Semiparametric Efficient Directional Tests

Juan Carlos Escanciano*
Indiana University

April 27th, 2012

Abstract

Consider a generic semiparametric model defined by an infinite number of moment conditions, including a finite dimensional parameter of interest and possibly including moment-specific nonparametric nuisance parameters. Within this setting, we propose tests for restrictions on the parameter of interest based on optimal functionals of the sample analog of the moments. The optimal functional takes the form of a Radom-Nikodym derivative or Likelihood Ratio in a nonparametric setting. This paper investigates the semiparametric efficiency and implementation of feasible versions of such directional tests. The paper provides four main contributions. First, it proves the semiparametric efficiency of the proposed tests. Second, it proposes and justifies feasible implementations of such directional tests based on a novel nonparametric estimator of the efficient score. Third, it establishes important and fruitful connections with the literature on generalized methods of moments. Finally, it applies 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 optimal inferences. Useful by-products of our analysis are optimal confidence sets by inverting our test statistic and a new algorithm for computing the efficiency bound for regular estimation of the parameter of interest.

Keywords: Neyman-Pearson lemma; Likelihood ratio; Semiparametric efficiency; 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 Xiaohong Chen, Miguel A. Delgado, Ulf Grenander, Roger Koenker, Ulrich Müller and Carlos Velasco. Research funded by the Spanish Plan Nacional de I+D+I, reference number SEJ2007-62908.

1 Introduction

Consider a generic semiparametric model defined by an infinite number of moment conditions, each moment depending on a finite dimensional parameter of interest and possibly including moment-specific nonparametric nuisance parameters. Within this setting, suppose we are interesting in testing restrictions on the finite dimensional parameter. Omnibus consistent tests based on continuous functionals of the sample analog of the moments, say $\hat{R}_n$, have been extensively investigated in the literature, albeit for particular versions of this model. These consistent omnibus tests do not have optimality properties, beyond being asymptotically admissible (cf. Bierens and Ploberger, 1997), and they are often difficult to interpret (see Escanciano, 2009). Here, we investigate *directional tests*, rather than omnibus tests, corresponding to *optimal* functionals of $\hat{R}_n$, or equivalently, optimal tests under an average power function criteria. The optimal functional is shown to be the Radom-Nikodym derivative of the limiting distribution of $\hat{R}_n$ under local alternatives with respect to the distribution under the null. This nonparametric Likelihood Ratio (LR) principle has been already applied to several important semiparametric settings, but the semiparametric efficiency of the resulting inferences in these settings, or in more general settings like ours, remains unknown. This paper proves the semiparametric efficiency of the procedure in a generic framework, proposes and justifies feasible (and easy to implement) versions of such optimal directional tests and makes fruitful connections with existing econometric methods.

Semiparametric efficient inference is discussed extensively in the econometrics and statistics literatures, see Newey (1990) for an excellent review, and Bickel, Klaasen, Ritov, and Wellner (1993) for a comprehensive treatment. The bulk of the literature focuses on the estimation theory using the concepts of parametric submodels and tangent spaces. The main insight was given by Stein (1956), and involves an infinite number of applications of the basic Neyman-Pearson lemma and considering a worst case scenario. In this paper we take a different approach based on functional versions of the Neyman-Pearson lemma. This approach was first suggested by Grenander (1950), and it provides a better fit with the extensive nonparametric testing literature based on moment restrictions (i.e. functionals of $\hat{R}_n$). Applying this principle to the setting described above requires the following steps: (i) first, prove that $\hat{R}_n$ converges weakly, in a functional sense, to some limit process under the null hypothesis of interest as well as under local alternatives; (ii) then, compute a feasible version of the LR test, i.e. the Radom-Nikodym derivative, via the Functional Neyman-Pearson lemma for the limiting problem. The Functional Neyman-Pearson Test (FNPT) or optimal directional test is then given by the LR evaluated at $\hat{R}_n$, and corresponds to the optimal functional of $\hat{R}_n$, say $\phi^*(\hat{R}_n)$. Some examples below illustrate the procedure.

As mentioned earlier, this directional testing approach has already been applied to several semiparametric models in econometrics and statistics. Sowell (1996) proposed a FNPT for parameter instability in a Generalized Method of Moments (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) applied the FNPT to test the correct specification of regression models. Extensions to conditional distributions are 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, consider applications to stationary and non-stationary diffusion processes, respectively. Recently, Watson and Müller (2008) construct a finite-dimensional approximation of a FNPT 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. Song (2010) suggests applications of the FNPT to a general class of semiparametric conditional moment models. However, despite the extensive list of applications of this principle, the semiparametric efficiency of the resulting tests remains unknown. This paper proves that the FNPT is asymptotically efficient in a large class of regular semiparametric models under Local Asymptotic Normality (LAN), where efficiency is defined in a "classical" sense, as formalized in Choi, Hall and Schick (1996). See also Section 3 for a brief review of efficient semiparametric tests.

Our results on efficiency complement alternative efficiency results recently obtained by Müller (2011). He has shown that the FNPT is also optimal in a class of tests that control asymptotic size for all data generating processes under which $\hat{R}_n$ 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 Choi et al. (1996), and it provides a sense of robustness of the FNPT. 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 FNPT. The wide applicability and optimality properties of the FNPT suggest that it should be a useful and powerful testing procedure in econometrics.

To prove the semiparametric efficiency of the FNPT we first obtain a generic asymptotic representation of this test as a score-type process (i.e. as a quadratic form of a sample mean). We characterize the score function in terms of certain covariance operator and shift function. Then, we prove that the resulting score coincides with the so-called semiparametric *efficient score* in the semiparametric model defined by the moment restrictions, thereby establishing the semiparametric efficiency of the procedure. Feasible implementations of the FNPT generally require the estimation of the spectrum of some unknown covariance operator, which hampers the general applicability of the method.[1] Our second main contribution is the development of implementations of the FNPT that do not require knowledge of the spectrum, thereby widening the scope of applications of these methods. We combine our characterization of the score function with well-known results from ill-posed problems to construct a novel nonparametric estimator of the efficient score. The proposed feasible FNPT is a classical LM test with the estimated score, and it is quite simple to compute. 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 are either not feasible or require rather complicated arguments.

Our efficiency results are related to the recent literature on efficient estimation of semiparametric

[1] Interestingly enough, in many non-regular settings there is a closed form expression for the FNPT, so estimation of the spectrum is not necessary; see e.g. the examples in Müller (2011). However, in most regular problems this is not generally the case, and we believe this has considerably hampered the practical application of this nonparametric LR approach (cf. Stute, 1997).

models by GMM estimators employing an infinite number of moments, see e.g. Ai and Chen (2003), Newey (1988, 2004) and Carrasco and Florens (2000, 2008). Our paper differs from these works in several aspects. First, the GMM literature has been focused on estimation, with rather few results on testing available. Carrasco and Florens (2000) discussed tests based on the optimal GMM in parametric moments, but their tests were omnibus rather than directional. The bulk of our paper deals with the testing problem, but our results have direct implications on computation of efficiency bounds, computation of optimal confidence sets and the construction of one-step efficient estimators, as shown below. Second, we use a LR approach in a nonparametric sense, as in Grenander (1950), or more recently 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 approach is closely related to a Lagrange Multiplier (LM) test based on a modified optimal GMM objective function. The modification is needed to account 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.

The rest of the paper is organized as follows: Section 2 introduces notation, the semiparametric model, the testing problem and the FNPT. It then provides an asymptotic representation of the FNPT 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 FNPT. We first discuss the case of known spectrum, with the leading example of the so-called martingale-transform-based tests. We then consider implementations that do not require knowledge or estimation of the unknown spectrum. The new estimator for the efficient score is introduced here. Section 5 contains an application to a linear quantile regression model, which illustrates the utility of our results. Other examples such as partially identified models have a similar structure, and they are briefly mentioned in this section using as an illustration the application in Altonji, Elder and Taber (2005), who study the effect of attending a Catholic school on educational attainment. Section 7 concludes with some final remarks. Appendix A provides some preliminary results as well as sufficient conditions for a uniform expansion that can be used to establish our main assumptions. Mathematical proofs of our results are gathered in Appendix B.

[2] Carrasco and Florens (2000, 2008) 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 FNPT

2.1 Notation

This section contains notation that will be used throughout the paper. Henceforth, A' and $|A|$ denote the transposition and the Euclidean norm $|A| := (tr(A'A))^{1/2}$ for 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 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 $L_2(\mu)$. 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). 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 $\ker(K) := \{f \in L_2(\mu_1) : Kf = 0\}$. Finally, for a subspace $V \subset L_2(\mu)$, $V^\perp$ and $\overline{V}$ denote, respectively, its orthogonal complement and closure in $L_2(\mu)$. We will extensively use basic results from operator theory and Hilbert spaces. The reader is referred to 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. The level of generality in (1) allows us to handle simultaneously standard models such as semiparametric conditional moment restrictions as well as less standard situations in which nuisance parameters change with the moment, as in semiparametric quantile regressions or partially identified semiparametric models.[3] The following example helps to fix ideas.

[3] For potential applications of our results to partially identified semiparametric models see Scharfstein, Rotnitzky and Robins (1999), Song, Kosorok and Fine (2009), Chen, Tamer and Torgovitsky (2010), Chernozhukov, Rigobon and Stoker (2010), Arellano, Hansen and Sentana (2011), Bontemps, Magnac and Maurin (2011) and Escanciano and Zhu (2012), among many others.

