Semiparametric Efficient Tests

Juan Carlos Escanciano*
Indiana University

July 24th, 2012

Abstract

This paper proposes efficient tests for restrictions on finite-dimensional parameters in regular semiparametric models. Our theory overcomes the main limitation of the existing theory, which requires explicit computation and estimation of certain projections onto infinite-dimensional tangent spaces and a case-by-case analysis. We consider generic semiparametric models defined by an infinite number of moment conditions, including a finite-dimensional parameter of interest and possibly containing moment-specific nonparametric nuisance parameters. We investigate tests based on functionals of the sample analog of the moments, and show that the optimal functional takes the form of a Radon-Nikodym derivative or nonparametric Likelihood Ratio (LR). We first show that the resulting LR test is efficient in our general semiparametric setting. The LR is generally infeasible, as it assumes knowledge of a certain spectrum. We then propose and justify feasible efficient tests based on a novel nonparametric estimator of the so-called efficient score, without requiring direct computation of projections onto tangent spaces or sample splitting techniques. Thus, the proposed efficient tests are widely applicable, while being straightforward to implement. Finally, to illustrate the benefits of the approach, we apply the new methods to a semiparametric linear quantile regression model with a continuum of quantiles. Optimal inferences in this model were not available because classical efficiency arguments are difficult to apply. In contrast, our methods deliver relatively simple efficient tests in this example.

Keywords: Neyman-Pearson lemma; Likelihood ratio; Semiparametric efficiency; Efficient score; Quantile Regression; Empirical processes theory.

JEL classification: C12, C14.

*Department of Economics, Indiana University, 349 Wylie Hall, 100 South Woodlawn Avenue, Bloomington, IN 47405–7104, USA. E-mail: jescanci@indiana.edu. Web Page: http://mypage.iu.edu/~jescanci/. Part of this paper was circulated under the title: "On the Efficiency of the Functional Neyman-Pearson Test in Semiparametric Models". I would like to thank helpful comments from Manuel Arellano, Xiaohong Chen, Miguel A. Delgado, Ulf Grenander, Roger Koenker, Ignacio Lobato, Ulrich Müller and Carlos Velasco. Research funded by the Spanish Plan Nacional de I+D+I, reference number SEJ2007-62908.

1 Introduction

Semiparametric models are a compromise between tight parametric specifications and more flexible, but rather imprecise, nonparametric models. Nowadays, these models are widely used in applications in the social sciences, and a well developed theory of estimation and testing has been established for many classes of semiparametric models; see e.g. Robinson (1988), Powell (1994) and Wellner, Klaasen and Ritov (2006) for surveys on the topic, and Bickel, Klaasen, Ritov, and Wellner (1993) for a comprehensive treatment. Unfortunately, although there has been significant progress in the formalization of an efficiency theory in this setting, this theory is far from complete.[1] Specifically, feasible optimal procedures for inferences about a finite-dimensional parameter of interest need to account for the potential information loss derived from not knowing nuisance parameters, see Neyman (1959) for the formulation of the problem in fully parametric models. In semiparametric models this loss is quantified by an orthogonal projection of the score of interest onto an infinite-dimensional space, the tangent space of nuisance parameters, leading to the so-called *efficient score*, and its computation and estimation often requires complicated arguments and a case-by-case analysis. Moreover, there are many examples for which these projections, and hence optimal inferences, are unknown; see some examples below. This paper provides a general and unified theory of efficient tests in semiparametric models, which does not require direct computation of projections and is, therefore, simple to implement.

Semiparametric efficient inference is extensively discussed in econometrics and statistics. Newey (1990) provides an excellent introduction to the topic. The bulk of the literature focuses on the estimation theory, using the concepts of parametric submodels and tangent spaces. Efficient tests for restrictions on a finite-dimensional parameter in regular semiparametric models are formally defined in Choi, Hall and Schick (1996, henceforth CHS); see also Section 3 and Appendix A for a review. Optimal tests are shown to be asymptotically equivalent to semiparametric versions of the $C(\alpha)-$test of Neyman (1959). CHS propose feasible efficient procedures, provided a suitable estimator of the efficient score is available. However, it is not known how such estimators are obtained for general semiparametric models. Broadly speaking, achieving feasible efficient inferences in semiparametric models is, to a large extent, equivalent to estimating the efficient score. We provide a generic efficient score's estimate in this paper.

Our approach to efficient testing is different from the standard one used in the literature. Rather than looking at least favorable parametric submodels, reducing the semiparametric problem to a parametric one, we deal directly with the semiparametric model and use functional versions of the Neyman-Pearson lemma. We consider generic models defined by an infinite number of moment conditions, depending on a finite-dimensional parameter of interest and possibly containing moment-specific nonparametric nuisance parameters. This framework is quite general, as any semiparametric model can be written in this form. Then, within this setting, suppose we are interested in testing restrictions on

[1] In their introduction, Bickel et al. (1993) write: "*What are general methods and techniques for constructing asymptotically efficient estimates for such (semiparametric) models? Although we have made some progress on this question, the overall picture is somewhat disappointing. There are a number of methods that heuristically should yield procedures with the properties we want. But which approach works best or can most easily be proved to work depends on the example or class of examples.*"

the finite-dimensional parameter. A large class of tests can be based on continuous functionals of the sample analog of the moments, say $\hat{R}_n$, evaluated at a restricted estimator. The subscript n in $\hat{R}_n$ denotes the sample size. Typical functionals, such as the Kolmogorov-Smirnov or Cramér-von-Mises functionals, are not optimal for this problem. We show below that the optimal functional is the Radon-Nikodym derivative of the limit distribution of $\sqrt{n}\hat{R}_n$ under local alternatives with respect to the limit distribution under the null. This functional is monotone in an asymptotically sufficient test statistic $L(\hat{R}_n)$, and we denote the test rejecting for "large" values of $L(\hat{R}_n)$ the *Functional Likelihood Ratio Test* (FLRT). This nonparametric Likelihood Ratio (LR) principle was first suggested by Grenander (1950) in a different context, and it has been already applied to some semiparametric settings, see Section 2 for a review of this literature. Yet, the semiparametric efficiency of the resulting inferences in these specific applications, or in more general settings like ours, remains completely unknown.

The first main contribution of this paper proves the efficiency of the FLRT in general semiparametric models. To that end, we first obtain a generic asymptotic representation of the test statistic $L(\hat{R}_n)$ as a score-type process (i.e. as a sample mean of a score function). We characterize the score function in terms of the limiting covariance and mean of $\sqrt{n}\hat{R}_n$. Then, we show that the resulting score coincides with the efficient score in the semiparametric model defined by the moment restrictions, which, by virtue of CHS's results, establishes the semiparametric efficiency of the FLRT. The functional $L(\cdot)$ is in general not feasible, and feasible implementations require the estimation of the spectrum of the limiting covariance operator of $\sqrt{n}\hat{R}_n$, which has hampered the practical applicability of this functional LR method. The problem involved is essentially equivalent to constructing feasible versions of the semiparametric $C(\alpha)-$test (i.e. estimating projections onto tangent spaces). Our second main contribution is the development of feasible implementations of the FLRT that do not require knowledge of the spectrum. We combine our characterization of the efficient score function, which is much like a *generalized information equality* formula, with well-known results from ill-posed problems to construct a novel nonparametric estimator of the efficient score. The proposed feasible FLRT uses the estimated score, and it is quite simple to compute. Thus, our test can be viewed as a semiparametric version of the celebrated Neyman's (1959) $C(\alpha)-$test. To illustrate the benefits of our implementation, we consider an example in quantile regression with a continuum of quantiles. In this example, standard methods to efficiency require rather complicated arguments and efficient inferences were unknown.

Our results complement alternative efficiency results recently obtained by Müller (2011). He has shown that the FLRT is also optimal in a class of tests that control asymptotic size for all data generating processes under which $\sqrt{n}\hat{R}_n$ satisfies a weak convergence requirement; see Section 3 for a more formal discussion. This efficiency concept can be potentially different from the "classical" semiparametric efficiency concept in CHS, and it provides a sense of robustness of the FLRT. An appealing property of Müller's efficiency concept is that it applies to regular and non-regular settings, whereas extensions of the classical semiparametric efficiency theory to non-regular problems are generally difficult. Hence, our results complement rather than substitute Müller's (2011) results, and together they imply a broad sense of optimality of the FLRT.

Our efficiency results are also related to the recent literature on efficient estimation of semiparametric models by Generalized Method of Moments (GMM) employing potentially infinite number of moments, see e.g. Ai and Chen (2003), Newey (1988, 2004) and Carrasco and Florens (2000, 2011). Our paper differs from these works in several aspects. Firstly, the GMM literature has been focused on estimation, with rather few results on testing available. Carrasco and Florens (2000) proposed tests based on the optimal GMM objective function in parametric moments, but their tests are not efficient in our setting. Secondly, we use a functional LR approach, as in e.g. Müller (2011). The LR approach has some additional benefits, such as allowing the researcher to compute, otherwise complicated, probabilities under the local alternatives via Lecam's third Lemma. See, for instance, the local power analysis carried out in Escanciano (2009). Nevertheless, we show below that our LR test is closely related to a Lagrange Multiplier (LM) test based on a modified optimal GMM objective function. The modification accounts for the presence and impact of nuisance parameters.[2] To the best of our knowledge, the connection between GMM and our LR approach is new and leads to mutual benefits for these two approaches. For instance, it implies that some modifications of GMM-based tests will share the optimality properties of our LR test, including Müller's (2011) optimality in non-regular problems. This connection also opens the door for new implementations of the GMM-based tests and estimators, which are not available in the general semiparametric setting discussed here.

Summarizing, this paper proposes a general and unified method to derive semiparametric feasible efficient tests using a functional LR principle. Although the bulk of our paper deals with the testing problem, we obtain some by-products pertaining to other aspects of inference. Most notably, we provide: (i) a new general formula for the efficiency bound of semiparametric estimation; (ii) optimal confidence sets by inverting our test statistics; and (iii) a candidate for a semiparametric efficient one-step estimator. These by-products are of independent interest.

The rest of the paper is organized as follows: Section 2 introduces notation, the semiparametric model, the testing problem and the FLRT. It then provides an asymptotic representation of the FLRT as a score-type test. Section 3 establishes the semiparametric efficiency of the procedure and connections with the GMM literature. Section 4 investigates the implementation of the FLRT. The new estimator for the efficient score is introduced here. Section 5 contains applications to quantile and mean regression, respectively, which illustrate the utility of our results. Other applications, such as to partially identified models, are briefly mentioned at the end of this section. Section 6 concludes with some final remarks. Appendix A provides some preliminary results and sufficient conditions for a uniform expansion that can be used to establish some of our assumptions in the text. Mathematical proofs of our results are gathered in Appendix B. Finally, Appendix C provides an algorithm for implementation of the feasible test. It is intended to facilitate the application of our methods to readers less interested in the technical background.

[2] Carrasco and Florens (2000, 2011) do not consider nuisance parameters and in Newey (2004) it is assumed that they do not affect the asymptotic variance of estimates; see Newey (2004, p. 1879). In this paper we allow for nuisance parameters to have an impact on the asymptotic variance, and that possibility complicates to a large extent our theory.

2 Setting and the FLRT

2.1 Notation

This section contains notation that will be used throughout the paper. Henceforth, A', $tr(A)$ and $|A| := (tr(A'A))^{1/2}$ denote the transpose, trace and the Euclidean norm of a matrix A, respectively. The symbol $:=$ denotes definitional relation. Let Γ be a set and let $\mu(\cdot)$ be a positive measure on Γ, with support identical to Γ. Let $L_2(\mu) \equiv L_2(\Gamma, \mu)$ be the Hilbert space of all real-valued functions f such that $\int_\Gamma |f(x)|^2 \mu(dx) < \infty$. If μ is a probability measure P with a cumulative distribution function (cdf) F, we also denote $L_2(F) := L_2(\mu)$ and $\|f\|_{2,P}^2 := \int f^2 dP$. As usual, equality of functions is understood almost surely with respect to μ. With some abuse of notation, for a p-dimensional function f, we write $f \in L_2(\mu)$ if all its components belong to $L_2(\mu)$ (similarly for other functional spaces). In $L_2(\mu)$ we define the inner product $\langle f, g \rangle := \int_\Gamma f(x)g(x)\mu(dx)$. As usual, $L_2(\mu)$ is endowed with the natural Borel σ-field induced by the norm $\|\cdot\| := \langle \cdot, \cdot \rangle^{1/2}$. Let $\Longrightarrow$ denote weak convergence in the Hilbert space $L_2(\mu)$; see e.g. Chapter 1.8 in van der Vaart and Wellner (1996). Unless otherwise stated, all limits are taken as $n \to \infty$. For a linear operator $K : L_2(\mu_1) \to L_2(\mu_2)$, denote the subspaces $Im(K) := \{f \in L_2(\mu_2) : \exists s \in L_2(\mu_1), Ks = f\}$ and $\ker(K) := \{f \in L_2(\mu_1) : Kf = 0\}$. Let $\mathcal{D}(K)$ denote the domain of definition of K. For a subspace $V \subset L_2(\mu)$, $V^\perp$ and $\overline{V}$ denote, respectively, its orthogonal complement and closure in $L_2(\mu)$. Henceforth, for a closed subspace V, Π_V denotes its orthogonal projection operator. We will extensively use basic results from operator theory and Hilbert spaces. See Carrasco, Florens and Renault (2006) for an excellent review of these results.

2.2 Semiparametric Model and Testing Problem

We describe now the model and our general testing problem, introducing the null hypothesis of interest and some further notation. Assume we observe a sample of size $n \geq 1$, $\{Z_i\}_{i=1}^n$, of independent and identically distributed (iid) random vectors in $\mathbb{R}^d$, distributed as Z, and satisfying the set of moment conditions

$$E[\psi(Z, x, \beta, \eta_0(Z, x))] = 0 \text{ for all } x \in \Gamma, \tag{1}$$

where $\beta \in \Theta_\beta \subset \mathbb{R}^p$ is a finite-dimensional parameter of interest, and $\eta_0(\cdot, x) \in \Theta_{\eta x}$ (of arbitrary dimension) is an unknown nuisance parameter for each $x \in \Gamma$. Without loss of generality (w.l.g), we take ψ to be real-valued. Although not explicit in the notation, we allow for $\eta_0(\cdot, x)$ to depend on β, i.e. $\eta_0(\cdot, x) \equiv \eta_0(\cdot, x, \beta)$. Set $\theta_0 := (\beta_0, \eta_0) \in \Theta := \Theta_\beta \times \Theta_\eta$, where β_0 is fixed and known and Θ_η denotes the parameter space for η_0. Let F denote the cdf of Z, with probability measure P. Unless otherwise stated, all expectations are with respect to F. Note that any semiparametric model can be written as (1).[3] In particular, our setting includes standard models such as semiparametric conditional moment restrictions, as well as less standard models with nuisance parameters that change with the moment, as in semiparametric quantile regression models. The following example illustrates this point.

[3] To see the generality, if Z follows the semiparametric cdf $F_{\beta,\eta}$, we can construct moments as $E[1(Z \leq x) - F_{\beta,\eta}(x)] = 0$ for all $x \in \mathbb{R}^d$. If the model is defined through conditional moment restrictions of the form $E[\rho(Z, \beta, \eta_0)| X] = 0$ a.s., where X is a subvector of Z of dimension d_x, then take $E[\rho(Z, \beta, \eta_0)1(X \leq x)] = 0$ for all $x \in \mathbb{R}^{d_x}$.

Example 1: **Linear Quantile Regression (QR) with a continuum of quantiles**. Consider the infinite number of moment restrictions

$$E[\{1(Y \leq \beta' X_1 + \eta_0(\tau)' X_2) - \tau\} 1(X \leq w)] = 0 \text{ for all } x = (\tau, w')' \in \mathcal{T} \times \mathbb{R}^{d_x}, \tag{2}$$

where $\mathcal{T}$ is a generic compact subset of $[0,1]$, $\mathcal{T} \subseteq [0,1]$, $X = (X_1', X_2')'$, $Z = (Y, X')'$, $d_x = d-1$, and $1(A)$ denotes the indicator function of the event A. Under some mild smoothness condition, these moments identify $\beta' X_1 + \eta_0(\tau)' X_2 \equiv X' \theta_0(\tau)$ as the conditional τth quantile of Y given X, for all $\tau \in \mathcal{T}$. This model includes as special case the classical pure-location regression model, with $X_2 \equiv 1$ and $\eta_0(\tau)$ the unknown (unconditional) error quantile function with $\mathcal{T} \equiv [0,1]$, or semiparametric extensions where the independence between errors and covariates only occurs at certain parts of the distribution defined by the set of quantiles $\mathcal{T}$. In this model the nuisance parameter η_0 varies with x (specifically with τ). Although our results are applicable to generalizations or variations of this model, such as location-scale models with unknown conditional scale or partially linear quantile regressions, we prefer to keep the exposition simple. The model in (2) is a special case of the classical linear quantile regression model of Koenker and Bassett (1978), which is extensively studied and applied in the literature. It is a model for which semiparametric efficient inference is unknown, beyond the special case of pure-location model or the case of a single quantile $\mathcal{T} = \{\tau_0\}$, see Komunjer and Vuong (2010) for the latter. As it turns out, standard efficiency theory is not easily applicable to this model when $\mathcal{T}$ includes an infinite number of quantiles, whereas our methods provide relatively simple procedures. This model is investigated in detail in Section 5. We note there that similar structures appear in semiparametric models that are partially identified, see Escanciano and Zhu (2012) and references therein. $\square$

We introduce our testing problem. We aim to find an asymptotically optimal procedure for testing

$$H_0 : \beta = \beta_0, \tag{3}$$

against the local alternatives

$$H_n : \beta_n = \beta_0 + n^{-1/2} c_\beta,$$

for some $c_\beta \in \mathbb{R}^p \backslash \{0\}$. The nuisance parameter η_0 is unknown under both, the null and the alternative hypotheses, and we assume that a consistent, but not necessarily efficient, estimator $\widehat{\eta}_n$ under the null is available. For a more formal definition of the local alternatives considered see Appendix A. In the main text we keep a simpler description for ease of exposition. Henceforth, to simplify the notation, we drop the dependence of $\widehat{\eta}_n$ on (Z_i, x) and write $\widehat{\eta}_n \equiv \widehat{\eta}_n(Z_i, x)$, and similarly for η_0. Reciprocally, when we want to emphasize the dependence on x we write $\theta_0(x) := (\beta_0, \eta_0(Z_i, x)) \in \Theta_x := \Theta_\beta \times \Theta_{\eta x}$.

2.3 Weak Convergence

Under our setting in (1), and given a random sample $\{Z_i\}_{i=1}^n$ and the hypothesis of interest H_0, it is natural to consider the sample analog of the moments with estimated parameters, i.e.

$$\hat{R}_n(x) := \frac{1}{n} \sum_{i=1}^n \psi(Z_i, x, \beta_0, \widehat{\eta}_n), \tag{4}$$

as a "sufficient" statistic for the testing problem. Tests can be based on continuous functionals of $\hat{R}_n$, such as the classical Kolmogorov-Smirnov test statistic $\sup_{x\in\Gamma}\left|\hat{R}_n(x)\right|$. See Bickel, Ritov and Stoker (2006) for a recent proposal in a general semiparametric setting. As we show below, typically used functionals are not optimal in our problem. In this paper, we propose optimal functionals.

The general discussion here is organized around a few "high-level" assumptions. More primitive conditions are shown in the Appendix and in the examples below. Our first "high-level" assumption requires the weak convergence of $\sqrt{n}\hat{R}_n$ in a suitable Hilbert space. Specifically, the process $\hat{R}_n$ is viewed here as a random element taking values in $L_2(\mu)$, for a suitable probability measure $\mu(\cdot)$ on Γ. For some discussion on the impact of $\mu(\cdot)$ on our theory see Remark 3 below.

