Econometric Theory, **41**, 2025, 709–738.
doi:10.1017/S0266466624000100

ENCOMPASSING TESTS FOR NONPARAMETRIC REGRESSIONS

ELIA LAPENTA
CREST and ENSAE, Institut Polytechnique de Paris

PASCAL LAVERGNE
Toulouse School of Economics, Université Toulouse Capitole

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. We empirically study the use of a data-driven bandwidth and illustrate the attractive features of our tests for small and moderate samples.

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. Our 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

We thank the co-editor Michael Jansson and two referees for their comments that helped to improve the paper significantly. Elia Lapenta acknowledges funding from the French National Research Agency (ANR) under grant ANR-23-CE26-0008-01. Pascal Lavergne acknowledges funding from ANR under grant ANR-17-EURE-0010 (Investissements d'Avenir program). Address correspondence to Elia Lapenta, CREST, 5 Avenue Le Chatelier, 91120 Palaiseau, France; e-mail: elia.lapenta@ensae.fr.

© The Author(s), 2024. Published by Cambridge University Press.

https://doi.org/10.1017/S0266466624000100 Published online by Cambridge University Press

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 integrated conditional moment (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 resulting statistic 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 property. 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 et al., 2022 for a general approach). We here adapt the two approaches to our empirical process of interest, and we show that these 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 et al., 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; 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,

https://doi.org/10.1017/S0266466624000100 Published online by Cambridge University Press

e.g., Escanciano, Jacho-Chávez, and Lewbel (2014), 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 data-generating process (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. We check that bootstrapping allows to correctly control size and that our tests have good power. We evaluate the benefits of our bias-reducing approaches, and we investigate thoroughly the influence of the bandwidth as well as of the trimming parameter, which is theoretically necessary. We also provide an empirical illustration. Section 8 contains the proofs of our main results. A Supplementary Material contains the proof of a technical lemma.

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 α_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 $\mathrm{E}_f\left[\log\left(f(\cdot)/g(\cdot,\alpha)\right)\right]$, with expectation taken with respect to $f(\cdot)$. The binding 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 Model $\mathcal{M}_W$ 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 L^2 distance.

https://doi.org/10.1017/S0266466624000100 Published online by Cambridge University Press

That is, $\mathcal{M}_W$ and $\mathcal{M}_X$ are, respectively, defined as $L^2(W)$, the space of the square integrable functions of W, and $L^2(X)$. The "pseudo-true functions" are then the regression functions

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

The binding function is similarly defined as

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

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

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

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 $\mathrm{E}\,(\mathrm{E}(Y|X)|W) = \mathrm{E}(Y|W)$ almost surely.

In what follows, we explore the implications of the definition of nonparametric encompassing and relate it to previous work on the comparison of regression models. In particular, 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.

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

$$H_0 : \mathrm{E}[Y - m(W)]^2 - \mathrm{E}[Y - m_X(X)]^2 = 0,$$
$$H_W : \mathrm{E}[Y - m(W)]^2 - \mathrm{E}[Y - m_X(X)]^2 < 0,$$
$$H_X : \mathrm{E}[Y - m(W)]^2 - \mathrm{E}[Y - m_X(X)]^2 > 0.$$

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.

Comparing nonparametric regressions through their fit seems natural. 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 $\mathrm{E}(Y|W)$ is at least as good as the one of $\mathrm{E}(Y|X)$, and is strictly better except when $\mathrm{E}(Y|W) = \mathrm{E}(Y|X)$ a.s.

https://doi.org/10.1017/S0266466624000100 Published online by Cambridge University Press

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

(a) *H_X cannot hold,*
(b) *H_0 holds iff $\mathrm{E}(Y|W) = \mathrm{E}(Y|X)$ a.s. iff $\mathcal{M}_X$ encompasses $\mathcal{M}_W$.*

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

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

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

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

Therefore, except in the case where the two regressions are equal, encompassing implies a strictly better fit for the encompassing model. Conversely, one model can have a better fit than another without encompassing it (see Lavergne and Vuong, 1996 for some examples where this happens). Hence, the encompassing concept is not tailored for model selection.

The test statistic proposed by Lavergne and Vuong (1996) to compare models' fit is asymptotically degenerate under H_0 when the two models are "generalized nested regressions," that is, when either $\mathrm{E}(\mathrm{E}(Y|W)|X) = \mathrm{E}(Y|X)$ a.s. or $\mathrm{E}(\mathrm{E}(Y|X)|W) = \mathrm{E}(Y|W)$ a.s. Hence, their test cannot be used when there is encompassing, and the tests we develop below are complementary to theirs. The latter might also be used as preliminary tests to check whether their test can be entertained. This would imply (i) testing whether Model $\mathcal{M}_X$ encompasses Model $\mathcal{M}_W$, and (ii) testing whether Model $\mathcal{M}_W$ encompasses Model $\mathcal{M}_X$. One rejection among two would imply (we cannot reject) that the encompassing model has a strictly better fit than the encompassed one; no rejection would mean (we cannot reject) that $\mathrm{E}(Y|W) = \mathrm{E}(Y|X)$ a.s. Only if we rejected twice would we need to entertain the test of Lavergne and Vuong (1996) to determine whether one model has a better fit. A more direct alternative to this involved procedure is to apply the test proposed by Liao and Shi (2020), which is universally valid. We leave the study of the respective merits of these competing procedures for future work.

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

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

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

https://doi.org/10.1017/S0266466624000100 Published online by Cambridge University Press

regressions has been a focus of extensive work (see Fan and Li, 1996; Lavergne and Vuong, 2000; Ait-Sahalia, Bickel, and Stoker, 2001; Delgado and Manteiga, 2001; Lavergne, 2001).

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

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

By conditioning again with respect to X, one obtains that

$$\mathrm{E}(Y|X) = \mathrm{E}[\mathrm{E}(Y|W)|X] \text{ a.s.},$$

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*

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

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

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

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

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 almost surely equal to zero, $\mathcal{M}_W$ encompasses $\mathcal{M}_X$, but X is a significant covariate in $\mathrm{E}(Y|W,X)$. The following provides a concrete example.

Example. Let W be continuous univariate, symmetrically distributed around 0, with density $f_W(\cdot)$. Consider $p(\cdot) \in (0,1)$, $\varphi(\cdot)$ and $\psi(\cdot)$ two univariate densities with mean 0. Define

$$f_{X|W}(x|w) = p(w)\varphi(x) + (1-p(w))\psi(x)\,,$$

so that X has a mixture distribution conditionally on W. Then

$$f_X(x) = \left[\int p(w) f_W(w)\,dw\right]\varphi(x) + \left[\int (1-p(w)) f_W(w)\,dw\right]\psi(x)\,,$$

$$f_{W|X}(w|x) = \frac{f_W(w)}{f_X(x)}\left(p(w)\varphi(x) + (1-p(w))\psi(x)\right).$$

https://doi.org/10.1017/S0266466624000100 Published online by Cambridge University Press

Let

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

Then

$$\mathrm{E}(X|W) = p(W)\int x\varphi(x)\,dx + (1-p(W))\int x\psi(x)\,dx = 0.$$

One can select $h(\cdot)$ and $p(\cdot)$ so that $\mathrm{E}[h(W)|X] = 0$. Consider first a case where $h(\cdot)$ is an odd function with mean 0 and $p(\cdot)$ is even. Then by the symmetry of $f_W(\cdot)$, $\mathrm{E}[h(W)|X] = 0$. Another setup used in our simulations below is for $h(\cdot)$ even with mean 0 and $p(-w) = 1-p(w)$. Again, it is easy to show that $\mathrm{E}[h(W)|X] = 0$.

