop a bootstrap procedure for obtaining critical and p-values.

[4] We note however that the optimal nonparametric bandwidth is likely not optimal for estimation of $\mathbb{E}\left[(Y - \mathbb{E}(Y|W))\,\varphi(s'X)\right]$.

# 5 Bootstrap Tests

We use a wild bootstrap procedure that imposes the null hypothesis $H_0$ when resampling the observations. The bootstrap data generating process (DGP) writes as

$$Y_i^* = \widehat{m}(W_i) + \xi_i \widehat{\varepsilon}_i \qquad \widehat{\varepsilon}_i = Y_i - \widehat{m}(W_i)\,,$$

where $\{\xi_i, i = 1, \dots n\}$ is a sequence of independent bootstrap weights with $\mathbb{E}\,\xi = 0$ and $\mathbb{E}\xi^2 = 1$. From each bootstrap sample $\{(Y_i^*, W_i, X_i) : i = 1, .., n\}$, we proceed as above to obtain $\widehat{\varepsilon}^* = Y_i^* - \widehat{m}^*(W_i)$ and $\widetilde{\varepsilon}_i^* = Y_i^* - \widetilde{m}^*(W_i)$, where

$$\widetilde{m}^*(w) = \widehat{m}^*(w) - \widehat{B}^*(w) = \widehat{m}^*(w) - \left( \frac{\overline{\widehat{m}^*}(w)}{\widehat{f}(w)} - \widehat{m}^*(w) \right), \qquad \widehat{m}^* = \overline{Y^*}/\widehat{f}\,.$$

The kernel smoothers $\overline{Y^*}$ and $\overline{\widehat{m}^*}$ are constructed in the same way as in Equation (6) to control for the random denominators in $Y^*$ and $\widehat{m}^*$. The bias correction could be the one used in the original statistic, as done by Xia et al. (2004), however we noted in simulations that recomputing the bias correction with bootstrap data yields a better behavior. We thus consider the bootstrap statistics

$$n \int_{\mathcal{S}} |\mathbb{P}_n \left( \widetilde{\varepsilon}^* \phi_s \widehat{t} \right)|^2 d\mu(s) \quad \text{and} \quad n \int_{\mathcal{S}} |\mathbb{P}_n \left( \widehat{\varepsilon}^* (\phi_s - \widehat{\iota}_s) \widehat{t} \right)|^2 d\mu(s)\,. \tag{8}$$

In practice, one can compute many bootstrap statistics and obtain their $1-\alpha$ quantiles, denoted as $q_{1-\alpha}^{BC}$ and $q_{1-\alpha}^{LR}$. The bootstrap test rejects $H_0$ whenever $S_n^j > q_{1-\alpha}^j$, for $j = BC$ or $LR$.

**Proposition 5.1.** *Let $\{\xi_i, i = 1, \dots n\}$ be an i.i.d. sequence with $\mathbb{E}\,\xi = 0$, $\mathbb{E}\xi^2 = 1$, and $\mathbb{E}|\xi|^3 < \infty$ independent of the sample. Under Assumptions A-D,*

*(a) $\sqrt{n}\mathbb{P}_n(\widetilde{\varepsilon}^* \phi_s \widehat{t}) = \sqrt{n}\mathbb{P}_n(\xi \varepsilon(\phi_s - \iota_s)) + o_p(1)$ and $\sqrt{n}\mathbb{P}_n(\widehat{\varepsilon}^*(\phi_s - \widehat{\iota}_s)\widehat{t}) = \sqrt{n}\mathbb{P}_n(\xi\varepsilon(\phi_s - \iota_s)) + o_p(1)$ uniformly in $s \in \mathcal{S}$.*[5]*

*(b) Under $H_0$, $\Pr\left[S_n^j > q_{1-\alpha}^j\right] \to \alpha$, $j = BC$ or $LR$.*

*(c) Under $H_1$, $\Pr\left[S_n^j > q_{1-\alpha}^j\right] \to 1$, $j = BC$ or $LR$.*

# 6 Small Sample Behavior

We used a DGP in line with our example of Section 2.2. Specifically,

$$Y = m(W) + h(W)X + \gamma\delta(X) + \eta\,, \qquad \mathbb{E}\,(\eta | W, X) = 0\,,$$

[5] Here the probability space is the joint probability on random bootstrap weights and sample data.

where $\eta \sim N(0,1)$ is independent of $(W,X)$, $W \sim N(0,\sigma^2 = (1/2)^2)$,

$$X|W \sim N(0, \sigma^2 = (1/4 + \mathbb{I}(W < 0)1/2)^2)\,,$$
$$m(w) = w^3 - 2w\,, \qquad h(w) = w^2 - (1/2)^2\,, \qquad \delta(x) = x^4 - 3x^2\,.$$

When $\gamma = 0$, $\mathcal{M}_W$ encompasses $\mathcal{M}_X$, while it does not when $\gamma \neq 0$.

For the implementation of the test, the weighting function was $a(x) = \text{sinc}(\pi x)$, which corresponds to a uniform $\mu$ with complex exponential function $\varphi(\cdot)$. The observations on $X$ were transformed by the logistic c.d.f. then standardized before being passed as arguments to $a(\cdot)$. For bootstrapping, we used the two-point distribution defined through $\Pr(\xi = \frac{3-\sqrt{5}}{2}) = \frac{5+\sqrt{5}}{10}$ and $\Pr(\xi = \frac{3+\sqrt{5}}{2}) = \frac{5-\sqrt{5}}{10}$. We chose this simple distribution with third central moment equal to one in the hope to better approximate the distribution of the statistic, as is the case in simpler setups (Mammen, 1992). For nonparametric estimation, we employed Gaussian kernels of order 2 and the bandwidth rule $h = C\widehat{\sigma}_W n^{-1/5}$, where $\widehat{\sigma}_W$ is the estimated standard deviation of $W$. To check the performance of our tests under different bandwidth choices, we let the constant $C$ vary. As there is no clear way to choose the amount of trimming, we trimmed the 2% more extreme observations in a first step, and we subsequently investigated the influence of trimming. We ran 10000 simulations with $n = 200$ and 400. To speed up computations, we used the warp-speed method proposed by Davidson and MacKinnon (2007) and studied by Giacomini, Politis, and White (2013). Specifically, we drew one bootstrap sample for each simulated data, and we used the whole set of bootstrap statistics to compute the bootstrap p-values associated with each original statistic.

We compared our tests based on $S_n^{BC}$ and $S_n^{LR}$ to the test based on the uncorrected ICM statistic $S_n$. We let the bandwidth constant $C$ vary to analyze its effect on the tests' size. We report actual rejection probabilities for 5% and 10% nominal sizes in Table 1. We also report in Figures 1 and 2 errors in rejection probability (ERP), the difference between the empirical rejection proportion and the nominal size under the null hypothesis $H_0$. A perfect test would exhibit an ERP of zero for any nominal size. This gives us a visual way to evaluate whether the null distribution of the test statistic is well approximated by its bootstrap approximation. For $C = .5$ or $C = 1$, though the ERP curves lie mostly slightly above zero, the three tests have good size control for levels up to 10%, and size controls improve when increasing the sample size. When $C$ increases to 1.5 then 2, the size control deteriorates for the uncorrected ICM test, while our two tests keep size well under control as expected.

We also report in Figure 3 the power of each of our tests when $\gamma$ varies for $C = 1$. Our two tests basically have the same power, and power increases with sample size. We complemented our study with an investigation of the trimming influence. The most striking results are reported as ERP curves with no trimming. As can be seen in Figures 4 and 5,

| C | 0.5 | | 1 | | 1.5 | | 2 | |
|---|---|---|---|---|---|---|---|---|
| $n = 200$ | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% |
| Bias Corr. | 5.79 | 11.81 | 5.31 | 10.58 | 4.71 | 10.03 | 5.02 | 10.51 |
| Locally Rob. | 5.74 | 11.75 | 5.47 | 10.73 | 4.78 | 10.17 | 5.16 | 10.57 |
| ICM | 6.03 | 11.92 | 5.75 | 10.85 | 6.61 | 12.82 | 9.99 | 18.49 |
| $n = 400$ | | | | | | | | |
| Bias Corr. | 5.61 | 10.13 | 5.45 | 9.67 | 5.00 | 9.80 | 5.14 | 10.58 |
| Locally Rob. | 5.59 | 10.17 | 5.21 | 9.65 | 5.04 | 9.41 | 5.26 | 10.46 |
| ICM | 5.65 | 10.39 | 5.54 | 10.59 | 6.62 | 12.99 | 11.30 | 19.58 |

Table 1: Empirical rejection percentages under $H_0$.

this change greatly affects the behavior of the locally robust test for large bandwidths. This phenomenon is even more pronounced with a larger sample size. We suspect this could be due to the first-step estimation of $\iota_s(\cdot)$: this function has varying frequency depending on $s$ and may not be very accurately estimated at every frequency at the boundary of the support of (the transformed) $X$. By contrast, the absence of trimming does not affect significantly the size of the bias-corrected test.

# 7 Concluding Remarks

Another test statistic could be considered for testing the encompassing hypothesis. Indeed, it can be easily checked that

$$\mathbb{E}\left[Y\left(\varphi(s'X) - \mathbb{E}\left(\varphi(s'X)|W\right)\right)\right] = 0 \quad \forall s \in \mathcal{S}$$

is equivalent to our null hypothesis. We ran simulations using a test statistic based on the related empirical process. Unfortunately, the test exhibited much lower power than its competitors.

We have studied two general approaches to obtain a small bias property for our encompassing test statistic. These approaches could potentially be used in other estimation and testing problems involving a first-step nonparametric estimation. Our limited simulation experiment seems to indicate that while the boosting bias correction is relatively robust to choices related to bandwidth and trimming, the locally robust approach is quite sensitive to trimming or its absence. Further study is needed to determine why and in which conditions this is expected.