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 in 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, the classical linear quantile regression model of Koenker and Bassett (1978) or partially linear quantile regressions, we prefer to keep the exposition simple. We choose the model in (2) for illustrative purposes, because it is a model for which semiparametric efficiency 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. $\square$

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

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

against the local (directional) alternatives

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

for some $c_\beta \in \mathbb{R}^p$. The nuisance parameter η_0 is unknown under both, the null and the alternative, and we assume that a consistent, but not necessarily efficient, estimator $\widehat{\eta}_n$ is available, satisfying some conditions below. For a more formal description of the local alternatives considered see Appendix A. In the main text we keep a simpler description for simplicity 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 empirical process with estimated parameters

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

as a "sufficient" statistic for the testing problem. Omnibus tests based on continuous functionals of $\hat{R}_n$, such as classical Kolmogorov-Smirnov tests based on $\sup_{x\in\Gamma}\left|\hat{R}_n(x)\right|$, abound in the literature. See e.g. Bickel, Ritov and Stoker (2006) for a recent proposal in a general semiparametric setting. As we show below, typical functionals used in omnibus tests are not optimal. 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 $\hat{R}_n$ in (4) 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 the examples and Remark 2 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 the 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) in the context of partially identified semiparametric models. Useful Functional Central Limit Theorems (FCLT) in Hilbert spaces can be found in van der Vaart and Wellner (1996) and Politis and Romano (1994) for independent observations and in Jakubowski (1980) and Chen and White (1998) for dependent heterogeneous arrays.

Example 1 (cont.): **Linear QR with a continuum of quantiles**. In this example the natural estimate for the nuisance parameters $\eta_0(\tau)$ is the QR estimator proposed by Koenker and Bassett (1978) applied to the "dependent" variable $Y_i - \beta_0' X_{1i}$, with covariates X_{2i}, and denoted by $\widehat{\eta}_n(\tau)$. We shall provide primitive conditions under which the following expansion holds uniformly in $x = (\tau, w')' \in \Gamma := \mathcal{T}\times\mathbb{R}^{d_x}$,

$$\begin{aligned}\hat{R}_n(x) &= \frac{1}{n}\sum_{i=1}^{n}\{1(Y_i \le \beta_0' X_{1i} + \widehat{\eta}_n'(\tau)X_{2i}) - \tau\}1(X_i \le w)\\ &= \frac{1}{n}\sum_{i=1}^{n}\zeta_i(\tau)w(X_i,x) + o_P(n^{-1/2}),\end{aligned}$$

where $\zeta_i(\tau) := 1(Y_i \le X_i'\theta_0(\tau)) - \tau$, $w(X_i,x) := 1(X_i \le w) - A(x)B^{-1}(\tau)X_{2i}f_{2i\tau}$, $A(x) := E[X_{2i}f_{2i\tau}1(X_i \le w)]$, $B(\tau) := E[X_{2i}X_{2i}'f_{2i\tau}^2]$ and $f_{2i\tau}$ is the conditional density of $Y_i - \beta_0' X_{1i}$ given X_{2i}, evaluated at

$\eta_0(\tau)'X_{2i}$. Thus, in this example $\psi(Z,x,\theta_0(\tau)) = \zeta_i(\tau)1(X_i \le w)$, $m(Z_i,x,\theta_0(\tau)) = \zeta_i(\tau)w(X_i,x)$, and Assumption W follows under some mild conditions with $D(x) = -E[X_{1i}f_{i\tau}w(X_i,x)]$, where $f_{i\tau}$ is the conditional density of Y_i given $X_{i,}$ evaluated at $X_i'\theta_0(\tau)$, see Section 5. Note that the weight $w(X_i,x)$ depends on infinite dimensional nuisance parameters that are different from η_0, namely $f_{2i\tau}$. To deal with this issue we assume, w.l.g but with some abuse of notation, that η_0 is enlarged to include these additional nuisance parameters appearing in m. For instance, in this example we redefine η_0 as the QR coefficients, say $\gamma(\tau)$, and the conditional density $f_{2i\tau}$. See Bickel, Ritov and Stoker (2006) for a similar implicit assumption. $\square$

2.4 Limiting Problem and the FNPT

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 associated to the limiting distributions of R_n in $L_2(\mu)$ under the null H_0 and under the local alternative hypotheses H_n, 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 covariance matrix 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 Radom-Nikodym derivative or LR of $\mathbb{P}_1$ with respect to $\mathbb{P}_0$. In the current setting, it is known (see Skorohod, 1974, Chapter 16, Theorem 2), that $\mathbb{P}_1$ will be absolute continuous with respect to $\mathbb{P}_0$ provided the following condition holds

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

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

$$\frac{d\mathbb{P}_1}{d\mathbb{P}_0}(h) = \exp\left(c_\beta' L(h) - \frac{1}{2}\sum_{j=1}^\infty \lambda_j^{-1}\left(c_\beta'\delta_j\right)^2 \right), \qquad h \in L_2(\mu), \tag{8}$$

where

$$L(h) := \sum_{j=1}^\infty \lambda_j^{-1}\langle h,\varphi_j\rangle\langle D,\varphi_j\rangle, \qquad h \in L_2(\mu), \tag{9}$$

and $\delta_j := \langle D, \varphi_j \rangle$, $j \in \mathbb{N}$. Useful intuition about the expression of the LR can be obtained from a finite dimensional approximation of the problem. Let $\{\varepsilon_j := \lambda_j^{-1/2} \langle R_\infty, \varphi_j \rangle\}_{j=1}^{\infty}$ be the so-called principal components of R_∞, which are iid standard normal under the null hypothesis. Then, (8) can be obtained as the limit of the (standard) LR of the distribution of $\{\varepsilon_j\}_{j=1}^{k}$ under the null and under local alternatives, when $k \to \infty$; see e.g. Stute (1997). This intuition is formalized in e.g Skorohod (1974). The existing literature does not provide primitive conditions for the key "contiguity" assumption (7). Below, we show that this assumption is intimately related to the assumption of finite Fisher information, see Section 3.

As evidenced from (8), $L(R_\infty)$ is a sufficient statistic for our testing problem. In terms of this sufficient statistic, the testing 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. Choi et al. (1996), suggest that an optimal test for testing H_0 against H_1 is given by $\phi^*(R_\infty)$, where

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

and where $\chi^2_{1-\alpha,p}$ is the $(1-\alpha)$-quantile of the chi-squared distribution with p degrees of freedom, $\alpha \in (0,1)$. The FNPT uses the finite sample analog of R_∞ and is given by $\phi_n^* := \phi^*(\sqrt{n}\hat{R}_n)$, and the first purpose of this paper is to study the efficiency properties of the test ϕ_n^* and related tests.

In some applications the FNPT has a closed form as a functional of $\sqrt{n}\hat{R}_n$, i.e. L is fully known, see e.g. Akritas and Johnson (1982), Luschgy (1991), Sowell (1996) and Müller (2011) for examples. However, in most regular problems a closed form expression for the FNPT is not available and estimation (regularization) of the operator L is often needed. We deal with the implementation of feasible versions of the FNPT in Section 4. There, we show that the 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 for the time being the infeasible test ϕ_n^*.

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

The objective of this section is to provide an asymptotic representation for the FNPT 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 FNPT follows. Later in Section 4, we will use the characterization of the score to implement a feasible FNPT.

With this objective in mind, we introduce the singular value decomposition of K; see Kress (1999). Henceforth, to simplify notation set $m(Z_i, x) \equiv m(Z_i, x, \theta_0(x))$. 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

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

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

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

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 $r > 0$, 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: $D \in \Psi_1$.

As previously mentioned, Assumption D is equivalent to the absolute continuity of $\mathbb{P}_1$ with respect to $\mathbb{P}_0$. 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 now 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 in 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 development. 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 the FNPT statistic $L(\sqrt{n}\hat{R}_n)$ and the sample mean S_n^*. Henceforth, for a closed subspace M, Π_M denotes its orthogonal projection operator.

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 FNPT 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 FNPT based on the standard empirical process boils down to the classical Rao-Score test in finite samples. Theorem 1(ii) offers an alternative way to compute the score s^* in (13) that does not require knowledge of 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 into $\ker^\perp(T')$ or $\overline{Im}(T)$. 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. In all the examples we have considered solving $T's = D$ was a trivial task, as the following classical example illustrates.

Example 2: **Regression model checks**. Stute (1997) proposed a FNPT for testing the significance of additional variables in homoskedastic linear-in-parameters regressions. He considered models such as 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 direction, e.g. $a(X) = X^2$, and the conditional distribution of ε given X is unknown. Defining $Z = (Y, X)'$, 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 $\widehat{\eta}_n$, and the interest is in testing $H_0 : \beta = 0$, against local deviations. Stute (1997) provided sufficient conditions for the asymptotic uniform (in $x \in \mathbb{R}$) representation under $H_n : \beta_n = n^{-1/2}c_\beta$,

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

where $\varepsilon_{i0} := Y_i - \eta_{01} - \eta_{02}X_i$, $w(X_i, x) := 1(X_i \leq x) - E[\tilde{X}_i' 1(X_i \leq x)]E[\tilde{X}_i\tilde{X}_i']^{-1}\tilde{X}_i$, and $\tilde{X}_i = (1, X_i)'$. Thus, Assumption W holds with $m(Z, x, \theta_0) = \varepsilon_{i0}w(X_i, x)$, $D(x) = E[a(X_i)w(X_i, x)]$ and μ the probability measure of X. Stute (1997) assumed homoskedasticity, i.e. $\sigma^2(X) := E\left[\varepsilon^2 \middle| X\right] = \sigma^2$. We first discuss the implications of our results for this case, and then we discuss extensions to heteroskedastic regressions. Assumption D plays a fundamental role, so we start discussing sufficient conditions for this assumption in the context of this example. First, the law of iterated expectations implies that the equation $T's(x) = D(x)$, that is,

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

is trivially solved by $s(Z) = \sigma^{-2}\varepsilon_{i0}a(X)$. Note that the solution does not depend on the form of w; see Remark 2 below. Second, Parseval's identity, (11) and Theorem 1 yield

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

Hence, a sufficient (and also necessary) condition for Assumption D is $E[a^2(X)] < \infty$. Similarly, it can be shown that $\ker(T') = span\{\varepsilon_{i0}\tilde{X}_i\}$. Hence, from Theorem 1(ii) s^* is simply the least squares errors in a regression of $s(Z)$ against X, i.e.