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

3. TESTS STATISTICS

3.1. ICM Statistic

We want to test

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

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

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

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 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 $\mathrm{E}(Y|W)$, then Bierens' ICM statistic for testing H_0 is

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

https://doi.org/10.1017/S0266466624000100 Published online by Cambridge University Press

where μ 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 Cramér–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 $\mathrm{E}(Y|W)$. Let p be the dimension of W, $K(\cdot)$ be a kernel on $\mathbb{R}^p$, $h(=h_n)$ a bandwidth, and $K_h(u) = K\left(u_1/h, \dots u_p/h\right)$.[1] Define

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

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

To control for the random denominator in the kernel estimator, we introduce a trimming factor $\widehat{t}(w) = \mathbf{1}\left(\widehat{f}(w) \geq \tau_n\right)$, where τ_n converges to zero. Let $\widehat{\varepsilon}_i = Y_i - \widehat{m}(W_i)$, $\phi_s(\cdot) = \varphi(s'\cdot)$, and $\mathbb{P}_n(g) = n^{-1}\sum_{i=1}^{n} g(Z_i)$ denote 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 μ to be symmetric around the origin. Define

$$a(z) = \int_{\mathcal{S}} \exp(is'z)\, d\mu(s) = \int_{\mathcal{S}} \cos(s'z)\, d\mu(s),$$

due to the symmetry of μ, and let $\widehat{\boldsymbol{\varepsilon}} = \left(\widehat{\varepsilon}_j, j = 1, \dots n\right)'$ and $\widehat{t}_i = \widehat{t}(W_i)$. Then the statistic becomes

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

$$= n^{-1} \sum_{j=1}^{n}\sum_{m=1}^{n} \widehat{\varepsilon}_j\widehat{\varepsilon}_m\widehat{t}_j\,\widehat{t}_m \int_{\mathcal{S}} \cos(s'(X_j - X_m))\, d\mu(s) = \widehat{\boldsymbol{\varepsilon}}'\boldsymbol{A}\widehat{\boldsymbol{\varepsilon}},$$

where $\boldsymbol{A}$ 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 (up to a constant) the Fourier transform of μ, 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

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

https://doi.org/10.1017/S0266466624000100 Published online by Cambridge University Press

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 to allow for an optimal nonparametric bandwidth. This is convenient because there are well-known methods for approximating 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.

L^2 boosting starts with an initial nonparametric estimator and builds a bias-corrected updated estimator. The bias correction is roughly based on nonparametric residuals. The method can be iterated, but we will restrict to a single boosting step, which is sufficient for our purpose. In our context, the bias correction to be applied to $\widehat{m}(w)$ is

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

where

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

and the trimming controls for the random denominator of $\widehat{m}$. The bias-corrected estimator thus is

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

We then 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, Sect. 3) who considered density-weighted nonparametric estimators. To use their technique in our setup would necessitate applying 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

https://doi.org/10.1017/S0266466624000100 Published online by Cambridge University Press

a general Generalized Method of Moments (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 better 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

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

Because for any s, $\mathrm{E}\left[(Y-\mathrm{E}(Y|W))\,\mathrm{E}(\varphi(s'X)|W)\right]=0$, 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 $\iota_s(W)=\mathrm{E}(\phi_s(X)|W)$ and its estimator

$$\widehat{\iota}_s(w)=\overline{\phi}_s(w)/\widehat{f}(w), \qquad \overline{\phi}_s(w)=(nh^p)^{-1}\sum_{i=1}^{n}\phi_s(X_i)K_h(W_i-w)\,.$$

Our locally robust approach thus delivers the statistic

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

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 previous section does not yield the same statistic as the one based on the locally robust process. In practice, there is no need to compute $\widehat{\iota}_s(W)$ for each $s\in\mathcal{S}$ and to integrate. Simple algebra reveals that, if $\boldsymbol{K}$ denotes the $n\times n$ matrix of generic element $K_h(W_i-W_j)/[nh^p\widehat{f}(W_j)]$, then

$$S_n^{LR}=\widehat{\boldsymbol{\varepsilon}}'\boldsymbol{A}\widehat{\boldsymbol{\varepsilon}}-2\widehat{\boldsymbol{\varepsilon}}'\boldsymbol{A}\boldsymbol{K}\widehat{\boldsymbol{\varepsilon}}+\widehat{\boldsymbol{\varepsilon}}'\boldsymbol{K}'\boldsymbol{A}\boldsymbol{K}\widehat{\boldsymbol{\varepsilon}},$$

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 assuming some undersmoothing.

https://doi.org/10.1017/S0266466624000100 Published online by Cambridge University Press

We first introduce some definitions. Define the differential operator

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

DEFINITION 4.1. *(a)* $\mathcal{G}_\lambda(\mathcal{A}) = \left\{ g : \mathcal{A} \mapsto \mathbb{R} : \sup_{a \in \mathcal{A}} |\partial^l g(a)| < M \text{ for all } |l| \le \lambda \right\}$. *(b)* $\mathcal{K}^r_\lambda$ *is the class of product univariate kernels* $k(\cdot)$ *such that* $k(\cdot)$ *is of order r,* λ *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) The density of W $f(\cdot) \in \mathcal{G}_r(\mathbb{R}^p)$ and $m(\cdot) \in \mathcal{G}_r(\mathcal{W})$, with $r \ge \lceil (p+1)/2 \rceil$.[2] (ii) $\iota_s(\cdot) \in \mathcal{G}_r(\mathcal{W})$ uniformly in $s \in \mathcal{S}$, with $r \ge \lceil (p+1)/2 \rceil$. (iii) $K(\cdot) \in \mathcal{K}^r_\lambda$ with $\lambda \ge \lceil (p+1)/2 \rceil + 1$.

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) $d_n h_n^{-|l|} \tau_n^{-(2+|l|)} = o(1)$ for $|l| = \lceil (p+1)/2 \rceil$.

Assumption D. (i) $p_n = \Pr(f(W) \le 3\tau_n/2)$ 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 $n \ge N$, $\mathcal{W}_n = \{w : f(w) \ge \tau_n/2\}$ is convex.