Assumption W: *Under the local alternatives H_n,*

$$\sqrt{n}\hat{R}_n \Longrightarrow R_\infty \equiv R_\infty^0 + c_\beta' D, \tag{5}$$

where $D(\cdot) := -\partial E\left[m(Z,\cdot,\theta_0(\cdot))\right]/\partial\beta \in L_2(\mu)$ and R_∞^0 is a Gaussian process with zero mean and covariance function

$$C(x,y) := E[m(Z,x,\theta_0(x))m(Z,y,\theta_0(y))], \qquad (x,y)\in\Gamma\times\Gamma.$$

In Appendix A we provide relatively "simple" sufficient conditions on the model and data generating process for Assumption W to hold. It is shown there how the influence function $m(Z,x,\theta_0)$ depends on the moment $\psi(Z,x,\theta_0)$ and generally on the impact of estimation of nuisance parameters, see (27). The uniform expansion in Appendix A is of independent interest. Related primitive conditions can be found in the literature, see e.g. Chen and Fan (1999) and Song (2010) for semiparametric conditional moment restrictions and Escanciano and Zhu (2012) for partially identified semiparametric models. Functional Central Limit Theorems (FCLT) in Hilbert spaces for independent observations can be found in Politis and Romano (1994) and van der Vaart and Wellner (1996).

2.4 Limiting Problem and the FLRT

We aim to find the asymptotically optimal functional of $\hat{R}_n$ for testing H_0 vs H_n. Let $\mathbb{P}_0$ and $\mathbb{P}_1$ be the probability measures in $L_2(\mu)$ associated to R_∞^0 and $R_\infty^0 + c_\beta' D$, respectively. For a general treatment of probability measures of random elements in Hilbert spaces see Parthasarathy (1967). In terms of the limiting random element R_∞, the testing problem can be written as

$$H_0 : R_\infty \sim \mathbb{P}_0 \qquad vs \qquad H_1 : R_\infty \sim \mathbb{P}_1.$$

To construct an optimal test, we need to introduce some further notation. Let K be the covariance operator associated to R_∞ (cf. Assumption W), i.e.

$$K(h)(x) := \int_\Gamma C(x,y)h(y)\mu(dy), \qquad \text{for all } h\in L_2(\mu). \tag{6}$$

The operator K extends the notion of asymptotic covariance matrix (viewed as a linear operator) in the finite-dimensional case. Since K is a compact, linear and positive operator, it has a countable spectrum $\{\lambda_j, \varphi_j\}_{j=1}^{\infty}$, where $\{\lambda_j\}_{j=1}^{\infty}$ are real-valued, positive, with $\lambda_j \downarrow 0$, and $\{\varphi_j\}_{j=1}^{\infty}$ forms a complete orthonormal basis for $\overline{Im}(K)$ such that $K\varphi_j = \lambda_j \varphi_j$, for all $j \in \mathbb{N}$.

By the functional version of the Neyman-Pearson lemma, the optimal test is given by the Radon-Nikodym derivative of $\mathbb{P}_1$ with respect to $\mathbb{P}_0$. To introduce this LR, let $\varepsilon_j := \lambda_j^{-1/2} \langle R_\infty, \varphi_j \rangle$, $j \in \mathbb{N}$, be the so-called principal components of R_∞, and let $\delta_j := \lambda_j^{-1/2} \langle D, \varphi_j \rangle$, $j \in \mathbb{N}$, be the standardized Fourier coefficients of D. Noting that $\{\varepsilon_j\}_{j=1}^{k}$ are iid standard normal under $\mathbb{P}_0$, and have mean $(c_\beta' \delta_1, ..., c_\beta' \delta_k)$ under $\mathbb{P}_1$, it seems intuitive to use the approximation, for large k,

$$\frac{d\mathbb{P}_1}{d\mathbb{P}_0}(h) \approx \exp\left(c_\beta' \sum_{j=1}^{k} \lambda_j^{-1/2} \langle h, \varphi_j \rangle \delta_j - \frac{1}{2} \sum_{j=1}^{k} \left(c_\beta' \delta_j \right)^2 \right).$$

Indeed, this intuition is formalized in e.g. Skorohod (1974, Chapter 16, Theorem 2), who shows that $\mathbb{P}_1$ is absolute continuous with respect to $\mathbb{P}_0$, provided the following condition holds

$$\sum_{j=1}^{\infty} \lambda_j^{-1} \langle D_l, \varphi_j \rangle^2 < \infty, \qquad \text{for all } l = 1, ..., p, \tag{7}$$

where D_l denotes the lth component of D (cf. (5)). In that case, the functional LR is given by

$$\frac{d\mathbb{P}_1}{d\mathbb{P}_0}(h) = \exp\left(c_\beta' L(h) - \frac{1}{2} \sum_{j=1}^{\infty} \left(c_\beta' \delta_j \right)^2 \right), \tag{8}$$

where L is the linear operator

$$L(h) := \sum_{j=1}^{\infty} \lambda_j^{-1} \langle h, \varphi_j \rangle \langle D, \varphi_j \rangle, \qquad h \in \mathcal{D}(L). \tag{9}$$

As evidenced from (8), $L(R_\infty)$ is a sufficient statistic for our limiting testing problem. In terms of this sufficient statistic, the problem can be equivalently characterized as the familiar $H_0 : L(R_\infty) \sim N(0, \Sigma)$ against $H_1 : L(R_\infty) \sim N(\Sigma c_\beta, \Sigma)$, where $\Sigma := Var(L(R_\infty))$. The Neyman-Pearson lemma and some standard testing arguments, see e.g. CHS, suggest that an optimal α-level test for testing H_0 against H_1 is given by $\phi^*(R_\infty)$, with

$$\phi^*(h) := 1\left(L(h) \Sigma^{-1} L(h) \geq \chi^2_{1-\alpha,p} \right),$$

where $\chi^2_{\tau,p}$ denotes the τ-quantile of the chi-squared distribution with p degrees of freedom. We define now the FLRT, which uses the finite sample analog of R_∞ in the limiting optimal functional.

Definition 1 *The FLRT is defined by* $\phi_n^* := \phi^*(\sqrt{n}\hat{R}_n)$.

Notice that ϕ_n^* involves a quadratic form in a linear combination of the sample principal components,

$$L(\sqrt{n}\hat{R}_n) = \sum_{j=1}^{\infty} \delta_j \hat{\varepsilon}_{nj},$$

where $\hat{\varepsilon}_{nj} := \lambda_j^{-1/2} \left\langle \sqrt{n}\hat{R}_n, \varphi_j \right\rangle$ is the sample analog of ε_j, $j \in \mathbb{N}$. The main purpose of this paper is to study the efficiency properties of ϕ_n^* and related tests.

As mentioned above, this functional LR approach has already been applied to several semiparametric models in econometrics and statistics. Sowell (1996) proposed a FLRT for testing parameter instability in a GMM setting; see also Elliot and Müller (2009). Stute (1997), Stute, Thies and Zhu (1998), Boning and Sowell (1999), Bischoff and Miller (2000) and Escanciano (2009) used this principle to test the correct specification of, possibly dynamic, regression models. Applications to conditional distributions were given in Delgado and Stute (2008), and to tests for correct specification of the covariance structure of a linear process in Delgado, Hidalgo and Velasco (2005). Akritas and Johnson (1982) and Luschgy (1991), among others, investigated optimal tests in stationary and non-stationary diffusion processes, respectively. Recently, Watson and Müller (2008) construct a finite-dimensional approximation of a FLRT for testing low-frequency variability in persistent time series. Müller (2011) considers applications to unit root testing, weak instruments and parameter instability, among many others, while Song (2010) suggests applications to a general class of semiparametric conditional moment models. None of the aforementioned papers, however, have shown the semiparametric efficiency, in the sense of CHS, of the resulting tests.

It is worth mentioning that in some applications the FLRT has a closed form, as a functional of $\sqrt{n}\hat{R}_n$, i.e. L is fully known, see e.g. Luschgy (1991) and Müller (2011) for examples. However, in most regular problems such a closed form expression is in general not available, and estimation (regularization) of the operator L, the drift D and the matrix Σ is often needed. We deal with the implementation of feasible versions of the FLRT in Section 4. There, we show that a feasible test based on a quadratic form of

$$\widehat{L}_n = \frac{1}{n}\sum_{i=1}^{n} \widehat{s}^*(Z_i),$$

for a suitable estimated score $\widehat{s}^*(Z_i)$, is asymptotically equivalent to ϕ_n^*. Thus, for asymptotic efficiency purposes it suffices to consider the infeasible test ϕ_n^* for the time being.

2.5 Asymptotic Representation of the FLRT as a Score-Type Test

The objective of this section is to provide an asymptotic representation of the FLRT as a score-type test. This result is instrumental for other results in the paper. In Section 3, it will be shown that the resulting score coincides with the *efficient score* for the corresponding semiparametric problem, so the optimality of the FLRT follows. Later in Section 4, we will use the characterization of the score to propose an estimate of it and to implement a feasible FLRT.

With this objective in mind, we introduce the singular value decomposition of K; see Kress (1999). Henceforth, to simplify notation, when we evaluate moments at the true values, we remove the dependence on these values, e.g. $m(Z_i, x) \equiv m(Z_i, x, \theta_0(x))$. The covariance operator $K(h)(x) = E[\langle m(Z,\cdot), h\rangle\, m(Z,x)]$ can be written as $K = T'T$, where T' and T are compact linear operators defined, respectively, by

$$Th(z) := \langle m(z,\cdot), h\rangle \qquad z \in \mathbb{R}^d, h \in L_2(\mu)$$

and

$$T'a(x) := E[a(Z)m(Z,x)] \qquad x \in \Gamma, a \in L_2(F).$$

Also note that T' is the adjoint (dual) operator of T, that is, for all $h \in L_2(\mu)$ and $a \in L_2(F)$,

$$E[a(Z)Th(Z)] = \langle T'a, h \rangle . \tag{10}$$

The abuse of notation in T' is motivated from the equivalence of transposition and adjoint operators for matrices.

In addition to the sequence $\{\lambda_j, \varphi_j\}_{j=1}^{\infty}$, there exists a complete orthonormal basis for $\overline{Im}(T) = \ker^{\perp}(T')$, say $\{\psi_j\}_{j=1}^{\infty}$, satisfying, for all $j \in \mathbb{N}$, (cf. Kress, 1999, Theorem 15.16)

$$T\varphi_j = \lambda_j^{1/2}\psi_j \qquad \text{and} \qquad T'\psi_j = \lambda_j^{1/2}\varphi_j. \tag{11}$$

For $r > 0$, we introduce the subspace of $L_2(\mu)$,

$$\Psi_r := \left\{ h \in L_2(\mu) \text{ such that } \|h\|_r^2 := \sum_{j=1}^{\infty} \lambda_j^{-r} \langle h, \varphi_j \rangle^2 < \infty \right\},$$

with the corresponding inner product $\langle h, g \rangle_r := \sum_{j=1}^{\infty} \lambda_j^{-r} \langle h, \varphi_j \rangle \langle g, \varphi_j \rangle$. It is well-known that Ψ_1 is the so-called Reproducing Kernel Hilbert space associated to K and that $\Psi_1 = Im(T') \supset Im(K)$. We now introduce two assumptions that are needed for our representation.

Assumption D: *The function D is such that $D \in \Psi_1$.*

As previously mentioned, Assumption D is equivalent to the absolute continuity of $\mathbb{P}_1$ with respect to $\mathbb{P}_0$. Intuitively, this assumption requires D not to be too "large", relative to R_{∞}^0, for continuity to hold. Below, we show that this "contiguity" assumption is intimately related to the assumption of finite efficient Fisher information, see Section 3.

Define the process with "known" parameters

$$M_n(x) := \frac{1}{n}\sum_{i=1}^{n} m(Z_i, x, \theta_0). \tag{12}$$

We require the asymptotic equivalence of $L(\sqrt{n}\hat{R}_n)$ and $L(\sqrt{n}M_n)$. In view of Assumption W, this can be understood as a continuity assumption of $L(\cdot)$ with respect to $\|\cdot\|$.

Assumption C: *Under H_n, $L(\hat{R}_n) = L(M_n) + o_P(n^{-1/2})$.*

There are at least two ways to prove the high-level Assumption C. Since the operator L is continuous in Ψ_1 with the Reproducing Kernel Hilbert space norm $\|\cdot\|_1$, one possibility is to strengthen Assumption W so that $\left\|\hat{R}_n - M_n\right\|_1 = o_P(n^{-1/2})$. A second approach is to keep Assumption W but require continuity of L with respect to $\|\cdot\|$, as assumed in e.g. Müller (2011). This is the case, for instance, if $D \in \Psi_2$. A sufficient condition for the latter is that $Im(T)$ is closed (see Lemma 3.4 in var der Vaart, 1991). This assumption imposes further smoothness on the model, as shown below. See also Chen, Chernozhukov, Lee and Newey (2011) for related discussion.

Note that Assumption D is equivalent to the following random vector being well defined in $L_2(F)$,

$$s^*(Z_i) := \sum_{j=1}^{\infty} \lambda_j^{-1/2} \langle D, \varphi_j \rangle \psi_j(Z_i). \tag{13}$$

The score function s^* will play a crucial role in our theory. Define the standardized sample mean

$$S_n^* := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} s^*(Z_i). \tag{14}$$

Our next result proves the asymptotic equivalence of $L(\sqrt{n}\hat{R}_n)$ and the sample mean S_n^*.

Theorem 1: *Let Assumptions W, D and C hold. Then,*

(i) $L(\sqrt{n}\hat{R}_n) = S_n^* + o_P(1)$, *under* H_n.

(ii) *Moreover,* s^* *satisfies* $T's^* = D$, *and for any other* $s \in L_2(F)$ *satisfying* $T's = D$, *it holds that* $s^* = \Pi_{\ker^{\perp}(T')} s$.

Remark 1: Theorem 1(i) proves the asymptotic equivalence of the FLRT with a score-type test. Its proof only uses elementary considerations, but that does not vitiate its utility. In a model with no nuisance parameters, the equivalence is also in finite samples. For instance, it can be shown that in fully parametric models with no nuisance parameters, the FLRT based on the standard empirical process boils down to the classical Rao-Score test in finite samples. In the general case, Theorem 1(ii) can be viewed as a form of *generalized information equality*. It offers an alternative way to compute the score s^* in (13) without explicitly using the spectrum. This is practically important, since expressions for $\{\lambda_j, \varphi_j, \psi_j\}_{j=1}^{\infty}$ are only available for very special situations. Thus, Theorem 1(ii) offers the following algorithm for computing s^*: (i) First, find a solution to the integral equation $T's = D$, then (ii) compute the projection of s onto $\ker^{\perp}(T')$. An immediate consequence of Theorem 1(ii) is that, among all possible solutions s of $T's = D$, the one with minimum variance corresponds to s^*. Note that the existence of one solution of $T's = D$ in $L_2(F)$ implies Assumption D. Later we shall show that s^* is the so-called efficient score, and hence, the previous algorithm provides a new method to compute efficiency bounds for regular estimation of β_0, a result of independent interest. In all the examples we have considered, solving $T's = D$ was a trivial task, as the following example illustrates.

Example 2: **Significance in mean regression**. Consider the linear semiparametric regression model

$$Y = \eta_{01} + \eta_{02}X + \beta a(X) + \varepsilon, \qquad E[\varepsilon | X] = 0 \text{ almost surely (a.s.)},$$

where Y and X are random variables, $\eta_0 = (\eta_{01}, \eta_{02})'$, $a(X)$ is a known function, e.g. $a(X) = X^2$, and the conditional distribution of ε given X is unknown. This semiparametric model can be characterized by the infinite number of moments (cf. Stute, 1997)

$$E[\{Y - \eta_{01} - \eta_{02}X - \beta a(X)\}1(X \leq x)] = 0 \text{ for all } x \in \mathbb{R}. \tag{15}$$

In this example η_0 is parametric and estimated by the Ordinary Least Squares (OLS) estimator $\widehat{\eta}_n$ of Y on $\tilde{X}_i = (1, X_i)'$, and the interest is in testing $H_0 : \beta = 0$ against $H_n : \beta_n = n^{-1/2}c_\beta$. Stute (1997) has shown the asymptotic uniform (in $x \in \mathbb{R}$) representation under H_n,

$$\begin{aligned}\hat{R}_n(x) &= \frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{\eta}_1 - \widehat{\eta}_2 X_i)1(X_i \leq x), \\ &= \frac{1}{n}\sum_{i=1}^{n}\varepsilon_{i0}q(X_i, x) + o_P(n^{-1/2}),\end{aligned} \tag{16}$$

where $\varepsilon_{i0} := Y_i - \eta_{01} - \eta_{02}X_i$ and $q(X_i, x) := 1(X_i \leq x) - E[\tilde{X}_i' 1(X_i \leq x)]E[\tilde{X}_i\tilde{X}_i']^{-1}\tilde{X}_i$. Thus, Assumption W holds under some mild moment assumptions, with $m(Z, x, \theta_0) = \varepsilon_{i0}q(X_i, x)$, $D(x) = E[a(X_i)q(X_i, x)]$ and μ the probability measure of X. Define $\sigma^2(X) := E\left[\varepsilon^2 \middle| X\right]$ and $Z := (Y, X)'$. The integral equation $T's(x) = D(x)$, which is given here by

$$E[\varepsilon_{i0}q(X, x)s(Z)] = E[q(X, x)a(X)],$$

is trivially solved by the score $s(Z_i) := \sigma^{-2}(X_i)\varepsilon_{i0}a(X_i)$. It is also easy to show that $\ker(T') = span\{\sigma^{-2}(X_i)\varepsilon_{i0}\tilde{X}_i'\}$. It then follows from our Theorem 1 that

$$L(\hat{R}_n) = \frac{1}{n}\sum_{i=1}^{n}\varepsilon_{i0}\sigma^{-2}(X_i)a^*(X_i) + o_P(n^{-1/2}),$$

where

$$a^*(X_i) := a(X_i) - E[a(X_i)\sigma^{-2}(X_i)\tilde{X}_i']E[\sigma^{-2}(X_i)\tilde{X}_i\tilde{X}_i']^{-1}\tilde{X}_i.$$

The resulting score $s^*(Z_i) \equiv \Pi_{\ker^\perp(T')}s = \varepsilon_{i0}\sigma^{-2}(X_i)a^*(X_i)$ is indeed the efficient score, see e.g. Chamberlain (1987). For further discussion on this example, see Section 5. $\square$

3 On the Efficiency of the FLRT

We show in this section that the conclusion from the mean regression example holds more generally in our semiparametric setting. That is, we show that the FLRT is an efficient test in the class of semiparametric models defined by (1). Specifically, we use the efficiency concept of *asymptotically uniformly most powerful and invariant* test of level α, in short AUMPI(α), defined formally in CHS (Section 5). When $p = 1$, alternative definitions of efficiency, which do not require invariance, are typically used. For completeness, these and related definitions of efficiency are reviewed in Appendix A. For a thorough discussion see CHS. Let $\mathcal{P} := \{P_{(\beta,\eta)} : \beta \in \Theta_\beta, \eta \in \Theta_\eta\}$ be the semiparametric model satisfying (1). Note that indexing the semiparametric model by (β, η) does not entail a loss of generality, see e.g. Bickel, Ritov and Stoker (2006) for a similar approach. Define the marginal class with β fixed at β_0 by $\mathcal{P}_2 := \{P_{(\beta_0,\eta)} : \eta \in \Theta_\eta\}$, and let $\dot{\mathcal{P}}_2$ be the tangent space of $\mathcal{P}_2$ at $P_{(\beta_0,\eta_0)}$, i.e. the closed linear span of scores passing through the semiparametric model $P \equiv P_{(\beta_0,\eta_0)}$. Given the score $\dot{\ell}_\beta$ in the marginal family $\mathcal{P}_1 = \{P_{(\beta,\eta_0)} : \beta \in \Theta_\beta\}$, we define the efficient score $\dot{\ell}_\beta^*$ as the orthogonal projection of the score $\dot{\ell}_\beta$ onto the orthocomplement of $\dot{\mathcal{P}}_2$, i.e., $\dot{\ell}_\beta^* := \dot{\ell}_\beta - \Pi_{\dot{\mathcal{P}}_2}\dot{\ell}_\beta$. Let $B^* := Var(\dot{\ell}_\beta^*)$ be the efficient information, and assume B^* is positive definite and finite. Write $\xi_n(\eta_0) := (nB^*)^{-1/2} \sum_{i=1}^n \dot{\ell}_\beta^*(Z_i, \eta_0)$. An efficient test statistic T_n satisfies $T_n = \xi_n(\eta_0) + o_P(1)$, for every η_0. CHS (Corollary 3) show that the test $\phi_n^* := 1\left(T_n' T_n \geq \chi^2_{1-\alpha,p}\right)$ is AUMPI(α).