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

Stute (1997) suggested a FNPT 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 parameter η_0, and hence, there is no need to estimate the spectrum since much simpler asymptotically equivalent implementations of the FNPT based on our Theorem 1(ii) exist. Namely, the classical t-test, which is known to be optimal in the homoskedastic case, does not require spectrum estimates. Similar simplifications apply to other applications of the FNPT considered in the literature, see e.g. Delgado and Stute (2008). In these applications the problem $T's = D$ is not ill-posed and there is no need to regularize the problem by introducing tuning parameters, such as the number of principal components k.

Consider now the conditionally heteroskedastic case. Using the same arguments above, it can be shown that the equation $T's = D$ is solved by $s(X_i) := \sigma^{-2}(X_i)\varepsilon_{i0}a(X_i)$, Assumption D holds provided $E[\sigma^{-2}(X)a^2(X)] < \infty$, and $\ker(T') = span\{\sigma^{-2}(X_i)\varepsilon_{i0}\tilde{X}_i'\}$. 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

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

The resulting score $s^*(Z_i) = \varepsilon_{i0}\sigma^{-2}(X_i)a^*(X_i)$ is the efficient score corresponding to the semiparametric problem $H_0 : \beta_0 = 0$ against H_n, see e.g. Chamberlain (1987). The efficient score is now unknown, and implementations of the FNPT as suggested in e.g. Stute, Thies and Zhu (1998) and others use implicitly a series estimator for this score based on the basis $\{\psi_j\}_{j=1}^{\infty}$, see (13). A more common approach is to estimate nonparametrically $\sigma^2(X_i)$ and plug in this estimate in $s^*(Z_i)$, as suggested in Robinson (1988). In Section 4.2 we propose an alternative approach.

Thus, the application of Theorem 1 to this example, jointly with well-known efficiency theory, implies that the tests proposed in Stute (1997), Stute, Thies and Zhu (1998), and Escanciano (2009) are approximately semiparametrically efficient. They are not fully efficient because the number of components used (the number of summands in L) was fixed in these applications. In Section 4 we construct feasible versions of such procedures that are fully semiparametrically efficient. Our results in this example also show that the FNPT test in Boning and Sowell (1999) is not efficient. These authors further assume ε_i to be independent of X_i, but they do not use this information in the moment restrictions. Note that Boning and Sowell's (1999) test is still efficient in the sense of Müller (2011), so this example illustrates the differences between Müller's (2011) efficiency and the classical semiparametric efficiency. Note also that a simple modification of the moments used can account for the independence between errors and regressors, so that to deliver an efficient test under the independence assumption of Boning and Sowell (1999). □

Remark 2: All our results go through in the previous example if we replace the indicator function in $w(X_i, x)$ by other comprehensively revealing class of functions. See Bierens (1982), Stinchcombe and White (1998) and Escanciano (2006) for examples of such classes. For instance, we could use the class $\{\exp(x\phi(X)) : x \in \Gamma \in \mathbb{R}\}$, where Γ is an interval containing zero, and ϕ is a one-to-one bounded mapping, see Bierens and Ploberger (1997). As mentioned earlier, 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 a 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

$$E[\exp(x\phi(X_i))b(X)] = 0, \text{ for all } x \in \Gamma,$$

which in turn, implies that $b(X) = 0$ a.s. Since any function can be decomposed as $b(X) = c_0 + c_1X + c_2b^{\perp}(X)$, 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.

3 On the Efficiency of the FNPT

We show in this section that the FNPT is a semiparametric efficient test in the class of semiparametric models defined by (1). Specifically, we use the concept of *asymptotically uniformly most powerful and invariant* test of level α, in short AUMPI (α), defined formally in Choi et al. (1996, Section 5). When $p = 1$ alternative definitions of efficiency that do not require invariance are typically used. For definitions of standard concepts used in semiparametric estimation theory, such as regular parametric submodels or tangent spaces the reader is referred to Bickel et al. (1993). Let $\mathcal{P} := \{P_{(\beta,\eta)} : \beta \in \Theta_\beta, \eta \in \Theta_\eta\}$ be a 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 $\dot{\ell}_1$ in the marginal family $\mathcal{P}_1 = \{P_{(\beta,\eta_0)} : \beta \in \Theta_\beta\}$, we define the efficient score ℓ_1^* as the orthogonal projection of the score $\dot{\ell}_1$ onto the orthocomplement of $\dot{\mathcal{P}}_2$, i.e., $\ell_1^* := \dot{\ell}_1 - \Pi_{\dot{\mathcal{P}}_2}\dot{\ell}_1$, where $\Pi_{\dot{\mathcal{P}}_2}h$ denotes the orthogonal projection in $L_2(F)$ of h onto $\dot{\mathcal{P}}_2$. Let $I^* := Var(\ell_1^*)$, and assume I^* is positive definite. Write $\xi_n(\eta_0) := n^{-1/2}I^{*-1/2}\sum_{i=1}^n \ell_1^*(Z_i, \eta_0)$. An efficient test statistic T_n must satisfy $T_n = \xi_n(\eta_0) + o_P(1)$, for every η_0. Choi et al. (Corollary 3, 1996) show that the test $\phi_n^* := 1\left(T_n'T_n \geq \chi^2_{1-\alpha,p}\right)$ is AUMPI (α). Hence, the FNPT will be AUMPI (α) if we prove that, for every η_0,

$$L(\hat{R}_n) = \frac{1}{n}\sum_{i=1}^n \ell_1^*(Z_i, \eta_0) + o_P(n^{-1/2}).$$

In view of Theorem 1, this is the case if and only if $s^* = \ell_1^*$. Our next result proves that this is indeed the case. Define $\ker^0(T') := \{h \in \ker(T') : E[h(Z)] = 0\}$. Standard regularity conditions that imply LAN, among other things, and that are required for the definition of efficiency are gathered in the Appendix A.

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

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

(ii) $s^* \equiv \ell_1^*$*, and hence the FNPT is AUMPI (α).*

The result in 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 large class of semiparametric models. Theorem 2 can be used to obtain efficient inference in models such as the quantile regression model or in semiparametric models with partial identification, as shown in Section 5. Theorem 2(ii) shows the semiparametric efficiency of the FNPT.

For completeness, we discuss an alternative sense of efficiency of the FNPT. Müller (2011) has recently shown that the FNPT is optimal in a class of tests that control asymptotic size for all data generating processes for which the underlying random element, $\hat{R}_n$, has the corresponding limiting distribution. We particularize Müller's results to our framework, and discuss connections with the semiparametric efficiency results established here. 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 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 FNPT is the most efficient test in the class $\mathcal{C}$, and for any other test in $\mathcal{C}$ with higher asymptotic average 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 FNPT. Our paper complements Müller's efficiency results by proving that in regular semiparametric problems the FNPT is also semiparametric efficient in the "classical" sense of Choi et al. (1996).

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, 2008). 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. Our discussion here is intentionally informal. Some formal results are provided in Section 4, but a complete set of results is beyond the scope of this paper. We modify Carrasco and Florens (2000, 2008) 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

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

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) efficient concept even in non-regular settings.

4 Implementation of the Feasible FNPT

We have investigated so far the efficiency properties of the infeasible FNPT. The test is not feasible because the Fisher information matrix Σ and the operator L are in general unknown. The implementation of the feasible FNPT greatly depends on whether or not the spectrum of K is known. Here, we suggest different implementations for these two exhaustive alternatives. Henceforth, $\widehat{m}(z,x) := m(z,x,\beta_0,\widehat{\eta}_n)$ and $\widehat{m}_i(x) := m(Z_i,x,\beta_0,\widehat{\eta}_n)$.

4.1 Known Spectrum

If the spectrum $\{\lambda_j,\varphi_j\}$ is known, L and Σ can be easily estimated 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 FNPT considered here replaces $L(\hat{R}_n)$ by $L_k(\hat{R}_n)$ and Σ by $\widehat{\Sigma}_k$. For instance, when the moment function is smooth in β a natural estimate for D is

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

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. Note that Stute (1997) and Escanciano (2009) have used similar approximations to (17), but they have not investigated the properties of the resulting tests as $k_n \to \infty$. Therefore, their tests are only approximately efficient. Full efficiency requires $k_n \to \infty$, and it is developed in this section.

It turns out that, under suitable conditions provided below, the feasible FNPT behaves asymptotically as the infeasible test, i.e.

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

The following assumption restricts the rate of divergence of k_n and requires further smoothness in the model.