Assumption A ensures that the encompassing hypothesis can be written as the continuum of moment conditions (4). In particular, Bierens (2017, Thm. 2.2) builds on previous results by Bierens (1982) and Stinchcombe and White (1998), and shows that Assumption A(iii) is sufficient. Intuitively, if $\varphi(\cdot)$ is analytic non-polynomial, Equation (4) implies that $Y - \mathrm{E}(Y|W)$ is uncorrelated with any polynomial in $s'X$, and thus yields the required equivalence. 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

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

https://doi.org/10.1017/S0266466624000100 Published online by Cambridge University Press

class of sufficiently smooth functions with limited entropy/complexity. This is needed for the asymptotic stochastic 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. 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; 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.[3] We could extend our theoretical results to stochastic bandwidths to allow for the use of data-driven methods, as was done in other contexts (Mammen, 1992; Andrews, 1995; Lavergne, 2001). The details would be more involved here because of the stochastic trimming, so we do not pursue this issue further.

Assumption D as a whole is used to ensure that trimming has a negligible effect overall, see e.g. Lavergne and Vuong (1996) or Escanciano et al. (2014) for similar assumptions. It may be difficult to check, but our simulations seem to indicate that trimming is actually not crucial: when a data-driven bandwidth is used, results with no trimming are comparable, and mostly better, than results with trimming.

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

PROPOSITION 4.1. *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, say $\mathbb{G}_s$, of both empirical processes is the limiting one of $\sqrt{n}\mathbb{P}_n(\varepsilon(\phi_s - \iota_s))$. This $\mathbb{G}_s$ is a zero-mean tight Gaussian process valued in $L^\infty(\mathcal{S})$, the space of uniformly bounded functionals over $\mathcal{S}$, and characterized by the collection of covariances $\{\mathrm{E}(Y - m(W))^2(\phi_s(X) - \iota_s(X))(\phi_t(X) - \iota_t(W)) : s,t \in \mathcal{S}\}$. 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.

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

https://doi.org/10.1017/S0266466624000100 Published online by Cambridge University Press

5. BOOTSTRAP TESTS

We use a wild bootstrap procedure that imposes the null hypothesis H_0 when resampling the observations. The bootstrap 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 $\mathrm{E}\,\xi = 0$ and $\mathrm{E}\,\xi^2 = 1$. From each bootstrap sample $\{(Y_i^*, W_i, X_i) : i = 1, \dots, n\}$, we proceed as above to obtain $\widehat{\varepsilon}_i^* = Y_i^* - \widehat{m}^*(W_i)$ and $\widetilde{\varepsilon}_i^* = Y_i^* - \widetilde{m}^*(W_i)$, where

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

The kernel smoothers $\overline{Y^*}$ and $\widehat{\overline{m^*}}$ 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 at asymptotic level α whenever $S_n^j > q_{1-\alpha}^j$, for $j = BC$ or LR.

PROPOSITION 5.1. *Let* $\{\xi_i, i = 1, \dots n\}$ *be a bounded i.i.d. sequence with* $\mathrm{E}\,\xi = 0$, $\mathrm{E}\,\xi^2 = 1$ *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* $s \in \mathcal{S}$.[4]

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

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

6. NUMERICAL RESULTS

6.1. Small Sample Behavior

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

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

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

$$f_{X|W}(x|w) = p(w)\varphi(x; 1/4) + (1 - p(w))\varphi(x; 3/4),$$

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

https://doi.org/10.1017/S0266466624000100 Published online by Cambridge University Press

with $\varphi(\cdot;s)$ the density of a $N(0,\sigma^2 = s^2)$ and $p(\cdot)$ the cumulative distribution function of W. We set $m(w) = w^3 - 2w$, $h(w) = w^2 - (1/2)^2$, $\delta(x) = x^4 - 3x^2$. When $\gamma = 0$, $\mathcal{M}_W$ encompasses $\mathcal{M}_X$, while it does not when $\gamma \neq 0$.

To study the behavior of our tests based on S_n^{BC} and S_n^{LR} and compare them to the test based on the uncorrected ICM statistic S_n, we ran 10,000 simulations for different values of γ and for sample sizes n =200 and 400. For our implementation, the weighting function was $a(x) = \text{sinc}(\pi x)$, which corresponds to a uniform μ 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 $a(\cdot)$. 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. 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). 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 employed Gaussian kernels of order 2 for nonparametric estimation. In a first step, we used 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. We report actual rejection probabilities for 5% and 10% nominal sizes in Table 1. For $C = 0.5$ or $C = 1$, the three tests have equal size control and the empirical size becomes closer to the nominal level when increasing the sample size. When C increases to 1.5 then 2, size control deteriorates for the uncorrected ICM test, while it improves for our two tests. We also report in Figure 1 errors in rejection probability (ERP), that is, 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. These graphs clearly show the high sensitivity of the ICM test to the bandwidth choice, and the relative robustness of our two procedures. We complemented our study with an investigation of the trimming influence. While we do not report details, the most striking result we obtained was that the absence of trimming adversely affected the behavior of the locally robust test for large bandwidths.

Following the suggestion of a referee, we investigated the issue of bandwidth selection. Since there is no role for the bandwidth in the first-order expansion of our test statistics, we decided to implement a data-driven bandwidth that should be asymptotically optimal for nonparametric estimation. We used an improved version of the Akaike information criterion (AIC) proposed by Hurvich, Simonoff, and Tsai (1998), which has been found to have excellent small sample

https://doi.org/10.1017/S0266466624000100 Published online by Cambridge University Press

TABLE 1. Empirical rejection percentages under H_0 at 10% and 5% nominal levels.

	$h = C\widehat{\sigma}_W n^{-1/5}$				Data-driven h	
	$C = 0.5$	$C = 1$	$C = 1.5$	$C = 2$	Trimming	No trimming
$n = 200$						
Bias corrected	12.15	11.13	10.35	9.83	11.34	11.43
	6.86	5.64	5.56	5.08	5.74	5.79
Locally robust	12.08	11.05	10.39	9.71	11.48	11.27
	6.77	5.67	5.38	5.36	5.77	5.67
ICM	12.27	11.09	11.51	13.51	11.49	11.31
	6.65	5.93	6.02	7.18	6.05	6.13
$n = 400$						
Bias corrected	11.41	10.74	10.31	9.43	10.70	10.64
	5.90	5.60	5.33	4.99	5.73	5.79
Locally robust	11.32	10.73	10.17	9.70	10.81	10.24
	5.91	5.64	5.21	4.91	5.58	5.75
ICM	11.13	10.19	10.93	13.24	10.46	10.42
	6.03	5.69	5.71	6.69	5.85	5.80

performances by Li and Racine (2004). The criterion is

$$AIC_c = \log\left(\frac{1}{n}\sum_{i=1}^{n}[Y_i - \widehat{m}(W_i)]^2\right) + \frac{1 + tr(H)/n}{1 - [tr(H) + 2]/n},$$

where H is the "hat matrix" such that $(\widehat{m}(W_1), \ldots, \widehat{m}(W_n))' = H\,(Y_1, \ldots, Y_n)'$. As seen from Table 1, with this data-driven bandwidth, the three tests have good size control. We also report errors in rejection probability in Figure 2, which shows the excellent adequacy of the bootstrap approximation. We then repeated our experiment without any trimming and observed that ERP was closer to zero for our tests compared to the case where trimming is implemented.