As an example, we consider the fully parametric case. Let $\dot{\ell}_\eta$ denote the score of $\mathcal{P}_2$, which spans the finite-dimensional tangent space $\dot{\mathcal{P}}_2$. Then, $\Pi_{\dot{\mathcal{P}}_2} \dot{\ell}_\beta = E[\dot{\ell}_\beta \dot{\ell}_\eta'] E[\dot{\ell}_\eta \dot{\ell}_\eta']^{-1} \dot{\ell}_\eta$, which can be easily estimated by the sample analog using $\widehat{\eta}_n$. Neyman (1959) proposed the efficient test ϕ_n^* with $T_n = \hat{\xi}_n(\widehat{\eta}_n)$, where $\hat{\xi}_n$ is defined as ξ_n but with the Fisher information and co-information estimated by their sample analog. This is the so-called $C(\alpha)-$test of Neyman (1959). He showed that $T_n = \xi_n(\eta_0) + o_P(1)$, for every η_0, thereby proving the optimality of the test.[4]

The situation in the semiparametric case is more complicated, as there is no general expression for $\Pi_{\dot{\mathcal{P}}_2} \dot{\ell}_\beta$ (here $\dot{\mathcal{P}}_2$ is infinite-dimensional). In this paper we overcome this limitation by using a functional LR approach. In the general case, the FLRT will be AUMPI(α) if we prove that, for every η_0,

$$L(\hat{R}_n) = \frac{1}{n} \sum_{i=1}^n \dot{\ell}_\beta^*(Z_i, \eta_0) + o_P(n^{-1/2}).$$

In view of Theorem 1, this is the case if and only if $s^* = \dot{\ell}_\beta^*$ a.s. This is proved in the next theorem. Define $\ker^0(T') := \{h \in \ker(T') : E[h(Z)] = 0\}$. Standard regularity conditions that imply Local Asymptotic Normality (LAN), among other things, and which are required for the definition of efficiency are gathered in Appendix A.

THEOREM 2: *Let Assumptions C in the text and A1 and A2 in Appendix A hold. Then,*

(i) $\dot{\mathcal{P}}_2 = \ker^0(T')$.

(ii) $s^* = \dot{\ell}_\beta^*$ *a.s., and hence, the FLRT is AUMPI(α).*

Theorem 2(i) is of independent interest. This result characterizes in simple mathematical terms the tangent space of nuisance parameters in a general class of semiparametric models defined by moment restrictions. It extends related results by Bickel et al. (1993, Section 6.2) to a larger class of semiparametric models. Theorem 2(ii) shows the semiparametric efficiency of the FLRT.

For completeness, we discuss an alternative sense of efficiency of the FLRT. Müller (2011) has recently shown that the FLRT is optimal in a class of tests that control asymptotic size for all data generating processes for which the underlying random element, $\sqrt{n}\hat{R}_n$, has the corresponding limiting distribution in Assumption W. We particularize Müller's results to our framework. He defines the class of statistical models $\mathcal{M}$ as the class of models for which Assumption W holds. Then, he defines the class of tests $\mathcal{C}$ as those tests that have asymptotic level $\alpha \in (0,1)$ for all models in $\mathcal{M}$. That is, the class of models is defined through a weak convergence requirement. Then, Müller's main finding is as follows. Assuming that the mapping L in (9) is continuous with respect to $\|\cdot\|$, the FLRT is the most

[4] Neyman (1959) only considered the case $p = 1$. His results were extended to $p > 1$ by Bühler and Puri (1966). For the relationship between the $C(\alpha)$-test and our FLRT in fully parametric models see Akritas (1988).

efficient test in the class $\mathcal{C}$, and for any other test in $\mathcal{C}$ with higher asymptotic power for any model in $\mathcal{M}$, there exits a model in $\mathcal{M}$ for which the test has asymptotic null rejection probability larger than the nominal level α. Thus, this new concept of efficiency provides a sense of robustness of the FLRT. Our paper complements Müller's efficiency results by proving that, in regular semiparametric problems, the FLRT is also semiparametrically efficient in the "classical" sense of CHS.

We also relate our results to the recent literature in econometrics proving that efficient estimation of semiparametric models can be achieved by GMM estimators employing an infinite number of moments, see e.g. Ai and Chen (2003), Newey (2004) and Carrasco and Florens (2000, 2011). We establish here an important connection between the GMM literature and our LR approach. This connection is mutually beneficial, both in theory and implementation of the procedures. We modify Carrasco and Florens (2000, 2011) and Newey (2004) to properly account for the presence of estimated, possibly infinite-dimensional, nuisance parameters and suggest a candidate for an optimal GMM estimator as the minimizer of the following objective function

$$\left\| K_n^{-1/2} \hat{M}_n(\cdot, \beta) \right\|^2,$$

where $K_n^{-1/2}$ is some consistent estimator of the operator $K^{-1/2}$, $\hat{M}_n$ is defined as M_n but with $\widehat{\eta}_n$ replacing η_0, and where we emphasize the dependence of $\hat{M}_n$ on β, see (12). Implementations vary according to the estimator (regularization) $K_n^{-1/2}$ used. Note that the estimator should use $\hat{M}_n$ rather than the original $\hat{R}_n$ for our arguments below to hold. Under some regularity conditions that allow us to replace $K_n^{-1/2}$ by $K^{-1/2}$, see Section 4.2, it can be shown that the feasible optimal GMM will be asymptotically equivalent to the minimizer of

$$Q_n(\beta) := \frac{1}{2} \sum_{j=1}^{\infty} \lambda_j^{-1} \left\langle \hat{M}_n(\cdot, \beta), \varphi_j \right\rangle^2 .$$

The GMM testing theory is well known in the standard setting – we can construct Wald, LM or LR tests based on $Q_n(\beta)$; see Newey and West (1987). Similar ideas apply here. If we consider an LM approach and assume smoothness in β for simplicity, the LM test for H_0 involves a quadratic form in

$$\sqrt{n} \frac{\partial Q_n(\beta_0)}{\partial \beta} = \sum_{j=1}^{\infty} \lambda_j^{-1} \left\langle \sqrt{n} \hat{M}_n(\cdot), \varphi_j \right\rangle \left\langle \frac{\partial \hat{M}_n(\cdot)}{\partial \beta}, \varphi_j \right\rangle,$$

which resembles the asymptotic expression for $L(\sqrt{n}\hat{R}_n)$. Hence, the LM test based on the modified GMM objective function can be interpreted as a LR test in our semiparametric context. This connection has important theoretical implications. It implies that extensions of GMM-based tests will be semiparametric efficient in our general semiparametric context, and more generally will share Müller's (2011) efficiency concept even in non-regular settings.

4 Implementation of the Feasible FLRT

We have investigated so far the efficiency properties of the infeasible FLRT. The test is not feasible because Σ, D and the operator L are in general unknown. The implementation of a feasible FLRT critically depends on whether or not the spectrum of K is known. Here, we suggest different implementations for these two exhaustive alternatives.

4.1 Known Spectrum

If the spectrum $\{\lambda_j, \varphi_j\}$ is known, then L and Σ can be estimated, respectively, by

$$L_k(h) = \sum_{j=1}^{k} \lambda_j^{-1} \langle h, \varphi_j \rangle \left\langle \hat{D}, \varphi_j \right\rangle \tag{17}$$

and

$$\widehat{\Sigma}_k = \sum_{j=1}^{k} \lambda_j^{-1} \left\langle \hat{D}, \varphi_j \right\rangle \left\langle \hat{D}, \varphi_j \right\rangle',$$

for a suitable consistent estimate $\hat{D}$ of D and $k \equiv k_n \geq 1$, with $k_n \to \infty$ as $n \to \infty$. The feasible FLRT considered here replaces $L(\hat{R}_n)$ by $L_k(\hat{R}_n)$ and Σ by $\widehat{\Sigma}_k$. When the moment function is smooth in β a natural estimate for D is

$$\hat{D}(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{-\partial \widehat{m}_i(x)}{\partial \beta}.$$

If moments are not smooth in β, an estimator for $\hat{D}$ may be constructed using our Theorem 1 as $\hat{D} = \hat{T}' \hat{s}(x)$, for a suitable estimate $\hat{T}'$ of T' and score estimate $\hat{s}$ (not necessarily the efficient one). A related idea is used in the quantile regression example below.

The assumption of known spectrum is justified, not because it holds generally, but because often general transformations of $\hat{R}_n$ exist with known spectrum representations; see the so-called Khmaladze or martingale transformations (cf. Khmaladze, 1981). There is an extensive literature on this transformation in econometrics and statistics. Khmaladze (1981) first considered such transformations for classical parametric problems, but recently Song (2010) has substantially extended it to a general class of semiparametric models, thereby widening the scope of applications of the feasible versions that we discuss here. When Khmaladze's transformation is used, it remains to justify that our efficiency and asymptotic results do not change, and we provide some insights showing that this is indeed the case.

As it turns out, under suitable conditions provided below, the feasible FLRT behaves asymptotically as the infeasible test, i.e.

$$L_k(\hat{R}_n) = L(M_n) + o_P(n^{-1/2}) \qquad \text{and} \qquad \widehat{\Sigma}_k = \Sigma + o_P(1). \tag{18}$$

The following assumption restricts the rate of divergence of k_n. Define the norm $\|\cdot\|_{r,k}$ as $\|h\|_{r,k}^2 := \sum_{j=1}^{k} \lambda_j^{-r} \langle h, \varphi_j \rangle^2$.

Assumption R: *(i)* $k_n \to \infty$; *(ii) Under* H_n, $\left\| \hat{R}_n - M_n \right\|_{1,k} = o_P(n^{-1/2})$ *and* $\left\| \hat{D} - D \right\|_{1,k} = o_P(k_n^{-1/2})$.

We will provide specific restrictions that R(ii) imposes for a regression example in Section 5. Assumption R(ii) can be replaced by $\left\| \hat{R}_n - M_n \right\| = o_P(n^{-1/2})$, $\left\| \hat{D} - D \right\|_{2,k} = o_P(1)$ and $\|D\|_2 < \infty$. As mentioned earlier, the latter assumption implies the continuity of L with respect to $\|\cdot\|$, and it can be understood in terms of further smoothness in the sense of a fast decay of the Fourier coefficients for the score s^*. To see this, note that

$$\sum_{j=1}^{\infty} \lambda_j^{-2} \langle D_l, \varphi_j \rangle^2 = \sum_{j=1}^{\infty} \lambda_j^{-1} \left(E[s_l^*(Z)\psi_j(Z)] \right)^2 .$$

In fact, $\|D\|_2 < \infty$ is equivalent to $s_l^* \in Im(T)$ for all $l = 1, ..., p$. If there are no nuisance parameters, then Assumption R can be simplified to $k_n \to \infty$ and $\left\| \hat{D} - D \right\|_{1,k} = o_P(1)$.

PROPOSITION 1: *Let Assumptions D, W and R hold. Then, under H_n (18) holds.*

A corollary of Proposition 1 is that the feasible FLRT is an AUMPI(α) test. Proposition 1 is applicable to cases where the asymptotic limit distribution $\hat{R}_n$ has a known spectrum. We discuss now a generic approach that leads to that case, and justify the efficiency in this generic example. For simplicity of the exposition, we restrict our analysis here to conditional moment restrictions of the form

$$E\left[\rho(Z, \beta, \eta_0) | \, X \right] = 0 \text{ a.s.}$$

where X is a subvector of Z of dimension d_x. A standard way to characterize this conditional moment model is through the moment restrictions

$$E[\rho(Z, \beta, \eta_0) 1(X \leq x)] = 0 \text{ for all } x \in \mathbb{R}^{d_x}.$$

However, as proved in Appendix A, sample feasible versions of the moments are generally not asymptotic distribution-free, leading to the so-called Durbin problem (see Koenker and Xiao, 2002). An approach that has been suggested in the literature to overcome this problem is to consider moments

$$E[\rho(Z, \beta, \eta_0) \mathcal{M} 1(X \leq x)] = 0 \text{ for all } x \in \mathbb{R}^{d_x},$$

where $\mathcal{M}$ is a linear operator satisfying certain properties, specifically, it is an isometry projecting onto the orthocomplement of the tangent space of nuisance parameters, see Song (2010) for details. It can be shown that our results applied to the moment function $\psi(Z_i, x, \beta, \eta_0) = \rho(Z, \beta, \eta_0)\mathcal{M}1(X \leq x)$ deliver a semiparametric efficient test. The set of solutions of $T's = D$ does not change by the presence of $\mathcal{M}$. Note that the orthogonality of $\mathcal{M}$ with the tangent space of nuisance parameters implies that $m \equiv \psi$. See Song (2010) for a formal proof. By the same orthogonality, $\ker(T')$ does not change by the presence of the transformation $\mathcal{M}$. Thus, from Theorem 1 the resulting score is the same with or without the transformation, and by Theorem 2 this is the efficient score.

4.2 Unknown Spectrum

In most applications the spectrum $\{\lambda_j, \varphi_j\}$ is, however, unknown, and for most models distribution-free transformations, such as Khmaladze's transformation, are not available. One possible solution to this problem, as suggested by Carrasco and Florens (2000), is to estimate nonparametrically the spectrum. Escanciano (2009) discussed this approach for regression model checks. Here, we propose an alternative method based on the characterization of the efficient score in Theorem 1(ii) and on well-known results from the theory of linear inverse problems, see Carrasco, Florens and Renault (2006) for a review of this theory. Our estimator for the efficient score seems to be new in the literature.

Theorem 1 shows that

$$L(\hat{R}_n) = \frac{1}{n}\sum_{i=1}^{n} s^*(Z_i) + o_P(n^{-1/2}),$$

where $s^*(Z_i)$ is characterized as the solution of $T's = D$ with minimum norm, i.e. a Moore-Penrose generalized inverse of T'. The idea is simple, we write the equation as $TT's = TD$, and solve the sample analog of this equation using estimates for T, T' and D to obtain a nonparametric estimate of s^*, say $\widehat{s}^*$. Then, we propose a feasible FLRT replacing $L(\hat{R}_n)$ by

$$\widehat{L}_n = \frac{1}{n}\sum_{i=1}^{n} \widehat{s}^*(Z_i). \tag{19}$$

Since the inverse problem $TT's = TD$ is in general ill-posed, we need to apply some regularization technique. We choose Tikhonov regularization, as it is simple to apply. This method is based on solving the perturbed equation

$$(\alpha_n I + TT')s^*_{\alpha_n} = TD,$$

where $s^*_{\alpha_n}$ is implicitly defined, α_n is a regularization (tuning) parameter, such that $\alpha_n \downarrow 0$ at a suitable rate, and I is the identity operator. Note that such solution $s^*_{\alpha_n}$ always exists under Assumption D, and it is given by

$$s^*_{\alpha_n}(Z) := \sum_{j=1}^{\infty} \frac{\sqrt{\lambda_j}}{\lambda_j + \alpha_n} \langle D, \varphi_j \rangle \, \psi_j(Z).$$

In practice, T and T' are unknown and estimated by

$$\hat{T}h(z) := \frac{1}{n}\sum_{j=1}^{n} \widehat{m}(z, x_j)h(x_j) \qquad z \in \mathbb{R}^d, h \in L_2(\mu)$$

and

$$\hat{T}'s(x) := \frac{1}{n}\sum_{i=1}^{n} \widehat{m}_i(x)s(Z_i) \qquad x \in \Gamma, s \in L_2(F),$$

where $\{x_j\}_{j=1}^n$ is a random sample from μ. For instance, in many applications, such as conditional moment restrictions, μ can be chosen as the probability measure of X, so we can take $\{x_j\}_{j=1}^n \equiv \{X_j\}_{j=1}^n$. Note that there is some abuse of notation here because $\hat{T}'$ is not the adjoint of $\hat{T}$, but this notation is justified asymptotically. Then, simple algebraic arguments show that the finite sample version $(\alpha_n I + \hat{T}\hat{T}')\widehat{s}^* = \hat{T}\hat{D}$ has a closed form solution given by

$$\widehat{s}^*(z) := \frac{1}{n\alpha_n}\sum_{j=1}^{n} \tilde{D}(x_j)\widehat{m}(z, x_j), \tag{20}$$

where

$$\tilde{D}(x_j) := \hat{D}(x_j) - \frac{1}{n}\sum_{h=1}^{n} p_h \widehat{m}_h(x_j)$$

and the vector $p=(p_1,...,p_n)'$ satisfies the system of linear equations $(\alpha_n I + A)p = b$, where A is an $n\times n$ matrix with principal element

$$a_{j,l} = \frac{1}{n^2}\sum_{h=1}^{n}\widehat{m}_j(x_h)\widehat{m}_l(x_h)$$

and $b=(b_1,...,b_n)'$, with

$$b_j = \frac{1}{n}\sum_{h=1}^{n}\hat{D}(x_h)\widehat{m}_j(x_h).$$

The estimated score $\widehat{s}^*(z)$ is used in (19) and in estimating the Fisher information matrix by

$$\widehat{\Sigma}_{\alpha_n} := \frac{1}{n}\sum_{h=1}^{n}\widehat{s}^*(Z_i)\left(\widehat{s}^*(Z_i)\right)'.$$

Finally, the asymptotic $\alpha-$th level feasible FLRT is given by

$$\hat{\phi}_n^* := 1(n\widehat{L}_n'\widehat{\Sigma}_{\alpha_n}^{-1}\widehat{L}_n \geq \chi^2_{1-\alpha,p}).$$

The test only requires estimates $\{\widehat{m}(Z_i,x_j),\hat{D}(x_j)\}_{i,j=1}^n$, and it is extremely easy to implement; see Appendix C for an algorithm to compute the test statistic. We show below that $\hat{\phi}_n^*$ is asymptotically equivalent to the infeasible ϕ_n^*, i.e.

$$\widehat{L}_n = L(M_n) + o_P(n^{-1/2}), \qquad \widehat{\Sigma}_{\alpha_n} = \Sigma + o_P(1).$$

The following assumption plays the role of Assumption R in the current context. For a bounded linear operator define (with some abuse of notation) $\|B\|_b := \sup_{\|h\|_a\leq 1}\|Bh\|_b$, where the norms $\|\cdot\|_a$ and $\|\cdot\|_b$ are the norms in the domain and range of definition of B, respectively. Define the operator $\hat{T}^*: L_2(F)\to L_2(\mu)$ by

$$\hat{T}^* a := \int a(z)\widehat{m}(z,x)F(dz).$$

Assumption RE: *(i)* $n\alpha_n^{5/2}\to\infty$ *and* $\alpha_n\to 0$ *as* $n\to\infty$; *(ii)* $\left\|\hat{D}-\hat{T}'s^*\right\| = O_P(n^{-1/2})$ *and* $\left\|\hat{T}^*-T'\right\|_{2,P} = O_P(n^{-1/2})$; *and (iii)* $D\in\Psi_2$.

The conditions in RE(ii) can be checked using our results in the Appendix. When the estimator $\widehat{\eta}_n$ is $\sqrt{n}$-consistent and the moments are smooth in η_0, RE(ii) follows from standard Taylor arguments. A sufficient condition for $\left\|\hat{T}^*-T'\right\| = O_P(n^{-1/2})$ is

$$\int\int\left|\widehat{m}(z,x)-m(z,x)\right|^2 F(dz)\mu(dx) = O_P(n^{-1}),$$

which is easy to check in many applications. Needless to say that for a given example, our assumptions can be relaxed by exploiting the specifics of the model.

Theorem 3: *Let the assumptions of Theorem 2 and Assumption RE hold. Then, the feasible $\alpha-$level FLRT based on (19) with $\widehat{s}^*$ as in (20) is AUMPI(α).*

Remark 2: If $P\left(\sqrt{n}(\hat{D}-\hat{T}'s^*)\in \mathrm{Im}(T')\right)\to 1$ as $n\to\infty$ then Assumption RE(i) can be relaxed to $n\alpha_n^2\to\infty$. Note that $\mathrm{Im}(T')$ is dense in $L_2(\mu)$, which suggests that the previous condition is not strong. Similar simplifications can be obtained if $\hat{D}=\hat{T}'\hat{s}$ for some, possibly estimated, score $\hat{s}$ satisfying some regularity conditions.