Assumption R: *(i)* $k \equiv k_n \to \infty$, *(ii)* $\left\|\hat{R}_n - M_n\right\|_1 = o_P(n^{-1/2})$ *and* $\left\|\hat{D} - D\right\|_1 = o_P(k_n^{-1})$.

We will provide specific restrictions that R(ii) imposes for a generic example below. Assumption R(ii) can be replaced by $\left\|\hat{R}_n - M_n\right\| = o_P(n^{-1/2})$, $\left\|\hat{D} - D\right\|_2 = o_P(1)$ 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 $D_l \in Im(K)$ or $s_l^* \in Im(T)$ for all $l = 1, ..., p$. If there are no nuisance parameters, then Assumption R can be simplified to $k_n \to \infty$ and $\left\|\hat{D} - D\right\|_1 = o_P(1)$.

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

A corollary of Proposition 1 is that the feasible FNPT is an AUMPI(α) test. Proposition 1 is applicable to cases where the asymptotic limit distribution $\hat{R}_n$ has 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

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

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 into the space orthogonal to the tangent space of nuisance parameters, see Song (2010) for details. It can be shown that our results applied to the moment function $\psi(Z_i, x, \beta_0, \eta_0) = \rho(Z, \beta, \eta_0) \mathcal{M} 1(X \leq x)$ 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 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.

EXAMPLE 2 (CONT.): **Regression model checks**. Stute, Thies and Zhu (1998) investigated omnibus asymptotic distribution-free tests based on the Khmaladze's transformation applied to the process $\hat{R}_n$ in (16) under conditional heteroskedasticity. The limiting Gaussian process after the transformation (including the integral transformation) is a standard Brownian motion, whose spectrum is 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). As mentioned earlier, a sufficient and necessary condition for Assumption D is that $E[\sigma^{-2}(X)a^2(X)] < \infty$. Some simple algebra shows that $\left\|\hat{R}_n - M_n\right\|_1 = O_p(n^{-1/2}k^{1/2}(\widehat{\eta}_n - \eta_0))$. Hence, a sufficient condition for $\left\|\hat{R}_n - M_n\right\|_1 = o_P(n^{-1/2})$ is $nk_n^{-1} \to \infty$. Similarly, it can be shown that $\left\|\hat{D} - D\right\|_1 = o_P(k_n^{-1})$, provided $nk_n^{-3} \to \infty$.

4.2 Unknown Spectrum

As mentioned earlier, in most applications the spectrum $\{\lambda_j, \varphi_j\}$ is unknown. One possible approach, as suggested by Carrasco and Florens (2000), is to estimate nonparametrically the spectrum. Here, we propose an alternative method that is 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). 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 analogue 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 FNPT 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 regularize the problem. 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 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 are 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, when μ is the probability measure of X, 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 arguments show that the finite sample version $(\alpha_n I + \hat{T}T')s^*_{\alpha_n} = \hat{T}\hat{D}$ 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

$$b_j = \frac{1}{n^2}\sum_{h=1}^{n}\hat{D}(x_h)\widehat{m}_j(x_h).$$

Finally, the Fisher information matrix is estimated by

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

The αth level feasible FNPT is then given by

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

The test only requires estimates $\{\widehat{m}(Z_i,x_j), \hat{D}(x_j)\}_{i,j=1}^n$ and is quite easy to implement. We show below that $\hat{\phi}^*_n$ is asymptotically equivalent to the infeasible ϕ^*_n, by showing that under suitable conditions

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

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.

Assumption RE: *(i)* $n\alpha_n^4 \to \infty$ *and* $\alpha_n \to 0$ *as* $n \to \infty$; *(ii)* $\sqrt{n}(\hat{D} - D)$ *is asymptotically tight in* $L_2(\mu)$ *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. A sufficient condition for asymptotic tightness is weak convergence, as implied by Prohorov's theorem, see van der Vaart and Wellner (1996). 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 and the FCLT.

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

Remark 3: 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 $\hat{D} = \hat{T}'d$ for some d.

5 Application to a Semiparametric Linear Quantile Regression

We implement the efficient feasible FNPT for the quantile regression example. At the end of this section we discuss other potential applications of our methods for which available methods are hard to apply. We modify the notation in the QR example to account for the presence of additional infinite dimensional nuisance parameters in the limiting distribution, so the model is defined by the moment restrictions

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

where $\zeta_i(\tau) = 1(Y_i \le \beta' X_{1i} + \gamma_0'(\tau)X_{2i}) - \tau$. Define $\delta_0(\tau) := (\beta_0', \gamma_0'(\tau))'$ and $\theta_0(\tau) := (\delta_0'(\tau), f_{2i\tau})'$, where $f_{2i\tau}$ is the conditional density of $Y_i - \beta_0' X_{1i}$ given X_{2i}, evaluated at $\gamma_0(\tau)'X_{2i}$. A natural estimator for $\gamma_0(\tau)$ is the QR estimator, initially proposed by Koenker and Basset (1978), 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 $\rho_\tau(u) = u\left(\alpha - 1\left\{u \le 0\right\}\right)$ 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. Our algorithm for computing the efficient score suggests first to solve $T's(x) = D$ and then find the projection $s^* = \Pi_{\ker^\perp(T')}s$. The first step is straightforward in this example – a solution is $s(Z) = X_1 \dot{f}\left(Y|\, X\right)/f\left(Y|\, X\right)$, 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 $\Pi_{\ker^\perp(T')}s$ seems to be a rather complicated task. This difficulty does not stop us from implementing a feasible FNPT 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

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

and

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

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

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

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

and $\hat{f}_{i\tau}$ ($\hat{f}_{2i\tau}$) is a nonparametric estimator for the conditional density of Y_i ($Y_i - \beta_0' X_{1i}$) given X_i (X_{2i}) evaluated at $X_i'\theta_0(\tau)$ ($\gamma_0(\tau)' X_{2i}$). We follow Escanciano and Goh (2012) and construct estimators for these quantities 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 estimators for $f_{2i\tau}$ and $f_{2i\tau}$ are, respectively, $\hat{f}_{i\tau} := \hat{f}\left(X_i'\widehat{\delta}(\tau) \middle| X_i\right)$ and $\hat{f}_{2i\tau} := \hat{f}_2\left(X_{2i}'\widehat{\gamma}_n(\tau) \middle| X_{2i}\right)$, where

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

and

$$\hat{f}_2\left(y | X_{2i}\right) \equiv \hat{f}_2\left(y | X_{2i}, \widehat{\gamma}_n\right) := \frac{1}{nh}\sum_{j=1}^{n} K\left(\frac{y - X_{2i}'\widehat{\gamma}_n(\tau)}{h}\right)$$

and where $\widehat{\delta}(\tau)$ is a QR estimator that does not impose the null, $h > 0$ is a scalar smoothing parameter and $K(\cdot)$ is a smoothing kernel satisfying some conditions below. To simplify the notation denote $\tilde{f}_{i\tau} := \hat{f}\left(X_i'\delta_0(\tau) | X_i, \delta_0\right)$ and $\tilde{f}_{2i\tau} := \hat{f}_2\left(X_{2i}'\gamma_0(\tau) | X_{2i}, \gamma_0\right)$ the kernel estimates using the true QR parameters. See Escanciano and Goh (2012) for motivation of the nonparametric estimates $\hat{f}_{i\tau}$ and $\hat{f}_{2i\tau}$. These estimators posses several appealing properties over more classical kernel estimates (cf. Rosenblatt, 1969).

For a fixed τ and τ_j in $\mathcal{T}$, let $g_{(\tau,\tau_j)}(u, v)$ and $g_{2(\tau,\tau_j)}(u, v)$ be the densities of $(X'\delta_0(\tau_j), X'\delta_0(\tau))$ and $(X_2'\gamma_0(\tau_j), X_2'\gamma_0(\tau))$ evaluated at (u, v), and define the functions ($\otimes$ denotes the Kronecker product)

$$\begin{aligned}
r_{1(\tau,\tau_j)}(u, v, w) &:= E\left[X_1 1(X_i \le w) | X'\delta_0(\tau) = u, X'\delta_0(\tau_j) = v\right], \\
r_{2(\tau,\tau_j)}(u, v, w) &:= E\left[X_1 \otimes X_2 1(X_i \le w) | X_2'\gamma_0(\tau) = u, X_2'\gamma_0(\tau_j) = v\right], \\
q_{1(\tau,\tau_j)}(u, v, w) &:= r_{1(\tau,\tau_j)}(u, v, w) g_{(\tau,\tau_j)}(u, v)
\end{aligned}$$

and

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

Then, regularity conditions that are sufficient for our high-level assumptions in the quantile regression example are given as follows.