# 8 Proofs

For any real or complex valued function $g(\cdot)$, we denote with $||g||_\infty$ the supremum norm taken over the support of its argument and $Pg = \int g(z)\, dP(z)$. We define the population counterpart of $\widehat{t}$ as $t(w) = \mathbb{I}(f(w) \geq \tau_n)$. By convention, $(\overline{Y}/\widehat{f})(w)$ will be set to 0 whenever $\widehat{f}(w) = 0$. The same notation will hold for $\overline{\widehat{m}}/\widehat{f}$, $\overline{\phi}_s/\widehat{f}$, $\overline{Y^*}/\widehat{f}$, and $\overline{\widehat{m}^*}/\widehat{f}$. $C$ denotes a generic constant that may vary from line to line.

## 8.1 Proof of Proposition 1

(a) We study the bias corrected empirical process

$$\begin{aligned}\sqrt{n}\mathbb{P}_n(\widetilde{\varepsilon}\phi_s\widehat{t}) &= \sqrt{n}\mathbb{P}_n\left(\left(Y - \widehat{m} + \widehat{B}\right)\phi_s\widehat{t}\right) \\ &= \sqrt{n}\mathbb{P}_n(\varepsilon\phi_s\widehat{t}) + \sqrt{n}\mathbb{P}_n\left((m-\widehat{m})\phi_s\widehat{t}\right) + \sqrt{n}\mathbb{P}_n(\widehat{B}\phi_s\widehat{t})\,. \end{aligned} \tag{9}$$

Since $|\varepsilon|$ is bounded and $|\phi_s(\cdot)|$ is uniformly bounded,

$$\sup_s \left|\sqrt{n}\mathbb{P}_n(\varepsilon\phi_s(\widehat{t}-1))\right| \leq C\sqrt{n}\mathbb{P}_n|\widehat{t}-1| = o_p(1)$$

by Lemma 8.1-(i). Hence, $\sqrt{n}\mathbb{P}_n(\varepsilon\phi_s\widehat{t}) = \sqrt{n}\mathbb{P}_n(\varepsilon\phi_s) + o_p^{\mathcal{S}}(1)$, where $o_p^{\mathcal{S}}(1)$ denotes a uniform in $s \in \mathcal{S}$ $o_p(1)$. For the second term, as $(\widehat{t}-t) = \widehat{t}^2 - t^2 = (\widehat{t}+t)(\widehat{t}-t)$, and $\|(m-\widehat{m})(\widehat{t}+t)\|_\infty = o_p(n^{-1/4})$ by Lemma 8.1-(ii),

$$\sup_s \left|\sqrt{n}\mathbb{P}_n(m-\widehat{m})\phi_s(\widehat{t}-t)\right| \leq C\|(m-\widehat{m})(\widehat{t}+t)\|_\infty\sqrt{n}\mathbb{P}_n|\widehat{t}-t| = o_p(1)\,.$$

We now show that the third term is such that

$$\sqrt{n}\mathbb{P}_n(\widehat{B}\phi_s\widehat{t}) = -\sqrt{n}\mathbb{P}_n(\varepsilon\iota_s) - \sqrt{n}\mathbb{P}_n\left((m-\widehat{m})\iota_s t\right) + o_p^{\mathcal{S}}(1)\,.$$

First, use $\|\widehat{B}(\widehat{t}+t)\|_\infty = o_p(n^{-1/4})$, see Lemma 8.1-(iii), and similar arguments as above to show that $\sqrt{n}\mathbb{P}_n(\widehat{B}\phi_s\widehat{t}) = \sqrt{n}\mathbb{P}_n(\widehat{B}\phi_s t) + o_p^{\mathcal{S}}(1)$. Lemma 8.3-(i)-(ii) ensures in addition that $\widehat{B} \in \mathcal{G}_l(\mathcal{W}_n)$ with probability approaching one, where $l = \lceil (p+1)/2 \rceil$ and $\mathcal{W}_n := \{w : f(w) \geq \tau_n/2\}$. So, Lemma 8.4-(i) yields

$$\sqrt{n}\mathbb{P}_n(\widehat{B}t\phi_s) = \sqrt{n}P(\widehat{B}t\phi_s) + o_p^{\mathcal{S}}(1)\,, \tag{10}$$

$$P(\widehat{B}t\phi_s) = \int \widehat{B}(w)t(w)\iota_s(w)f(w)\,dw = P(\widehat{B}t\iota_s)\,. \tag{11}$$

Since $\|[\widehat{B} - (\widehat{\overline{m}} - \overline{Y})/f]t\|_\infty = o_p(n^{-1/2})$ from Lemma 8.1-(iv) and $\iota_s$ is uniformly bounded,

$$\begin{aligned}\sqrt{n}P(\widehat{B}t\iota_s) &= \sqrt{n}\int(\widehat{\overline{m}}(w) - \overline{Y}(w))t(w)\iota_s(w)\,dw + o_p^{\mathcal{S}}(1)\\ &= -\sqrt{n}\mathbb{P}_n\left[(Y - \widehat{t}\widehat{m})\int K(u)(t\iota_s)(W+uh)\,du\right] + o_p^{\mathcal{S}}(1)\\ &= -\sqrt{n}\mathbb{P}_n\left[\left(\varepsilon - (\widehat{m}-m)\widehat{t} - m(\widehat{t}-1)\right)\int K(u)(t\iota_s)(W+uh)\,du\right] + o_p^{\mathcal{S}}(1)\,,\end{aligned}$$

where the second equality follows from a change of variable. Let us deal with each term separately. Since $\varepsilon$ is bounded and $|t(w+uh)-1| = \mathbb{I}\{f(w+uh) < \tau_n\}$,

$$\left|\sqrt{n}\mathbb{P}_n\varepsilon\int K(u)\iota_s(W+uh)[t(W+uh)-1]\,du\right| \le C\int|K(u)|\sqrt{n}\mathbb{P}_n\mathbb{I}\left\{f(W+uh)<\tau_n\right\}\,du\,.$$

By a mean-value expansion, for all $(u,w) \in \mathrm{Supp}(K)\times\mathbb{R}^p$ we have $f(w+hu) - f(w) = \partial^T f(\widetilde{w})uh$, for some $\widetilde{w}$. By Assumptions A-B, $\left|\partial^T f(w)uh\right| \le Ch$ for all $(u,w)\in\mathrm{Supp}(K)\times\mathbb{R}^p$. Since $h\tau_n^{-1} = o(1)$, for $n$ large enough $\mathbb{I}\{f(w+uh)<\tau_n\} \le \mathbb{I}\{f(w)\le 3\tau_n/2\}$. Now use $\mathbb{P}_n\mathbb{I}\Big\{f(W)\le 3\tau_n/2\Big\} = O_P(p_n) = o_P(n^{-1/2})$, from Assumption D(i), to obtain

$$\sqrt{n}\mathbb{P}_n\left[\varepsilon\int K(u)(t\iota_s)(W+uh)\,du\right] = \sqrt{n}\mathbb{P}_n\left[\varepsilon\int K(u)\iota_s(W+uh)\,du\right] + o_p^{\mathcal{S}}(1)\,.$$

The same reasoning allows the same replacement in the second and third term of the decomposition, using that $\|(\widehat{m}-m)\widehat{t}\|_\infty$ is $o_p(1)$ by Lemma 8.1-(ii) and that $m(\widehat{t}-1)$ is bounded.

Then

$$\begin{aligned}\sqrt{n}P(\widehat{B}t\iota_s) = &-\sqrt{n}\mathbb{P}_n\left[\varepsilon\int K(u)\iota_s(W+uh)\,du\right] + \sqrt{n}\mathbb{P}_n\left[(\widehat{m}-m)\widehat{t}\int K(u)\iota_s(W+uh)\,du\right]\\ &+\sqrt{n}\mathbb{P}_n\left[m(\widehat{t}-1)\int K(u)\iota_s(W+uh)\,du\right] + o_p^{\mathcal{S}}(1)\,.\end{aligned} \tag{12}$$

Lemma 8.4(ii) yields

$$\sqrt{n}\mathbb{P}_n\left(\varepsilon\int K(u)\iota_s(W+uh)\,du\right) = \sqrt{n}\mathbb{P}_n(\varepsilon\iota_s) + o_p^{\mathcal{S}}(1)\,,$$

as $P\left(\varepsilon\int K(u)\iota_s(W+uh)\,du\right) = P\left(\varepsilon\iota_s\right) = 0$. A Taylor expansion of order $r$ guarantees that $\int K(u)\iota_s(w+uh)du = \iota_s(w) + O(h^r)$ uniformly in $\mathcal{S}\times\mathcal{W}$ as $\iota_s(\cdot)$ has uniformly bounded

derivatives of order $r$. Since $\|(\widehat{m}-m)\widehat{t}\|_\infty = o_p(n^{-1/4})$ and $nh^{4r} = o(1)$,

$$\sqrt{n}\mathbb{P}_n\left((\widehat{m}-m)\widehat{t}\int K(u)\iota_s(W+uh)\,du\right) = \sqrt{n}\mathbb{P}_n((\widehat{m}-m)\widehat{t}\iota_s) + o_p^{\mathcal{S}}(1)\,.$$

Now use similar arguments as before to replace $\widehat{t}$ by $t$. The last term in (12) is negligible. Indeed, the argument of $\mathbb{P}_n$ is uniformly bounded, and $\sqrt{n}\mathbb{P}_n|t-1| = o_P(1)$ from Lemma 8.1-(i). Hence,

$$\sqrt{n}P(\widehat{B}t\iota_s) = -\sqrt{n}\mathbb{P}_n(\varepsilon\iota_s) + \sqrt{n}\mathbb{P}_n((\widehat{m}-m)t\iota_s) + o_p^{\mathcal{S}}(1)\,.$$

Gathering results,

$$\sqrt{n}\mathbb{P}_n(\widetilde{\varepsilon}\phi_s\widehat{t}) = \sqrt{n}\mathbb{P}_n(\varepsilon(\phi_s-\iota_s)) + \sqrt{n}\mathbb{P}_n\left((m-\widehat{m})(\phi_s-\iota_s)t\right) + o_p^{\mathcal{S}}(1)\,.$$

From arguments similar to those used for (10), since $\|(m-\widehat{m})t\|_\infty = o_p(n^{-1/4})$ and $\widehat{m} \in \mathcal{G}_l(\mathcal{W}_n)$ with probability approaching one,

$$\sqrt{n}\mathbb{P}_n\left((m-\widehat{m})(\phi_s-\iota_s)t\right) = \sqrt{n}P\left((m-\widehat{m})(\phi_s-\iota_s)t\right) + o_p^{\mathcal{S}}(1)\,.$$

But $\sqrt{n}P\left((m-\widehat{m})(\phi_s-\iota_s)t\right) = 0$ as $\mathbb{E}\left(\phi_s(X)|W\right) = \iota_s(W)$. Finally, we obtain $\sqrt{n}\mathbb{P}_n(\widetilde{\varepsilon}\phi_s\widehat{t}) = \sqrt{n}\mathbb{P}_n(\varepsilon(\phi_s-\iota_s)) + o_p^{\mathcal{S}}(1)$ as expected.