Finally, for a data-driven bandwidth and no trimming, we report in Figure 3 the power of the three tests when γ varies. Our two tests basically have the same power, and power increases with sample size, while the uncorrected ICM test is significantly less powerful.

6.2. Empirical Illustration

We apply our encompassing tests to competing models of consumption behavior, in the spirit of Gaver and Geisel (1974) and Greene (2003, Chap. 8). The first model assumes that consumption only depends on income, and relates consumption to current and past income. Formally, it writes

$$C_i = g_1(I_i, PI_i) + \varepsilon_i \qquad \mathrm{E}[\varepsilon_i | I_i, PI_i] = 0, \tag{9}$$

https://doi.org/10.1017/S0266466624000100 Published online by Cambridge University Press

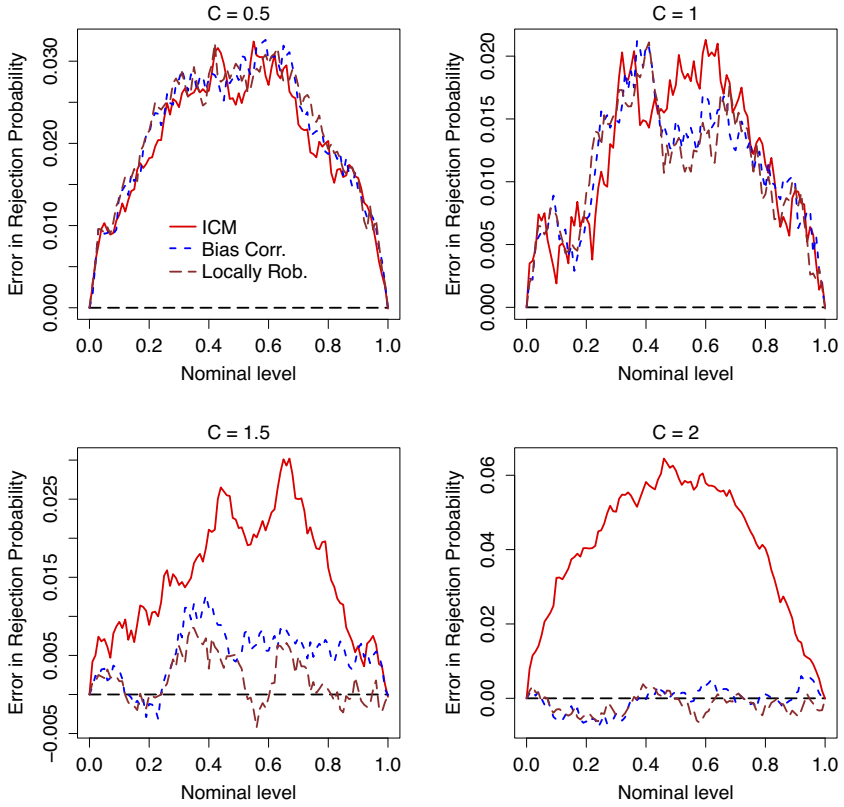


FIGURE 1. Errors in rejection probabilities for $n = 400$ using a rule-of-thumb bandwidth $h = C\widehat{\sigma}_W n^{-1/5}$.

where C_i denotes (log of) consumption of household i, I_i denotes (log of) income of household i, and PI_i denotes (log of) past income of household i. The second model assumes the presence of habits in consumption behavior, and relates consumption to current income and past consumption. It writes

$$C_i = g_2(I_i, PC_i) + \eta_i \qquad \mathrm{E}[\eta_i | I_i, PC_i] = 0, \tag{10}$$

where PC_i denotes (log of) past consumption of household i.

We used data compiled by Arellano, Blundell, and Bonhomme (2017) from the Panel Study of Income Dynamics, which concerns $n = 792$ households. We focus on the first two time periods, namely the years 1999 and 2001, so that C_i and I_i correspond to variables in 2001, while PC_i and PI_i correspond to 1999. We first tested if Model (9) encompasses Model (10), that is,

$$H_0 : \mathrm{E}[g_1(I_i, PI_i) \,|\, I_i, PC_i] = g_2(I_i, PC_i)\,. \tag{11}$$

https://doi.org/10.1017/S0266466624000100 Published online by Cambridge University Press

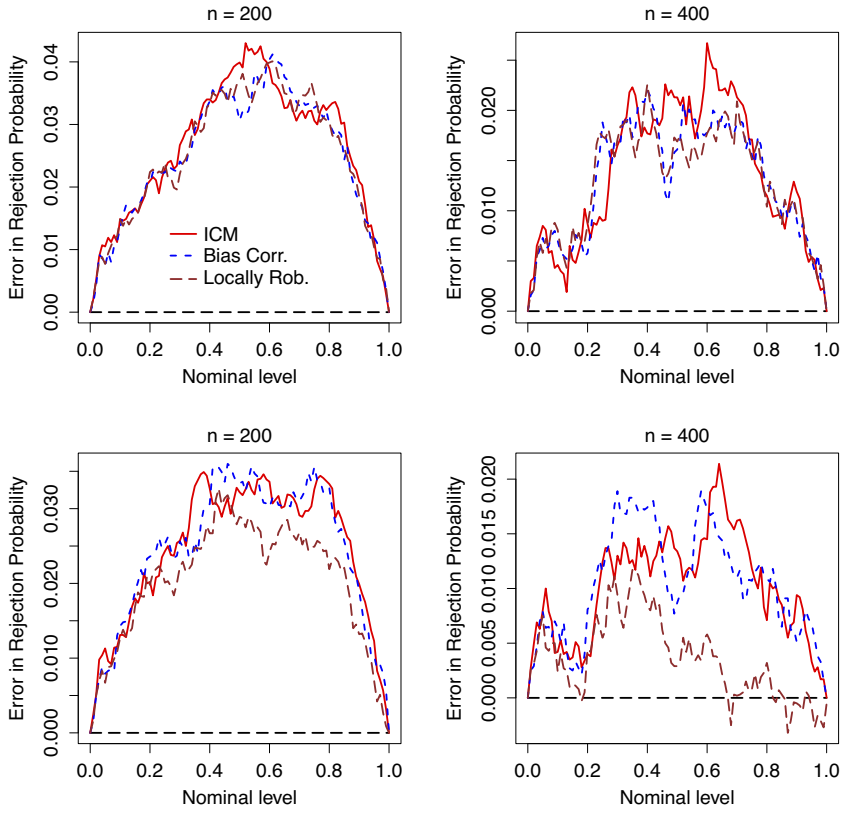


FIGURE 2. Errors in rejection probabilities using a data-driven bandwidth: results with trimming on first row, results with no trimming on second row.