5 Applications

5.1 Linear Quantile Regression

We implement the efficient feasible FLRT for the quantile regression example. We modify the original notation to account for the presence of additional infinite-dimensional nuisance parameters in the limiting distribution, so the model is defined by the moment restrictions

$$E[\zeta_i(\tau)1(X_i\leq w)]=0 \text{ for all } x=(\tau,w')'\in\Gamma:=\mathcal{T}\times\mathbb{R}^{d_x},$$

where $\zeta_i(\tau)=1(Y_i\leq\beta'X_{1i}+\gamma_0'(\tau)X_{2i})-\tau$. Define $\delta_0(\tau):=(\beta_0',\gamma_0'(\tau))'$ and $\theta_0(\tau):=(\delta_0'(\tau),f_{\cdot\tau})'$, where $f_{i\tau}$ is the conditional density of Y_i given X_i, evaluated at $X_i'\delta_0(\tau)$. A natural estimator for $\gamma_0(\tau)$ is the QR estimator, initially proposed by Koenker and Basset (1978), and defined as any solution $\widehat{\gamma}_n(\tau)$ minimizing

$$\gamma\longmapsto\sum_{i=1}^{n}\rho_\tau\left(Y_i-\beta_0'X_{1i}-\gamma'X_{2i}\right),$$

where $\rho_\tau(u)=u\left(\tau-1\left\{u\leq 0\right\}\right)$ is the so-called "check" function.

Standard efficiency theory is difficult to apply to this model. In contrast, our results can be easily applied. Theorem 1 suggests that the efficient score solves $T's(x)=D$, and among all solutions is the one with minimum variance. Finding a solution of the integral equation $T's(x)=D$ is relatively straightforward in this example – namely, $s_f(Z):=X_1\dot{f}\left(Y|\,X\right)/f\left(Y|\,X\right)$, where $\dot{f}\left(y|\,x\right):=\partial f\left(y|\,X=x\right)/\partial y$, and $f\left(y|\,X=x\right)$ is the conditional density of Y_i given X_i. However, computing the projection onto the orthocomplement of the tangent space, $\Pi_{\ker^\perp(T')}s_f$, seems to be a rather complicated task. This difficulty does not stop us from implementing a feasible FLRT as suggested in the previous section.

Hence, we proceed to estimate the efficient score in (20). To that end, we need consistent estimates for m and D. These are given by

$$\widehat{m}_i(x)=\widehat{\zeta}_i(\tau)\widehat{q}(X_i,x)$$

and

$$\hat{D}(x)=-\frac{1}{n}\sum_{i=1}^{n}X_{1i}\hat{f}_{i\tau}\widehat{q}(X_i,x),$$

where $\widehat{\zeta}_i(\tau) = 1(Y_i \leq X_i'\widehat{\delta}_0(\tau)) - \tau$, $\widehat{\delta}_0(\tau) := (\beta_0, \widehat{\gamma}_n(\tau))$, $\widehat{q}(X_i, x) := 1(X_i \leq w) - A_n(x)B_n^{-1}(\tau)X_{2i}\hat{f}_{i\tau}$,

$$A_n(x) := \frac{1}{n}\sum_{i=1}^{n} X_{2i}\hat{f}_{i\tau}1(X_i \leq w), \tag{21}$$

$$B_n(\tau) := \frac{1}{n}\sum_{i=1}^{n} X_{2i}X_{2i}'\hat{f}_{i\tau}^2 \tag{22}$$

and $\hat{f}_{i\tau}$ is a nonparametric estimator for $f_{i\tau}$. We follow Escanciano and Goh (2012), and construct an estimator for this nonparametric nuisance parameter as follows. Let $\mathcal{A}_n \equiv \{\tau_j\}_{j=1}^n$ be a random sample from a uniform distribution in $\mathcal{T}$, independent of the original sample $\mathcal{Z}_n \equiv \{Z_i\}_{i=1}^n$. The proposed estimator for $f_{i\tau}$ is $\hat{f}_{i\tau} := \hat{f}\left(X_i'\widehat{\delta}_0(\tau) \middle| X_i \right)$, where

$$\hat{f}\left(y | X_i \right) \equiv \hat{f}\left(y | X_i, \widehat{\delta}_0 \right) := \frac{1}{nh}\sum_{j=1}^{n} K\left(\frac{y - X_i'\widehat{\delta}_0(\tau_j)}{h} \right), \tag{23}$$

where $h > 0$ is a scalar smoothing parameter and $K(\cdot)$ is a smoothing kernel satisfying some conditions below. See Escanciano and Goh (2012) for motivation on the nonparametric estimate $\hat{f}_{i\tau}$. With this estimate, we then calculate the matrix M with entries $\{\widehat{m}(Z_i, x_j)\}_{i,j=1}^n$, and the $n \times p$ matrix D with entries $\{\hat{D}(x_j)\}_{j=1}^n$, and apply the algorithm of Appendix C to compute the test.

For a fixed τ and τ_j in $\mathcal{T}$, let $g_{(\tau,\tau_j)}(u, v)$ be the density of $(X'\delta_0(\tau), X'\delta_0(\tau_j))$ evaluated at (u, v), and let $h_{(\tau,\tau_j)}(z, u, v)$ be the conditional density of Z given $(X'\delta_0(\tau) = u, X'\delta_0(\tau_j) = v)$. Define the functions

$$s_{g(\tau,\tau_j)}(u) := \frac{\partial g_{(\tau,\tau_j)}(u, v)/\partial v\Big|_{v=u}}{g_{(\tau,\tau_j)}(u, u)} \text{ and } s_{h(\tau,\tau_j)}(z, u) := \frac{\partial h_{(\tau,\tau_j)}(z, u, v)/\partial v\Big|_{v=u}}{g_{(\tau,\tau_j)}(u, u)}.$$

Then, primitive conditions that are sufficient for our high-level assumptions in the quantile regression example are given as follows. Let $\mathcal{X}$ be the support of X, and define $\mathcal{X}_Q := \{x'\delta_0(\tau) : x \in \mathcal{X}, \tau \in \Gamma\}$.

Assumption E1: *(i) $\{Z_i\}_{i=1}^n$ is a sequence of iid d-dimensional random vectors; (ii) the conditional densities $\left\{ f\left(\cdot | x \right) : x \in \mathbb{R}^{d_x} \right\}$ are uniformly bounded, from above and below (from zero) on $\mathcal{X}_Q$, with uniformly bounded derivative with respect to y, and such that $E[s_f^2(Z)] < \infty$; (iii) for each $y \in \mathcal{X}_Q$, the density $f\left(y | x \right)$ is twice continuously differentiable in x, with uniformly bounded derivatives; (iv) $E\left[XX'\right]$ is nonsingular and finite, and $E\left[|X_2|^4\right] < \infty$; (v) for each fixed z, τ and τ_j in $\mathcal{T}$, $u \in \mathbb{R}$ and $w \in \mathbb{R}^{d_x}$, the functions $g_{(\tau,\tau_j)}(u, v)$* and *$h_{(\tau,\tau_j)}(z, u, v)$ are twice continuously differentiable in v at u, with uniformly (in τ, τ_j, u and z) bounded derivatives, $E\left[\left|s_{g(\tau,\tau_j)}(X'\delta_0(\tau))X_2\right|^2\right] < \infty$ and $E\left[\left|s_{h(\tau,\tau_j)}(Z, X'\delta_0(\tau))X_2\right|^2\right] < \infty$.*

Assumption E2: *For all $\tau \in \mathcal{T}$, $\delta_0(\tau)$ belongs to the interior of $\Theta_\delta \subset \mathbb{R}^{d_x}$, with Θ_δ compact.*

Assumption E3: *(a) The kernel function $K\left(t\right) : \mathbb{R} \to \mathbb{R}$ is symmetric, bounded, three times continuously differentiable and satisfies the following conditions: $\int K\left(t\right)dt = 1$, $\int \left|t^2K\left(t\right)\right| dt < \infty$,*

$\left|\partial^{(j)}K(t)/\partial t^j\right| \leq C$ *and for some* $v > 1$, $\left|\partial^{(j)}K(t)/\partial t^j\right| \leq C\left|t\right|^{-v}$ *for* $|t| > L_j$, $0 < L_j < \infty$, *for* $j = 1, 2$; *(b) the possibly data dependent bandwidth* h *satisfies* $P(a_n \leq h \leq b_n) \to 1$ *as* $n \to \infty$, *for deterministic sequences of positive numbers* a_n *and* b_n *such that* $b_n \to 0$, $b_n^4 n \to 0$ *and* $a_n^2 n/\log n \to \infty$.

Most of these assumptions are standard in the literature of quantile regression. Assumption E1 implies that a solution of $T's(x) = D$ is well-defined, so Assumption D holds. Our next result shows the optimality of the feasible FLRT applied to this example.

THEOREM 4: *Suppose that Assumptions E1-E3, RE(i) and RE(iii) hold. Then, the* $\alpha-$*level feasible FLRT in (19) with* $\widehat{s}^*$ *as in (20) is* $AUMPI(\alpha)$ *for the quantile regression example.*

5.2 Significance in Mean Regression

We provide here further details on the mean regression example. Primitive conditions for our results to hold in this example can be easily found. For instance, Assumption D holds if and only if $E[\sigma^{-2}(X)a^2(X)] < \infty$. Stute (1997) proposed a FLRT for testing the significance of additional variables in homoscedastic linear-in-parameters regressions. In this special case, $\sigma^2(X) \equiv \sigma^2$ and our previous computations yield the efficient score

$$s^*(Z_i, \eta_0) := \sigma^{-2}\varepsilon_{i0}\{a(X_i) - E[a(X_i)\tilde{X}_i']E[\tilde{X}_i\tilde{X}_i']^{-1}\tilde{X}_i\}.$$

Stute (1997) proposed a FLRT approximation using certain estimates $\{\hat{\lambda}_j, \hat{\varphi}_j\}$ of $\{\lambda_j, \varphi_j\}$ and truncating the operator L in (9). However, notice that in this example s^* is known, up the parameters η_0 and σ^2, which suggests that simpler than Stute's (1997) FLRT efficient feasible tests exist. Indeed, the classical t-test is efficient, in a semiparametric sense, and it does not require spectrum estimates. Hence, in the homoscedastic case there is no need to regularize the problem by introducing tuning parameters, such as the number of principal components k.

Returning to the general conditionally heteroskedastic case, our results imply that the tests proposed in Stute, Thies and Zhu (1998) and Escanciano (2009) are approximately efficient. They are not fully efficient because the number of components used (the number of summands in L) was fixed in these applications. Moreover, it is not clear how estimation effects of $\{\hat{\lambda}_j, \hat{\varphi}_j\}$ can be justified when $k \to \infty$. Our results of Section 4 suggest two alternative ways to implement feasible efficient inference in this example. Firstly, we can use Khmaladze's transformation, as proposed by Stute, Thies and Zhu (1998), and the results of Section 4.1. The limiting Gaussian process after the transformation (including the integral transformation) is a standard Brownian motion, whose spectrum is known and given by

$$\lambda_j = \frac{1}{(j-0.5)^2\pi^2} \qquad \varphi_j(x) = \sqrt{2}\sin\left((j-0.5)\pi x\right),$$

where $x \in [0,1]$. Sufficient conditions for Assumption W are provided in Stute, Thies and Zhu (1998). Some simple algebra shows that

$$\left\|\hat{R}_n - M_n\right\|_{1,k} = O_p(n^{-1/2}k^{1/2}(\widehat{\eta}_n - \eta_0)) \qquad \text{and} \qquad \left\|\hat{D} - D\right\|_{1,k} = O_P(n^{-1/2}).$$

Hence, a sufficient condition for Assumption R in Section 4.1 is $nk_n^{-1} \to \infty$.

Secondly, we could use the results of Section 4.2 and construct a feasible test based on a nonparametric estimator of the efficient score. The inputs for the feasible test are

$$\widehat{m}_i(x) = \widehat{\varepsilon}_{i0}\widehat{q}(X_i, x)$$

and

$$\hat{D}(x) = \frac{1}{n}\sum_{i=1}^{n} a(X_i)\widehat{q}(X_i, x),$$

where $\widehat{\varepsilon}_{i0} = Y_i - \widehat{\eta}_n' \tilde{X}_i$ and $\widehat{q}(X_i, x) := 1(X_i \leq x) - \left(\sum_{i=1}^{n} \tilde{X}_i' 1(X_i \leq x)\right) \left(\sum_{i=1}^{n} \tilde{X}_i \tilde{X}_i'\right)^{-1} \tilde{X}_i$. This second approach does not require Khmaladze's transformation and it is simpler to implement. The feasible test rejects for large absolute values of

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \widehat{\varepsilon}_{i0}\widehat{q}^*(X_i),$$

where

$$\widehat{q}^*(X_i) := \frac{1}{n\alpha_n}\sum_{j=1}^{n} \tilde{D}(X_j)\widehat{q}(X_i, X_j)$$

and $\tilde{D}$ is given after (20). Note that $\{\widehat{q}(X_i, X_j)\}_{i=1}^{n}$ are simply the residuals from an OLS projection of $\{1(X_i \leq X_j)\}_{i=1}^{n}$ on $\{\tilde{X}_i\}_{i=1}^{n}$, so they can be easily computed with standard statistical packages.

A more traditional approach to efficient inference in this model is to estimate nonparametrically $\sigma^2(X_i)$ and to plug in this estimate in $s^*(Z_i)$, as suggested by Robinson (1988). The finite sample comparison of these alternative efficient tests will be investigate in future research.

This example also serves to illustrate an important point: our results on efficiency are invariant to equivalent parametrizations of the model as an infinite number of moment restrictions, as the following remark shows.

Remark 3: All our results go through in the previous example if we replace the indicator function in $q(X_i, x)$ by other comprehensively revealing class of functions. See Bierens and Ploberger (1997) and Stinchcombe and White (1998) for examples of such classes. For instance, we could use the class $\{\exp(x\phi(X)) : x \in \Gamma \subset \mathbb{R}\}$, where Γ is an interval containing zero, and ϕ is a one-to-one bounded mapping, see Bierens and Ploberger (1997). It can be shown that the solution $s(X_i) := \sigma^{-2}(X_i)\varepsilon_{i0}a(X_i)$ does not depend on the class used. It is also straightforward to prove that $\sigma^{-2}(X_i)\varepsilon_{i0}\tilde{X}_i' \in \ker(T')$. In fact, it can be shown that for any comprehensively revealing class $\ker(T') = \{\sigma^{-2}(X_i)\varepsilon_{i0}\tilde{X}_i'\}$. To see this, by Lemma 3.4 in Newey (1990) it suffices to consider scores of the form $\sigma^{-2}(X_i)\varepsilon_{i0}b(X_i)$, for some function $b(\cdot)$. First, consider the case where $b(X_i)$ is orthogonal to $\tilde{X}_i$. In that case,

$$E[\{\exp(x\phi(X_i)) - E[\tilde{X}_i' \exp(x\phi(X_i))]E[\tilde{X}_i\tilde{X}_i']^{-1}\tilde{X}_i\}b(X_i)] \equiv 0$$

is equivalent to

$$E[\exp(x\phi(X_i))b(X)] \equiv 0,$$

which in turn, implies that $b(X) = 0$ a.s. Since any function can be decomposed as $b(X) = c_0 + c_1 X + c_2 b^{\perp}(X)$, where $b^{\perp}(X_i)$ is orthogonal to $\tilde{X}_i$, we conclude that $\ker(T') = \{\sigma^{-2}(X_i)\varepsilon_{i0}\tilde{X}_i'\}$. Note that the measure μ plays no role in this argument.

5.3 Further Examples

There are many examples for which standard efficiency theory can be hard to apply, but for which our results are directly applicable. Newey (2004) discussed two such examples: censored regression and transformation models. Further examples include parametric copulas, semiparametric models with nonparametric generated regressors, or partially identified models, among many others. Here we discuss in more detail the general class of semiparametric partially identified models investigated in Escanciano and Zhu (2012); see also Arellano, Hansen and Sentana (2011) for parametric moments. Efficiency within a class of GMM estimates has been discussed in Arellano et al. (2011) for the parametric setting, but in the semiparametric setting this issue remains completely unexplored. Our results provide here the first feasible optimal tests in both the semiparametric and parametric frameworks. The structure of the problem is similar to the quantile regression example. The model satisfies the moment restrictions

$$E[\xi(Z, x, \eta_0(Z, x))] = 0 \text{ for all } x \in \Gamma,$$

where $\eta_0(Z, x)$ contains parametric, say $\delta_0(x)$, and possibly nonparametric components $\zeta_0(Z, x)$. The model is not identified because x or a subvector of it is not identified. The model is still partially identified in the sense that for each $x \in \Gamma$ there is a unique solution $\eta_0(Z, x)$ of the moment restrictions. Suppose the parameter of interest is $\beta_0 = \delta_0(x_0)$ for a given $x_0 \in \Gamma$, as in Arellano et al. (2011). Then, this model fits our setting if we define

$$\psi(Z, x, \beta, \eta_0(Z, x)) = \begin{cases} \xi(Z, x_0, \beta, \zeta_0(Z, x_0)) & \text{if } x = x_0 \\ \xi(Z, x, \eta_0(Z, x)) & \text{if } x \neq x_0. \end{cases} \tag{24}$$

A complete analysis of this generic class of examples is beyond the scope of this paper, and it is deferred to future research. An interesting application within this class of models is considered in Altonji, Elder and Taber (2005), who study the effect of attending a Catholic school on educational attainment.

6 Final Remarks

In this paper, we have investigated the efficiency, in a classical semiparametric sense, and implementation of the FLRT in a general class of semiparametric models. We have shown that under quite general conditions, the FLRT is asymptotically equivalent to a semiparametric $C(\alpha)-$test. We have suggested a general algorithm for computing the associated score function in terms of certain covariance operator and shift function resulting under local alternatives. The semiparametric efficiency of the FLRT has been established by showing that the score function is the efficient score associated to the model. We have proposed and justified feasible versions of the FLRT when the spectrum is known and when is unknown. Finally, an application to a semiparametric quantile regression model has highlighted the benefits of our approach. Our investigation complements Müller's (2011) optimality results, and shows that the functional Neyman-Pearson approach advocated by Grenander (1950) can lead to semiparametric efficient inference. In sum, this paper provides the first generally applicable approach to efficient testing parametric restrictions in regular semiparametric models.

Although the main focus of the paper has been on efficient tests, our results have important implications for efficient estimation. Our results show that the semiparametric efficiency bound of regular estimators of β_0 is $\Sigma = \|D\|_1$, and we have provided consistent estimators for this bound. Similarly, a simple one-step efficient estimator for β_0 can be constructed as follows,

$$\widehat{\beta}_n = \widehat{\beta}_0 - \widehat{\Sigma}_{\alpha_n}^{-1} \frac{1}{n} \sum_{i=1}^{n} \widehat{s}^*(Z_i),$$

where $\widehat{\beta}_0$ is an initial $\sqrt{n}-$consistent estimator of β_0 that is also used in the computation of $\widehat{\Sigma}_{\alpha_n}$ and $\widehat{s}^*$. After our results, the efficiency and asymptotic distribution theory for $\widehat{\beta}_n$ can be easily obtained combining our methods here with those well established in the literature, see Lecam (1956). Formalizing these are related estimation results is a priority in our research agenda. Efficient estimation can be also achieved by GMM estimators, along the lines of Carrasco and Florens (2000, 2011) and Newey (2004). The results of this paper can be useful to extend existing GMM theory to our semiparametric setting. Similarly, allowing for simultaneous estimation of parametric and nonparametric components, as in Ai and Chen (2003) or Chen and Pouzo (2012), should be important.

There are also other open questions that remain for future research. We have not addressed the issue of "bandwidth" choice. Note that in our setting this is a very complicated matter, since our problem is one of testing, and a general theory of bandwidth choice for testing is not available, even in much simpler settings than ours. Developing this theory is beyond the scope of this paper. It seems reasonable to first obtain such theory for the estimation problem, for which related results are available for comparison. Monte Carlo experiments will be carried out to evaluate the finite sample performance of the proposed tests and estimators.

Many applications involve time series data, so it would be important to allow for dependence. The main difficulty in extending our results to time series is the lack of an efficiency theory in the general semiparametric setting considered here. For specific models and dependence structures, for instance, Markov processes, efficiency results are available and our results can be straightforwardly extended; see Carrasco and Florens (2011) for important results in this direction. See also Hallin and Werker (2003) for a general theory of efficiency allowing for dependent data. We have applied the FLRT to finite-dimensional parameters, but it could be also applied to infinite-dimensional parameters. It is unknown whether or not the FLRT delivers in this case optimal inference. This extension would have important applications in partially identified models and quantile regression models, among many others.