(b) We now study the locally robust empirical process

$$\begin{aligned}\sqrt{n}\mathbb{P}_n(\widehat{\varepsilon}(\phi_s-\widehat{\iota}_s)\widehat{t}) &= \sqrt{n}\mathbb{P}_n\left((Y-\widehat{m})\left(\phi_s-\widehat{\iota}_s\right)\widehat{t}\right)\\ &= \sqrt{n}\mathbb{P}_n(\varepsilon(\phi_s-\widehat{\iota}_s)\widehat{t}) + \sqrt{n}\mathbb{P}_n\left((m-\widehat{m})(\phi_s-\widehat{\iota}_s)\widehat{t}\right)\,. \qquad (13)\end{aligned}$$

A similar reasoning as in Part (a) allows replacing $\widehat{t}$ by $t$ and yields

$$\sqrt{n}\mathbb{P}_n(\widehat{\varepsilon}(\phi_s-\widehat{\iota}_s)\widehat{t}) = \sqrt{n}\mathbb{P}_n(\varepsilon(\phi_s-\widehat{\iota}_s)t) + \sqrt{n}\mathbb{P}_n\left((m-\widehat{m})(\phi_s-\widehat{\iota}_s)t\right) + o_p^{\mathcal{S}}(1)\,,$$

based on the boundedness of $\varepsilon$, $\sqrt{n}\mathbb{P}_n|\widehat{t}-t| = o_p(1)$, and the boundedness in probability of $\sup_s\|(\phi_s-\widehat{\iota}_s)t\|_\infty$, $\sup_s\|(\phi_s-\widehat{\iota}_s)\widehat{t}\|_\infty$, $\|(\widehat{m}-m)t\|_\infty$, and $\|(\widehat{m}-m)\widehat{t}\|_\infty$, see Lemma 8.1. From Lemma 8.1, $\sup_s\|(\widehat{\iota}_s-\iota_s)t\|_\infty = o_p(n^{-1/4})$ and $\|(m-\widehat{m})t\|_\infty = o_P(n^{-1/4})$, hence

$$\sqrt{n}\mathbb{P}_n\left((m-\widehat{m})(\phi_s-\widehat{\iota}_s)t\right) = \sqrt{n}\mathbb{P}_n\left((m-\widehat{m})(\phi_s-\iota_s)t\right) + o_p^{\mathcal{S}}(1) = o_p^{\mathcal{S}}(1)\,,$$

where the last equality is established in Part (a). Now,

$$\mathbb{P}_n\varepsilon(\phi_s-\widehat{\iota}_s)t = \mathbb{P}_n\varepsilon(\phi_s-\iota_s)t - \mathbb{P}_n\varepsilon(\widehat{\iota}_s-\iota_s)t\,,$$

and $\sup_s |\sqrt{n}\mathbb{P}_n(\varepsilon(\phi_s - \iota_s)(t-1))| \leq C\sqrt{n}\mathbb{P}_n|t-1| = o_P(1)$ by Lemma 8.1-(i). Last, $\sqrt{n}\mathbb{P}_n\varepsilon(\widehat{\iota}_s - \iota_s)t = o_p^{\mathcal{S}}(1)$ by Lemma 8.4-(i). Gathering results, $\sqrt{n}\mathbb{P}_n(\widehat{\varepsilon}(\phi_s - \widehat{\iota}_s)\widehat{t}) = \sqrt{n}\mathbb{P}_n(\varepsilon(\phi_s - \iota_s)) + o_p^{\mathcal{S}}(1)$.

## 8.2 Proof of Proposition 2

We here consider statements relative to $P^{\xi} \otimes P$, the joint probability measure of both the bootstrap weights and the sample data.

(a) We study the bootstrap version of the bias corrected empirical process

$$\begin{aligned}\sqrt{n}\mathbb{P}_n(\widetilde{\varepsilon}^*\phi_s\widehat{t}) &= \sqrt{n}\mathbb{P}_n\left(\left(Y^* - \widehat{m}^* + \widehat{B}^*\right)\phi_s\widehat{t}\right) \\ &= \sqrt{n}\mathbb{P}_n(\xi\varepsilon\phi_s\widehat{t}) + \sqrt{n}\mathbb{P}_n\left(\xi(m-\widehat{m})\phi_s\widehat{t}\right) + \sqrt{n}\mathbb{P}_n\left((\widehat{m}-\widehat{m}^*)\phi_s\widehat{t}\right) + \sqrt{n}\mathbb{P}_n(\widehat{B}^*\phi_s\widehat{t})\,.\end{aligned}$$

From here we only stress the differences with the proof of Proposition 1. Proceeding as in the latter proof and using results in Lemma 8.1,

$$\begin{aligned}\sqrt{n}\mathbb{P}_n(\xi\varepsilon\phi_s\widehat{t}) &= \sqrt{n}\mathbb{P}_n(\xi\varepsilon\phi_s t) + o_P^{\mathcal{S}}(1)\,, \\ \sqrt{n}\mathbb{P}_n\left(\xi(m-\widehat{m})\phi_s\widehat{t}\right) &= \sqrt{n}\mathbb{P}_n\left(\xi(m-\widehat{m})\phi_s t\right) + o_P^{\mathcal{S}}(1)\,, \\ \sqrt{n}\mathbb{P}_n\left((\widehat{m}-\widehat{m}^*)\phi_s\widehat{t}\right) &= \sqrt{n}\mathbb{P}_n\left((\widehat{m}-\widehat{m}^*)\phi_s t\right) + o_P^{\mathcal{S}}(1)\,, \\ \sqrt{n}\mathbb{P}_n(\widehat{B}^*\phi_s\widehat{t}) &= \sqrt{n}\mathbb{P}_n(\widehat{B}^*\phi_s t) + o_P^{\mathcal{S}}(1)\,.\end{aligned}$$

For the last term, use $\|\widehat{B}^*t\|_\infty = o_P(n^{-1/4})$, Lemma 8.3-(ii)-(iv)-(v), and Lemma 8.4-(i) to show that

$$\sqrt{n}\mathbb{P}_n(\widehat{B}^*t\phi_s) = \sqrt{n}P(\widehat{B}^*t\phi_s) + o_P^{\mathcal{S}}(1)\,. \tag{14}$$

From the law of iterated expectations, $P(\widehat{B}^*t\phi_s) = P(\widehat{B}^*t\iota_s)$. Since $\|[\widehat{B}^* - (\overline{\widehat{m}^*} - \overline{Y^*})/f]t\|_\infty = o_P(n^{-1/2})$ from Lemma 8.1-(viii),

$$\begin{aligned}\sqrt{n}P(\widehat{B}^*t\iota_s) &= \sqrt{n}\int(\overline{\widehat{m}^*}(w) - \overline{Y^*}(w))t(w)\iota_s(w)\,dw + o_P^{\mathcal{S}}(1) \\ &= -\sqrt{n}\mathbb{P}_n\left[(Y^* - \widehat{m}^*)\widehat{t}\int K(u)(t\iota_s)(W+uh)\,du\right] + o_P^{\mathcal{S}}(1)\,,\end{aligned}$$

where the second equality follows from a change of variable. Replace $(Y^* - \widehat{m}^*)\widehat{t}$ by $\xi\varepsilon t + \xi\varepsilon(\widehat{t}-t) + \xi(m-\widehat{m})\widehat{t} + (\widehat{m}-\widehat{m}^*)\widehat{t}$ and proceed as in Proposition 1 to obtain

$$\sqrt{n}P(\widehat{B}^*t\phi_s) = -\sqrt{n}\mathbb{P}_n(\xi\varepsilon\iota_s t) - \sqrt{n}\mathbb{P}_n((\widehat{m}-\widehat{m}^*)t\iota_s) - \sqrt{n}\mathbb{P}_n\xi(m-\widehat{m})t\iota_s + o_P^{\mathcal{S}}(1)\,.$$

Gathering results,

$$\begin{aligned}\sqrt{n}\mathbb{P}_n(\widehat{\varepsilon}^*\phi_s\widehat{t}) =&\sqrt{n}\mathbb{P}_n(\xi\varepsilon(\phi_s-\iota_s)t)+\sqrt{n}\mathbb{P}_n\left(\xi(m-\widehat{m})(\phi_s-\iota_s)t\right)\\&+\sqrt{n}\mathbb{P}_n((\widehat{m}-\widehat{m}^*)t(\phi_s-\iota_s))+o_P^{\mathcal{S}}(1)\,.\end{aligned}$$

A reasoning similar to Proposition 1 then yields $\sqrt{n}\mathbb{P}_n(\xi\varepsilon(\phi_s-\iota_s))t=\sqrt{n}\mathbb{P}_n(\xi\varepsilon(\phi_s-\iota_s))+o_P^{\mathcal{S}}(1)$, $\sqrt{n}\mathbb{P}_n\left(\xi(m-\widehat{m})(\phi_s-\iota_s)t\right)=o_P^{\mathcal{S}}(1)$, and $\sqrt{n}\mathbb{P}_n((\widehat{m}-\widehat{m}^*)t(\phi_s-\iota_s))=o_P^{\mathcal{S}}(1)$. Hence

$$\sqrt{n}\mathbb{P}_n\widetilde{\varepsilon}^*\widehat{t}\phi_s=\sqrt{n}\mathbb{P}_n\xi\varepsilon(\phi_s-\iota_s)+o_P^{\mathcal{S}}(1).$$

For the locally robust empirical process,

$$\begin{aligned}\sqrt{n}\mathbb{P}_n(\widehat{\varepsilon}^*(\phi_s-\widehat{\iota}_s)\widehat{t}) =&\sqrt{n}\mathbb{P}_n(\xi\varepsilon(\phi_s-\widehat{\iota}_s)\widehat{t})\\&+\sqrt{n}\mathbb{P}_n\left(\xi(m-\widehat{m})(\phi_s-\widehat{\iota}_s)\widehat{t}\right)+\sqrt{n}\mathbb{P}_n\left((\widehat{m}-\widehat{m}^*)(\phi_s-\widehat{\iota}_s)\widehat{t}\right)\,.\end{aligned}$$

The proof proceeds along similar lines to show that the first term equals $\sqrt{n}\mathbb{P}_n(\xi\varepsilon(\phi_s-\iota_s))+o_P^{\mathcal{S}}(1)$ and the other terms are both $o_P^{\mathcal{S}}(1)$.