We then tested the reversed hypothesis that Model (10) encompasses Model (9), that is,

$$H_0' : \mathrm{E}[g_2(I_i, PC_i)|I_i, PI_i] = g_1(I_i, PI_i)\,. \tag{12}$$

We applied our encompassing tests with the bandwidth selected by the AIC_c criterion and 999 bootstrap samples drawn to obtain critical values. Table 2 reports the values of our test statistics, together with their 95% bootstrap quantiles and bootstrap p-values. The first encompassing hypothesis H_0 is clearly rejected, hence Model (9) does not encompass Model (10). The encompassing hypothesis H_0', however, cannot be rejected by any of our tests, even at a 10% nominal level. This suggests that consumption may be adequately modeled as a function of income and past consumption, in a way coherent with the consumption habit formation theory.

https://doi.org/10.1017/S0266466624000100 Published online by Cambridge University Press

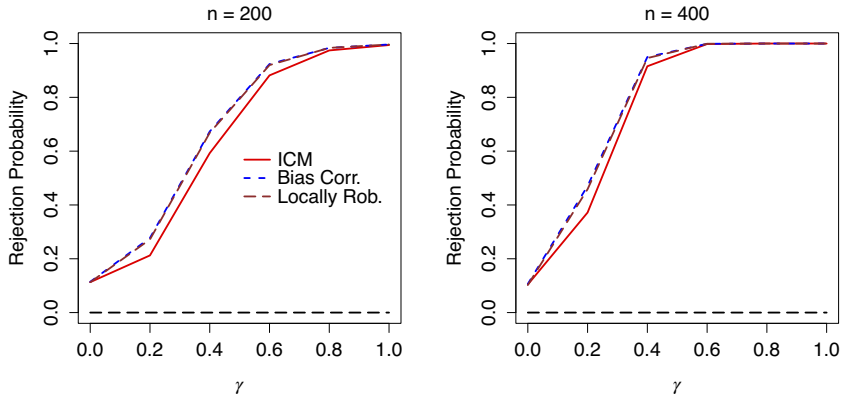


FIGURE 3. Power curves for tests at 10% level using a data-driven bandwidth and no trimming.

TABLE 2. Test of the encompassing hypotheses H_0 in (11) and H_0' in (12).

	Statistics	5% critical value	P-value
Test of H_0			
Bias corrected	3.0848	0.1820	0.0000
Locally robust	3.1125	0.1968	0.0000
ICM	3.3948	0.2499	0.0000
Test of H_0'			
Bias corrected	0.0069	0.0294	0.8799
Locally robust	0.0355	0.0447	0.1912
ICM	0.0249	0.0539	0.5485

7. CONCLUDING REMARKS

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 simulation experiment seems to indicate that when coupled with a data-driven bandwidth, these two approaches are robust to trimming choices and deliver good performances.

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 $t(w) = \mathbf{1}(f(w) \geq \tau_n)$. By convention, $(\overline{Y}/\widehat{f})(w)$ will

https://doi.org/10.1017/S0266466624000100 Published online by Cambridge University Press

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.

Our main proofs rely on auxiliary lemmas stated in Section 8.3.

8.1. Proof of Proposition 4.1

(a) We study the bias corrected empirical process

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

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

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

by Lemma 8.2(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 8.2(ii),

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

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 8.2(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) and (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 $\mathcal{W}_n := \{w : f(w)\ge\tau_n/2\}$. Lemma 8.4(i) yields

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

$$P(\widehat{B}t\phi_s) = \int \widehat{B}(w)t(w)\iota_s(w)f(w)\,dw = P(\widehat{B}t\iota_s)\,. \tag{15}$$

Since $\|[\widehat{B} - (\overline{\widehat{m}}-\overline{Y})/f]t\|_\infty = o_p(n^{-1/2})$ from Lemma 8.2(iv) and ι_s is uniformly bounded,

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

https://doi.org/10.1017/S0266466624000100 Published online by Cambridge University Press

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

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

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 and 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 $\mathbf{1}\{f(w+uh) < \tau_n\} \le \mathbf{1}\{f(w) \le 3\tau_n/2\}$. Now use $\mathbb{P}_n\mathbf{1}\left\{f(W) \le 3\tau_n/2\right\} = O_P(p_n) = o_P(n^{-1/2})$, 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 8.2(ii) and that $m(\widehat{t}-1)$ is bounded. Then

$$\begin{aligned}\sqrt{n}P(\widehat{B}t\iota_s) = &-\sqrt{n}\mathbb{P}_n\left[\varepsilon\int K(u)\iota_s(W+uh)\,du\right]\\ &+\sqrt{n}\mathbb{P}_n\left[(\widehat{m}-m)\widehat{t}\int K(u)\iota_s(W+uh)\,du\right]\\ &+\sqrt{n}\mathbb{P}_n\left[m(\widehat{t}-1)\int K(u)\iota_s(W+uh)\,du\right] + o_p^{\mathcal{S}}(1). \qquad \textbf{(16)}\end{aligned}$$

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 $P\left(\varepsilon\int K(u)\iota_s(W+uh)\,du\right) = P(\varepsilon\iota_s) = 0$. A Taylor expansion of order r guarantees that $\int K(u)\iota_s(w+uh)du = \iota_s(w) + O(h^r)$ uniformly in $(s,w) \in \mathcal{S}\times\mathcal{W}$ 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 (16) 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 8.2(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,

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

https://doi.org/10.1017/S0266466624000100 Published online by Cambridge University Press

From arguments similar to those used for (14), 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 $\mathrm{E}(\phi_s(X)|W) = \iota_s(W)$. 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

$$\begin{aligned}\sqrt{n}\mathbb{P}_n(\widehat{\varepsilon}(\phi_s-\widehat{\iota}_s)\widehat{t}) &= \sqrt{n}\mathbb{P}_n\left((Y-\widehat{m})(\phi_s-\widehat{\iota}_s)\widehat{t}\right)\\ &= \sqrt{n}\mathbb{P}_n(\varepsilon(\phi_s-\widehat{\iota}_s)\widehat{t}) + \sqrt{n}\mathbb{P}_n\left((m-\widehat{m})(\phi_s-\widehat{\iota}_s)\widehat{t}\right). \qquad \textbf{(17)}\end{aligned}$$

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 ε, $\sqrt{n}\mathbb{P}_n|\widehat{t}-t| = o_p(1)$, 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 8.2. From Lemma 8.2, $\sup_s\|(\widehat{\iota}_s-\iota_s)t\|_\infty = o_p(n^{-1/4})$ 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 $\sup_s\left|\sqrt{n}\mathbb{P}_n(\varepsilon(\phi_s-\iota_s)(t-1))\right| \le C\sqrt{n}\mathbb{P}_n|t-1| = o_P(1)$ by Lemma 8.2(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 5.1