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), with uniformly* *bounded derivative with respect to $y \in \mathbb{R}$; (iii) the density $f(y|x)$ is twice continuously differentiable in x, with uniformly bounded derivatives; (iv) $E[XX']$ is nonsingular and finite; (v) for each fixed τ and τ_j in $\mathcal{T}$, $u \in \mathbb{R}$ and $w \in \mathbb{R}^{d_x}$, $k = 1, 2$, the function $q_{k(\tau,\tau_j)}(u, v, w)$ is well-defined and twice continuously differentiable in v with uniformly (in τ, τ_j, u and w) bounded derivatives and the first derivative of $q_{1(\tau,\tau_j)}(u, v, w)$ with respect to u is Liptschitz in (τ, w) for each τ_j, u and v.*

Assumption E2: *For all $\tau \in \mathcal{T}$, the parameter $\delta_0(\tau) \in \Theta \subset \mathbb{R}^{d_x}$, Θ is compact and $\delta_0(\tau)$ belongs to its interior.*

Assumption E3: *(a) The kernel function $K(t) : R \to R$ is symmetric, bounded, three times continuously differentiable and satisfies the following conditions: $\int K(t)\,dt = 1$, $\int tK(t)\,dt = 0$, and $\int \left|t^2 K(t)\right| dt < \infty$, $\left|\partial^{(j)} K(t)/\partial t^j\right| \leq C$ and for some $v > 1$, $\left|\partial^{(j)} K(t)/\partial t^j\right| \leq C\,|t|^{-v}$ 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 $s(Z) = X_1 \dot{f}(Y|X)/f(Y|X)$ of $T's(x) = D$ is well-defined, so Assumption D holds. Note that bounded and smoothness conditions on the conditional density also imply similar conditions on the conditional density of $Y_i - \beta_0' X_{1i}$ given X_{2i}. Our next result shows the optimality of the FNPT applied to this example.

Theorem 4: *Suppose the conditions of Assumptions E1-E3, Assumption RE(i) and RE(iii) hold. Then, the $\alpha-$level feasible FNPT in (19) with $\widehat{s}^*$ as in (20) is AUMPI (α) for the quantile regression model.*

We briefly mention other 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, Powell's (1986) censored regression quantile estimators and transformation models. A general class of models for which our results have applications is the class of semiparametric partially identified models investigated in Escanciano and Zhu (2012); see also Arellano et al. (2011) for parametric versions. Efficiency within a class of GMM estimates has been discussed in Arellano et al. (2011) for the parametric setting, but in the semiparametric setting remains 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 $x_0 \in \Gamma$. 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. To illustrate the application to partially identified models we consider an example from Altonji et al. (2005).

Example 3: **Assessing the effectiveness of Catholic Schools**. Altonji et al. (2005) investigated the effect of attending a Catholic school on educational attainment. In this empirical study Y is e.g. college attendance, CH is a dummy for Catholic school attendance and X is a vector of individual characteristics, including family background, demographics, etc. They use a bivariate probit model

$$\begin{aligned} CH &= 1(X'\gamma_0 + u), \\ Y &= 1(X'\delta_0 + \beta CH + v), \end{aligned}$$

where (u,v) is jointly normal with correlation τ. Their approach to deal with the lack of exclusion restriction is to consider τ as an unidentified parameter, so effectively assuming that the model is partially identified (i.e. conditional on τ, the rest of parameters are identified). Altonji et al. (2005) were particularly interested in testing significance of CH, and they proposed pointwise inferences for several choices of τ. A more efficient approach can be based on the methods developed in this paper. For instance, suppose we would like to test the individual hypothesis $H_0 : \beta(\tau_0) = 0$ against $H_0 : \beta(\tau_0) \neq 0$ for some $\tau_0 \in \mathcal{T}$ and a certain set $\mathcal{T}$, or suppose we want to test the more stringent hypothesis $H_0 : \beta(\tau) = 0$ for all $\tau \in \mathcal{T}$ against $H_0 : \beta(\tau_0) \neq 0$ for some $\tau_0 \in \mathcal{T}$. Both testing problems can be handled by our methods using as moments the set of score equations or transformations of them as in (24). For instance, in the second testing problem we can consider an average power criteria, and consider the model as

$$E[\psi(Z,\tau,\beta,\eta_0(\tau))] = 0 \text{ for all } \tau \in \mathcal{T},$$

where ψ is the set of scores from the bivariate probit model, $Z = (Y, X', CH)'$, $\eta_0(\tau) = (\gamma_0'(\tau), \delta_0'(\tau))'$ and the choice of $\mathcal{T}$ can be based on the effect of selection. See Altonji et al. (2005) for details on the choice of $\mathcal{T}$. Our test, applied with $\beta_0 = 0$, would exploit information from cross-equation restrictions, and would lead to more efficient inferences than the pointwise results considered in Altonji et al. (2005). Again, for this class of examples existing efficiency theory might be hard to apply, whereas our methods lead to relatively simple efficient inferences. □

6 Final Remarks

In this paper, we have investigated the efficiency, in a classical semiparametric sense, and implementation of the FNPT in a general class of semiparametric models. We have shown that under quite general conditions the FNPT is asymptotically equivalent to a score-type test. We have suggested a general algorithm for computing the associated score function in terms of the covariance operator and the shift function resulting under local alternatives. The semiparametric efficiency of the FNPT has been established by showing that the score function is the efficient score function associated to the model. We have proposed and justified feasible versions of the FNPT when the spectrum is known and when is unknown. Finally, we have applied our results to a semiparametric quantile regression model. Our investigation complements the optimality results for the FNPT found in Müller (2011), and shows that the functional Neyman-Pearson approach advocated by Grenander (1950) can lead to semiparametric efficient inference. In sum, this paper extends the FNPT to general semiparametric models, establishes its semiparametric efficiency in regular models and justifies simple practical implementations of these procedures.

In addition to the efficiency, the main appealing property of the FNPT is its wide applicability. It can be applied to any of the myriad of papers where omnibus tests have been proposed, and which use continuous functionals of the sample analog of the moment restrictions. It can be also applied to non-regular problems, and in these problems it also possesses optimality properties as shown by Müller (2011).

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 $\Sigma = \|D\|_1 = Var(s^*(Z_i))$, and we have provided consistent estimators for it and a new algorithm for computing 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 using similar methods to those well established in the literature, see Lecam (1956). Obtaining more general estimation results is a priority in our research agenda. Efficient estimation can be achieved by GMM estimators, along the lines of Carrasco and Florens (2000, 2008) and Newey (2004). The results developed of this paper can be useful to extend the existing GMM theory to our semiparametric setting; see our proposal in Section 3.

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 for bandwidth choice for testing is not available, even in much simpler settings. 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. 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, e.g. Markov processes, efficiency results are available and our results can be straightforwardly extended; see Carrasco and Florens (2000, 2008) for important results in this direction. We have applied the FNPT to finite dimensional parameters, but it can be also applied to infinite dimensional parameters. It is unknown whether or not the FNPT delivers in this case optimal inference. This extension would have important applications in efficient inference in partially identified models. Finally, our theory has been restricted to situations where the LAN holds with a limit distribution of the form $R_\infty \equiv R_\infty^0 + c_\beta' D$, for a Gaussian process R_∞^0. There are, however, instances where the impact of the local parameter c_β in the limiting distribution is nonlinear, such as in unit-root testing based on partial-sum processes, or the limiting distribution is non-Gaussian. It should be of interest to investigate the semiparametric optimality properties of the resulting FNPT in these non-standard cases, in comparison with those already established by Müller (2011).

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 $\Theta_\beta(\delta) := \{\beta \in \Theta_\beta : |\beta - \beta_0| \leq \delta\}$ and $\Theta_\eta(\delta) := \left\{\eta \in \Theta_\eta : \|\eta - \eta_0\|_\eta \leq \delta\right\}$ 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 $\{\eta + \lambda\left(\overline{\eta}-\eta\right) : \lambda \in [0,1]\} \subset \Theta_\eta$ and

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

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

Assumption A1: *Suppose that:*

(i) (Smoothness in η) for each $x\in\Gamma$, the pathwise derivative $V_\eta\left(x,\beta_0,\eta_0\right)[\eta-\eta_0]$ 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

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

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

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

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

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

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

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 semiparametric literature. Related assumptions are given in Escanciano and Zhu (2012) for 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

$$m(z,x,\beta,\eta_0) := \psi(z,x,\beta,\eta_0)+\phi(z,x,\beta,\eta_0), \tag{27}$$

where ϕ is as in A1(iv).

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

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

Proof of Theorem A1: Henceforth, to simplify the notation when we evaluate β at β_0 we remove the dependence on β_0 from all arguments. Define the linear approximation

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

First, by Assumption A1(ii),(iii),(iv), uniformly in $x\in\Gamma$,

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

Hence, we conclude 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.

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 Choi et al. (1996). Define the local parameters, $t \in [0,\infty)$,

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

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 $\|\cdot\|_{l\eta}$. Note that $c = (c_\beta, c_\eta)$ denotes the direction in which the local parameter $\theta_t(c) \triangleq (\beta_t(c_\beta), \eta_t(c_\eta))$ 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 $f\left(z|\,\theta\right)$ 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 if the density of P_t, say f_t, is mean-square differentiable

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

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

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

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

$$E_t\left[q(Z)\right] := \int q(z) f_t(z) dz,$$

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: It is well known that an important implication of (29) is the LAN property

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

see e.g. van der Vaart (1998, Theorem 7.2). 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 apply Lecam's third lemma to $\left\langle \sqrt{n}\hat{R}_n, h \right\rangle$ with $h \in L_2(\mu)$, 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 $T'g_\eta = 0$ 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 (integration by parts), Lemma 7.2 in Ibragimov and Hasminskii (1981), under Assumption A2. ■

7.1.2 Preliminary results for the QR example

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 Euclidean and function-valued parameters. In this connection, define a generic function class

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

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

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

Other definitions of concepts from empirical processes theory may be found in e.g., van der Vaart and Wellner (1996).