(b) Since the class $\{(y,w,x)\mapsto(y-m(w))(\phi_s(x)-\iota_s(w)):s\in\mathcal{S}\}$ is Donsker, $\sqrt{n}\mathbb{P}_n\varepsilon(\phi_s-\iota_s)$ converges weakly to a tight zero-mean Gaussian process $\mathbb{G}_s$ under $H_0$, see the main text. By the continuity of the Cramer-Von Mises functional and Proposition 1, $S_n^{BC}$ and $S_n^{LR}$ weakly converge to $\int|\mathbb{G}_s|^2\mu(ds)$. From Bentkus, Götze, and Zitikis (1993), this distribution is continuous, so pointwise convergence implies uniform convergence.

Using van der Vaart and Wellner (2000, Theorem 2.9.6), we get weak convergence of $\sqrt{n}\mathbb{P}_n\xi\varepsilon(\phi_s-\iota_s)$ in probability conditionally upon the initial sample to a tight zero-mean Gaussian process $\mathbb{G}'_s$ with the same covariance function as $\mathbb{G}_s$. From (a), the bootstrap statistics (8) weakly converge to $\int|\mathbb{G}'_s|^2\mu(ds)$ in probability conditionally upon the initial sample. The desired result then follows.

(c) From (a) and a Glivenko-Cantelli property of the class $\{(y,w,x)\mapsto(y-m(w))(\phi_s(x)-\iota_s(w)):s\in\mathcal{S}\}$, $\mathbb{P}_n\widehat{\varepsilon}\widehat{t}\phi_s=P\varepsilon(\phi_s-\iota_s)+o_P^{\mathcal{S}}(1)$ and similarly for $\mathbb{P}_n\widehat{\varepsilon}(\phi_s-\widehat{\iota})\widehat{t}$. Hence, $S_n^{LR}/n$ and $S_n^{BC}/n\xrightarrow{p}\int|\mathbb{E}\varepsilon(\phi_s-\iota_s)|^2d\mu(s)$. From Bierens (2017, Theorem 2.2), under $H_1$, $\mathbb{E}\varepsilon(\phi_s-\iota_s)\neq0$ for almost all $s$ and $\int|\mathbb{E}\varepsilon(\phi_s-\iota_s)|^2\,d\mu(s)>0$. From (b), it holds that the bootstrap statistics are bounded in probability, and the result follows.

## 8.3 Auxiliary Lemmas

**Lemma 8.1.** *Under Assumptions A-D,*

*(i)* $\mathbb{P}_n|t-1|=o_p(n^{-1/2})$ *and* $\mathbb{P}_n|\widehat{t}-t|=o_p(n^{-1/2})$,

*(ii)* $\|(\widehat{m}-m)t\|_\infty=o_p(n^{-1/4})$ *and* $\|(\widehat{m}-m)\widehat{t}\|_\infty=o_p(n^{-1/4})$,

*(iii)* $\|\widehat{B}t\|_\infty = o_p(n^{-1/4})$ *and* $\|\widehat{B}\widehat{t}\|_\infty = o_p(n^{-1/4})$,

*(iv)* $\|[\widehat{B} - (\overline{\widehat{m}} - \overline{Y})/f]t\|_\infty = o_p(n^{-1/2})$,

*(v)* $\sup_s \|(\widehat{\iota}_s - \iota_s)t\|_\infty = o_p(n^{-1/4})$ *and* $\sup_s \|(\widehat{\iota}_s - \iota_s)\widehat{t}\|_\infty = o_p(n^{-1/4})$,

*(vi)* $\|(\widehat{m}^* - m)t\|_\infty = o_P(n^{-1/4})$ *and* $\|(\widehat{m}^* - m)\widehat{t}\|_\infty = o_P(n^{-1/4})$,

*(vii)* $\|\widehat{B}^* t\|_\infty = o_P(n^{-1/4})$ *and* $\|\widehat{B}^*\widehat{t}\|_\infty = o_P(n^{-1/4})$,

*(viii)* $\|[\widehat{B}^* - (\overline{\widehat{m}^*} - \overline{Y^*})/f]t\|_\infty = o_P(n^{-1/2})$.

*Proof.* $(i)$ Notice first that $|t-1| = \mathbb{I}(f(\cdot) < \tau_n)$ and

$$\mathbb{E}\mathbb{P}_n\mathbb{I}(f(W) < \tau_n) \leq \Pr(f(W) \leq 3\tau_n/2) = p_n = o(n^{-1/2})$$

where the last equality follows from Assumption D(i). So, by Markov's inequality, $\mathbb{P}_n|t-1| = o_P(n^{-1/2})$.

To prove the second part of $(i)$, define the event $\mathcal{A}_C := \{\|\widehat{f} - f\|_\infty \leq Cd_n\}$. From Lemma 8.2 $\|\widehat{f} - \mathbb{E}\widehat{f}\|_\infty = O_P(\sqrt{(\log n)/(nh^p)})$. Standard bias manipulations ensure that $\|\mathbb{E}\widehat{f} - f\|_\infty = O(h^r)$. So, $\|\widehat{f} - f\|_\infty = O(d_n)$ and by choosing $C$ large enough $\Pr(\mathcal{A}_C)$ can be made arbitrarily close to 1 for each large $n$. Over such event, since $d_n\tau_n^{-1} = o(1)$ (see Assumption C), for each $n$ large enough $1 - \tau_n^{-1}(\widehat{f}(w) - f(w)) \leq 3/2$ for all $w \in \mathcal{W}$. Thus,

$$\mathbb{I}\Big(\widehat{f}(w) \geq \tau_n\Big) = \mathbb{I}\Big(f(w) \geq \tau_n\Big[1 - \frac{\widehat{f}(w) - f(w)}{\tau_n}\Big]\Big) \geq \mathbb{I}\Big(f(w) \geq 3\tau_n/2\Big)$$

so that $f(w) \geq (3/2)\tau_n$ implies $t(w) = \widehat{t}(w) = 1$. Accordingly, over $\mathcal{A}_C$ for large enough $n$

$$|\widehat{t}(w) - t(w)| \leq \mathbb{I}\Big(f(w) < 3\tau_n/2\Big) \text{ for all } w \in \mathcal{W}$$

and

$$\mathbb{P}_n|\widehat{t} - t| \leq \mathbb{P}_n\mathbb{I}\Big(f(W) < 3\tau_n/2\Big) = O_P(p_n) = o_P(n^{-1/2})\,,$$

where the last two equalities follow from Markov's inequality and Assumption D(i). Conclude by recalling that for $C$ large enough $\Pr(\mathcal{A}_C)$ can be made arbitrarily close to 1 for each large $n$.

$(ii)$ Consider the event $\mathcal{A}_C$ previously defined and fix $\delta \in (0, 1]$. Since $d_n\tau_n^{-1} = o(1)$, over such event for each large $n$ we have $1 + 2[f(w) - \widehat{f}(w)]/(\delta\tau_n) \leq 2$ for all $w \in \mathcal{W}$. So,

$$\mathbb{I}\Big(\widehat{f}(w) \geq \delta\tau_n/2\Big) = \mathbb{I}\Big(f(w) \geq \frac{\delta\tau_n}{2}\Big[1 + 2\frac{f(w) - \widehat{f}(w)}{\delta\tau_n}\Big]\Big) \geq \mathbb{I}\Big(f(w) \geq \delta\tau_n\Big).$$

As already seen, by choosing $C$ large enough $\Pr(\mathcal{A}_C)$ can be made arbitrarily close to 1 for each large $n$. So, we obtain that

$$\text{for each } \delta \in (0,1] \text{ wpa1} : \mathbb{I}\Big(f(w) \geq \delta\tau_n\Big) \leq \mathbb{I}\Big(\widehat{f}(w) \geq \delta\tau_n/2\Big) \text{ for all } w \in \mathcal{W}, \tag{15}$$

where wpa1 stands for "with probability approaching one". Switching the roles of $\widehat{f}$ and $f$, by a similar argument we obtain that

$$\text{for each } \delta \in (0,1] \text{ wpa1} : \mathbb{I}\Big(\widehat{f}(w) \geq \delta\tau_n\Big) \leq \mathbb{I}\Big(f(w) \geq \delta\tau_n/2\Big) \text{ for all } w \in \mathcal{W}. \tag{16}$$

Now, for any fixed $\delta \in (0,1]$ (15) implies that with probability approaching one

$$\begin{aligned}(\widehat{m}-m)\,\mathbb{I}(f(\cdot) \geq \delta\tau_n) =& \frac{\overline{Y}-m\widehat{f}}{f}\,\mathbb{I}(f(\cdot) \geq \delta\tau_n) \\ &+ \frac{\overline{Y}-m\widehat{f}}{f}\,\frac{f-\widehat{f}}{\widehat{f}}\,\mathbb{I}(f(\cdot) \geq \delta\tau_n)\,\mathbb{I}(\widehat{f}(\cdot) \geq \delta\tau_n/2)\,.\end{aligned}$$

To bound the RHS, from Lemma 8.2 $\|\overline{Y}-\mathbb{E}\overline{Y}\|_\infty = O_P(\sqrt{(\log n)/(nh^p)})$, while standard bias computations yield $\|\mathbb{E}\overline{Y}-mf\|_\infty = O(h^r)$. Hence, $\|\overline{Y}-mf\|_\infty = O_P(d_n)$. Similarly, $\|\widehat{f}-f\|_\infty = O_P(d_n)$ . Using these rates and the above display gives

$$\|(\widehat{m}-m)\,\mathbb{I}(f(\cdot) \geq \delta\tau_n)\|_\infty = O_P\left(\frac{d_n}{\tau_n}+\frac{d_n^2}{\tau_n^2}\right) = o_P(n^{-1/4}) \tag{17}$$

where the last equality follows from Assumption C. Since $\delta \in (0,1]$, the LHS of the above display is an upper bound for $\|(\widehat{m}-m)t\|_\infty$. So $\|(\widehat{m}-m)t\|_\infty = o_P(n^{-1/4})$. To obtain the rate with $\widehat{t}$, an application of (16) with $\delta = 1$ gives that with probability approaching one $\|(\widehat{m}-m)\widehat{t}\|_\infty \leq \|(\widehat{m}-m)\,\mathbb{I}(f(\cdot) \geq \tau_n/2)\|_\infty$. Applying (17) with $\delta = 1/2$ gives the $n^{-1/4}$ rate for the RHS of the latter inequality.