7 Appendix

7.1 Appendix A

7.1.1 Sufficient conditions for Assumption W

In this section, we establish the weak convergence of $\hat{R}_n$ in (4) as a random element in $L_2(\mu)$. The function space Θ_η is endowed with a pseudo-metric $\|\cdot\|_\eta$, which is a sup-norm with respect to x, and a pseudo-metric with respect to Z. An example is $\|\eta\|_\eta = \sup_{z\in\mathcal{Z},x\in\Gamma}|\eta(z,x)|$. Define a δ-enlargement of the parameter sets $\Theta_\beta(\delta) := \{\beta \in \Theta_\beta : |\beta-\beta_0| \le \delta\}$ and $\Theta_\eta(\delta) := \left\{\eta\in\Theta_\eta : \|\eta-\eta_0\|_\eta \le \delta\right\}$ for $\delta > 0$. Define $R(x,\beta,\eta) := E\left[\psi(Z,x,\beta,\eta)\right]$ and

$$R_n(x,\beta,\eta) := \frac{1}{n}\sum_{i=1}^{n}\psi(Z_i,x,\beta,\eta).$$

We first introduce the definition of pathwise functional derivative to deal with the estimation effects of $\widehat{\eta}_n$. For each $(x,\beta,\eta)\in\Gamma\times\Theta$, we say that $R(x,\beta,\eta)$ is pathwise differentiable at $\eta\in\Theta_\eta$ in the direction $[\overline{\eta}-\eta]$ if $\{\eta+\lambda(\overline{\eta}-\eta) : \lambda\in[0,1]\}\subset\Theta_\eta$ and

$$\lim_{\lambda\to 0}\frac{R(x,\beta,\eta+\lambda(\overline{\eta}-\eta))-R(x,\beta,\eta)}{\lambda} \text{ exists;}$$

the derivative is denoted as $V_\eta(x,\beta,\eta)[\overline{\eta}-\eta]$. For the weak convergence we need the following assumptions. Henceforth, C is a generic constant.

Assumption A1: *Suppose that:*

(i) (Smoothness in η) for each $x\in\Gamma$, the pathwise derivative $V_\eta(x,\beta_0,\eta_0)[\eta-\eta_0]$ of $R(x,\beta_0,\eta)$ at $\eta=\eta_0$ exists in all directions $[\eta-\eta_0]\in\Theta_\eta$; and for all $(x,\eta)\in\Gamma\times\Theta_\eta(\delta_n)$ with a positive sequence $\delta_n\to 0$, it holds that

$$\sup_{x\in\Gamma}|R(x,\beta_0,\eta)-R(x,\beta_0,\eta_0)-V_\eta(x,\beta_0,\eta_0)[\eta-\eta_0]| \le C\|\eta-\eta_0\|_\eta^2. \tag{25}$$

(ii) $P(\widehat{\eta}\in\Theta_\eta)\to 1$, and $\|\widehat{\eta}-\eta_0\|_\eta = o_P\left(n^{-1/4}\right)$.

(iii) (Stochastic Equicontinuity) for all sequences of positive numbers $\delta_n\to 0$,

$$\sup_{(x,\eta)\in\Gamma\times\Theta_\eta(\delta_n)}|R_n(x,\beta_0,\eta)-R(x,\beta_0,\eta)-R_n(x,\beta_0,\eta_0)+R(x,\beta_0,\eta_0)| = o_P\left(n^{-1/2}\right). \tag{26}$$

(iv) $\sqrt{n}V_\eta(x,\beta_0,\eta_0)[\widehat{\eta}-\eta_0]$ admits an asymptotic expansion (uniformly in x):

$$\sqrt{n}V_\eta(x,\beta_0,\eta_0)[\widehat{\eta}-\eta_0] = \frac{1}{\sqrt{n}}\textstyle\sum_{i=1}^{n}\phi(Z_i,x,\beta_0,\eta_0)+o_P(1).$$

Assumptions A1(i)-(iv) are uniform versions (in x) of related assumptions in Chen, Linton and Van Keilegom (2003). These assumptions are discussed extensively in the literature. Related assumptions are given in Escanciano and Zhu (2012) for the analysis of semiparametric partially identified models. For a fixed x, the results in Newey (1994) can be applied to find the expression for ϕ. Define

$$m\left(z, x, \beta, \eta\right) := \psi\left(z, x, \beta, \eta\right) + \phi\left(z, x, \beta, \eta\right), \tag{27}$$

where ϕ is as in A1(iv).

THEOREM A1: *Under Assumption A1 and H_0, the following expansion holds:*

$$\sup_{x\in\Gamma}\left|\sqrt{n}\hat{R}_n(x) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n} m(Z_i, x, \beta_0, \eta_0)\right| = o_P\left(1\right).$$

PROOF OF THEOREM A1: Define the linear approximation

$$\mathcal{L}_n\left(x, \eta_0\right) := R_n(x, \eta_0) + V_\eta\left(x, \eta_0\right)[\widehat{\eta} - \eta_0].$$

First, by Assumption A1(i-iii), uniformly in $x \in \Gamma$,

$$\begin{aligned}
&\left|\hat{R}_n(x) - \mathcal{L}_n\left(x, \eta_0\right)\right| \\
&\leq \left|\hat{R}_n(x) - R(x, \widehat{\eta}) - R_n(x, \eta_0) + R(x, \eta_0)\right| \\
&+ \left|R(x, \widehat{\eta}) + R_n(x, \eta_0) - R(x, \eta_0) - \mathcal{L}_n\left(x, \eta_0\right)\right| \\
&\leq \left|\hat{R}_n(x) - R(x, \widehat{\eta}) - R_n(x, \eta_0) + R(x, \eta_0)\right| \\
&+ \left|R(x, \widehat{\eta}) - R(x, \eta_0) - V_\eta\left(x, \eta_0\right)[\widehat{\eta} - \eta_0]\right| \\
&= o_P\left(n^{-1/2}\right).
\end{aligned}$$

Hence, we conclude from Assumption A1(iv) that, uniformly in $x \in \Gamma$,

$$\hat{R}_n(x) = M_n(x, \beta_0) + o_P\left(n^{-1/2}\right),$$

where

$$M_n(x, \beta) := \frac{1}{n}\sum_{i=1}^{n} m(Z_i, x, \beta, \eta_0).$$

■

We obtain the following corollary, whose proof is omitted; see Politis and Romano (1994).

COROLLARY A1: *Under Assumption A1, $E[\|m(Z_i, \cdot)\|^2] < \infty$ and H_0:*

$$\sqrt{n}\hat{R}_n \Longrightarrow R^0_\infty, \text{ in } L_2(\mu)$$

where R^0_∞ is as in Assumption W.

We now introduce a formal description of the local alternatives considered, and the limiting distribution of $\sqrt{n}\hat{R}_n$ under local alternatives. We follow CHS. Define the local parameters, $t \in [0, \infty)$,

$$\begin{aligned}
\beta_t &:= \beta_0 + tc_\beta + r_{\beta t} \text{ and} \\
\eta_t &:= \eta_0 + tc_\eta + r_{\eta t},
\end{aligned} \tag{28}$$

where $c_\beta \in \mathcal{H}_\beta$, $c_\eta \in \mathcal{H}_\eta$, $|r_{\beta t}| = o(t)$, $\|r_{\eta t}\|_{l\eta} = o(t)$, as $t \downarrow 0$. Here $\mathcal{H}_\beta$ is a local parameter space that is a subset of $\mathbb{R}^p$ containing zero and $\mathcal{H}_\eta$ is the local nuisance parameter space that is assumed to be a Hilbert space with norm $\|\cdot\|_{l\eta}$. With some abuse of notation, denote by $\langle \cdot, \cdot \rangle$ the inner product in $\mathcal{H} := \mathcal{H}_\beta \times \mathcal{H}_\eta$. Note that $c = (c_\beta, c_\eta)$ denotes the direction in which the local parameter $\theta_t(c) := (\beta_t(c_\beta), \eta_t(c_\eta))$ deviates from the point (β_0, η_0). We think of the parameter $\theta_t(c)$ as the parameter corresponding to a smooth regular parametric submodel passing through $P \equiv P_{\theta_0}$. We define this important concept as follows. Let $\mathcal{P} := \{P_\theta : \theta \equiv (\beta, \eta),\ \beta \in \Theta_\beta,\ \eta \in \Theta_\eta\}$ be the semiparametric model satisfying (1). Let ν be a $\sigma-$finite measure dominating P_θ, and let $f(z|\,\theta)$ be the corresponding density. $\mathcal{P}_0 := \{P_t : t \in [0, \infty)\}$ is a smooth regular parametric submodel passing through $P \equiv P_{\theta_0}$ if $\mathcal{P}_0 \subset \mathcal{P}$, $P_0 = P$ and the density of P_t, say f_t, is mean-square differentiable, i.e.,

$$\int \left| \frac{f_t^{1/2} - f_0^{1/2}}{t} - \frac{1}{2} g f_0^{1/2} \right| d\nu \to 0 \text{ as } t \to 0, \tag{29}$$

where g is a measurable function, that necessarily satisfies $E[g(Z)] = 0$ and $E[g^2(Z)] < \infty$. We then define formally the local alternatives as

$$H_n : P \sim P_{\theta_{nc}},$$

where $\theta_{nc} := \theta_{n^{-1/2}}(c)$ and $P_t \equiv P_{\theta_t(c)}$ is a smooth parametric submodel with fixed c and $c_\beta \neq 0$. Henceforth, define for a measurable function q

$$E_{\theta_t(c)}\left[q(Z)\right] \equiv E_t\left[q(Z)\right] := \int q(z) f_t(z) dz,$$

Then, it is well known that an important implication of (29) is the LAN property,

$$L_n(c) := \log \prod_{i=1}^{n} \frac{dP_{\theta_{nc}}}{dP_{\theta_0}}(Z_i) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(Z_i) - \frac{1}{2} E[g^2(Z)] + o_p(1); \tag{30}$$

see e.g. van der Vaart (1998, Theorem 7.2). Obviously, the score g depends on the direction c (cf. van der Vaart (1990)), and we write

$$g =: g_\beta' c_\beta + g_\eta c_\eta \text{ a.s.},$$

where g_β and g_η are the scores for β and η, respectively (here g_η is a linear bounded operator from $\mathcal{H}_\eta$ to $\mathbb{R}$). We need the following regularity condition:

Assumption A2: *For all smooth parametric submodels and each $x \in \Gamma$, the map $t \to E[m(Z, x, \beta_t, \eta_t)]$ is continuously differentiable at $t = 0$ and $\sup_{t \in \mathcal{N}} E_t\left[m^2(Z, x, \beta_0, \eta_0)\right] < \infty$, where $\mathcal{N}$ is a neighborhood of 0. The parameter β_0 belongs to the interior of Θ_β.*

Theorem A2: *Under Assumptions A1 and A2, Assumption W holds.*

Proof of Theorem A2: To establish the limiting distribution of $\sqrt{n}\hat{R}_n$ under H_n we apply Lecam's third lemma in van der Vaart and Wellner (1996, Theorem 3.10.7). To characterize the limit, we first use (30), Assumption A1, and apply Lecam's third lemma to $\left\langle \sqrt{n}\hat{R}_n, h \right\rangle$ with a fixed $h \in L_2(\mu)$, which yields that under H_n

$$\left\langle \sqrt{n}\hat{R}_n, h \right\rangle \to_d N(\tau, \langle h, Kh \rangle),$$

where

$$\tau := E[\langle m, h \rangle (Z) g(Z)].$$

By the adjoint property $\tau = \langle h, T'g \rangle$. Since this is true for all $h \in L_2(\mu)$, we conclude that under H_n,

$$\sqrt{n}\hat{R}_n \Longrightarrow R^0_\infty + T'g, \text{ in } L_2(\mu).$$

It remains to prove that $T'g = c'_\beta D$. The part of the score corresponding to the nuisance parameter satisfies $T'g_\eta \equiv 0$ by Theorem 2 below, and hence it suffices to prove that $T'g_\beta = c'_\beta D$, where g_β is the score corresponding to β with η_0 fixed. But this follows from the classical information equality, see Lemma 7.2 in Ibragimov and Hasminskii (1981), under Assumption A2. ■

7.1.2 Semiparametric efficient tests

For completeness, we review some concepts on efficient testing in semiparametric models. We follow the notation of CHS, where further details can be found. Write (30) in terms of linear functionals as

$$L_n(c) = S_n c - \frac{1}{2}\sigma^2(c) + r_n(c),$$

where $S_n = (S_{n\beta}, S_{n\eta})'$ is a random linear functional which is asymptotically centered Gaussian with covariance operator B under the null hypothesis, and $r_n(c) = o_p(1)$ for every c under the null hypothesis. Hence the variance $\sigma^2(c)$ of $S_n c$ is equal to $\langle c, Bc \rangle$.

A test ψ_n takes values in $\{0, 1\}$, where 1 represents rejection and 0 acceptance of the null hypothesis, respectively. Consider first the one-sided testing problem with scalar β, i.e.

$$H_0 : c_\beta = 0 \text{ against } H_1 : c_\beta > 0.$$

Fix $c_0 = (0, c_\eta)$. A test ψ_n is of *asymptotic level* $\alpha \in (0, 1)$ if

$$\limsup_n E_{\theta_{nc_0}} \psi_n \leq \alpha,$$

for every c_η. Henceforth, we restrict the analysis to the class of asymptotic level α tests.

Then, using the LAN property, we can write

$$\begin{aligned} E_{\theta_{nc}} \psi_n &= E\psi_n \exp\left(L_n(c)\right) + o(1) \\ &= E\psi_n \exp\left(S_n c - \frac{1}{2}\sigma^2(c) + r_n(c)\right) + o(1). \end{aligned}$$

Fix $c_1 = (c_{1\beta}, c_{1\eta})$ with $c_{1\beta} > 0$, and consider testing the simple hypothesis $c_0 = (0, c_\eta)$ against c_1. Then, the Neyman-Pearson lemma gives an optimal test φ_n of asymptotic level α in the following form: $\varphi_n = 1$ if

$$S_n(c_1 - c_0) - \frac{1}{2}\{\sigma^2(c_1) - \sigma^2(c_0)\} + r_n(c_1) - r_n(c_0) > c_n;$$

and $\varphi_n = 0$ otherwise. And for this test, it is a straightforward to obtain the following bound for the (local) power of the test

$$\limsup E_{\theta_{nc_1}} \varphi_n \leq 1 - \Phi(z_\alpha - \sigma(c_1 - c_0)), \tag{31}$$

where z_α is the upper α-quantile of the standard normal distribution function Φ.

Now, we aim to devise a test that is uniformly most powerful at each point of $c_\eta \in \mathcal{H}_\eta$. The bound for the power of the test is attained by an optimal test against a simple alternative corresponding to the *least favorable direction*. Let $(B_{ij})_{i,j=1,2}$ denote the partition of B such that B_{11} is the information for β, B_{22} is the information for η, and B_{12} and B_{21} are co-informations. Obviously, from (31), the least favorable direction is obtained by minimizing $\sigma(c_1 - c_0)$ in c_η and is found to be $c_\eta^* = c_{1\eta} + B_{22}^{-1} B_{21} c_{1\beta}$. Hence, the point $(0, c_\eta^*)$ is the projection of c_1 onto the local null space under the inner product induced by B, namely, $\langle h, g \rangle_B = \langle h, Bg \rangle,\ h, g \in \mathcal{H}$. By plugging in this least favorable direction, we obtain

$$\limsup E_{\theta_{nc_1}} \psi_n \leq 1 - \Phi(z_\alpha - \sigma(B^{*1/2} c_{1\beta}))$$

where $B^* = B_{11} - B_{12} B_{22}^{-1} B_{21}$ is the *efficient information*. Let us define the efficient score S_n^* as $S_n^* a = S_{n\beta} a - S_{n\eta} B_{22}^{-1} B_{21} a$, $a \in \mathbb{R}$. Since c_β is a scalar, so are $S_{n\beta}$ and S_n^*. Note that S_n^* depends on η_0 and we write $S_n^*(\eta_0)$ explicitly. Define the standardized efficient score $\xi_n(\eta_0) := B^{*-1/2} S_n^*(\eta_0)$. Now, an optimal test is obtained by taking $\varphi_n = 1\left\{\xi_n(\eta_0) \geq z_\alpha\right\}$. The resulting test φ_n does not depend on $c_1 = (c_{1\beta}, c_{1\eta})$. Hence, the test is asymptotically uniformly most powerful (AUMP(α, η_0)) at the level α and at the nuisance parameter η_0.

The procedure easily applies to a two-sided test. A test ψ_n is *asymptotically unbiased* at η_0 if $\limsup_n E_{\theta_{nc_0}} \psi_n \leq \liminf_n E_{\theta_{nc_1}} \psi_n$ for every $c_0 = (0, c_\eta)$ and $c_1 = (c_{1\beta}, c_{1\eta})$ with $c_{1\beta} \neq 0$. Then, Theorem 2 of CHS gives the following bound for the local power:

$$\limsup E_{\theta_{nc}} \psi_n \leq \Phi(|B^{*1/2} c_\beta| - z_{\alpha/2}) + \Phi(-|B^{*1/2} c_\beta| - z_{\alpha/2})$$

for all $c = (c_\beta, c_\eta) \in \mathcal{H}$. The two-sided test that is AUMP and unbiased, in short AUMPU(α, η_0), among the asymptotically unbiased tests is given by

$$1\left\{|\xi_n(\eta_0)| \geq z_{\alpha/2}\right\}.$$

For the multivariate case $p \geq 1$, the class of tests is restricted to satisfy an asymptotic rotation invariance property, see CHS (p. 851) for details, and the AUMPI(α, η_0) test is given by

$$1\left\{\xi_n'(\eta_0)\xi_n(\eta_0) \geq \chi^2_{1-\alpha,p}\right\}.$$

In the above efficiency concepts, when the test does not depend on η_0, the reference to the nuisance parameter is dropped.

7.1.3 Preliminary results

We collect in this section a number of known results that will be instrumental in proving Theorem 4. We refer to references for the proofs. We begin with an important result of Chen, Linton and van Keilegom

(2003) that allows for the bounding of entropy numbers and the verification of stochastic equicontinuity for processes indexed by both parametric and nonparametric parameters. In this connection, define a generic class

$$\mathcal{H} = \{z \to m(z, \theta, g) : \theta \in \Theta, g \in \mathcal{G}\},$$

where Θ and $\mathcal{G}$ are Banach spaces with associated norms $\|\cdot\|_\Theta$ and $\|\cdot\|_\mathcal{G}$, respectively. Recall that the covering number $N(\epsilon, \Theta, \|\cdot\|_\Theta)$ of Θ is the minimal number N for which there exist ϵ-neighborhoods $\{\{\theta : \|\theta - \theta_j\|_\Theta \leq \epsilon\}, \|\theta_j\|_\Theta < \infty, j = 1, \ldots, N\}$ covering Θ. A bracket $[l_j, u_j]$ is the set of elements $\theta \in \Theta$ such that $l_j \leq \theta \leq u_j$. The covering number with bracketing $N_{[\cdot]}(\epsilon, \Theta, \|\cdot\|_\Theta)$ is the minimal N for which there exist ϵ-brackets $\{[l_j, u_j] : \|l_j - u_j\|_\Theta \leq \epsilon, \|l_j\|_\Theta, \|u_j\|_\Theta < \infty, j = 1, \ldots, N\}$ covering Θ. An envelope function G for the class $\mathcal{G}$ is a measurable function such that $G(x) \geq \sup_{g \in \mathcal{G}} |g(x)|$. Define the entropy number

$$J(\delta, \mathcal{G}, \|\cdot\|_{2,P}) := \int_0^\delta \sqrt{\log N(\varepsilon, \mathcal{W}, \|\cdot\|_{2,P})} d\varepsilon.$$

Henceforth, we abstract from measurability issues in some of the expectations involved. The interested reader can check van der Vaart and Wellner (1996) for solutions to potential lack of measurability, and for basic definitions in empirical processes theory.