Lemma Q1. *Assume that*

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

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

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

A typical application of Lemma Q1 implies that $J(\delta, \mathcal{G}, \|\cdot\|_{2,P}) < \infty$ 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 $\delta_n = o(1)$,

$$\sup_{\|\theta_1-\theta_2\|_{\Theta}\leq\delta_n,\, \|g_1-g_2\|_{\mathcal{G}}\leq\delta_n} |M_n(\theta_1, g_1) - M_n(\theta_2, g_2) - M(\theta_1, g_1) + M(\theta_2, g_2)| = o_P(n^{-1/2}). \tag{30}$$

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*

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

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. Given a sequence $\{Z_{in}\}_{i=1}^{n}$ of iid arrays for each n, define the weighted empirical process

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

which is indexed by $\chi := (\delta, x) \in \mathcal{B} \times \Gamma$. Let F_X denote the cdf of X. Define the 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 w_n is such that for $\delta_n \downarrow 0$

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

and $W_n := \sup_{\chi}|w_n(\cdot, \chi)|$ 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 $\mathcal{W}_n := \{w_n(\cdot, \delta, x) : (\delta, x) \in \mathcal{B} \times \Gamma\}$ and require the following assumption:

Assumption Q1. *The class $\mathcal{W}_n$ satisfies the previous conditions and is such that $J(\delta_n, \mathcal{W}_n, \|\cdot\|_2) \to 0$ for every $\delta_n \downarrow 0$.*

Theorem Q1. *Under Assumptions E1, E2 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. ■

The following lemma is needed to justify the Bahadur representation of the QR estimator. Recall $f_2\left(y|\, X_2 = x_2\right)$ denotes the conditional density of $Y_i - \beta_0' X_{1i}$ given $X_{2i} = x_2$ evaluated at y.

Lemma Q3. *If $f\left(y|\, X = x\right)$ satisfies Assumption E1, then $f_2\left(y|\, X_2 = x_2\right)$ also satisfies Assumption E1.*

Proof of Lemma Q3. Let $g_{12}\left(x_1|\, X_2 = x_2\right)$ denote the conditional density of X_{1i} given $X_{2i} = x_2$ evaluated at x_1. Then, correct specification and simple algebra implies the relation

$$f_2\left(y|\, X_2 = x_2\right) = \int f\left(y + \beta_0' x_1 \middle|\, x\right) g_{12}\left(x_1|\, X_2 = x_2\right) dx_1.$$

The result then follows from this expression and simple arguments. ■

Lemma Q4. *Under Assumption E1,*

(i)

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

(ii) *The estimators $\widehat{\delta}(\cdot)$ and $\widehat{\gamma}_n(\cdot)$ satisfy that $\widehat{\delta}, \widehat{\gamma}_n \in \mathcal{B}$ with probability tending to one and $\delta_0 \in \mathcal{B}$, with $\mathcal{B}$ a class of Lipschitz functions from $\mathcal{T}$ to Θ.*

Proof of Lemma Q4. For (i) see for instance Gutenbrunner and Jurecková (1992). Part (ii) follows from (i) and our Assumptions E1 in a routine fashion. ■

Our next result is related to the uniform convergence rates for kernel estimators $\hat{f}_{i\tau}$ and $\hat{f}_{2i\tau}$. We view $\hat{f}_{i\tau}$ as a function of $\widehat{\delta}$ and write the Taylor approximation around the true value δ_0 as

$$\hat{f}_{i\tau} = \tilde{f}_{i\tau} + \dot{f}_{i\tau}(\widehat{\Delta}) + \ddot{f}_{i\tau}(\widehat{\Delta}) + r_{i\tau}, \tag{31}$$

where $\widehat{\Delta}(\cdot) := \sqrt{n}(\widehat{\delta}(\cdot) - \delta_0(\cdot))$ and

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

and

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

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. A similar expansion to (31) holds for $\hat{f}_{2i\tau}$,

$$\hat{f}_{2i\tau} = \tilde{f}_{2i\tau} + \dot{f}_{2i\tau}(\widehat{\Delta}) + \ddot{f}_{2i\tau}(\widehat{\Delta}) + r_{2i\tau}, \tag{32}$$

with obvious definitions for the terms in the expansion.

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, for $j = 1$ and 2*

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

and

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

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 and E2,*

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

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 $\mathcal{Q} := \{z \to 1(y \leq \beta_0' x_1 + \gamma' z_{2i}) - \tau : \gamma \in \Theta, \tau \in \mathcal{T}\}$. The proof of the following result is standard, and hence omitted.

Lemma Q7. *Let Assumption E1 hold. Then, the class $\mathcal{Q}$ of functions is VC, and hence satisfies Assumption Q1.*

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 $s(Z)$, and satisfying the semiparametric 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

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

The first term is just the pathwise derivative of $\gamma(t) := E\left[\psi(Z, x, \beta_0, \eta_t(Z, x))\right]$, which by our Assumption 1 satisfies

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

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

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

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, Section 25.3). 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) without changes 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 $E_t\left[m(Z, x, \beta_t, \eta_0)\right] = 0$. Differentiating this equation with respect to t at 0 we get

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

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

$$
\partial E_t\left[m(Z, x, \beta_0, \eta_0)\right]\big|_{t=0} = E\left[m(Z, x, \beta_0, \eta_0)\dot{\ell}_1(Z)\right],
$$

where $\dot{\ell}_1$ is the score with respect to β at β_0. Hence, we conclude using our notation that

$$D \equiv T'\dot{\ell}_1. \tag{33}$$

Hence, by part (i), the zero mean property of scores and Theorem 1(ii) $\ell_1^* = \dot{\ell}_1 - \Pi_{\dot{\mathcal{P}}_2}\dot{\ell}_1 = \Pi_{\ker^{\perp}(T')}\dot{\ell}_1 = s^*$. ■

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

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

By Cauchy-Schwarz's inequality and Assumptions D and R the absolute value of all terms is $o_P(1)$, where for the last term we use that $\{\lambda_j^{-1/2} \left\langle M_n, \varphi_j \right\rangle\}_{j=1}^{\infty}$ are uncorrelated and with unit variance. This also shows that $L_k(M_n) = L(M_n) + o_P(n^{-1/2})$, since under Assumption D,

$$nVar\left(L_k(M_n) - L(M_n)\right) = \sum_{j=k+1}^{\infty} \lambda_j^{-1} \left\langle D, \varphi_j \right\rangle^2 \to 0,$$

as $k \to \infty$. 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 $B_n^* = o_P(1)$. Using well-known expansions for $s^*_{\alpha_n}$ and s^* we can 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 $b(\alpha,\lambda) = \alpha/(\alpha+\lambda)$ and $\{\varepsilon_{nj}\}_{j=1}^{\infty}$ are implicitly defined. Note that $\{\varepsilon_{nj}\}_{j=1}^{\infty}$ are uncorrelated, with zero mean and unit variance. Hence, $E[B_n^*] = 0$ 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 $L_2(F)$ to $L_2(F)$ and h_n is a random element of $L_2(F)$, then

$$\left|\frac{1}{n}\sum_{i=1}^{n} B_n h_n(Z_i)\right| = O_P\left(\|B_n\|_{2,P}\left(E[h_n^2(Z)]\right)^{1/2}\right) \tag{34}$$

The proof of the last equality follows from Chebyshev and Cauchy-Schwarz inequalities.

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, 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 $\|\cdot\|_n$. A well known result in operator theory shows that

$$\left\|\hat{T} - T\right\|_n = \left\|\hat{T}' - T'\right\|_{2,n}$$

By Markov's inequality

$$\begin{aligned}
\left\|\hat{T}' - T'\right\|_{2,n}^2 &= O_P\left(\left\|\hat{T}' - T'\right\|_{2,P}^2\right) \\
&= O_P\left(n^{-1}\right).
\end{aligned}$$

Hence, $\left\|\hat{T} - T\right\|_n^2 = O_P\left(n^{-1}\right)$. But a simple inequality implies $\left\|\hat{T} - T\right\|^2 \leq E\left[\left\|\hat{T} - T\right\|_n^2\right]$, where the expectation is with respect to empirical distribution used in $\|\cdot\|_n$, and hence

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

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 $\widehat{A}_{\alpha_n} = (\alpha_n I + \hat{T}T')^{-1}$ 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 prove that

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

By Prohorov's theorem we can assume w.l.g that $h_n(\cdot) := \sqrt{n}(\hat{D} - T's^*)(\cdot)$ converges weakly to h_∞. Then, applying (34) with $B_n = \widehat{A}_{\alpha_n}\hat{T} - A_{\alpha_n}T$, we conclude

$$
\begin{aligned}
\int \sqrt{n}\Delta_{1n}(z)F_n(dz) &= \frac{1}{n}\sum_{i=1}^{n} A_{\alpha_n}Th_n(Z_i) + o_P(1) \\
&= \frac{1}{n}\sum_{i=1}^{n} A_{\alpha_n}Th_\infty(Z_i) + o_P(1) \\
&= o_P(1),
\end{aligned}
$$

where the second equality follows from another application of (34) and the last equality follows from a law of large numbers for arrays, after noting that

$$
\frac{\sup_n E\left[(A_{\alpha_n}Th_\infty(Z_i))^2\right]}{n} \to 0
$$

and

$$
E\left[A_{\alpha_n}Th_\infty(Z_i)\right] = 0.
$$

As for Δ_{2n}, we write,

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

Thus,

$$
\int \sqrt{n}\Delta_{2n}(z)F_n(dz) = \int \sqrt{n}\Delta_{21n}(z)F_n(dz) + \int \sqrt{n}\Delta_{22n}(z)F_n(dz),
$$

where $\Delta_{21n}(z) = (\widehat{A}_{\alpha_n} - A_{\alpha_n})(\hat{T}T' - TT')(s^* - s^*_{\alpha_n})(z)$ and $\Delta_{22n}(z) = A_{\alpha_n}(\hat{T}T' - TT')(s^* - s^*_{\alpha_n})(z)$. Using the equalities that

$$
\left\|\widehat{A}_{\alpha_n}\right\| = \left\|A_{\alpha_n}\right\| = O_P(\alpha_n^{-1})
$$