$(iii)$ Since $\widehat{B} = \overline{\widehat{m}}/\widehat{f} - \widehat{m}$, in view of $(ii)$ it suffices to obtain a rate for $\overline{\widehat{m}}/\widehat{f}$. To this end, notice that

$$\overline{\widehat{m}}(w) = \overline{m}(w) + \overline{m(\widehat{t}-1)}(w) + \overline{(\widehat{m}-m)\widehat{t}}(w)\,. \tag{18}$$

Combining Lemma 8.2 with standard bias computations gives $\|\overline{m}-mf\|_\infty = O_P(d_n)$. For the second term, the boundedness of $K$ and $m$ implies $|\overline{m(\widehat{t}-1)}(w)| \leq Ch^{-p}\mathbb{P}_n|\widehat{t}-1|$. Using the same arguments as in the proof of $(i)$, $\mathbb{P}_n|\widehat{t}-1| \leq \mathbb{P}_n|\widehat{t}-t| + \mathbb{P}_n|t-1| = O_P(p_n)$. So, $\|\overline{m(\widehat{t}-1)}\|_\infty = O_P(p_nh^{-p})$. For the third term on the RHS of (18)

$$\overline{(\widehat{m}-m)\widehat{t}}(w) \leq \|(\widehat{m}-m)\widehat{t}\|_\infty h^{-p}\mathbb{P}_n\Big|K\Big(\frac{W-w}{h}\Big)\Big| = O_P\Big(\frac{d_n}{\tau_n}\Big)\cdot h^{-p}\mathbb{P}_n\Big|K\Big(\frac{W-w}{h}\Big)\Big|$$

where in the last equality we have used $\left\|(\widehat{m}-m)\widehat{t}\right\|_\infty = O_P(d_n/\tau_n)$ from the proof of $(ii)$. An application of Lemma 8.2 and standard bias computations yield that the last term on the RHS is $O_P(1)$ uniformly in $w \in \mathcal{W}$. Gathering results,

$$\|\overline{\widehat{m}} - mf\|_\infty = O_P\left(d_n + p_n h^{-p} + \frac{d_n}{\tau_n}\right). \tag{19}$$

Combining the above display with arguments similar to the proof of $(ii)$ gives

$$\left\|\left(\frac{\overline{\widehat{m}}}{\widehat{f}} - m\right)\widehat{t}\right\|_\infty = O_P\left(\frac{d_n}{\tau_n^2} + \frac{p_n}{h^p\tau_n}\right) \text{ and } \left\|\left(\frac{\overline{\widehat{m}}}{\widehat{f}} - m\right)t\right\|_\infty = O_P\left(\frac{d_n}{\tau_n^2} + \frac{p_n}{h^p\tau_n}\right). \tag{20}$$

Using Assumption C and D(i) gives the desired result.

$(iv)$ Since $\widehat{B} = \overline{\widehat{m}}/\widehat{f} - \overline{Y}/\widehat{f}$, Equation (15) implies that with probability approaching one

$$\left(\widehat{B} - \frac{\overline{\widehat{m}} - \overline{Y}}{f}\right) t = (\overline{\widehat{m}} - \overline{Y})\left(\frac{f - \widehat{f}}{\widehat{f}f}\right) t\, \mathbb{I}(\widehat{f}(\cdot) \geq \tau_n/2)\,.$$

Using (19), $\|\overline{Y} - mf\|_\infty = O_P(d_n)$, and $\|\widehat{f} - f\|_\infty = O_P(d_n)$ (see the proof of $(ii)$), the RHS of the above display is

$$O_P\left(\left(\frac{d_n}{\tau_n^2} + \frac{p_n}{h^p\tau_n}\right)\frac{d_n}{\tau_n}\right) = o_P(n^{-1/2})$$

uniformly in $w \in \mathcal{W}$, where the last equality follows from Assumptions C and D(i).

$(v)$ The proof follows from arguments similar to the proof of $(ii)$, so it is omitted.

$(vi)$ First, notice that

$$\widehat{m}^* = \frac{\overline{Y^*}}{\widehat{f}} = \frac{\overline{\widehat{m}}}{\widehat{f}} + \frac{\overline{\xi\widehat{\varepsilon}}}{\widehat{f}}. \tag{21}$$

The uniform convergence rate of $\overline{\widehat{m}}/\widehat{f}$ has already been obtained in (20), so it suffices to show that the second addendum is negligible with a suitable rate. Now, $\overline{\xi\widehat{\varepsilon}} = \overline{\xi\varepsilon} + \overline{\xi\varepsilon(\widehat{t}-1)} + \overline{\xi(m-\widehat{m})\widehat{t}}$. Using this decomposition and proceeding as in the proof of $(iii)$ gives

$$\left\|\frac{\overline{\xi\widehat{\varepsilon}}}{\widehat{f}}\widehat{t}\right\|_\infty = O_P\left(\frac{d_n}{\tau_n^2} + \frac{p_n}{h^p\tau_n}\right) \text{ and } \left\|\frac{\overline{\xi\widehat{\varepsilon}}}{\widehat{f}}t\right\|_\infty = O_P\left(\frac{d_n}{\tau_n^2} + \frac{p_n}{h^p\tau_n}\right) \tag{22}$$

with $d_n\tau_n^{-2} + p_n h^{-p}\tau_n^{-1} = o(n^{-1/4})$.

$(vii)$ Recall that

$$\widehat{B}^* = \frac{\overline{\widehat{m}^*}}{\widehat{f}} - \widehat{m}^*\,.$$

In view of $(iv)$, it suffices to obtain a suitable convergence rate for the first addendum. To

this end, from (20), (21), and (22) we have

$$\|(\widehat{m}^* - m)\widehat{t}\|_\infty = O_P\left(\frac{d_n}{\tau_n^2} + \frac{p_n}{h^p \tau_n}\right) . \tag{23}$$

Also, $\overline{\widehat{m}^*} = \overline{m} + \overline{m(\widehat{t}-1)} + \overline{(\widehat{m}^* - m)\widehat{t}}$. Using this decomposition, (23), and proceeding similarly as in the proof of $(iii)$ gives

$$\left\|\left(\frac{\widehat{m}^*}{\widehat{f}} - m\right)\widehat{t}\right\|_\infty = O_P\left(\frac{d_n}{\tau_n^3} + \frac{p_n}{h^p\tau_n^2}\right) \text{ and } \left\|\left(\frac{\widehat{m}^*}{\widehat{f}} - m\right)t\right\|_\infty = O_P\left(\frac{d_n}{\tau_n^3} + \frac{p_n}{h^p\tau_n^2}\right) .$$

By Assumptions C and D(i) we obtain the desired result.

$(viii)$ The proof combines arguments already used previously. For completeness, we also provide it here. Using the definition of $\widehat{B}^*$ and (15), with probability approaching one

$$\left(\widehat{B}^* - \frac{\overline{\widehat{m}^*} - \overline{Y^*}}{f}\right) t = (\overline{\widehat{m}^*} - \overline{Y^*})\left(\frac{f - \widehat{f}}{f\widehat{f}}\right) t\, \mathbb{I}(\widehat{f}(\cdot) \geq \tau_n/2) . \tag{24}$$

As noticed in the proof of $(vii)$, $\overline{\widehat{m}^*} = \overline{m} + \overline{m(\widehat{t}-1)} + \overline{(\widehat{m}^* - m)\widehat{t}}$. By this decomposition, (23), and reasoning as in the proof of $(iii)$ we get

$$\|\overline{\widehat{m}^*} - mf\|_\infty = O_P\Big(\frac{d_n}{\tau_n^2} + \frac{p_n}{h^p\tau_n}\Big). \tag{25}$$

Combining $\overline{\xi\widehat{\varepsilon}} = \overline{\xi\varepsilon} + \overline{\xi\varepsilon(\widehat{t}-1)} + \overline{\xi(m-\widehat{m})\widehat{t}}$ with arguments used in the proof of $(iii)$ gives

$$\|\overline{\xi\widehat{\varepsilon}}\|_\infty = O_P\left(\frac{d_n}{\tau_n} + \frac{p_n}{h^p}\right).$$

By the previous display, $\overline{Y^*} = \overline{\widehat{m}} + \overline{\xi\widehat{\varepsilon}}$, and (19), we get

$$\|\overline{Y^*} - mf\|_\infty = O_P\Big(\frac{d_n}{\tau_n} + \frac{p_n}{h^p}\Big). \tag{26}$$

Finally, plugging $\|\widehat{f} - f\|_\infty = O_P(d_n)$, (25), and (26) into (24) and then using Assumptions C and D(i) gives the desired result. ■