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

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

From here, we only stress the differences with the proof of Proposition 4.1. Proceeding as in the latter proof and using results in Lemma 8.2,

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

https://doi.org/10.1017/S0266466624000100 Published online by Cambridge University Press

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

$$\sqrt{n}\mathbb{P}_n(\widehat{B}^* t\phi_s) = \sqrt{n}P(\widehat{B}^* t\phi_s) + o_P^{\mathcal{S}}(1). \tag{18}$$

From the law of iterated expectations, $P(\widehat{B}^* t\phi_s) = P(\widehat{B}^* t\iota_s)$. Since $\|[\widehat{B}^* - (\widehat{\overline{m}^*} - \overline{Y^*})/f]t\|_\infty = o_P(n^{-1/2})$ from Lemma 8.2(viii),

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

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 4.1 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,

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

A reasoning similar to Proposition 4.1 then yields $\sqrt{n}\mathbb{P}_n(\xi\varepsilon(\phi_s - \iota_s))t = \sqrt{n}\mathbb{P}_n(\xi\varepsilon(\phi_s - \iota_s)) + o_P^{\mathcal{S}}(1)$, $\sqrt{n}\mathbb{P}_n(\xi(m - \widehat{m})(\phi_s - \iota_s)t) = o_P^{\mathcal{S}}(1)$, 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 Cramér–von Mises functional and Proposition 4.1, 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, Thm. 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.

https://doi.org/10.1017/S0266466624000100 Published online by Cambridge University Press

(c) From Proposition 4.1 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 $S_n^{BC}/n \xrightarrow{p} \int |\mathrm{E}\,\varepsilon(\phi_s-\iota_s)|^2 d\mu(s)$. From Bierens (2017, Thm. 2.2), under H_1, $\mathrm{E}\,\varepsilon(\phi_s-\iota_s) \neq 0$ for almost all s and $\int |\mathrm{E}\,\varepsilon(\phi_s-\iota_s)|^2 d\mu(s) > 0$. From (b), it holds that the bootstrap statistics are bounded in probability, and the result follows.

8.3. Auxiliary Lemmas

LEMMA 8.1. *Let K be as in Assumption B(iii). 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 $\mathcal{A} \subset \mathbb{R}^p$*

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

The same result holds when $K(\cdot)$ is replaced by $|K(\cdot)|$, $\partial^l K(\cdot)$, or $|\partial^l K(\cdot)|$ with $|l| \leq \lceil (p+1)/2 \rceil$.

Proof. Follows from minor changes in the proof of Li and Racine (2006, Thm. 1.4). □

LEMMA 8.2. *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) $\|[\widehat{B}-(\widehat{\overline{m}}-\overline{Y})/f]t\|_\infty = o_p(n^{-1/2})$,
(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) $\|[\widehat{B}^*-(\widehat{\overline{m}^*}-\overline{Y^*})/f]t\|_\infty = o_P(n^{-1/2})$.

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

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

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 8.1, $\|\widehat{f}-\mathrm{E}\widehat{f}\|_\infty = O_P(\sqrt{(\log n)/(nh^p)})$. Standard bias manipulations ensure that $\|\mathrm{E}\widehat{f}-f\|_\infty = O(h^r)$. 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

https://doi.org/10.1017/S0266466624000100 Published online by Cambridge University Press

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,

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

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

$$|\widehat{t}(w) - t(w)| \leq \mathbf{1}\Big(f(w) < 3\tau_n/2\Big) \text{ for all } w \in \mathcal{W}$$

and

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

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,

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

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

$$\text{for each } \delta \in (0, 1] \text{ wpa1} : \mathbf{1}\Big(f(w) \geq \delta\tau_n\Big) \leq \mathbf{1}\Big(\widehat{f}(w) \geq \delta\tau_n/2\Big) \text{ for all } w \in \mathcal{W}, \tag{19}$$

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

$$\text{for each } \delta \in (0, 1] \text{ wpa1} : \mathbf{1}\Big(\widehat{f}(w) \geq \delta\tau_n\Big) \leq \mathbf{1}\Big(f(w) \geq \delta\tau_n/2\Big) \text{ for all } w \in \mathcal{W}. \tag{20}$$

Now, for any fixed $\delta \in (0, 1]$ (19) implies that with probability approaching one

$$\begin{aligned}(\widehat{m} - m)\ \mathbf{1}(f(\cdot) \geq \delta\tau_n) = {} & \frac{\overline{Y} - m\widehat{f}}{f}\ \mathbf{1}(f(\cdot) \geq \delta\tau_n) \\ & + \frac{\overline{Y} - m\widehat{f}}{f}\,\frac{f - \widehat{f}}{\widehat{f}}\ \mathbf{1}(f(\cdot) \geq \delta\tau_n)\ \mathbf{1}(\widehat{f}(\cdot) \geq \delta\tau_n/2).\end{aligned}$$

To bound the RHS, from Lemma 8.1, $\|\overline{Y} - \mathrm{E}\,\overline{Y}\|_\infty = O_P(\sqrt{(\log n)/(nh^p)})$, while standard bias computations yield $\|\mathrm{E}\,\overline{Y} - mf\|_\infty = O(h^r)$. Hence,

https://doi.org/10.1017/S0266466624000100 Published online by Cambridge University Press

$\|\overline{Y} - mf\|_\infty = O_P(d_n)$. Similarly, $\|\widehat{f} - f\|_\infty = O_P(d_n)$. Using these rates and the above display gives

$$\|(\widehat{m} - m)\ \mathbf{1}(f(\cdot) \geq \delta\tau_n)\|_\infty = O_P\left(\frac{d_n}{\tau_n} + \frac{d_n^2}{\tau_n^2}\right) = o_P(n^{-1/4}), \tag{21}$$

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 (20) with $\delta = 1$ gives that with probability approaching one $\|(\widehat{m} - m)\widehat{t}\|_\infty \leq \|(\widehat{m} - m)\ \mathbf{1}(f(\cdot) \geq \tau_n/2)\|_\infty$. Applying (21) with $\delta = 1/2$ gives the $n^{-1/4}$ rate for the RHS of the latter inequality.