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

$$\widehat{A}_{\alpha_n} - A_{\alpha_n} = \widehat{A}_{\alpha_n}(TT' - \hat{T}\hat{T}')A_{\alpha_n}$$

and

$$E[\left|s^*(Z) - s^*_{\alpha_n}(Z)\right|^2] = O(\alpha_n^2),$$

we obtain

$$\begin{aligned}\left(E[\left|\sqrt{n}\Delta_{21n}(z)\right|^2]\right)^{1/2} &\leq \sqrt{n}\left\|\hat{T}\hat{T}' - TT'\right\|^2 \left(E[\left|s^*(Z) - s^*_{\alpha_n}(Z)\right|^2]\right)^{1/2} \left\|\widehat{A}_{\alpha_n}\right\| \|A_{\alpha_n}\| \\ &= o_P(n^{-1/2}\alpha_n^{-1}) \\ &= o_P(1).\end{aligned}$$

Hence,

$$\int \sqrt{n}\Delta_{21n}(z)F_n(dz) = o_P(1).$$

Note that

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

Hence, our assumptions imply that $\sqrt{n}(\hat{T}T' - TT')\psi_j(z)$ is asymptotically tight, and by Prohorov's theorem we can assume it converges weakly, w.l.g. Now, the expansion for $(s^* - s^*_{\alpha_n})(z)$, $\|D\|_2 < \infty$ and the assumption on the weak convergence of $\sqrt{n}(\hat{T}T' - TT')\psi_j(z)$ implies the weak convergence of $\sqrt{n}(\hat{T}T' - TT')(s^* - s^*_{\alpha_n})(z)$ by Lemma 1 in Escanciano and Velasco (2006). Applying the same arguments as for Δ_{1n} to Δ_{22n}, we obtain

$$\begin{aligned}\int \sqrt{n}\Delta_{22n}(z)F_n(dz) &= \frac{1}{n}\sum_{i=1}^n A_{\alpha_n}G_\infty(Z_i) + o_P(1) \\ &= o_P(1),\end{aligned}$$

where G_∞ is the weak limit of $\sqrt{n}(\hat{T}T' - TT')(s^* - s^*_{\alpha_n})(z)$ and the second equality follows from a law of large numbers and the convergence

$$\frac{\sup_n E\left[\left(A_{\alpha_n}G_\infty(Z_i)\right)^2\right]}{n} \leq \frac{E\left[G^2_\infty(Z_i)\right]}{n\alpha_n^2} \to 0.$$

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-E4. To verify Assumption W, we apply Theorem Q1. Recall

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

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,

$$\sup_{x \in \mathcal{T} \times \mathbb{R}^{d_x}} \left| \hat{R}_n(x) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \zeta_i(\tau) 1(X_i \le w) - A'(x) \sqrt{n} \left(\widehat{\gamma}_n(\tau) - \gamma_0(\tau) \right) \right| = o_P(1), \tag{35}$$

where $A(x) := E[X_{2i} f_{2i\tau} 1(X_i \le w)]$. To obtain this expansion we apply Theorem Q1 with 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

$$\begin{aligned} \sup_{x \in \mathcal{T} \times \mathbb{R}^{d_x}} \Bigg| & \hat{R}_n(x) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \zeta_i(\tau) 1(X_i \le w) \\ & + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(E\left[\zeta_i(\tau) \middle| X_i \right] - E\left[\hat{\zeta}_i(\tau) \middle| X_i \right] \right) 1(X_i \le w) \Bigg| = o_P(1). \end{aligned} \tag{36}$$

Applying a mean value argument we obtain

$$\begin{aligned} \sup_{x \in \mathcal{T} \times \mathbb{R}^{d_x}} \Bigg| & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(E\left[\zeta_i(\tau) \middle| X_i \right] - E\left[\hat{\zeta}_i(\tau) \middle| X_i \right] \right) 1(X_i \le w) \\ & - \sqrt{n} \left(\widehat{\gamma}_n(\tau) - \gamma_0(\tau) \right)' \frac{1}{n} \sum_{i=1}^{n} f_2\left(X_{2i}' \widetilde{\gamma}_n(\tau) \middle| X_{2i} \right) X_{2i} 1(X_i \le w) \Bigg| = o_P(1), \end{aligned}$$

where $\widetilde{\gamma}_n(\tau)$ is such that $|\widetilde{\gamma}_n(\tau) - \gamma_0(\tau)| \le |\widehat{\gamma}_n(\tau) - \gamma_0(\tau)|$ a.s. for each $\tau \in \mathcal{T}$. By our assumptions, uniformly in $x \in \mathcal{T} \times \mathbb{R}^{d_x}$,

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^{n} f_2\left(X_{2i}' \widetilde{\gamma}_n(\tau) \middle| X_{2i} \right) 1(X_i \le w) &= \frac{1}{n} \sum_{i=1}^{n} f_2\left(X_{2i}' \gamma_0(\tau) \middle| X_{2i} \right) X_{2i} 1(X_i \le w) + o_P(1) \\ &= A(x) + o_P(1). \end{aligned}$$

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

$$\left\{ z \to f_2\left(x_2' \gamma_0(\tau) \middle| X_{2i} \right) 1(X_i \le w) : x \in \Gamma \right\}$$

is a Glivenko-Cantelli class by an application of Lemma Q1. Hence, we obtain the expansion (35). The null limiting distribution then follows from the expansion and Lemma Q4. 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 estimates $\tilde{f}_{i\tau}$ and $\tilde{f}_{2i\tau}$ only depend 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 functions as given. This simplifies substantially the arguments.

We first deal with $\sqrt{n}(T' - T')a(x)$, for $a \in L_2(F)$. We can assume w.l.g that a is bounded and continuous on the support of Z. We write

$$
\begin{aligned}
\sqrt{n}(\hat{T}' - T')a(x) &:= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\widehat{m}_i(x)a(Z_i) - E[m_i(x)a(Z_i)]\} \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\hat{\zeta}_i(\tau)1(X_i \le w)a(Z_i) - E[\zeta_i(\tau)1(X_i \le w)a(Z_i)]\right\} \\
&- A(x)B^{-1}(\tau)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\hat{\zeta}_i(\tau)X_{2i}\hat{f}_{i2\tau}a(Z_i) - E[\zeta_i(\tau)X_{2i}f_{i2\tau}a(Z_i)]\right\} \\
&+ \left(A(x)B^{-1}(\tau) - A_n(x)B_n^{-1}(\tau)\right)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\zeta}_i(\tau)X_{2i}\hat{f}_{i2\tau}a(Z_i) \\
&=: C_{11n}(x) - A(x)B^{-1}(\tau)C_{12n}(x) + C_{13n}(x).
\end{aligned}
$$

By Lemmas Q2 and Q7, and by standard empirical processes arguments the class

$$\{z \to q_1(z)1(x \le w)a(Z_i) : q_1 \in \mathcal{Q}, w \in [-\infty, \infty]^{d_x}\}$$

is P-Donsker. Hence, a stochastic equicontinuity argument and a uniform law of large numbers imply that, uniformly in $x \in \mathcal{T} \times \mathbb{R}^{d_x}$,

$$
\begin{aligned}
C_{11n}(x) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\zeta_i(\tau)1(X_i \le w)a(Z_i) - E[\zeta_i(\tau)1(X_i \le w)a(Z_i)]\} \\
&\quad + E[X_{2i}'a(X_i'\delta_0(\tau), X_i)1(X_i \le w)f_{i\tau}]\sqrt{n}\left(\widehat{\gamma}_n(\tau) - \gamma_0(\tau)\right) + o_P(1).
\end{aligned}
$$

Using the Taylor expansion in (32) we write

$$
\begin{aligned}
C_{12n}(x) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\hat{\zeta}_i(\tau)X_{2i}\tilde{f}_{2i\tau}a(Z_i) - E[\zeta_i(\tau)X_{2i}f_{i2\tau}a(Z_i)]\right\} \\
&+ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\zeta}_i(\tau)X_{2i}\dot{f}_{2i\tau}(\widehat{\Delta})a(Z_i) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\zeta}_i(\tau)X_{2i}\ddot{f}_{2i\tau}(\widehat{\Delta})a(Z_i) \\
&+ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\zeta}_i(\tau)X_{2i}r_{2i\tau}a(Z_i) \\
&=: C_{121n}(x) + C_{122n}(x) + C_{123n}(x) + C_{124n}(x).
\end{aligned}
$$