**Lemma 8.2.** *Let $\{U_i\}_{i=1}^n$ be a sequence of i.i.d. random variables taking values in $\mathbb{R}^q$. Assume that $\{\varphi_{n,s} : s \in \mathcal{S}\}$ is a sequence of classes of real-valued functions defined on the support of $U_1$ such that for any $n \in \mathbb{N}$ : $\sup_{s\in\mathcal{S}} \|\varphi_{n,s}\|_\infty < L_\varphi$ and $\|\varphi_{n,s_1} - \varphi_{n,s_2}\|_\infty \leq$*

$L_\varphi \|s_1 - s_2\|$ *for all* $s_1, s_2 \in \mathcal{S}$. *Then, for any compact set* $\mathcal{A} \subset \mathbb{R}^p$

$$\sup_{(w,\theta)\in\mathcal{A}\times\mathcal{S}} \left| h^{-p}(\mathbb{P}_n - P)\varphi_{n,s}(U) K\Big(\frac{W-w}{h}\Big)\right| = O_P\Big(\sqrt{\frac{\log n}{nh^p}}\Big)\,.$$

*Proof.* The result is a minor modification of the proof of Theorem 1.4 in Li and Racine (2006). ■

The following lemma provides the regularity features needed to apply stochastic equicontinuity results. Similar results can be found in Andrews (1994) and Andrews (1995). The differences with respect to these works are the presence of the bias correction components $\widehat{\overline{m}}/\widehat{f}$ and the different assumptions on the bandwidths. Recall that

$$\mathcal{W}_n := \{w : f_W(w) \geq \tau_n/2\}\ .$$

**Lemma 8.3.** *Under Assumptions B-D for* $l = \lceil (p+1)/2 \rceil$,

*(i)* $\Pr\Big(\widehat{m} \in \mathcal{G}_l(\mathcal{W}_n)\Big) \to 1$,

*(ii)* $\Pr\Big(\widehat{\overline{m}}/\widehat{f} \in \mathcal{G}_l(\mathcal{W}_n)\Big) \to 1$,

*(iii)* $\Pr\Big(\widehat{\iota}_s \in \mathcal{G}_l(\mathcal{W}_n)$ *for all* $s \in \mathcal{S}\Big) \to 1$,

*(iv)* $\Pr\Big(\widehat{\overline{\xi\varepsilon}}/\widehat{f} \in \mathcal{G}_l(\mathcal{W}_n)\Big) \to 1$,

*(v)* $\Pr\Big(\widehat{\overline{m^*}}/\widehat{f} \in \mathcal{G}_l(\mathcal{W}_n)\Big) \to 1$.

*Proof.* The proof of this result can be found in the Supplementary Material. ■

The following lemma provides a stochastic equicontinuity result. Similar results can be found in Andrews (1994) or Andrews (1995). Let us introduce some notation that will be used in the proof. For a generic space of functions $\mathcal{F}$ endowed with the $L_2(P)$ metric $||\cdot||_{2,P}$, we denote by $N(\epsilon, \mathcal{F}, ||\cdot||_{2,P})$ the $\epsilon$ covering number and with $N_{[\cdot]}(\epsilon, \mathcal{F}, ||\cdot||_{2,P})$ the $\epsilon$ bracketing number, see van der Vaart (1998) and van der Vaart and Wellner (2000).

**Lemma 8.4.** *Let Assumptions A-D hold and denote with Supp*$(\zeta)$ *the support of* $\zeta := (\xi, Y, W, X)$.

*(i) If* $\{\widehat{f}_s : s \in \mathcal{S}\}$ *is a collection of stochastic real-valued functions defined on* $\mathcal{W}$ *such that* $\sup_{s\in\mathcal{S}} \|\widehat{f}_s t\|_\infty = o_P(1)$ *and* $\Pr(\widehat{f}_s \in \mathcal{G}_l(\mathcal{W}_n)$ *for all* $s \in \mathcal{S}) \to 1$ *for* $l = \lceil (p+1)/2 \rceil$, *then*

$$\sup_{s\in\mathcal{S}} \left| \mathbb{G}_n t \widehat{f}_s \phi_s \right| = o_P(1)\,.$$

*The same result also holds when $\phi_s$ is replaced by $\phi_s - \iota_s$, $\xi(\phi_s - \iota_s)$, or a fixed bounded function.*

*(ii) If $g_n : Supp(\zeta) \mapsto \mathbb{R}$ is a sequence of functions such that $\|g_n\|_\infty = O(1)$,*

$$\mathbb{G}_n g_n \int K(u)\iota_s(\cdot + uh)\, du = \mathbb{G}_n g_n \iota_s + o_P^{\mathcal{S}}(1)\,.$$

*Proof.* $(i)$ Fix an arbitrary $\epsilon > 0$. The assumptions on $\{f_s : s \in \mathcal{S}\}$ ensure that with probability approaching one

$$\sup_{s\in\mathcal{S}} \left|\mathbb{G}_n t \widehat{f}_s \phi_s\right| \leq \sup_{g\in\mathcal{G}_n^\epsilon} |\mathbb{G}_n g| \tag{27}$$

with $\mathcal{G}_n^\epsilon := \{g \in \mathcal{G}_n : \|g\|_{2,P} < \epsilon\}$, $\mathcal{G}_n = t \cdot \mathcal{G}_l(\mathcal{W}_n) \cdot \Psi$, and $\Psi := \{\phi_s : s \in \mathcal{S}\}$. By van der Vaart (1998, Lemma 19.34), if

$$\log N_{[\cdot]}(\epsilon, \mathcal{G}_n, ||\cdot||_{2,P}) \leq C\epsilon^{-\gamma} \quad \text{for all } \epsilon \in (0,1) \text{ and some fixed } \gamma \in (0,2)\,,$$

then the expectation of the right-hand side of (27) can be made arbitrarily small asymptotically by choosing $\epsilon$ small enough. So, Markov's inequality would deliver the desired result.

Since $N_{[\cdot]}(2\epsilon, \mathcal{G}_n, ||\cdot||_{2,P}) \leq N(\epsilon, \mathcal{G}_n, ||\cdot||_\infty)$, we focus on the latter. Assumption D(ii) and van der Vaart and Wellner (2000, Theorem 2.7.1) ensure that for each $n$ large enough

$$\log N\Big(\epsilon\,, \mathcal{G}_l(\mathcal{W}_n)\,, \|\cdot\|_{\infty,\mathcal{W}_n}\Big) \leq C\varepsilon^{-\upsilon}, \quad \upsilon \in (0,2)$$

where $||g||_{\infty,\mathcal{W}_n} := \sup_{w\in\mathcal{W}_n} |g(w)|$ for any $g \in \mathcal{G}_l(\mathcal{W}_n)$. Since $\phi_s$ is Lipschitz on the compact $\mathcal{S}$, $N(\epsilon, \Psi, \|\cdot\|_\infty) \leq C\epsilon^{-q}$ by Kosorok (2008, Theorem 9.15). So, using the fact that $t$ is bounded,

$$N(\epsilon, \mathcal{G}_n, \|\cdot\|_\infty) \leq C\epsilon^{-q}\exp(\epsilon^{-\upsilon}) \text{ with } \upsilon \in (0,2)\ .$$

The same reasoning applies if $\phi_s - \iota_s$ replaces $\phi_s$, and if $\xi(\phi_s - \iota_s)$ replaces $\phi_s$, since $N_{[\cdot]}(\epsilon||\xi||_2, \xi \cdot \mathcal{G}_n, ||\cdot||_{2,P}) \leq N_{[\cdot]}(\epsilon, \mathcal{G}_n, ||\cdot||_{2,P})$, where $||\xi||_2^2 = \mathbb{E}|\xi|^2$.

$(ii)$ By a $r$th order Taylor expansion, $\int K(u)\iota_s(w + uh)\, du = \iota_s(w) + O(h^r)$ uniformly in $w \in \mathcal{W}$. Thus, the proof proceeds along similar lines as the proof of $(i)$.

■

# References


Ait-Sahalia, Y., P. J. Bickel, and T. M. Stoker (2001): "Goodness-of-Fit Tests for Kernel Regression with an Application to Option Implied Volatilities," *J. Econometrics*, 105, 363 – 412.

Andrews, D. W. K. (1994): "Empirical Process Methods in Econometrics," in *Handbook of Econometrics*, Elsevier, vol. 4, 2247 – 2294.

——— (1995): "Nonparametric Kernel Estimation for Semiparametric Models," *Economet Theor*, 11, 560–586.

Bentkus, V., F. Götze, and R. Zitikis (1993): "Asymptotic Expansions in the Integral and Local Limit Theorems in Banach Spaces with Applications to $w$-Statistics," *J Theor Probab*, 6, 54.

Bierens, H. J. (1982): "Consistent Model Specification Tests," *J Econometrics*, 20, 105–134.

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

——— (2017): *Econometric Model Specification*, World Scientific.

Bierens, H. J. and W. Ploberger (1997): "Asymptotic Theory of Integrated Conditional Moment Tests," *Econometrica*, 65, 1129–1152.

Bontemps, C., J.-P. Florens, and J.-F. Richard (2008): "Parametric and Non-Parametric Encompassing Procedures," *Oxford B Econ Stat*, 70, 751–780.

Bontemps, C. and G. E. Mizon (2008): "Encompassing: Concepts and Implementation," *Oxford B Econ Stat*, 70, 721–750.

Chernozhukov, V., J. C. Escanciano, H. Ichimura, W. K. Newey, and J. M. Robins (2022): "Locally Robust Semiparametric Estimation," *Econometrica*, forthcoming.

Davidson, R. and J. G. MacKinnon (2007): "Improving the Reliability of Bootstrap Tests with the Fast Double Bootstrap," *Comput. Statist. Data Anal.*, 51, 3259–3281.

Delgado, M. A. and W. G. Manteiga (2001): "Significance Testing in Nonparametric Regression Based on the Bootstrap," *Ann. Statist.*, 29, 1469–1507.

Delgado, M. A. and W. Stute (2008): "Distribution-Free Specification Tests of Conditional Models," *J Econometrics*, 143, 37–55.

Dhaene, G., C. Gourieroux, and O. Scaillet (1998): "Instrumental Models and Indirect Encompassing," *Econometrica*, 66, 673–688.

Di Marzio, M. and C. C. Taylor (2008): "On Boosting Kernel Regression," *J. Stat. Plan. Inference*, 138, 2483–2498.