Lemma Q1. *Assume that*

$$E\left[\sup_{\theta_2 : \|\theta_1 - \theta_2\|_\Theta < \delta} \sup_{g_2 : \|g_1 - g_2\|_\mathcal{G} < \delta} |m(Z, \theta_1, g_1) - m(Z, \theta_2, g_2)|^2\right] \leq C\delta^s$$

for some constant $s \in (0, 2]$. *Then for any* $\epsilon > 0$,

$$N_{[\cdot]}(\epsilon, \mathcal{H}, \|\cdot\|_{2,P}) \leq N\left(\left[\frac{\epsilon}{2C}\right]^{2/s}, \Theta, \|\cdot\|_\Theta\right) \times N\left(\left[\frac{\epsilon}{2C}\right]^{2/s}, \mathcal{G}, \|\cdot\|_\mathcal{G}\right).$$

A typical application of Lemma Q1 implies that $J(\delta, \mathcal{H}, \|\cdot\|_{2,P}) < \infty$, and hence, that the empirical process $\sqrt{n}(M_n - M)$, where $M_n(\theta, g) \equiv n^{-1}\sum_{i=1}^n m(Z_i, \theta, g)$ and $M(\theta, g) \equiv E[m(Z_i, \theta, g)]$, is asymptotically stochastically equicontinuous, i.e., for any sequence of positive constants $\delta_n = o(1)$,

$$\sup_{\|\theta_1 - \theta_2\|_\Theta \leq \delta_n, \|g_1 - g_2\|_\mathcal{G} \leq \delta_n} |M_n(\theta_1, g_1) - M_n(\theta_2, g_2) - M(\theta_1, g_1) + M(\theta_2, g_2)| = o_P(n^{-1/2}). \qquad (32)$$

The following lemma is implicit in Section 2.10.3 of van der Vaart and Wellner (1996).

Lemma Q2. *Let* $\mathcal{F}$ *and* $\mathcal{G}$ *be classes of functions with envelopes* F *and* G, *respectively, then, for any* $\epsilon > 0$,

$$N(2\epsilon\|FG\|_{2,P}, \mathcal{F}\cdot\mathcal{G}, \|\cdot\|_{2,P}) \leq N(\epsilon\|F\|_{2,P}, \mathcal{F}, \|\cdot\|_{2,P}) \times N(\epsilon\|G\|_{2,P}, \mathcal{G}, \|\cdot\|_{2,P}).$$

We now state a weak convergence theorem that is useful in dealing with estimation effects in test functionals involving the non-smooth summands $\zeta_i(\tau, \delta) = 1(Y_i \leq X_i'\delta(\tau)) - \tau$. Let $a(\cdot)$ be a bounded measurable function of Z_i, and let $\mathcal{B}$ be a class of Lipschitz and bounded functions from $\mathcal{T}$ to Θ_δ. Given a sequence $\{Z_{in}\}_{i=1}^n$ of iid arrays for each n, define the weighted empirical process

$$V_n(\delta, x) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(a(Z_{in})\zeta_{in}(\tau, \delta) - E\left[a(Z_{in})\zeta_{in}(\tau, \delta) \middle| X_{in}\right]\right) q_n(X_{in}, \delta, x),$$

which is indexed by $\chi := (\delta, x) \in \mathcal{B} \times \Gamma$, where $\Gamma := \mathcal{T} \times \mathbb{R}^{d_x}$. Let F_X denote the cdf of X. Define the pseudo-metric, for $\chi_1 := (\delta_1, \tau_1, w_1) \in \mathcal{B} \times \Gamma$,

$$\rho(\chi, \chi_1) := |\tau - \tau_1| + |F_X(w) - F_X(w_1)| + \|\delta - \delta_1\|_{\mathcal{T}},$$

where $\|\delta\|_{\mathcal{T}} := \sup_{\tau \in T} |\delta(\tau)|$, and assume that q_n is such that for $\epsilon_n \downarrow 0$

$$\sup_{\rho(\chi, \chi_1) < \epsilon_n} \|q_n(\cdot, \chi) - q_n(\cdot, \chi_1)\|_{2,P} = o(1)$$

and $W_n := \sup_\chi |q_n(\cdot, \chi)|$ satisfies the Lindeberg condition, for each $\varepsilon > 0$,

$$E[W_n^2] = O(1) \text{ and } E[W_n^2 1(W_n > \varepsilon\sqrt{n})] = o(1).$$

Furthermore, define the class $\mathcal{W}_n := \{q_n(\cdot, \delta, x) : (\delta, x) \in \mathcal{B} \times \Gamma\}$ and require the following assumption:

Assumption Q1. *The class $\mathcal{W}_n$ satisfies the previous conditions and is such that $J(\epsilon_n, \mathcal{W}_n, \|\cdot\|_{2,P}) \to 0$ for every $\epsilon_n \downarrow 0$.*

Theorem Q1. *Under Assumptions E1, E2 in Section 5.1 and Q1, the process V_n is $\rho-$stochastically equicontinuous.*

Proof of Theorem Q1. It follows from an application of Theorem 19.28 in van der Vaart (1998) and Lemma Q1. ■

Lemma Q3. *Under Assumption E1,*

$$\sup_{\tau \in \mathcal{T}} \left| \sqrt{n}\left(\widehat{\gamma}_n(\tau) - \gamma_0(\tau)\right) - \frac{1}{\sqrt{n}} B^{-1}(\tau) \sum_{i=1}^{n} \zeta_i(\tau) X_{2i} f_{i\tau} \right| = o_P(1).$$

Proof of Lemma Q3. See for instance Gutenbrunner and Jurecková (1992). ■

Lemma Q4. *Under Assumption E1, the estimator $\widehat{\gamma}_n(\cdot)$ satisfies that $\widehat{\gamma}_n \in \mathcal{B}$ with probability tending to one and $\gamma_0 \in \mathcal{B}$.*

Proof of Lemma Q4. Follows from Lemma Q3 and Assumption E1 in a routine fashion. ■

Our next result is related to the uniform convergence rates for the kernel estimator $\hat{f}_{i\tau}$. We view $\hat{f}_{i\tau}$ as a function of $\widehat{\delta}_0$ and write the Taylor approximation around the true value δ_0 as

$$\hat{f}_{i\tau} = \tilde{f}_{i\tau} + \dot{f}_{i\tau}(\widehat{\Delta}) + \ddot{f}_{i\tau}(\widehat{\Delta}) + r_{i\tau}, \tag{33}$$

where $\tilde{f}_{i\tau} := \hat{f}\left(X_i'\delta_0(\tau)\right| X_i, \delta_0\right)$, $\widehat{\Delta}(\cdot) := \sqrt{n}(\widehat{\delta}_0(\cdot) - \delta_0(\cdot))$,

$$\dot{f}_{i\tau}(\Delta) := \frac{1}{n^{3/2}h^2}\sum_{j=1}^{n}\dot{K}\left(\frac{X_i'\delta_0(\tau) - X_i'\delta_0(\tau_j)}{h}\right)\left\{X_i'(\Delta(\tau_j) + \Delta(\tau))\right\},$$

and

$$\ddot{f}_{i\tau}(\Delta) := \frac{1}{n^2h^3}\sum_{j=1}^{n}\ddot{K}\left(\frac{X_i'\delta_0(\tau) - X_i'\delta_0(\tau_j)}{h}\right)\left\{X_i'(\Delta(\tau_j) + \Delta(\tau))\right\}^2,$$

and where, henceforth, for a generic function K we denote $\dot{K}(t) := \partial^{(1)}K(t)/\partial t$ and $\ddot{K}(t) := \partial^{(2)}K(t)/\partial t^2$. The remainder term $r_{i\tau}$ is implicitly defined.

The proofs of the results below directly follow from Escanciano and Goh (2012). For a_n and b_n as in Assumption E3(b), define

$$d_n := \sqrt{\frac{\log a_n^{-1} \vee \log\log n}{na_n}} + b_n^2.$$

Lemma Q5. *Under Assumptions E1-E3,*

$$\sup_{a_n \le h \le b_n}\sup_{\tau\in\mathcal{T}}\max_{1\le i\le n}\left|\hat{f}_{i\tau} - f_{i\tau}\right| = O_p\left(n^{-1/2} + d_n\right);$$

and

$$\sup_{a_n \le h \le b_n}\sup_{\tau\in\mathcal{T}}\max_{1\le i\le n}\left|\tilde{f}_{i\tau} - f_{i\tau}\right| = O_p\left(d_n\right).$$

Similarly, we have the following uniform consistency results, see (21) and (22) for definitions of A_n and B_n.

Lemma Q6. *Under Assumptions E1-E3,*

$$\sup_{x\in\mathcal{T}\times\mathbb{R}^{d_x}}|A_n(x) - A(x)| = o_P(1)$$

and

$$\sup_{\tau\in\mathcal{T}}|B_n(\tau) - B(\tau)| = o_P(1).$$

Proof of Lemma Q6. It follows from a combination of Lemmas Q2, Q3 and Q5. ■

Define the class $\mathcal{Q} := \{z \to 1(y \le \beta_0'x_1 + \gamma'x_2) - \tau : \gamma \in \Theta_\gamma, \tau \in \mathcal{T}\}$, where $\Theta_\delta =: \Theta_\beta \times \Theta_\gamma \subset \mathbb{R}^{d_x}$. The proof of the following result is standard, and hence omitted.

Lemma Q7. *Let Assumption E1 hold. Then, the class $\mathcal{Q}$ of functions satisfies $J(\epsilon_n, \mathcal{Q}, \|\cdot\|_{2,P}) \to 0$ for every $\epsilon_n \downarrow 0$..*

7.2 Appendix B: Proofs of main results

Proof of Theorem 1: To prove (i), we use Assumption C and write

$$\begin{aligned}
L(\sqrt{n}\hat{R}_n) &= \sum_{j=1}^{\infty} \lambda_j^{-1} \left\langle \sqrt{n} M_n, \varphi_j \right\rangle \left\langle D, \varphi_j \right\rangle + o_P(1) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{\infty} \lambda_j^{-1} \left\langle m(Z_i, \cdot), \varphi_j \right\rangle \left\langle D, \varphi_j \right\rangle + o_P(1) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{\infty} \lambda_j^{-1/2} \left\langle D, \varphi_j \right\rangle \psi_j(Z_i) + o_P(1) \\
&= S_n^* + o_P(1).
\end{aligned}$$

As for (ii), note that by Kress (1999, Theorem 15.16) and Assumption D,

$$T' s^* = \sum_{j=1}^{\infty} \left\langle D, \varphi_j \right\rangle \varphi_j = \Pi_{\overline{Im(T')}} D = D,$$

and

$$\begin{aligned}
s^* &= \sum_{j=1}^{\infty} \lambda_j^{-1/2} \left\langle T' s, \varphi_j \right\rangle \psi_j \\
&= \sum_{j=1}^{\infty} E[s(Z)\psi_j(Z)]\psi_j \\
&= \Pi_{\overline{Im(T)}} s \equiv \Pi_{\ker^{\perp}(T')} s.
\end{aligned}$$

■

Proof of Theorem 2: Let $P_{(\beta_0,\eta_t)}$, $t \in [0,\varepsilon)$, $\varepsilon > 0$, be a regular parametric submodel passing through $P_{(\beta_0,\eta_0)}$, with score $s(Z)$. Hence, the model satisfies the restrictions

$$E_t \left[\psi(Z, x, \beta_0, \eta_t)\right] = 0.$$

Differentiating with respect to t and evaluating at $t = 0$, we obtain, by the chain rule,

$$\frac{\partial}{\partial t} E \left[\psi(Z, x, \beta_0, \eta_t(Z, x))\right]\bigg|_{t=0} + \frac{\partial}{\partial t} E_t \left[\psi(Z, x, \beta_0, \eta_0(Z, x))\right]\bigg|_{t=0} = 0.$$

The first term is just the derivative of $\gamma(t) := E\left[\psi(Z, x, \beta_0, \eta_t(Z, x))\right]$, which by our Assumption A1 satisfies

$$\frac{\partial \gamma(0)}{\partial t} = E\left[\phi\left(Z_i, x, \beta_0, \eta_0\right) s(Z)\right],$$

see e.g. (3.9) in Newey (1994). On the other hand, Lemma 7.2 in Ibragimov and Hasminskii (1981), under Assumption A2 implies that

$$\frac{\partial}{\partial t} E_t \left[\psi(Z, x, \beta_0, \eta_0(Z, x))\right]\bigg|_{t=0} = E\left[\psi(Z, x, \beta_0, \eta_0(Z, x)) s(Z)\right].$$

Hence, the score satisfies $s(Z) \in \ker(T')$, so that $\dot{\mathcal{P}}_2 \subset \ker^0(T')$.

We now prove that $\ker^0(T') \subset \dot{\mathcal{P}}_2$ holds. Define the map $\gamma : \mathcal{P} \to L_2(\mu)$ as

$$\gamma(P) := E_P[\psi(Z, x, \beta_0, \eta(P))].$$

The same arguments above show that γ is Frechet differentiable at P_0, viewed as a mapping on square roots of measures, with derivative $\dot{\gamma} = T'$; see e.g. van der Vaart (1998, p. 363). Then, for a given function $s \in \ker^0(T')$ we can use exactly the same arguments as in Bickel et al. (1993, pg. 54) to construct a parametric submodel with score s and passing through P_0. Thus, we conclude that $\dot{\mathcal{P}}_2 = \ker^0(T')$.

As for (ii), consider a parametric submodel satisfying $E_t\left[m(Z, x, \beta_t, \eta_0)\right] = 0$ with score $c'_\beta \dot{\ell}_\beta(Z)$. Differentiating this equation with respect to t at 0 we get

$$c'_\beta \frac{\partial E\left[m(Z, x, \beta_0, \eta_0)\right]}{\partial \beta} + \partial E_t\left[m(Z, x, \beta_0, \eta_0)\right]|_{t=0} = 0.$$

Regularity of the model and Assumption A2 imply, by Lemma 7.2 in Ibragimov and Hasminskii (1981),

$$\partial E_t\left[m(Z, x, \beta_0, \eta_0)\right]|_{t=0} = c'_\beta E\left[m(Z, x, \beta_0, \eta_0)\dot{\ell}_\beta(Z)\right],$$

where $\dot{\ell}_\beta$ is the score with respect to β at β_0. Since the previous equality holds for all c_β, we conclude using our notation that

$$D = T'\dot{\ell}_\beta. \tag{34}$$

Hence, by part (i), the zero mean property of scores and Theorem 1(ii) $\dot{\ell}^*_\beta = \dot{\ell}_\beta - \Pi_{\dot{\mathcal{P}}_2}\dot{\ell}_\beta = \Pi_{\ker^\perp(T')}\dot{\ell}_\beta = s^*$. ■

The following result is fundamental for many of the proofs that follow. Its proof is trivial, and hence omitted. Define $\{\varepsilon_{nj}\}_{j=1}^\infty$ as

$$\begin{aligned}\varepsilon_{nj} &:= n^{-1/2}\sum_{i=1}^n \psi_j(Z_i) \\ &= \lambda_j^{-1/2}\sqrt{n}\left\langle M_n, \varphi_j\right\rangle.\end{aligned}$$

Lemma Q8. *$\{\varepsilon_{nj}\}_{j=1}^\infty$ are uncorrelated and with unit variance.*

Proof of Proposition 1: We first prove that $L_k(\hat{R}_n) = L_k(M_n) + o_P(n^{-1/2})$. Note that

$$\begin{aligned}L_k(\hat{R}_n) - L_k(M_n) &= \sum_{j=1}^k \lambda_j^{-1}\left\langle \hat{R}_n - M_n, \varphi_j\right\rangle\left\langle \hat{D}, \varphi_j\right\rangle \\ &= \sum_{j=1}^k \lambda_j^{-1}\left\langle \hat{R}_n - M_n, \varphi_j\right\rangle\left\langle \hat{D} - D, \varphi_j\right\rangle \\ &+ \sum_{j=1}^k \lambda_j^{-1}\left\langle \hat{R}_n - M_n, \varphi_j\right\rangle\left\langle D, \varphi_j\right\rangle.\end{aligned}$$

By Cauchy-Schwarz's inequality and Assumptions D and R the absolute values of all terms are $o_P(n^{-1/2})$. Using Lemma Q8, we conclude that

$$
\begin{aligned}
\left|\sum_{j=1}^{k}\lambda_j^{-1}\left\langle M_n,\varphi_j\right\rangle\left\langle\hat{D}-D,\varphi_j\right\rangle\right| &\leq \left(\sum_{j=1}^{k}\lambda_j^{-1}\left\langle M_n,\varphi_j\right\rangle^2\right)^{1/2}\left(\sum_{j=1}^{k}\lambda_j^{-1}\left\langle\hat{D}-D,\varphi_j\right\rangle^2\right)^{1/2} \\
&= O_P(k^{1/2}n^{-1/2})o_P(k^{-1/2}) \\
&= o_P(n^{-1/2}).
\end{aligned}
$$

Hence, using Lemma Q8 one more time, it can be shown that $L_k(M_n) = L(M_n) + o_P(n^{-1/2})$. The proof that $\widehat{\Sigma}_k = \Sigma + o_P(1)$ is trivial, and hence, it is omitted. ■

Proof of Theorem 3: Write

$$
\begin{aligned}
\sqrt{n}\widehat{L}_n &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}s^*(Z_i) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\widehat{s}^*(Z_i) - s^*_{\alpha_n}(Z_i)\} \\
&+ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{s^*_{\alpha_n}(Z_i) - s^*(Z_i)\} \\
&\equiv S_n^* + C_n^* + B_n^*.
\end{aligned}
$$

We first prove that the bias term satisfies $B_n^* = o_P(1)$. Using well-known expansions for $s^*_{\alpha_n}$ and s^* we write

$$
\begin{aligned}
B_n^* &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{\infty}b(\alpha,\lambda_j)E[s^*(Z)\psi_j(Z)]\psi_j(Z_i) \\
&= \sum_{j=1}^{\infty}b(\alpha,\lambda_j)E[s^*(Z)\psi_j(Z)]\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_j(Z_i) \\
&\equiv \sum_{j=1}^{\infty}b(\alpha,\lambda_j)E[s^*(Z)\psi_j(Z)]\varepsilon_{nj},
\end{aligned}
$$

where $b(\alpha,\lambda) = \alpha/(\alpha+\lambda)$. By Lemma Q8, $E[B_n^*] = 0$ and

$$
E[(B_n^*)^2] = \sum_{j=1}^{\infty}b^2(\alpha,\lambda_j)\left(E[s^*(Z)\psi_j(Z)]\right)^2 \to 0,
$$

as $\alpha \to 0$, by dominated convergence.

We now prove that $C_n^* = o_P(1)$. We write

$$
\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\widehat{s}^* - s^*_{\alpha_n})(Z_i) = \int\sqrt{n}(\widehat{s}^* - s^*_{\alpha_n})(z)F_n(dz),
$$

where F_n is the empirical distribution of $\{Z_i\}_{i=1}^n$.

Henceforth, we will make use of the following basic result. If B_n is a possibly random operator from $L_2(\mu)$ to $L_2(F)$ and h_n is a random element taking values in $L_2(\mu)$, then

$$\left| \frac{1}{n} \sum_{i=1}^{n} B_n h_n(Z_i) \right| = O_P \left(\|B_n\|_{2,P} \|h_n\| \right). \tag{35}$$

Define the norms $\|\cdot\|_{2,n}$ and $\|\cdot\|_n$ as the norms $\|\cdot\|_{2,P}$ and $\|\cdot\|$ but with F and μ replaced by the empirical distribution functions of $\{Z_i\}_{i=1}^n$ and $\{x_i\}_{i=1}^n$, respectively. Introducing these norms is useful because although $\hat{T}$ is not the adjoint of $\hat{T}'$ with respect to the norms $\|\cdot\|_{2,P}$ and $\|\cdot\|$, they are duals with respect to $\|\cdot\|_{2,n}$ and $\|\cdot\|_n$. Define the inner product $\langle a(\cdot), b(\cdot)\rangle_{2,P} := E[a(Z)b(Z)]$.

Fix $a \in L_2(F)$ and $h \in L_2(\mu)$. By definition of $\hat{T}^*$ it holds that

$$\left\langle a(\cdot), \hat{T}h(\cdot) \right\rangle_{2,P} = \left\langle \hat{T}^* a(\cdot), h(\cdot) \right\rangle_n .$$

Define the operator

$$T_n^* h := \int h(x) m(z,x) \mu_n(dz).$$

Then, simple algebra shows

$$\left\langle a(\cdot), (\hat{T} - T)h(\cdot) \right\rangle_{2,P} = \left\langle (\hat{T}^* - T')a(\cdot), h(\cdot) \right\rangle_n + \langle a(\cdot), (T_n^* - T)h(\cdot)\rangle_{2,P} .$$

From this equality and Markov's inequality, we conclude that

$$\left\| \hat{T} - T \right\| \leq O_P \left(\left\| \hat{T}^* - T' \right\|_{2,P} \right) + \|T_n^* - T\| .$$

By a standard FCLT and since T' is bounded, it holds that

$$\|T_n^* - T\| = O_P \left(n^{-1/2} \right).$$

Hence, by Assumption RE(ii) we obtain

$$\left\| \hat{T} - T \right\| = O_P \left(n^{-1/2} \right).$$

Applying the same arguments to $\hat{T}' - T'$ and using the well-known equality $\|B\| = \|B'\|$ for a bounded linear operator B, we obtain $\left\| \hat{T}' - T' \right\| = O_P \left(n^{-1/2} \right)$.

