UNIVERSITY OF LEICESTER

School of Business
Economics Division

Testing a parametric transformation model versus a nonparametric alternative

Arkadiusz Szydłowski, University of Leicester

Working Paper No. 17/15

Testing a parametric transformation model versus a nonparametric alternative*

Arkadiusz Szydłowski†

University of Leicester

July 4, 2017

Abstract

Despite an abundance of semiparametric estimators of the transformation model, no procedure has been proposed yet to test the hypothesis that the transformation function belongs to a finite-dimensional parametric family against a nonparametric alternative. In this paper we introduce a bootstrap test based on integrated squared distance between a nonparametric estimator and a parametric null. As a special case, our procedure can be used to test the parametric specification of the integrated baseline hazard in a semiparametric mixed proportional hazard (MPH) model. We investigate the finite sample performance of our test in a Monte Carlo study. Finally, we apply the proposed test to Kennan's strike durations data.

JEL: C12, C14, C41

Keywords: Specification testing, Transformation model, Duration model, Bootstrap, Rank estimation

*I am grateful to Joel Horowitz and Elie Tamer for their encouragement and valuable suggestions. I would also like to thank Yu Zhu and participants at WIEM 2015 and ESEM 2016 conferences for their comments. This research used the ALICE High Performance Computing Facility at the University of Leicester and the Social Sciences Computing Cluster (SSCC) at Northwestern University.

†Division of Economics, University of Leicester, University Road, Leicester LE1 7RH, UK. *E-mail address*: ams102@le.ac.uk

1 Introduction

Consider a transformation model of the form:

$$\Lambda_0(Y) = X'\beta_0 + U \tag{1}$$

where Y is a scalar dependent variable, X is a vector of q nondegenerate explanatory variables, β_0 is a vector of coefficients belonging to a compact set $\Theta_\beta \subset \mathbb{R}^q$, $\Lambda_0(\cdot)$ is an increasing function and U is an unobserved error term with cumulative distribution function F that is independent of X. For the model to be identified, the following normalizations are used: $\Lambda_0(y_0) = 0$ for some finite y_0 and $\beta_{0,1} = 1$ (where $\beta_{0,1}$ denotes the first element of β_0). Note that the model belongs to the class of single index models, therefore β_0 can be estimated $\sqrt{n}$-consistently using, for example, maximum rank correlation estimator (Han (1987)) or semiparametric least squares (Ichimura (1993)). We assume that such estimator is available throughout our analysis.

Several nonparametric estimators have been proposed for the transformation function in this model. Let $Z \equiv X'b_n$ where b_n is a consistent estimator of β_0. Horowitz (1996) uses the fact that:

$$\begin{aligned}
\Psi(y|z) \equiv& P(Y \le y|Z = z) = F(\Lambda(y) - z)\\
\Psi_y(y|z) \equiv& \frac{\partial \Psi(y|z)}{\partial y} = \frac{d\Lambda(y)}{dy} f(\Lambda(y) - z)\\
\Psi_z(y|z) \equiv& \frac{\partial \Psi(y|z)}{\partial z} = -f(\Lambda(y) - z)
\end{aligned}$$

and suggests to estimate $\Lambda(\cdot)$ by:

$$\Lambda_n(y) = -\int_{y_0}^{y} \int_{S_w} w(z) \frac{\Psi_{ny}(v|z)}{\Psi_{nz}(v|z)} dz dv \qquad \text{(HJ)}$$

where Ψ_{ny}, Ψ_{nz} are kernel-based estimators of Ψ_y, Ψ_z and $w(\cdot)$ is a non-negative differentiable weight function, which integrates to one, with support $S_w \subset \mathbb{R}$ such that $z \in S_w \Rightarrow \Psi_z(y|z) > 0$. Integration over z using the weight function has a smoothing effect on the kernel estimators and allows obtaining $n^{-1/2}$ rate of convergence. This estimator has been extended to the case of censored Y's by Gørgens & Horowitz (1999).

Chen (2002) suggests a rank-based estimator. Define $d_{iy} = 1\{Y_i \ge y\}$ and $d_{jy_0} = 1\{Y_j \ge y_0\}$.

Using the normalization $\Lambda_0(y_0) = 0$, we have:

$$E[(d_{iy} - d_{jy_0})|Z_i, Z_j] = F[-Z_j] - F[\Lambda_0(y) - Z_i] \geq 0 \quad \Leftrightarrow \quad Z_i - Z_j \geq \Lambda_0(y),$$

which suggests estimating $\Lambda_0(y)$ by a version of the maximum rank correlation estimator:

$$\Lambda_n(y) = \arg\max_{\Lambda} \frac{1}{n(n-1)} \sum_{i \neq j} (d_{iy} - d_{jy_0}) \mathbb{1}\{Z_i - Z_j \geq \Lambda\}, \qquad \text{(CS)}$$

Chen (2002) proves consistency and derives asymptotic law for this estimator using much weaker conditions than Horowitz (1996). In particular, Λ_0 may be discontinuous and the pdf of U is required to have only two derivatives, compared to the requirement of at least nine derivatives in the latter paper. The estimator can be easily modified to accommodate random censoring.