Escanciano, J. C. (2006): "A Consistent Diagnostic Test for Regression Models Using Projections," *Econometric Theor*, 22, 1030–1051.

Escanciano, J. C., D. T. Jacho-Chávez, and A. Lewbel (2014): "Uniform Convergence of Weighted Sums of Non and Semiparametric Residuals for Estimation and Testing," *J Econometrics*, 178, 426–443.

Fan, Y. and Q. Li (1996): "Consistent Model Specification Tests: Omitted Variables and Semiparametric Functional Forms," *Econometrica*, 64, 865–890.

Florens, J.-P., D. F. Hendry, and J.-F. Richard (1996): "Encompassing and Specificity," *Economet Theor*, 12, 620–656.

Giacomini, R., D. N. Politis, and H. White (2013): "A Warp-Speed Method For Conducting Monte Carlo Experimens Involving Bootstrap Estimators," *Economet. Theor.*, 29, 567–589.

Gourieroux, C. and A. Monfort (1995): "Testing, Encompassing, and Simulating Dynamic Econometric Models," *Economet Theor*, 11, 195–228.

Gourieroux, C., A. Monfort, and A. Trognon (1983): "Testing Nested or Non-Nested Hypotheses," *J Econometrics*, 21, 83–115.

Hendry, D. F. and J.-F. Richard (1982): "On the Formulation of Empirical Models in Dynamic Econometrics," *J Econometrics*, 20, 3–33.

Kosorok, M. R. (2008): *Introduction to empirical processes and semiparametric inference*, New York: Springer.

Lavergne, P. (2001): "An Equality Test Across Nonparametric Regressions," *J Econometrics*, 103, 307–344.

Lavergne, P., S. Maistre, and V. Patilea (2015): "A Significance Test for Covariates in Nonparametric Regression," *Electron J Stat*, 9, 643–678.

Lavergne, P. and V. Patilea (2008): "Breaking the Curse of Dimensionality in Nonparametric Testing," *J Econometrics*, 143, 103–122.

Lavergne, P. and Q. H. Vuong (1996): "Nonparametric Selection of Regressors: The Nonnested Case," *Econometrica*, 64, 207–219.

——— (2000): "Nonparametric Significance Testing," *Economet Theor*, 16, 576–601.

Li, Q. and J. S. Racine (2006): *Nonparametric Econometrics: Theory and Practice*, Princeton University Press.

Mammen, E. (1992): *When Does Bootstrap Work?*, vol. 77 of *Lecture Notes in Statistics*, Springer, New York.

Mammen, E., C. Rothe, and M. Schienle (2016): "Semiparametric Estimation with Generated Covariates," *Economet Theor*, 32, 1140–1177.

Mizon, G. E. and J.-F. Richard (1986): "The Encompassing Principle and its Application to Testing Non-Nested Hypotheses," *Econometrica*, 54, 657–678.

Newey, W. K. (1990): "Semiparametric Efficiency Bounds," *J Appl Econom*, 5, 99–135, publisher: Wiley.

Newey, W. K., F. Hsieh, and J. M. Robins (2004): "Twicing Kernels and a Small Bias Property of Semiparametric Estimators," *Econometrica*, 72, 947–962.

Park, B. U., Y. K. Lee, and S. Ha (2009): "L2 boosting in kernel regression," *Bernoulli*, 15, 599–613.

STINCHCOMBE, M. B. AND H. WHITE (1998): "Consistent Specification Testing With Nuisance Parameters Present Only Under The Alternative," *Economet Theor*, 14, 295–325.

VAN DER VAART, A. W. (1998): *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.

VAN DER VAART, A. W. AND J. A. WELLNER (2000): *Weak Convergence and Empirical Processes: with Applications to Statistics*, New York: Springer.

XIA, Y., W. K. LI, H. TONG, AND D. ZHANG (2004): "A Goodness-of-Fit Test for Single-Index Models," *Stat Sinica*, 14, 1–28.

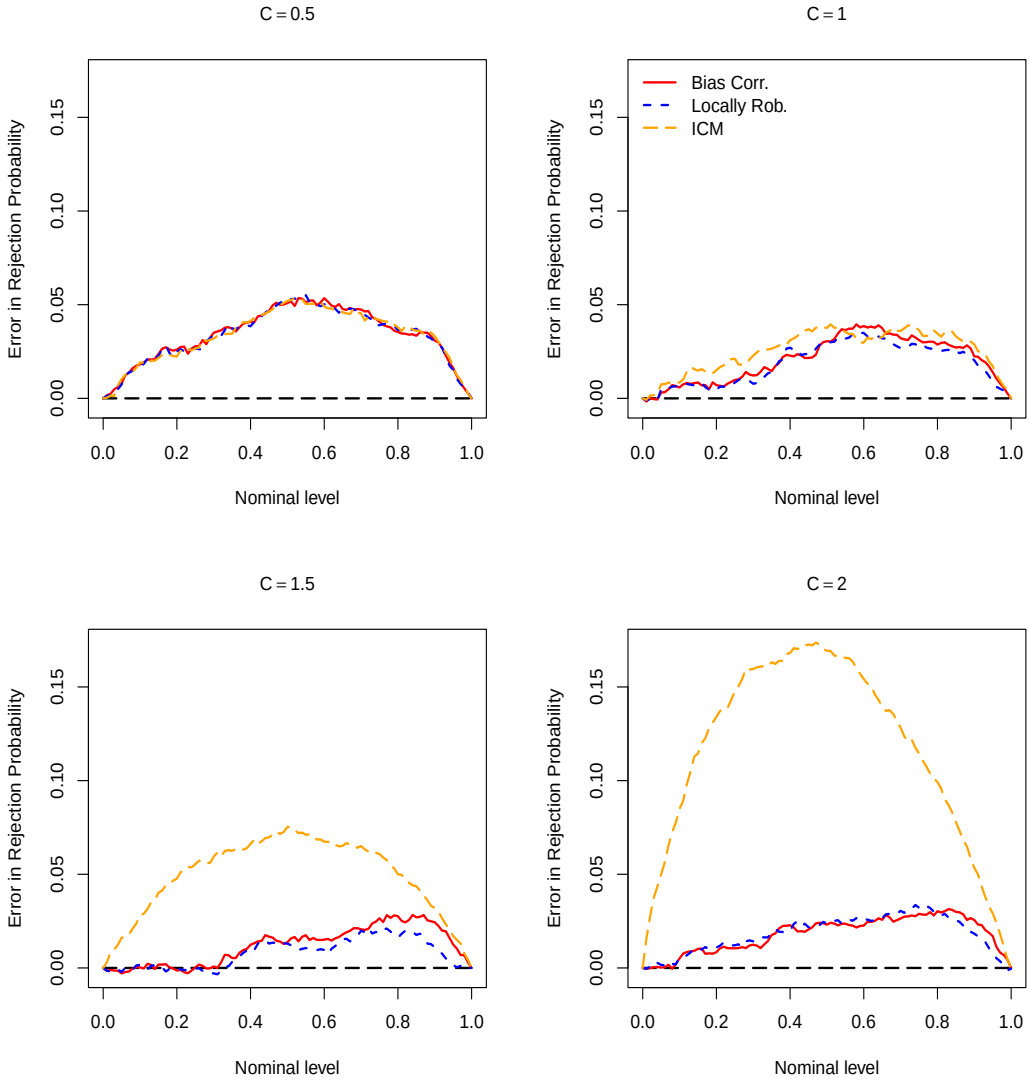


Figure 1: Error in Rejection Probabilities for $n = 200$.

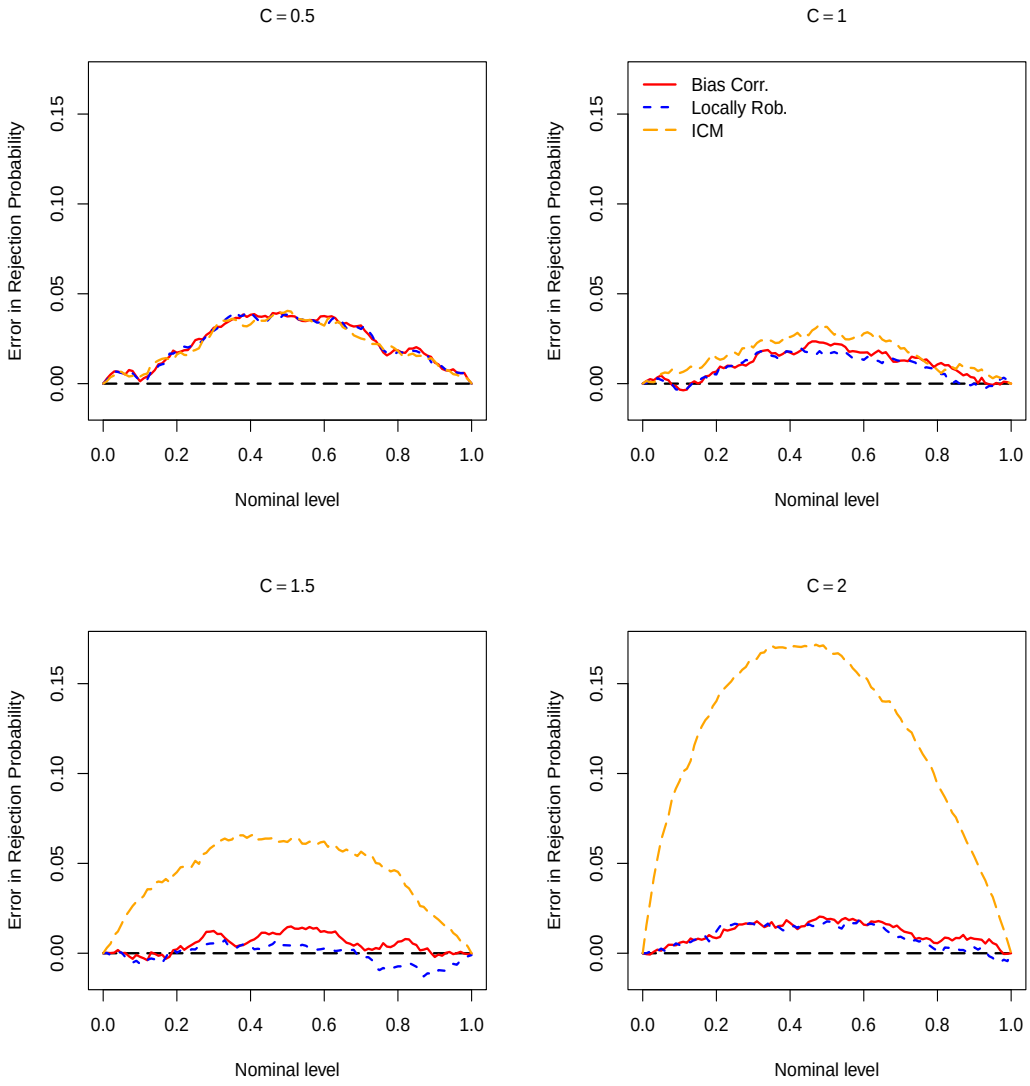


Figure 2: Error in Rejection Probabilities for $n = 400$.