Note that the previous rates imply

$$\begin{aligned} \left\| \hat{T}\hat{T}' - TT' \right\| &\leq \left\| \hat{T} - T \right\| \|T'\| + \left\| \hat{T} \right\| \left\| \hat{T}' - T' \right\| \\ &= O_P \left(n^{-1/2} \right). \end{aligned}$$

Using the definitions $\widehat{s}^*(Z_i) = \widehat{A}_{\alpha_n} \hat{T} \hat{D}$ and $s^*_{\alpha_n}(Z_i) = A_{\alpha_n} TD$, where $\widehat{A}_{\alpha_n} = (\alpha_n I + \hat{T}\hat{T}')^{-1}$ and $A_{\alpha_n} = (\alpha_n I + TT')^{-1}$, respectively, we write

$$\begin{aligned} \widehat{s}^* - s^*_{\alpha_n} &= \widehat{A}_{\alpha_n} \hat{T}(\hat{D} - \hat{T}' s^*) + \widehat{A}_{\alpha_n} \hat{T}\hat{T}' s^* - A_{\alpha_n} TD \\ &\equiv \Delta_{1n} + \Delta_{2n}. \end{aligned}$$

By the basic identity $(B^{-1} - C^{-1}) = B^{-1}(C - B)C^{-1}$, we can write

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Delta_{1n}(Z_i) = \frac{1}{n} \sum_{i=1}^{n} & \widehat{A}_{\alpha_n} (\hat{T} - T) h_n(Z_i) \\ & + \frac{1}{n} \sum_{i=1}^{n} A_{\alpha_n} T h_n(Z_i) \\ & + \frac{1}{n} \sum_{i=1}^{n} \widehat{A}_{\alpha_n} (\hat{T}\hat{T}' - TT') A_{\alpha_n} T h_n(Z_i), \end{aligned} \tag{36}$$

where, henceforth, $h_n(\cdot) := \sqrt{n}(\hat{D} - \hat{T}'s^*)(\cdot)$. Applying (35) and noting that $\left\| \widehat{A}_{\alpha_n} \right\|_{2,P} = O_P(\alpha_n^{-1})$, the first term of the last display is shown to be $o_P(1)$. As for the second, using the definition of A_{α_n}, Lemma Q8 and Cauchy-Schwarz's inequality,

$$\begin{aligned} \left| \frac{1}{n} \sum_{i=1}^{n} A_{\alpha_n} T h_n(Z_i) \right| &= \left| \sum_{j=1}^{\infty} \frac{\lambda_j^{1/2}}{\lambda_j + \alpha_n} \left\langle n^{-1/2} h_n, \varphi_j \right\rangle \varepsilon_{nj} \right| \\ &\leq \left(\sum_{j=1}^{\infty} \lambda_j \varepsilon_{nj}^2 \right)^{1/2} \left(\sum_{j=1}^{\infty} \frac{1}{(\lambda_j + \alpha_n)^2} \left\langle n^{-1/2} h_n, \varphi_j \right\rangle^2 \right)^{1/2} \\ &= O_P(1) O_P\left(\alpha_n^{-1} \left\| \hat{D} - \hat{T}'s^* \right\| \right) \\ &= o_P(1). \end{aligned}$$

To analyze the third term in (36), define $\Delta_{11n}(\cdot) := \widehat{A}_{\alpha_n} (\hat{T}\hat{T}' - TT') A_{\alpha_n} T h_n(Z_i)$. Since $\widehat{A}_{\alpha_n}$ is self-adjoint with respect to $\|\cdot\|_{2,n}$, we write by Cauchy-Schwarz's inequality

$$\begin{aligned} \left| \int \Delta_{11n}(z) F_n(dz) \right| &= \left| \langle \Delta_{11n}(\cdot), 1 \rangle_{2,n} \right| \\ &= \left| \left\langle \alpha_n^{-1} (\hat{T}\hat{T}' - TT') A_{\alpha_n} T h_n(Z_i), \alpha_n \widehat{A}_{\alpha_n} 1 \right\rangle_{2,n} \right| \\ &\leq \left\| \alpha_n^{-1} (\hat{T}\hat{T}' - TT') \right\|_{2,n} \| A_{\alpha_n} T h_n \|_{2,n} \left\| \alpha_n \widehat{A}_{\alpha_n} 1 \right\|_{2,n}. \end{aligned}$$

We analyze each of these terms. From previous arguments, we know that

$$\left\| \alpha_n^{-1} (\hat{T}\hat{T}' - TT') \right\|_{2,n} = O_P(\alpha_n^{-1} n^{-1/2}).$$

The expression for A_{α_n}, implies that

$$\begin{aligned} \alpha_n \| A_{\alpha_n} T h_n \|_{2,P}^2 &= \sum_{j=1}^{\infty} \frac{\lambda_j \alpha_n}{(\lambda_j + \alpha_n)^2} \langle h_n, \varphi_j \rangle^2 \\ &\leq \| h_n \|^2. \end{aligned}$$

Hence,

$$\| A_{\alpha_n} T h_n \|_{2,n} = O_P\left(\alpha_n^{-1/2} \right).$$

Finally,

$$
\begin{aligned}
\left\|\alpha_n \widehat{A}_{\alpha_n} 1\right\|_{2,n} &= O_P\left(\left\|\alpha_n \widehat{A}_{\alpha_n} 1\right\|_{2,P}\right) \\
&= O_P\left(\left\|\alpha_n \widehat{A}_{\alpha_n} - \alpha_n A_{\alpha_n}\right\|_{2,P}\right) \\
&= O_P\left(\alpha_n^{-1} n^{-1/2}\right).
\end{aligned}
$$

Then, we conclude

$$
\begin{aligned}
\left|\int \Delta_{11n}(z) F_n(dz)\right| &= O_P(\alpha_n^{-5/2} n^{-1}) + o_P(1) \\
&= o_P(1).
\end{aligned}
$$

As for Δ_{2n}, we write,

$$
\begin{aligned}
\Delta_{2n} &= \widehat{A}_{\alpha_n}(\hat{T}\hat{T}' - TT')s^* + (\widehat{A}_{\alpha_n} - A_{\alpha_n})TT's^* \\
&= \widehat{A}_{\alpha_n}(\hat{T}\hat{T}' - TT')(s^* - s^*_{\alpha_n}).
\end{aligned}
$$

Note that

$$
\begin{aligned}
(\hat{T}\hat{T}' - TT')\psi_j(z) &= (\hat{T}T' - TT')\psi_j(z) + (\hat{T}\hat{T}' - \hat{T}T')\psi_j(z) \\
&= \lambda_j^{1/2}(\hat{T} - T)\varphi_j(z) + \hat{T}(\hat{T}' - T')\psi_j(z).
\end{aligned}
$$

Hence, with $\bar{s}$ implicitly defined by $s^*(z) = T\bar{s}(z)$,

$$
\begin{aligned}
(\hat{T}\hat{T}' - TT')(s^* - s^*_{\alpha_n})(z) &= \alpha_n \sum_{j=1}^{\infty} \frac{\lambda_j}{\lambda_j + \alpha_n} \langle \bar{s}, \varphi_j \rangle (\hat{T} - T)\varphi_j(z) \\
&\quad + T\sum_{j=1}^{\infty} \frac{\alpha_n \lambda_j^{1/2}}{\lambda_j + \alpha_n} \langle \bar{s}, \varphi_j \rangle (\hat{T}' - T')\psi_j(z) \\
&\quad + o_P(n^{-1}).
\end{aligned}
\tag{37}
$$

Let Δ_{21n} denote the part of Δ_{2n} corresponding to the first summand in (37), and Δ_{22n} the one corresponding to the second summand. The latter part is asymptotically negligible, by the same arguments as for Δ_{1n}. Hence, we focus on Δ_{21n}. To that end, we define

$$
\bar{s}_\alpha(x) := \sum_{j=1}^{\infty} \frac{\lambda_j}{\lambda_j + \alpha} \langle \bar{s}, \varphi_j \rangle \varphi_j(x),
$$

and note that $\|\bar{s}_\alpha\| < \infty$ and $\|\bar{s}_\alpha - \bar{s}\|$ as $\alpha \to 0$ (w.l.o.g take $\bar{s} \in \overline{Im}(K)$). By an application of Lemma Q8, it suffices to work with $\sqrt{n}(\hat{T} - T)\bar{s}(\cdot)$. Using similar arguments as for Δ_{1n}, we write by Cauchy-Schwarz's inequality

$$
\begin{aligned}
\left|\int \sqrt{n}\Delta_{21n}(z) F_n(dz)\right| &= \left|\left\langle \sqrt{n}\Delta_{21n}(\cdot), 1\right\rangle_{2,n}\right| \\
&= \left|\left\langle \sqrt{n}(\hat{T} - T)\bar{s}(\cdot), \alpha_n \widehat{A}_{\alpha_n} 1\right\rangle_{2,n}\right| + o_P(1) \\
&\leq \left\|\sqrt{n}(\hat{T} - T)\bar{s}(\cdot)\right\|_{2,n} \left\|\alpha_n \widehat{A}_{\alpha_n} 1\right\|_{2,n} + o_P(1).
\end{aligned}
$$

Note that $\left\|\sqrt{n}(\hat{T}-T)\bar{s}(\cdot)\right\|_{2,n} = O_P(1)$ and $\left\|\alpha_n \widehat{A}_{\alpha_n} 1\right\|_{2,n} = o_P(1)$ by the arguments above. Hence, we conclude that $C_n^* = o_P(1)$. The proof of $\widehat{\Sigma}_{\alpha_n} = \Sigma + o_P(1)$ is simpler, and hence omitted. ■

Proof of Theorem 4: We shall apply Theorem 3. To that end, we need to check that Assumptions W and RE hold under Assumptions E1-E3. To verify Assumption W, we apply Theorem Q1 of Section 7.1.3. Recall

$$\hat{R}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \hat{\zeta}_i(\tau) 1(X_i \le w),$$

where $\hat{\zeta}_i(\tau) = 1(Y_i \le \beta_0' X_{1i} + \widehat{\gamma}_n'(\tau) X_{2i}) - \tau$ and $x = (\tau, w')' \in \mathcal{T} \times \mathbb{R}^{d_x}$. We shall prove that under our assumptions and H_0,

$$\sup_{x \in \mathcal{T} \times \mathbb{R}^{d_x}} \left| \hat{R}_n(x) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \zeta_i(\tau) 1(X_i \le w) - A'(x)\sqrt{n}\left(\widehat{\gamma}_n(\tau) - \gamma_0(\tau)\right) \right| = o_P(1), \tag{38}$$

where $A(x) := E[X_{2i} f_{i\tau} 1(X_i \le w)]$. To obtain this expansion we apply Theorem Q1 to the class $\mathcal{W} = \{x \to 1(x \le w) : w \in [-\infty, \infty]^{d_x}\}$, which satisfies the conditions of the theorem by Theorem 2.7.1 in van der Vaart and Wellner (1996). This yields

$$\begin{aligned} \sup_{x \in \mathcal{T} \times \mathbb{R}^{d_x}} \Bigg| & \hat{R}_n(x) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \zeta_i(\tau) 1(X_i \le w) \\ & + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(E\left[\zeta_i(\tau) \middle| X_i\right] - E\left[\hat{\zeta}_i(\tau) \middle| X_i\right] \right) 1(X_i \le w) \Bigg| = o_P(1). \end{aligned} \tag{39}$$

Applying a mean value argument we obtain

$$\begin{aligned} \sup_{x \in \mathcal{T} \times \mathbb{R}^{d_x}} \Bigg| & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(E\left[\zeta_i(\tau) \middle| X_i\right] - E\left[\hat{\zeta}_i(\tau) \middle| X_i\right] \right) 1(X_i \le w) \\ & - \sqrt{n}\left(\widehat{\gamma}_n(\tau) - \gamma_0(\tau)\right)' \frac{1}{n} \sum_{i=1}^{n} f\left(X_i' \widetilde{\delta}_0(\tau) \middle| X_i \right) X_{2i} 1(X_i \le w) \Bigg| = o_P(1), \end{aligned}$$

where $\widetilde{\delta}_0(\tau)$ is such that $\left|\widetilde{\delta}_0(\tau) - \delta_0(\tau)\right| \le \left|\widehat{\delta}_0(\tau) - \delta_0(\tau)\right|$ a.s. for each $\tau \in \mathcal{T}$. By our assumptions, uniformly in $x \in \mathcal{T} \times \mathbb{R}^{d_x}$,

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^{n} f\left(X_i' \widetilde{\delta}_0(\tau) \middle| X_i \right) 1(X_i \le w) &= \frac{1}{n} \sum_{i=1}^{n} f_{i\tau} X_{2i} 1(X_i \le w) + o_P(1) \\ &= A(x) + o_P(1). \end{aligned}$$

where the last equality follows from Glivenko-Cantelli's theorem, i.e. the class

$$\left\{ Z_i \to f\left(X_i' \delta_0(\tau) \middle| X_i \right) 1(X_i \le w) : x \in \Gamma \right\}$$

is a Glivenko-Cantelli class by an application of Lemma Q1. Hence, we obtain the expansion (38). The null limiting distribution then follows from the expansion and Lemma Q3. Combine this with Theorem

A2 to obtain Assumption W. Notice that Assumption A2 required in Theorem A2 holds under our conditions on the conditional density in E1.

We now check Assumption RE(ii). Throughout the proofs, we use the fact that the nonparametric estimate $\tilde{f}_{i\tau}$ only depends on the sample $\mathcal{A}_n \equiv \{\tau_j\}_{j=1}^n$ and X_i, and that $\mathcal{A}_n$ is independent of the original sample $\mathcal{Z}_n \equiv \{Z_i\}_{i=1}^n$. That means that in many of the probabilistic arguments we use, we can first condition on $\mathcal{A}_n$ and deal with conditional probabilities treating the nonparametric function as given. This simplifies substantially the arguments.

We prove that $\left\|\hat{T}^* - T'\right\| = O_P(n^{-1/2})$. For a fix $a \in L_2(F)$ with $\|a\|_{2,P} = 1$, set $\bar{x} = (\tau, w')' \in \Gamma$, and write,

$$
\begin{aligned}
\sqrt{n}(\hat{T}^* - T')a(\bar{x}) &= \sqrt{n}\int a(z)\left(\hat{\zeta}(\tau) - \zeta_i(\tau)\right)\widehat{q}(x,\bar{x})F(dz) \\
&- A(\bar{x})B^{-1}(\tau)\sqrt{n}\int a(z)\zeta(\tau)x_2(\hat{f}_{\cdot\tau} - f_{\cdot\tau})F(dz) \\
&+ \sqrt{n}\left(A_n(\bar{x})B_n^{-1}(\tau) - A(x)B^{-1}(\tau)\right)\int a(z)\zeta(\tau)x_2\hat{f}_{\cdot\tau}F(dz) \\
&=: C_{11n}(\bar{x}) - A(\bar{x})B^{-1}(\tau)C_{12n}(\tau) + C_{13n}(\bar{x}).
\end{aligned}
$$

Using a standard Taylor expansion, it holds, for each $\bar{x} \in \mathcal{T} \times \mathbb{R}^{d_x}$,

$$
C_{11n}(\bar{x}) = E[X_{2i}'a(X_i'\widetilde{\delta}_0(\tau), X_i)f\left(X_i'\widetilde{\delta}_0(\tau)\middle| X_i\right)\widehat{q}(X_i,x)]\sqrt{n}\left(\widehat{\gamma}_n(\tau) - \gamma_0(\tau)\right),
$$

where $\widetilde{\delta}_0(\tau)$ is such that $\left|\widetilde{\delta}_0(\tau) - \delta_0(\tau)\right| \leq \left|\widehat{\delta}_0(\tau) - \delta_0(\tau)\right|$ a.s. for each $\tau \in \mathcal{T}$. Lemma Q3, Cauchy-Schwarz's inequality and our moment conditions imply that $\|C_{11n}\| = O_P(1)$.

To deal with $C_{12n}(\tau)$, use the Taylor expansion in (33), and write

$$
\begin{aligned}
C_{12n}(\tau) &= \sqrt{n}\int a(z)\zeta(\tau)x_2(\tilde{f}_{\cdot\tau} - f_{\cdot\tau})F(dz) \\
&+ \sqrt{n}\int a(z)\zeta(\tau)x_2\dot{f}_{\cdot\tau}(\widehat{\Delta})F(dz) + \sqrt{n}\int a(z)\zeta(\tau)x_2\ddot{f}_{\cdot\tau}(\widehat{\Delta})F(dz) \\
&+ \sqrt{n}\int a(z)\zeta(\tau)xr_{i\tau}F(dz) \\
&=: C_{121n}(\tau) + C_{122n}(\tau) + C_{123n}(\tau) + C_{124n}(\tau).
\end{aligned}
$$

In turn, $C_{121n}(\tau)$ can be decomposed in a stochastic and bias part

$$
\begin{aligned}
C_{121n}(\tau) &= \sqrt{n}\int a(z)\zeta(\tau)x_2(\tilde{f}_{\cdot\tau} - E_{\tau_j}[f_{\cdot\tau}])F(dz) \\
&+ \sqrt{n}\int a(z)\zeta(\tau)x_2(E_{\tau_j}[f_{\cdot\tau}] - f_{\cdot\tau})F(dz).
\end{aligned}
$$

The stochastic part is a standardized sample mean of iid elements in $L_2(\mu)$, say $W_{j,h}(\tau)$, satisfying

$$
\begin{aligned}
E_{\tau_j}[\|W_{j,h}(\cdot)\|^2] &= E_{\tau_j}[\left\|E[a(Z)\zeta(\tau)X_2]K_h\left(X_i'\delta_0(\cdot) - X_i'\delta_0(\tau_j)\right)\right\|^2] \\
&\leq \int\left(\int \mu_a(u,\tau)g_{(\tau,\tau_j)}(u,u)du\right)^2 d\tau d\tau_j + O_P(h^2) \\
&\leq C,
\end{aligned}
$$

where $K_h(u) := h^{-1}K(u/h)$, $\mu_a(u,\tau) := E\left[a(Z)\zeta(\tau)X_2 \middle| X'\delta_0(\tau)=u\right]$, and where we have used that

$$\left|\int \mu_a(u,\tau)g_{(\tau,\tau_j)}(u,u)du\right| \leq E[|a(Z)X_2|].$$

Hence, by the FCLT, we conclude the stochastic part is $O_P(1)$. The bias part is also $O_P(1)$ by our assumptions on the conditional density and bandwidth. Hence, it holds $\|C_{121n}\|^2 = O_P(1)$.

We now turn into C_{122n}. Standard arguments show that

$$C_{122n}(\tau) = \frac{1}{n}\sum_{j=1}^{n}\int t\dot{K}(t)\,dt\int \dot{q}_{1a}(u,\tau,\tau_j)du + O\left(h^2\right),$$

where $\dot{q}_a(u,\tau,\tau_j) = \partial q_a(u,v,\tau,\tau_j)/\partial v|_{v=u}$,

$$\begin{aligned} q_a(u,v,\tau,\tau_j) &:= \mu_a(u,v,\tau,\tau_j)g_{(\tau,\tau_j)}(u,v) \text{ and} \\ \mu_a(u,v,\tau,\tau_j) &:= E\left[a(Z)\zeta(\tau)X_2 \middle| X'\delta_0(\tau)=u, X'\delta_0(\tau_j)=v\right]. \end{aligned}$$

Note that Assumption E1 implies that

$$\left|\int \dot{q}_a(u,\tau,\tau_j)du\right| \leq C.$$

The same arguments show that $\|C_{123n}\| = O_P(1)$. It is also straightforward to prove that $\|C_{124n}\| = O_P\left(n^{-2}h^{-2}\right) = o_P(1)$. Hence, $\|C_{12n}\| = O_P(1)$.

To analyze $C_{13n}(\bar{x})$, using some arguments below (see the analysis of D_{1n}), it can be shown that the vector process

$$\begin{pmatrix} \sqrt{n}(A_n(\cdot)-A(\cdot)) \\ \sqrt{n}(B_n(\cdot)-B(\cdot)) \end{pmatrix}$$

converges weakly in $L_2(\mu)$. Thus, by an application of the functional delta method, we obtain the weak convergence of

$$\sqrt{n}\left(A(\cdot)B^{-1}(\cdot) - A_n(\cdot)B_n^{-1}(\cdot)\right).$$

From this convergence and Lemma Q5, it easily follows that $\|C_{13n}\| = O_P(1)$. Together, these results imply $\left\|\hat{T}^* - T'\right\| = O_P(n^{-1/2})$.