Another estimator was proposed by Ye & Duan (1997). They impose a different normalization than the previous two approaches. Instead of fixing Λ_0 at a point they assume that median of U is equal to zero. Let $Me(z)$ denote the median of Y conditional on $Z = z$. First, they use the fact that:

$$\Lambda(Me(z)) = z;$$
$$\Psi(Me(Z + \Delta)|z) = F(\Delta),$$

where Ψ and Z are defined as above, to estimate F and then they use $\hat{F}$ to estimate Λ_0.[1] Yet another estimator, using a similar idea was suggested by Klein & Sherman (2002), who focus on the estimation of threshold points in an ordered response model. Their approach can also be used to estimate the values of the transformation function at a finite number of points. Both estimators can be extended to deal with censored values of Y.

Despite such an abundance of semiparametric estimation techniques, (to the best of our knowledge) so far there has been no practically appealing procedure that would allow testing parametric specification of the transformation model against an unrestricted one (cf. Horowitz (2009)). This

[1] Note that Λ_0^{-1} could be estimated in this model simply by running a nonparametric median regression of Y on Z since $M(Y|Z) = \Lambda_0^{-1}(Z)$. Thus, we could invert this estimator to obtain Λ_n (this may require imposing monotonicity on Λ_n^{-1}) and then use it to estimate F. However, the resulting estimator of F will depend on the first-stage estimator Λ_n. An advantage of the approach in Ye & Duan (1997) is that the estimator of F does not depend on Λ_n.

paper aims at filling this gap. The main objective is to provide a test that would distinguish between various parametric specifications of integrated baseline hazard function (e.g. Weibull hazard) and a fully nonparametric one. However, the procedures developed below can be used in specification testing in other context, e.g. testing the log-linear specification in wage regressions, testing the form of the marginal utility (profit) function in hedonic models (see e.g. Ekeland et al. (2004)).

In a related article, Fernandes & Grammig (2005) propose a specification test for the hazard function in the autoregressive conditional duration model (ACD) used in finance. Although their model does not fall in the i.i.d. setup we consider, it should be possible to extend their approach to our case. Nevertheless, their test is based on nonparametric estimates of the conditional density of the duration which requires choosing smoothing parameters. Since there is no guidance on how to choose these parameters in a finite sample and results of the test may be sensitive to this choice, this approach may not be appealing to practitioners. Instead, our bootstrap test is free of tuning parameters and therefore much easier to use in practice.

We propose a Cramer-von Mises type test for distinguishing between parametric and nonparametric transformations. The test uses the nonparametric estimator of the transformation function developed by Chen (2002) and compares it to the parametric specification using the L^2 norm. We chose to build our test on this estimator for three reasons. Firstly, CS estimator has a convenient linear asymptotic representation whereas no such representation is available for Klein & Sherman (2002) and Ye & Duan (1997), which makes the analysis of the test based on the latter estimators more complicated. Secondly, CS is much easier to compute than HJ (and similarly Gørgens & Horowitz (1999)) since using the latter would involve multiple computationally intensive numerical integrations. Finally, as shown in Chen (2002) CS generally performs better than the other estimators in terms of root mean-square error, especially in the tails of the data distribution.

In our model F is treated nonparametrically. As an alternative to our approach, one can assume a parametric distribution for F. If the data on Y is recorded on a finite grid, e.g. Y is unemployment duration and is recorded in weeks, then one can estimate Λ_0 at the points in the grid by maximum likelihood both with and without imposing parametric restriction on Λ_0 and run a likelihood ratio test to verify if the parametric model is valid.[2] The disadvantage of this

[2]See Meyer (1990) for estimation of a MPH model with nonparametric hazard, parametric distribution of U and discrete observations on Y.

approach is that misspecification of the parametric form of F may lead to invalid inference about the specification of Λ_0, whereas our approach will be robust to misspecifying F. Finally, our test can also be applied if F is restricted to a parametric class provided that the nonparametric estimator of Λ_0 satisfies the assumptions below.

The article is organized as follows. Section 2 discusses specification testing in the general transformation model given in (1). In this model the transformation function is identified only up to scale so the inference boils down to checking if the shape of Λ_0 is consistent with the parametric assumption.