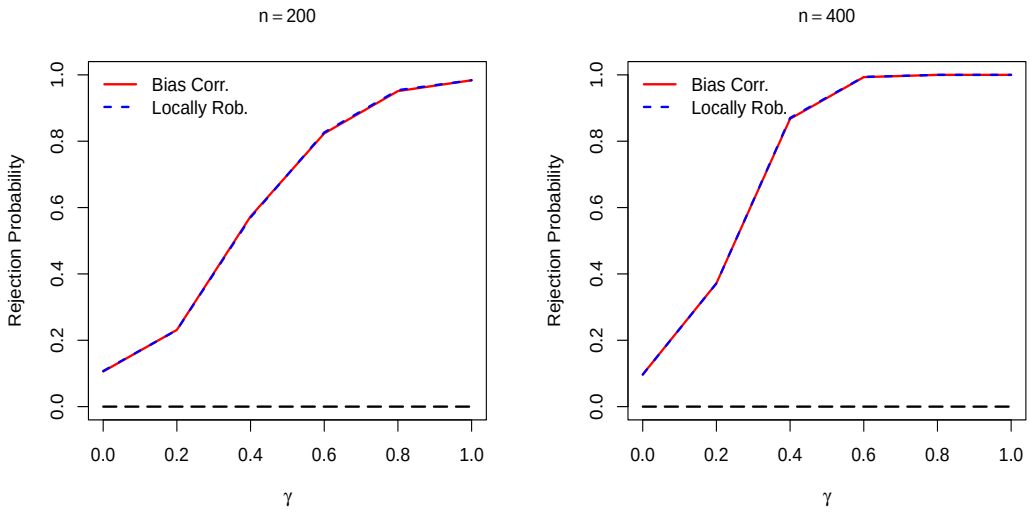


Figure 3: Power Curves for 10% level tests and $C = 1$.

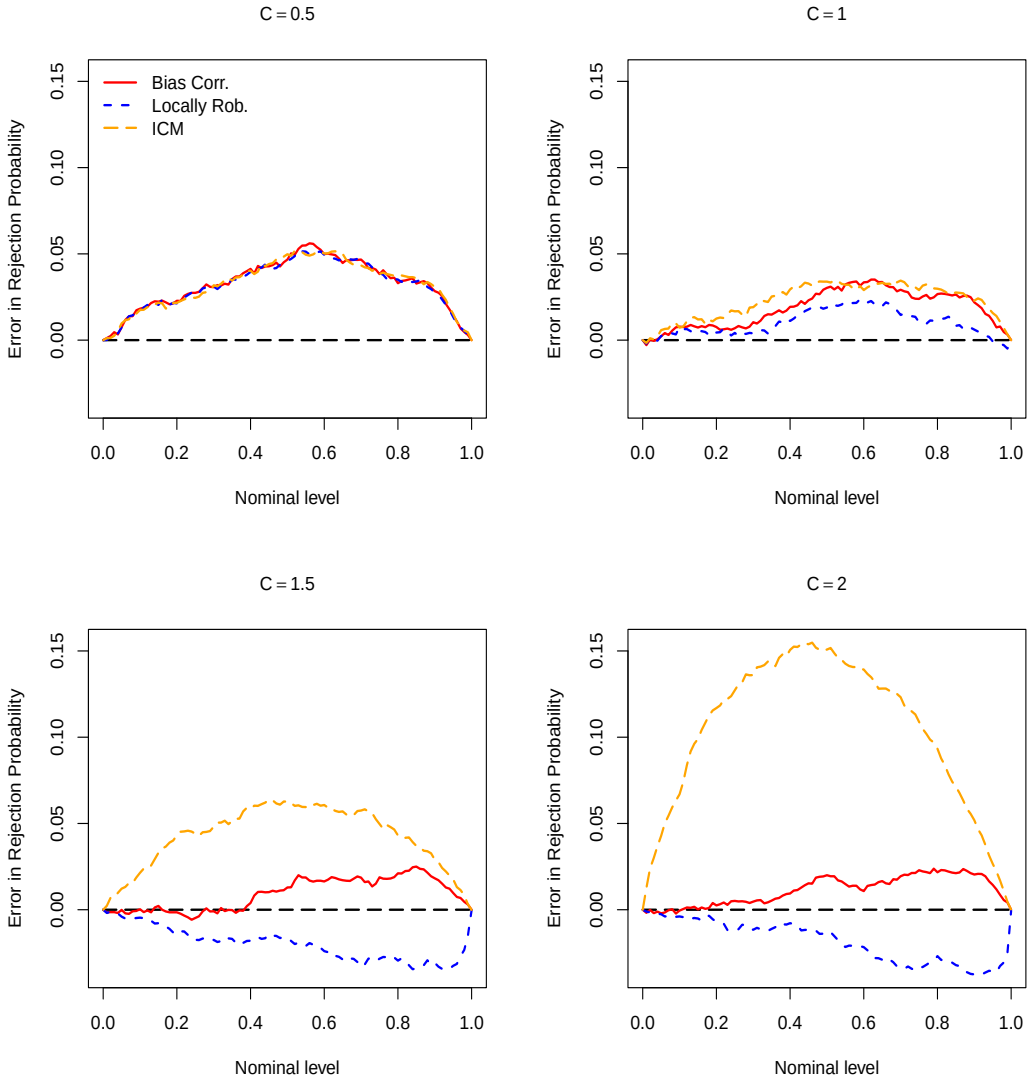


Figure 4: Error in Rejection Probabilities without trimming for $n = 200$.

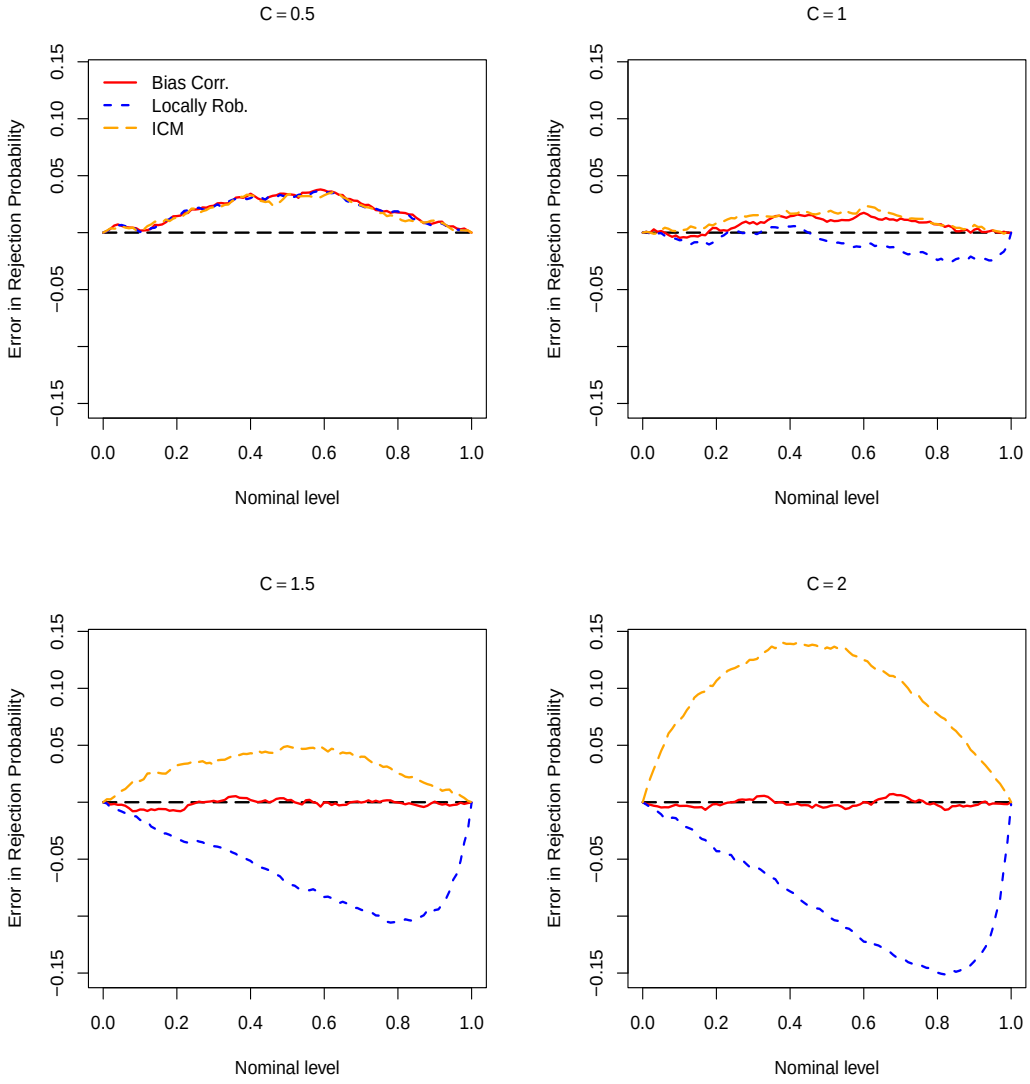


Figure 5: Error in Rejection Probabilities without trimming for $n = 400$.