Finally, to prove $\left\|\hat{D} - \hat{T}'s^*\right\| = O_P(n^{-1/2})$, it suffices to show that $\left\|\hat{D} - D\right\| = O_P(n^{-1/2})$, by the arguments in the proof of Theorem 3. For simplicity, we assume w.l.g that X_1 is a scalar. We write

$$\begin{aligned} \sqrt{n}(\hat{D}(x) - D(x)) &= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{X_{1i}\hat{f}_{i\tau}\widehat{q}(X_i,x) - D(x)\right\} \qquad (40) \\ &= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{X_{1i}\hat{f}_{i\tau}1(X_i\leq w) - E[X_{1i}f_{i\tau}1(X_i\leq w)]\right\} \\ &+ A(x)B^{-1}(\tau)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{X_{1i}X_{2i}\hat{f}_{i\tau}^2 - E[X_{1i}X_{2i}f_{i\tau}^2]\right\} \\ &- \left(A(x)B^{-1}(\tau) - A_n(x)B_n^{-1}(\tau)\right)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}X_{1i}X_{2i}\hat{f}_{i\tau}^2 \\ &=: -D_{1n}(x) + A(x)B^{-1}(\tau)D_{2n}(x) - D_{3n}(x). \end{aligned}$$

Long, but simple, algebra shows that D_{1n} is asymptotically equivalent to $D_{11n}+D_{12n}+D_{13n}$, uniformly in $x \in \Gamma$, where

$$
\begin{aligned}
D_{11n}(x) &:= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \{X_{1i} f_{i\tau} 1(X_i \le w) - E[X_{1i} f_{i\tau} 1(X_i \le w)]\}, \\
D_{12n}(x) &:= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_{1i}(\tilde{f}_{i\tau} - f_{i\tau})1(X_i \le w)
\end{aligned}
$$

and

$$
D_{13n}(x) := \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_{1i}\dot{f}_{i\tau}(\widehat{\Delta})1(X_i \le w).
$$

The process D_{12n}, centered at its expectation, is stochastic equicontinuous, and by the independence assumption between $\mathcal{A}_n$ and $\mathcal{Z}_n$, for each x,

$$
\begin{aligned}
Var\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_{1i}(\tilde{f}_{i\tau} - f_{i\tau})1(X_i \le w) \middle| \mathcal{A}_n \right) &\le E\left[X_{1i}^2(\tilde{f}_{i\tau} - f_{i\tau})^2 1(X_i \le w)\right] \\
&= o_P(1).
\end{aligned}
$$

Hence, uniformly in $x \in \Gamma$,

$$
\begin{aligned}
D_{12n}(x) &= \sqrt{n}E\left[X_{1i}(\tilde{f}_{i\tau} - f_{i\tau})1(X_i \le w)\right] + o_P(1) \\
&= \frac{1}{\sqrt{n}}\sum_{j=1}^{n} a_{1h}(x,\tau_j) + o_P(1),
\end{aligned}
$$

where

$$
\begin{aligned}
a_{1h}(x,\tau_j) &:= E\left[\mu_{1(\tau,\tau_j)}(X'\delta_0(\tau), X'\delta_0(\tau_j), w)\frac{1}{h^2}\dot{K}\left(\frac{X'\delta_0(\tau) - X'\delta_0(\tau_j)}{h}\right) \middle| \mathcal{A}_n \right] \\
\mu_{1(\tau,\tau_j)}(u,v,w) &:= E\left[X_1 1(X_i \le w) \middle| X'\delta_0(\tau) = u, X'\delta_0(\tau_j) = v \right].
\end{aligned}
$$

Define

$$
q_{1(\tau,\tau_j)}(u,v,w) := \mu_{1(\tau,\tau_j)}(u,v,w) g_{(\tau,\tau_j)}(u,v).
$$

Then,

$$
\begin{aligned}
\int q_{1(\tau,\tau_j)}(u,v,w)\frac{1}{h^2}\dot{K}\left(\frac{u-v}{h}\right) dudv &= -\int q_{1(\tau,\tau_j)}(u, u-th, w)\frac{1}{h}\dot{K}(t)\, dudt \\
&= \int t\dot{K}(t)\, dt \int \dot{q}_{1(\tau,\tau_j)}(u,u,w)du + O\left(h^2\right) \\
&=: \mu_3 a_1(x,\tau_j) + O\left(h^2\right),
\end{aligned}
$$

where $\dot{q}_{1(\tau,\tau_j)2}(u,v,w) = \partial q_{1(\tau,\tau_j)}(u,v,w)/\partial v$. Then,

$$
\max_{1\le j\le n} \sup_{x} |a_{1h}(x,\tau_j) - \mu_3 a_1(x,\tau_j)| = O\left(h^2\right).
$$

Using this, we conclude

$$D_{12n}(x) = \mu_3 \frac{1}{\sqrt{n}} \sum_{j=1}^{n} a_1(x, \tau_j) + o_P(1).$$

Then, $\|D_{12n}\| = O_P(1)$ follows easily from the last display and Assumption E1.

Using a stochastic equicontinuity argument, we obtain

$$\begin{aligned} D_{13n}(x) &= \sqrt{n} E[X_{1i} \dot{f}_{i\tau}(\widehat{\Delta}) 1(X_i \leq w)] + o_P(1) \\ &= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} a_{2h}(x, \tau_j) \{\widehat{\gamma}_n(\tau_j) - \gamma_0(\tau_j) + \widehat{\gamma}_n(\tau) - \gamma_0(\tau)\} + o_P(1), \end{aligned}$$

where

$$a_{2h}(x, \tau_j) := E\left[\mu_{2(\tau,\tau_j)}(X'\delta_0(\tau), X'\delta_0(\tau_j), w) \frac{1}{h^2} \dot{K}\left(\frac{X'\delta_0(\tau) - X'\delta_0(\tau_j)}{h} \right) \Big| \mathcal{A}_n \right]$$

and

$$\mu_{2(\tau,\tau_j)}(u, v, w) := E\left[X_1 X_2' 1(X_i \leq w) \big| X'\delta_0(\tau) = u, X'\delta_0(\tau_j) = v \right].$$

Define

$$q_{2(\tau,\tau_j)}(u, v, w) := \mu_{2(\tau,\tau_j)}(u, v, w) g_{2(\tau,\tau_j)}(u, v).$$

Then, with

$$a_2(x, \tau_j) := \int \dot{q}_{2(\tau,\tau_j)}(u, u, w) du,$$

and $\dot{q}_{2(\tau,\tau_j)2}(u, v, w) = \partial q_{2(\tau,\tau_j)}(u, v, w)/\partial v$, we obtain

$$\max_{1 \leq j \leq n} \sup_{x} |a_{2h}(x, \tau_j) - \mu_3 a_2(x, \tau_j)| = O\left(h^2\right).$$

Using this, we conclude

$$\begin{aligned} D_{13n}(x) &= \mu_3 \left(\frac{1}{n} \sum_{j=1}^{n} a_2(x, \tau_j) \right) \sqrt{n}(\widehat{\gamma}_n(\tau) - \gamma_0(\tau)) \\ &\quad + \frac{\mu_3}{\sqrt{n}} \sum_{j=1}^{n} a_2(x, \tau_j)(\widehat{\gamma}_n(\tau_j) - \gamma_0(\tau_j)) + o_P(1). \end{aligned}$$

By a standard law of large numbers and by Lemma 3.1 in Chang (1990), D_{13n} converges weakly to

$$\mu_3 \left\{ E[a_2(w, \cdot)] \alpha_\infty(\cdot) + \int_{\mathcal{T}} a_2(w, \tau) \alpha_\infty(\tau) d\tau \right\},$$

where $\alpha_\infty(\cdot)$ is the limiting distribution of $\sqrt{n}(\widehat{\gamma}_n(\cdot) - \gamma_0(\cdot))$. The analysis of D_{2n} is similar to that of D_{1n}, and that of $D_{3n}(x)$ is the same as for $C_{13n}(x)$, after noticing that uniformly in τ

$$\frac{1}{n} \sum_{i=1}^{n} \hat{\zeta}_i(\tau) X_{2i} \hat{f}_{i\tau} a(Z_i) = \frac{1}{n} \sum_{i=1}^{n} \zeta_i(\tau) X_{2i} f_{i\tau} a(Z_i) + o_P(1),$$

see Lemma Q5. Accounting for all the terms in the expansion (40), we conclude $\left\| \hat{D} - D \right\| = O_P(n^{-1/2})$. ■

7.3 Appendix C: Algorithm for implementation

This section summarizes the main steps in the implementation of the new efficient procedures. Recall the main ingredients of the problem:

- Model: $E[\psi(Z, x, \beta, \eta_0(Z, x))] = 0$ for all $x \in \Gamma$.
- Testing problem: $H_0 : \beta = \beta_0$ vs $H_1 : \beta \neq \beta_0$.
- The nuisance parameter η_0 is estimated consistently by $\widehat{\eta}_n$ under the null.
- Sample analog of the moments:

$$\hat{R}_n(x) := \frac{1}{n}\sum_{i=1}^{n}\psi(Z_i, x, \beta_0, \widehat{\eta}_n),$$

Implementing the proposed test requires the following steps:

- Derive the asymptotic expansion under H_0, in $L_2(\mu)$,

$$\sqrt{n}\hat{R}_n(x) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}m(Z_i, x, \beta_0, \eta_0) + o_P(1).$$

- Generate a random sample $\{x_j\}_{j=1}^{n}$ from μ.
- Compute the $n \times n$ matrix M with entries $\{\widehat{m}(Z_i, x_j) \equiv m(Z_i, x_j, \beta_0, \widehat{\eta}_n)\}_{i,j=1}^{n}$.
- Compute the $n \times p$ matrix D with entries $(\hat{D}(x_1), ..., \hat{D}(x_n))'$, where $\hat{D}(x_j)$ estimates $D(x_j) := -\partial E\left[m(Z, x_j, \theta_0(x_j))\right]/\partial\beta$.
- Given α_n, $\alpha_n \downarrow 0$, compute the $n \times n$ matrix $P_\alpha := I - M'(\alpha_n n^{-2} I + MM')^{-1}M$.
- Compute the $n \times p$ matrix $S := (\alpha_n n)^{-1} M P_\alpha D$.
- Reject H_0 at $\alpha-$th nominal level if $T_n = 1'S(S'S)^{-1}S'1 \geq \chi^2_{1-\alpha,p}$, where 1 is a $n \times 1$ vector of ones and $\chi^2_{\tau,p}$ is the τ-quantile of the chi-squared distribution with p degrees of freedom.

References

[1] Ai, C., and X. Chen (2003): "Efficient Estimation of Conditional Moment Restrictions Models Containing Unknown Functions," *Econometrica*, 71, 1795–1843.

[2] Altonji, J. G., T. E. Elder, and C. R. Taber, (2005): "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy*, 13, 151-84.

[3] Akritas, M. G. (1988): "An Asymptotic Derivation of Neyman's C(α) test," *Statistics & Probability Letters,* 6**,** 363–367.

[4] Akritas, M.G., and R.A. Johnson (1982): "Efficiencies of Tests and Estimators in Autoregressions under Nonnormal Error Distribution," *Annals of the Institute of Statistical Mathematics,* 34, Part A, 579-589.

[5] Arellano, M., L. P. Hansen, and E. Sentana (2011): "Underidentification?," *Journal of Econometrics*, forthcoming.

[6] Bickel, P.J., C.A. Klaasen, Y. Ritov and J.A. Wellner (1993): *Efficient and adaptive estimation for semiparametric models*. Johns Hopkins Univ. Press, Baltimore.

[7] Bickel, P. J., Y. Ritov and T. Stoker (2006): "Tailor-made Tests for Goodness-of-Fit to Semiparametric Hypotheses," *The Annals of Statistics,* 34, 721-741.

[8] Bierens, H. J., and W. Ploberger (1997): "Asymptotic Theory of Integrated Conditional Moment Tests," *Econometrica,* 65, 1129-1151.

[9] Bischoff, W. and Miller, F. (2000): "Asymptotically Optimal Tests and Optimal Designs for Testing the Mean in Regression Models with Applications to Change-Point Problems," *Annals of the Institute of Statistical Mathematics,* 52, 658-679.

[10] Boning, B., and F. Sowell (1999): "Optimality for the Integrated Conditional Moment Test," *Econometric Theory,* 15, 710-718.

[11] Bühler, W.J., and Puri, P.S. (1966): "On Optimal Asymptotic Tests of Composite Hypotheses With Several Constraints," *Z. Wahrsch. verw. Geb.*, 5, 71–88.

[12] Carrasco, M. and J. P. Florens (2000): "Generalization of GMM to a Continuum of Moment Conditions," *Econometric Theory*, 16, 797-834.

[13] Carrasco, M. and J. P. Florens (2011): "On the Asymptotic Efficiency of GMM," Working paper.

[14] Carrasco, M., J. P Florens and E. Renault (2006): "Linear Inverse Problems in Structural Econometrics: Estimation Based on Spectral Decomposition and Regularization," *Handbook of Econometrics*, Volume 6, ed. by J.J. Heckman and E.E. Leamer. New York: North-Holland.

[15] Chamberlain, G. (1987): "Asymptotic Efficiency in Estimation With Conditional Moment Restrictions," *Journal of Econometrics,* 34, 305-334.

[16] Chang, N. M. (1990): "Weak Convergence of a Self-Consistent Estimator of a Survival Function With Doubly Censored Data," *The Annals of Statistics,* 18, 391-404.

[17] Chen, X., V. Chernozhukov, S. Lee and W. K. Newey (2011): "Local Identification of Nonparametric and Semiparametric Models," working paper.

[18] Chen, X. and Y. Fan (1999): "Consistent Hypothesis Testing in Semiparametric and Nonparametric Models for Econometric Time Series," *Journal of Econometrics*, 91, 373-401.

[19] Chen, X., O. Linton and I. Van Keilegom (2003): "Estimation of Semiparametric Models When the Criterion Function is Not Smooth," *Econometrica*, 71, 1591-1608.

[20] Chen, X., and D. Pouzo (2012): "Estimation of Nonparametric Conditional Moment Models with Possibly Nonsmooth Moments," *Econometrica*, 80, 277-322.

[21] Choi, S., W.J. Hall and A. Schick (1996): "Asymptotically Uniformly Most Powerful Tests in Parametric and Semiparametric Models," *The Annals of Statistics,* 24, 841-861.

[22] Delgado, M.A., J. Hidalgo and C. Velasco (2005): "Distribution Free Goodness-of-fit Tests for Linear Processes," *The Annals of Statistics,* 33, 2568-2609.

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

[24] Elliot, G. and U.K. Müller (2008): "Pre and Post Break Parameter Inference," Working Paper, Princeton University.

[25] Escanciano, J. C. (2009): "On the Lack of Power of Omnibus Specification Tests," *Econometric Theory,* 25, 162-194.

[26] Escanciano, J. C., and C. Goh (2012): "Conditional Density Estimation in Linear Quantile Regression", working paper.

[27] Escanciano, J. C., and L. Zhu (2012): "Inferences in Semiparametric Partially Identified Models: An Empirical Process Approach," working paper.

[28] Grenander, U. (1950): "Stochastic Processes and Statistical Inference," *Arkiv for Matematik,* 1, 195-277.

[29] Gutenbrunner, C., and J. Jureckova (1992): "Regression Rank Scores and Regression Quantiles," *The Annals of Statistics,* 20, 305–330.

[30] Hallin, M. and Werker, B. J. M. (2003): "Semiparametric Efficiency, Distribution-Freeness and Invariance," *Bernoulli,* 9 137–165.

[31] Ibragimov, I.A., and R.Z. Hasminskii (1981): *Statistical Estimation.* Springer, New York.

[32] Khmaladze, E. V. (1981): "Martingale Approach to the Goodness of Fit Tests," *Theory of Probability and Its Applications,* 26 246-265.

[33] Koenker, R., and G. Bassett (1978): "Regression Quantiles," *Econometrica,* 46, 33-50.

[34] Koenker, R., and Z. Xiao (2002): "Inference on the Quantile Regression Process," *Econometrica*, 70, 1583–1612.

[35] Komunjer, I. and Q. Vuong (2010): "Efficient Estimation in Dynamic Conditional Quantile Models," *Journal of Econometrics*, 157, 272-285.

[36] Kress, R. (1999): *Linear Integral Equations*. Springer.

[37] Le Cam, L. (1956): "On the Asymptotic Theory of Estimation and Testing hypotheses," *Proc. 3rd Berkeley Symp. Math. Stat. and Prob.* 1, 129-156.

[38] Luschgy, H. (1991): "Testing One-sided Hypotheses for the Mean of a Gaussian Process," *Metrika,* 38, 179-194.

[39] Müller, U. (2011): "Efficient Tests Under a Weak Convergence Assumption," *Econometrica,* 79, 395–435.

[40] Newey, W. K. (1988): "Adaptive Estimation of Regression Models Via Moment Restrictions," *Journal of Econometrics,* 38, 301-339.,

[41] Newey, W. K. (1990): "Semiparametric Efficiency Bounds," *Journal of Applied Econometrics,* 5, 99-136.

[42] Newey, W.K. (1994): "The Asymptotic Variance of Semiparametric Estimators," *Econometrica*, 62, 1349–1382.

[43] Newey, W. K. (2004): "Efficient Semiparametric Estimation via Moment Restrictions," *Econometrica,* 72, 1877-1897.

[44] Newey, W. K., and West, K. D. (1987): "Hypothesis Testing with Efficient Method of Moments Estimation," *International Economic Review,* 28, 777-787.

[45] Neyman, J. (1959): "Optimal Asymptotic Tests of Composite Statistical Hypotheses," In Probability and Statistics, (ed. U. Grenander), Harald Cramer Volume, 212–234. Wiley, New York.

[46] Parthasarathy, K. R. (1967): *Probability Measures on Metric Spaces.* Academic, New York.

[47] Politis, D., and J. Romano (1994): "Limit Theorems for Weakly Dependent Hilbert Space Valued Random Variables With Application to the Stationary Bootstrap," *Statistica Sinica,* 4, 461-476.

[48] Powell, J.L. (1994): "Estimation in semiparametric models," in *The Handbook of Econometrics*, vol. IV. eds R.F. Engle and D.L. McFadden. North Holland.

[49] Robinson, P.M. (1987): "Asymptotically Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form," *Econometrica,* 55, 875–891.

[50] Robinson, P.M. (1988): "Semiparametric Econometrics: A Survey", *Journal of Applied Econometrics*, 3, 35-51.

[51] Skorohod, A. V. (1974): *Integration in Hilbert space.* Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 79, Springer-Verlag, New York.

[52] Song, K. (2010): "Testing Semiparametric Conditional Moment Restrictions Using Conditional Martingale Transforms," *Journal of Econometrics,* 154, 74-84.

[53] Sowell, F. (1996): "Optimal Tests for Parameter Instability in the Generalize Method of Moments Framework," *Econometrica,* 64, 1085-1107.

[54] Stinchcombe, M., and H. White (1998): "Consistent Specification Testing with Nuisance Parameters Present Only Under the Alternative," *Econometric Theory*, 14, 295-325.

[55] Stute, W. (1997): "Nonparametric Model Checks for Regression," *The Annals of Statistics,* 25, 613-641.

[56] Stute, W., S. Thies, and L.X. Zhu (1998): "Model Checks for Regression: An Innovation Process Approach," *The Annals of Statistics,* 26, 1916-1934.

[57] van der Vaart, A.W. (1990): "On Differentiable Functionals," *The Annals of Statistics,* 19, 178-204.

[58] van der Vaart, A.W. (1998): *Asymptotic Statistics.* Cambridge University Press. Cambridge, UK.

[59] van der Vaart, A.W., and J.A. Wellner (1996): *Weak Convergence and Empirical Processes.* Springer, New York.

[60] Watson, M., and U. Müller (2008): "Testing Models of Low-Frequency Variability," *Econometrica,* 76, 979–1016.

[61] Wellner, J. A., C. A. J. Klaassen, and Y. Ritov (2006): "Semiparametric Models: A Review of Progress Since BKRW (1993)," in Frontiers in Statistics: In Honor of Peter J. Bickel's 65th Birthday, ed. by J. Fan, and H. Koul, pp. 25–35. Imperial College Press, London.