Section 2.1 considers a special case of a mixed proportional hazard model. Thanks to additional structure, in this model both the shape and the scale of the transformation function are identified. We show that in order to test if the parametric specification of the integrated baseline hazard is correct it is enough to use the estimator up to scale. This has two advantages relative to simply comparing the estimated parametric and non-parametric integrated baseline hazards. Firstly, the scale of the integrated baseline hazard, whether in parametric or nonparametric model, can be estimated only at a rate slower than the standard $n^{-1/2}$ rate so by using estimates up to scale we still obtain a test that has power against alternatives that are $O(n^{-1/2})$ apart from the null hypothesis. Second, the available estimators of the scale (see Honoré (1990), Horowitz (1999)) are difficult to use in practice. For example, consider a mixed proportional hazard model with nonparametric Λ and F. Horowitz (1999) shows that the scale of Λ in this model can be estimated by:

$$\sigma_n = \frac{\sigma_n(t_{n1}) - n^{-\eta_1(1-\eta_2)}\sigma_n(t_{n2})}{1 - n^{-\eta_1(1-\eta_2)}}$$

where

$$\sigma_n(y) = -\frac{\int \Psi_{nz}(y|z)p_n(z)^2 dz}{\int \Psi_n(y|z)p_n(z)^2 dz}$$

and Ψ_{nz}, Ψ_n are estimators of Ψ_z, Ψ (defined above), p_n is a kernel estimator of the density of Z and $t_{n1} \to 0, t_{n2} \to 0$ at rates that depend on the tuning parameters η_1 and η_2. Thus, in order to implement this estimator the researcher needs to pick not only bandwidths for the estimation of

Ψ and its derivatives but also the tuning constants η_1 and η_2, which is troublesome given lack of prescriptions for how to pick these constants in a finite sample.

Our test statistic converges to a functional of a Gaussian process and we suggest using bootstrap to obtain the critical values. We show that bootstrap consistently estimates the asymptotic distribution of our statistic. As a by-product of our analysis we prove that nonparametric bootstrap can be used to obtain (pointwise) standard errors for the CS estimator. This is an important result by itself since previous approaches based on numerical derivatives or kernel smoothing proved to be quite unstable and hard to implement in practice. In Section 3 we investigate the finite sample performance of our test using a Monte Carlo study. Section 4 provides an application to Kennan's strike duration data.

2 General transformation model

We want to test:

$$H_0: \quad \Lambda_0(\cdot) \in \{\Lambda(\cdot, \gamma); \gamma \in \Theta_\gamma\} \quad \text{over} \quad [y_1, y_2]$$

where Θ_γ is an open subset of a d-dimensional Euclidean space. One needs to restrict oneself to a compact interval $[y_1, y_2]$ because $\Lambda_0(y)$ may not be bounded on the whole real line.[3] From now on we will refer to the model with parametric $\Lambda(\cdot, \gamma)$ as a 'parametric model' in contrast to a 'nonparametric model' in which Λ_0 is not restricted to lie in a parametric class, although both models leave the distribution of U unrestricted.

A natural way to construct a test is to take the L^2 distance between one of the estimators $\Lambda_n(\cdot)$ and the parametric estimator, e.g. the estimator of Box-Cox regression model proposed by Foster et al. (2001). However, as mentioned in the Introduction, the transformation function is only identified up to scale and location normalizations. We have two cases. Firstly, the same normalization may be imposed on both nonparametric and parametric model, i.e. $\Lambda_0(y_0) = \Lambda(y_0, \gamma)$ for some $y_0 \in \mathbb{R}$, and $\beta_1 = 1$. Secondly, often a parametric model for the transformation imposes a scale normalization by itself so we cannot restrict $\beta_1 = 1$ (for example, if the parametric specification

[3] One could expand the support of Λ with the sample size and as a result obtain a test over the whole support $\mathbb{R}$. We leave this extension for further research.

has a log-linear form: $\log Y = X'\beta + U$). Therefore, we have to normalize the nonparametric estimator so that the two transformation functions are comparable. This can be done by multiplying the nonparametric estimator by the estimator of the scale from the parametric model.[4]

Let $\hat{\beta}$ denote an estimator of the coefficient vector β in the parametric model and let $\hat{\beta}_1$ be its first element. Note that $\hat{\beta}_1 \Lambda_n(y)$ is equal to the estimator of the transformation function when the normalization $\beta_{0,1} = \hat{\beta}_1$ is imposed instead of $\beta_{0,1} = 1$. Thus, our test statistic is given by:

$$T_n = n \int_{y_1}^{y_2} \left[(a_n \Lambda_n(y) - \Lambda(y, \hat{\gamma})) w(y) \right]^2 dy. \tag{2}$$

where $\hat{\gamma}$ is an estimator of γ, $a_n = D + (1 - D)\hat{\beta}_1$ and

$$D = \begin{cases} 1 & \text{if both transformations are normalized at the same point} \\ 0 & \text{otherwise} \end{cases}$$

The weight function $w(y)$ may be used to redirect the power of the test over y. For example, an application may dictate that some region of y's is of particular interest.