Applying Theorem Q1 to the class $\mathcal{W}_n = \{w_n(\cdot, \tau, \Delta) = \dot{f}_{2i\tau}(\Delta) : \tau \in \mathcal{T}, \Delta \in \mathcal{B}\}$, see Lemma Q4 for definition of $\mathcal{B}$, and a stochastic equicontinuity argument we obtain the uniform expansion, in $x \in \Gamma$,

$$
\begin{aligned}
C_{122n}(x) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(E\left[\hat{\zeta}_i(\tau)a(Z_i)\middle| X_i\right] - E\left[\zeta_i(\tau)a(Z_i)\middle| X_i\right]\right)X_{2i}\dot{f}_{2i\tau}(\widehat{\Delta}) + o_P(1) \\
&= \sqrt{n}\left(\widehat{\gamma}_n(\tau) - \gamma_0(\tau)\right)'\frac{1}{n}\sum_{i=1}^{n}f\left(X_i'\widetilde{\delta}_n(\tau)\middle| X_{2i}\right)a(X_i'\widetilde{\delta}_n(\tau), X_i)X_{2i}X_{2i}'\dot{f}_{2i\tau}(\widehat{\Delta}) + o_P(1) \\
&= O_P\left(n^{-1/2}h^{-1}\right) \\
&= o_P(1),
\end{aligned}
$$

where we have used in the last equality that $\dot{f}_{2i\tau}(\widehat{\Delta}) = O_P\left(n^{-1/2}h^{-1}\right)$. The same arguments show that $C_{123n}(x) = o_P(1)$ uniformly in $x \in \Gamma$. It is also straightforward to prove that $C_{124n}(x) = O_P\left(n^{-2}h^{-2}\right) = o_P(1)$, uniformly in $x \in \Gamma$. Hence,

$$C_{12n}(x) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\hat{\zeta}_i(\tau)X_{2i}\tilde{f}_{2i\tau}a(Z_i) - E[\zeta_i(\tau)X_{2i}f_{i2\tau}a(Z_i)]\right\} + o_P(1).$$

An application of Theorem Q1 with $\mathcal{W}_n = \{w_n(X_i,\tau) = \tilde{f}_{2i\tau} : \tau \in \mathcal{T}\}$, and Lemma Q5 yield the uniform expansion

$$\begin{aligned}
C_{12n}(x) = {} & \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\zeta_i(\tau)X_{2i}f_{i2\tau}a(Z_i) - E[\zeta_i(\tau)X_{2i}f_{i2\tau}a(Z_i)]\} \\
& + E[X_{2i}X_{2i}'f_{i\tau}f_{i2\tau}a(X_i'\delta_0(\tau),X_i)]\sqrt{n}\left(\widehat{\gamma}_n(\tau) - \gamma_0(\tau)\right) + o_P(1).
\end{aligned}$$

To deal with $C_{13n}(x)$, it can be shown using our previous arguments 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

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

This together with the uniform (in τ) convergence

$$\frac{1}{n}\sum_{i=1}^{n}\hat{\zeta}_i(\tau)X_{2i}\hat{f}_{i2\tau}a(Z_i) = \frac{1}{n}\sum_{i=1}^{n}\zeta_i(\tau)X_{2i}f_{i2\tau}a(Z_i) + o_P(1),$$

see Lemma Q5, implies the weak convergence of C_{13n}. From these uniform expansions our conditions on $\sqrt{n}(T' - T')a(x)$ can be verified in a routine fashion.

Finally, we deal with $\sqrt{n}(\hat{D}(x) - D(x))$. For simplicity, we assume w.l.g that X_1 is a scalar. Similarly as for $\sqrt{n}(T' - T')a(x)$, we write

$$\begin{aligned}
\sqrt{n}(\hat{D}(x) - D(x)) &= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{X_{1i}\hat{f}_{i\tau}\widehat{w}(X_i,x) - D(x)\right\} \\
&= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{X_{1i}\hat{f}_{i\tau}1(X_i \le w) - E[X_{1i}f_{i\tau}1(X_i \le w)]\right\} \\
&+ A(x)B^{-1}(\tau)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{X_{1i}X_{2i}\hat{f}_{i\tau}\hat{f}_{i2\tau} - E[X_{1i}X_{2i}f_{i\tau}f_{i2\tau}]\right\} \\
&- \left(A(x)B^{-1}(\tau) - A_n(x)B_n^{-1}(\tau)\right)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}X_{1i}X_{2i}\hat{f}_{i\tau}\hat{f}_{i2\tau} \\
&=: -D_{1n}(x) + A(x)B^{-1}(\tau)D_{2n}(x) - D_{3n}(x).
\end{aligned}$$

Long but simple algebra shows that D_{1n} is asymptotically, uniformly in $x \in \Gamma$, equivalent to $C_{311n} + C_{312n} + C_{313n}$, where

$$C_{311n}(x) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ X_{1i} f_{i\tau} 1(X_i \leq w) - E[X_{1i} f_{i\tau} 1(X_i \leq w)] \right\},$$

$$C_{312n}(x) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_{1i} (\tilde{f}_{i\tau} - f_{i\tau}) 1(X_i \leq w)$$

and

$$C_{313n}(x) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_{1i} \dot{f}_{2i\tau}(\widehat{\Delta}) 1(X_i \leq w).$$

The process C_{312n}, centered at its expectation, is stochastic equicontinuous, and by the independence assumption, for each x,

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

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

$$\begin{aligned} C_{312n}(x) &= \sqrt{n} E\left[X_{1i} (\tilde{f}_{i\tau} - f_{i\tau}) 1(X_i \leq w) \right] + o_P(1) \\ &= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} a_{1h}(x, \tau_j) + o_P(1), \end{aligned}$$

where

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

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

Let $g_{(\tau,\tau_j)}(u, v)$ be the density of $(X'\delta_0(\tau_j), X'\delta_0(\tau))$ conditional on $\mathcal{A}_n$. Define

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

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,

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

Using this, we conclude

$$C_{312n}(x) = \mu_3 \frac{1}{\sqrt{n}} \sum_{j=1}^{n} a_1(x, \tau_j) + o_P(1).$$

The weak convergence of C_{312n} follows easily from the last display and Assumption E1.

The analysis of C_{313n} is quite similar to that of C_{312n}. Using a stochastic equicontinuity argument, we obtain

$$\begin{aligned} C_{312n}(x) &= \sqrt{n} E[X_{1i} \dot{f}_{2i\tau}(\widehat{\Delta}) 1(X_i \le w)] + o_P(1) \\ &= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} a_{2h}(x, \tau_j) \left\{ \widehat{\gamma}_n(\tau_j) - \gamma_0(\tau_j) + \widehat{\gamma}_n(\tau) - \gamma_0(\tau) \right\} + o_P(1), \end{aligned}$$

where

$$\begin{aligned} a_{2h}(x, \tau_j) &:= E\left[r_{2(\tau,\tau_j)}(X_2'\gamma_0(\tau), X_2'\gamma_0(\tau_j), w) \frac{1}{h^2} \dot{K}\left(\frac{X_2'\gamma_0(\tau) - X_2'\gamma_0(\tau_j)}{h} \right) \middle| \mathcal{A}_n \right] \\ r_{2(\tau,\tau_j)}(u, v, w) &:= E\left[X_1 X_2' 1(X_i \le w) \middle| X_2'\gamma_0(\tau) = u, X_2'\gamma_0(\tau_j) = v \right]. \end{aligned}$$

Let $g_{2(\tau,\tau_j)}(u, v)$ be the density of $(X_2'\gamma_0(\tau), X_2'\gamma_0(\tau_j))$ conditional on $\mathcal{A}_n$. Define

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

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

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

Using this, we conclude

$$\begin{aligned} C_{313n}(x) = \mu_3 \sqrt{n} (\widehat{\gamma}_n(\tau) - \gamma_0(\tau)) \frac{1}{n} \sum_{j=1}^{n} a_2(x, \tau_j) \\ + \frac{\mu_3}{\sqrt{n}} \sum_{j=1}^{n} a_2(x, \tau_j)(\widehat{\gamma}_n(\tau_j) - \gamma_0(\tau_j)) + o_P(1). \end{aligned}$$

By a standard law of large numbers and by Lemma 3.1 in Chang (1990), C_{313n} weakly converges to

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

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 similar to that of $C_{13n}(x)$. Details are omitted. ■

References

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

[43] 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., 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.

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

[5] 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.

[6] 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.

[7] Bierens, H. J. (1982): "Consistent Model Specification Tests," *Journal of Econometrics,* 20, 1982, 105-134.

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

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

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

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

[12] 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.

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

[43] 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.

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

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

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

[18] Chen, X. and H. White (1998): "Central Limit and Functional Central Limit Theorems for Hilbert-valued Dependent Heterogeneous Arrays With Applications," *Econometric Theory,* 14, 260-284.

[19] 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.

[20] 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.

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

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

[23] Escanciano, J. C. (2006): "Goodness-of-fit Tests for Linear and Non-linear Time Series Models," *Journal of the American Statistical Association,* 101, 531-541.

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

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

[26] Escanciano, J. C., and C. Velasco (2006): "Generalized Spectral Tests for the Martingale Difference Hypothesis," *Journal of Econometrics*, 134, 151-185.

[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.

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

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

[31] Jakubowski, A. (1980): "On Limit Theorems for Sums of Dependent Hilbert Space Valued Random Variables," *Lecture Notes in Statistics, 2* 178-187.

[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 Guantiles," *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] Koul, H. L., and W. Stute (1999): "Nonparametric model checks for time series," *The Annals of Statistics,* 27, 204-236.

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

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

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

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

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

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

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

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

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

[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. (1986): "Censored Regression Quantiles," *Journal of Econometrics,* 32, 143-155

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

[50] Rosenblatt, M. (1969): "Conditional Probability Density and Regression Estimators," *Multivariate Analysis II*, Academic Press, 25-31.

[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] Stein, C. (1956): "Efficient Nonparametric Testing and Estimation," in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, CA: University of California Press.

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

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

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

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

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

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

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