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

$$\widehat{\overline{m}}(w) = \overline{m}(w) + \overline{m(\widehat{t} - 1)}(w) + \overline{(\widehat{m} - m)\widehat{t}}(w). \tag{22}$$

Combining Lemma 8.1 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, $\|\overline{m(\widehat{t} - 1)}\|_\infty = O_P(p_n h^{-p})$. For the third term on the right-hand side of (22)

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

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 8.1 and standard bias computations yield that the last term on the right-hand side is $O_P(1)$ uniformly in $w \in \mathcal{W}$. Gathering results,

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

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

$$\left\|\left(\frac{\widehat{\overline{m}}}{\widehat{f}} - m\right)\widehat{t}\right\|_\infty = O_P\left(\frac{d_n}{\tau_n^2} + \frac{p_n}{h^p\tau_n}\right) \text{ and } \left\|\left(\frac{\widehat{\overline{m}}}{\widehat{f}} - m\right)t\right\|_\infty = O_P\left(\frac{d_n}{\tau_n^2} + \frac{p_n}{h^p\tau_n}\right). \tag{24}$$

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

(*iv*) Since $\widehat{B} = \widehat{\overline{m}}/\widehat{f} - \overline{Y}/\widehat{f}$, Equation (19) implies that with probability approaching one

$$\left(\widehat{B} - \frac{\widehat{\overline{m}} - \overline{Y}}{f}\right) t = (\widehat{\overline{m}} - \overline{Y})\left(\frac{f - \widehat{f}}{\widehat{f}f}\right) t\ \mathbf{1}(\widehat{f}(\cdot) \geq \tau_n/2).$$

https://doi.org/10.1017/S0266466624000100 Published online by Cambridge University Press

Using (23), $\|\overline{Y} - mf\|_\infty = O_P(d_n)$, 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(ii) and D(i).

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

(*vi*) First, notice that

$$\widehat{m}^* = \frac{\overline{Y^*}}{\widehat{f}} = \frac{\overline{\widehat{m}}}{\widehat{f}} + \frac{\overline{\xi\varepsilon}}{\widehat{f}}. \tag{25}$$

The uniform convergence rate of $\overline{\widehat{m}}/\widehat{f}$ has already been obtained in (24), so it suffices to show that the second addendum is negligible with a suitable rate. Now, $\overline{\xi\varepsilon} = \overline{\xi}\,\overline{\varepsilon} + \overline{\xi\varepsilon(\widehat{t} - 1)} + \overline{\xi(m - \widehat{m})\widehat{t}}$. Using this decomposition and proceeding as in the proof of (*iii*) gives

$$\left\| \frac{\overline{\xi\varepsilon}}{\widehat{f}} \widehat{t} \right\|_\infty = O_P\left(\frac{d_n}{\tau_n^2} + \frac{p_n}{h^p \tau_n}\right) \text{ and } \left\| \frac{\overline{\xi\varepsilon}}{\widehat{f}} t \right\|_\infty = O_P\left(\frac{d_n}{\tau_n^2} + \frac{p_n}{h^p \tau_n}\right) \tag{26}$$

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 (*vi*), it suffices to obtain a suitable convergence rate for the first addendum. To this end, from (24) to (26), we have

$$\|(\widehat{m}^* - m)\widehat{t}\|_\infty = O_P\left(\frac{d_n}{\tau_n^2} + \frac{p_n}{h^p \tau_n}\right). \tag{27}$$

Also, $\overline{\widehat{m}^*} = \overline{m} + \overline{m(\widehat{t} - 1)} + \overline{(\widehat{m}^* - m)\widehat{t}}$. Using this decomposition, (27), and proceeding similarly as in the proof of (*iii*) gives

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

By Assumptions C(ii) 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 (19), with probability approaching one

$$\left(\widehat{B}^* - \frac{\overline{\widehat{m}^*} - \overline{Y^*}}{f}\right) t = (\overline{\widehat{m}^*} - \overline{Y^*})\left(\frac{f - \widehat{f}}{f\widehat{f}}\right) t\, \mathbf{1}(\widehat{f}(\cdot) \geq \tau_n/2). \tag{28}$$

https://doi.org/10.1017/S0266466624000100 Published online by Cambridge University Press

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, (27), and reasoning as in the proof of (*iii*), we get

$$\|\overline{\widehat{m}^*} - mf\|_\infty = O_P\Big(\frac{d_n}{\tau_n^2} + \frac{p_n}{h^p \tau_n}\Big). \tag{29}$$

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 (23), we get

$$\|\overline{Y^*} - mf\|_\infty = O_P\Big(\frac{d_n}{\tau_n} + \frac{p_n}{h^p}\Big). \tag{30}$$

Finally, plugging $\|\widehat{f} - f\|_\infty = O_P(d_n)$, (29), and (30) into (28) and then using Assumptions C(ii) and D(i) gives the desired result. □

The next lemma provides the regularity features needed to apply stochastic equicontinuity results. Similar results can be found in Andrews (1994, 1995). The differences with respect to these works are the presence of the bias correction components and the different assumptions on the bandwidths. Recall that $\mathcal{W}_n = \{w : f(w) \geq \tau_n/2\}$.

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) $\Pr\Big(\overline{\widehat{m}}/\widehat{f} \in \mathcal{G}_l(\mathcal{W}_n)\Big) \to 1$,
(iii) $\Pr\Big(\widehat{\iota}_s \in \mathcal{G}_l(\mathcal{W}_n)$ *for all* $s \in \mathcal{S}\Big) \to 1$,
(iv) $\Pr\Big(\overline{\xi\widehat{\varepsilon}}/\widehat{f} \in \mathcal{G}_l(\mathcal{W}_n)\Big) \to 1$,
(v) $\Pr\Big(\overline{\widehat{m}^*}/\widehat{f} \in \mathcal{G}_l(\mathcal{W}_n)\Big) \to 1$.

Proof. The proof can be found in the Supplementary Material. □

The following lemma provides a stochastic equicontinuity result. Similar results can be found in Andrews (1994, 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 ϵ covering number and with $N_{[\cdot]}(\epsilon, \mathcal{F}, ||\cdot||_{2,P})$ the ϵ 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* $\mathrm{Supp}(\zeta)$ *the support of* $\zeta := (\xi, Y, W, X)$.

https://doi.org/10.1017/S0266466624000100 Published online by Cambridge University Press

(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, for* $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P)$,

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

The same result also holds when ϕ_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 $\{\widehat{f}_s : s \in \mathcal{S}\}$ ensure that with probability approaching one

$$\sup_{s\in\mathcal{S}} \left| \mathbb{G}_n t \widehat{f}_s \phi_s \right| \le \sup_{g\in\mathcal{G}_n^\epsilon} |\mathbb{G}_n g| \tag{31}$$

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, Lem. 19.34), if

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

then the expectation of the right-hand side of (31) can be made arbitrarily small asymptotically by choosing ϵ small enough. So, Markov's inequality would deliver the desired result.

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

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

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 ϕ_s is Lipschitz on the compact $\mathcal{S}$, $N(\epsilon, \Psi, \|\cdot\|_\infty) \le C\epsilon^{-q}$ by Kosorok (2008, Thm. 9.15). So, using the fact that t is bounded,

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

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

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

https://doi.org/10.1017/S0266466624000100 Published online by Cambridge University Press

SUPPLEMENTARY MATERIAL

Lapenta, E., and Lavergne, P. (2024): Supplement to "Encompassing Tests for Nonparametric Regressions," *Econometric Theory Supplementary Material*. To view, please visit https://doi.org/10.1017/S0266466624000100.

REFERENCES

Ait-Sahalia, Y., Bickel, P. J., & Stoker, T. M. (2001). Goodness-of-fit tests for kernel regression with an application to option implied volatilities. *Journal of Econometrics*, 105, 363–412.

Andrews, D. W. K. (1994). Empirical process methods in econometrics. In R.F. Engle and D.L. McFadden (eds), *Handbook of Econometrics* (pp. 2247–2294), vol. 4. Elsevier.

Andrews, D. W. K. (1995). Nonparametric kernel estimation for semiparametric models. *Econometric Theory*, 11, 560–586.

Arellano, M., Blundell, R., & Bonhomme, S. (2017). Earnings and consumption dynamics: A nonlinear panel data framework. *Econometrica*, 85, 693–734.

Bentkus, V., Götze, F., & Zitikis, R. (1993). Asymptotic expansions in the integral and local limit theorems in Banach spaces with applications to *w*-statistics. *Journal of Theoretical Probability*, 6, 54.

Bierens, H. J. (1982). Consistent model specification tests. *Journal of Econometrics*, 20, 105–134.

Bierens, H. J. (1990). A consistent conditional moment test of functional form. *Econometrica*, 58, 1443–1458.

Bierens, H. J. (2017). *Econometric model specification*. World Scientific.

Bierens, H. J., & Ploberger, W. (1997). Asymptotic theory of integrated conditional moment tests. *Econometrica*, 65, 1129–1152.

Bontemps, C., Florens, J.-P., & Richard, J.-F. (2008). Parametric and non-parametric encompassing procedures. *Oxford Bulletin of Economics and Statistics*, 70, 751–780.

Bontemps, C., & Mizon, G. E. (2008). Encompassing: Concepts and implementation. *Oxford Bulletin of Economics and Statistics*, 70, 721–750.

Chernozhukov, V., Escanciano, J. C., Ichimura, H., Newey, W. K., & Robins, J. M. (2022). Locally robust semiparametric estimation. *Econometrica*, 90, 1501–1535.

Davidson, R., & MacKinnon, J. G. (2007). Improving the reliability of bootstrap tests with the fast double bootstrap. *Computational Statistics & Data Analysis*, 51, 3259–3281.

Delgado, M. A., & Manteiga, W. G. (2001). Significance testing in nonparametric regression based on the bootstrap. *Annals of Statistics*, 29, 1469–1507.

Delgado, M. A., & Stute, W. (2008). Distribution-free specification tests of conditional models. *Journal of Econometrics*, 143, 37–55.

Dhaene, G., Gourieroux, C., & Scaillet, O. (1998). Instrumental models and indirect encompassing. *Econometrica*, 66, 673–688.

Di Marzio, M., & Taylor, C. C. (2008). On boosting kernel regression. *Journal of Statistical Planning and Inference*, 138, 2483–2498.

Escanciano, J. C. (2006). A consistent diagnostic test for regression models using projections. *Econometric Theory*, 22, 1030–1051.

Escanciano, J. C., Jacho-Chávez, D. T., & Lewbel, A. (2014). Uniform convergence of weighted sums of non and semiparametric residuals for estimation and testing. *Journal of Econometrics*, 178, 426–443.

Fan, Y., & Li, Q. (1996). Consistent model specification tests: Omitted variables and semiparametric functional forms. *Econometrica*, 64, 865–890.

Florens, J.-P., Hendry, D. F., & Richard, J.-F. (1996). Encompassing and specificity. *Econometric Theory*, 12, 620–656.

Gaver, K. M., & Geisel, M. S. (1974). Discriminating among alternative models: Bayesian and non-Bayesian methods. In Zarembka, P. (Ed.), *Frontiers of econometrics* (p. 80), 19. Academic Press.

https://doi.org/10.1017/S0266466624000100 Published online by Cambridge University Press

Giacomini, R., Politis, D. N., & White, H. (2013). A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators. *Econometric Theory*, 29, 567–589.

Gourieroux, C., & Monfort, A. (1995). Testing, encompassing, and simulating dynamic econometric models. *Econometric Theory*, 11, 195–228.

Gourieroux, C., Monfort, A., & Trognon, A. (1983). Testing nested or non-nested hypotheses. *Journal of Econometrics*, 21, 83–115.

Greene, W. H. (2003). *Econometric analysis*. (5th ed.) Prentice Hall.

Hendry, D. F., & Richard, J.-F. (1982). On the formulation of empirical models in dynamic econometrics. *Journal of Econometrics*, 20, 3–33.

Hurvich, C. M., Simonoff, J. S., & Tsai, C.-L. (1998). Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60, 271–293.

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

Lavergne, P. (2001). An equality test across nonparametric regressions. *Journal of Econometrics*, 103, 307–344.

Lavergne, P., Maistre, S., & Patilea, V. (2015). A significance test for covariates in nonparametric regression. *Electronic Journal of Statistics*, 9, 643–678.

Lavergne, P., & Patilea, V. (2008). Breaking the curse of dimensionality in nonparametric testing. *Journal of Econometrics*, 143, 103–122.

Lavergne, P., & Vuong, Q. H. (1996). Nonparametric selection of regressors: The nonnested case. *Econometrica*, 64, 207–219.

Lavergne, P., & Vuong, Q. H. (2000). Nonparametric significance testing. *Econometric Theory*, 16, 576–601.

Li, Q., & Racine, J. (2004). Cross-validated local linear nonparametric regression. *Statistica Sinica*, 14, 485–512.

Li, Q., & Racine, J. S. (2006). *Nonparametric econometrics: Theory and practice*. Princeton University Press.

Liao, Z., & Shi, X. (2020). A nondegenerate Vuong test and post selection confidence intervals for semi/nonparametric models. *Quantitative Economics*, 11, 983–1017.

Mammen, E. (1992). *When does bootstrap work?*. Lecture Notes in Statistics. Springer.

Mammen, E., Rothe, C., & Schienle, M. (2016). Semiparametric estimation with generated covariates. *Econometric Theory*, 32, 1140–1177.

Mizon, G. E., & Richard, J.-F. (1986). The encompassing principle and its application to testing nonnested hypotheses. *Econometrica*, 54, 657–678.

Newey, W. K. (1990). Semiparametric efficiency bounds. *Journal of Applied Econometrics*, 5, 99–135.

Newey, W. K., Hsieh, F., & Robins, J. M. (2004). Twicing kernels and a small bias property of semiparametric estimators. *Econometrica*, 72, 947–962.

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

Stinchcombe, M. B., & White, H. (1998). Consistent specification testing with nuisance parameters present only under the alternative. *Econometric Theory*, 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., & Wellner, J. A. (2000). *Weak convergence and empirical processes: With applications to statistics*. Springer.

Xia, Y., Li, W. K., Tong, H., & Zhang, D. (2004). A goodness-of-fit test for single-index models. *Statistica Sinica*, 14, 1–28.

https://doi.org/10.1017/S0266466624000100 Published online by Cambridge University Press