In principle, instead of using a Cramer-von-Mises type test, a Kolmogorov-Smirnov type test can be used. However, using the L^2 norm will be more convenient computationally than using a *sup* norm. Our preferred estimator Λ_n is non-smooth and $h(y) = (\Lambda_n(y) - \Lambda(y, \hat{\gamma}))^2$ may have multiple local maxima with respect to y. Thus, calculating the sup statistic would require using global optimization methods for discontinuous problems, which are usually very slow. It is much easier to integrate over the differences $(\Lambda_n(y) - \Lambda(y, \hat{\gamma}))^2$. If the integrand $h(y)$ is non-smooth, the numerical integration procedures may have trouble approximating the integral precisely. Still the Monte Carlo simulations reported in Section 3 show that they work pretty well in practice.

Frequently, especially in the context of duration models, the observations on Y_i are right-censored. Let C_i denote a random censoring threshold with cumulative distribution function G_0 and survival function $\bar{G}_0$, let $\tilde{Y}_i$ denote a latent (not censored) value of the dependent variable generated from (1) and let Y_i be a censored observation on $\tilde{Y}_i$, i.e. $Y_i = \min\{\tilde{Y}_i, C_i\}$. Additionally, define the censoring indicator $\delta_i = \mathbb{1}\{\tilde{Y}_i \leq C_i\}$.

[4]Throughout the article we will use hats to denote the estimators obtained using the parametric model and subscript n to denote estimators corresponding to the nonparametric model.

From now on, we will focus on the case in which Y's are censored. The case without censoring can be seen as a special case with $C_i = \infty$ for all i (i.e. $\bar{G}_0(y) = 1$ for all y in $[y_1, y_2]$) so all the arguments below will apply to this special case.

Define the Euclidean class of functions as in Pakes & Pollard (1989) and let $L^2(\mathcal{Y})$ denote a space of square integrable functions on $\mathcal{Y}$. We make the following assumptions:

Assumption 1. (DGP) *$\{X_i, Y_i, \delta_i : i = 1, \ldots, n\}$ is a random sample, U is independent of X, C is independent of (X, U) and $\bar{G}_0$ is bounded away from zero on $[y_1, y_2]$.*

Assumption 2. (Asymptotic linearity)

(a) There is a function $J : [y_1, y_2] \times \mathbb{R}^q \times [y_1, y_2] \times \Theta_\beta \to \mathbb{R}$ such that $E[J(Y, X; y, \beta_0)] = 0$, $E[J(Y, X; y, \beta_0) J(Y, X; y', \beta_0)]$ is finite for every $y, y' \in [y_1, y_2]$, $J(Y_i, X_i; \cdot, \beta_0) \in L^2([y_1, y_2])$ and, as $n \to \infty$:

$$\sqrt{n}(\Lambda_n(y) - \Lambda_0(y)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} J(Y_i, X_i; y, \beta_0) + o_p(1)$$

uniformly over $y \in [y_1, y_2]$. Moreover, the class of functions $\mathcal{J} = \{J(\cdot, \cdot; y, \beta_0), y \in [y_1, y_2]\}$ is Euclidean.

(b) Let γ be a probability limit of $\hat{\gamma}$. There exists a vector-valued function $\Omega_\gamma(Y_i, X_i; \gamma, \beta)$ with mean zero and finite covariance matrix such that, as $n \to \infty$:

$$\sqrt{n}(\hat{\gamma} - \gamma) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Omega_\gamma(Y_i, X_i, \delta_i; \gamma, \beta) + o_p(1).$$

(c) $\Lambda(y, \gamma)$ is twice differentiable in γ and the derivatives are bounded uniformly over $y \in [y_1, y_2]$.

(d) Let β_1 be a probability limit of $\hat{\beta}_1$. There exists a function $\Omega_1(Y_i, X_i, \delta_i; \gamma, \beta)$ with mean zero and finite variance such that, as $n \to \infty$:

$$\sqrt{n}(\hat{\beta}_1 - \beta_1) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Omega_1(Y_i, X_i, \delta_i; \gamma, \beta) + o_p(1).$$

Assumption 3. (Weight function) *The weight function $w(y)$ satisfies:*

$$\int_{y_1}^{y_2} w(y)^2 dy = 1.$$

Assumption 2(a) is satisfied by the CS and by the HJ estimator.[5] This assumption implies that $\sqrt{n}(\Lambda_n(y) - \Lambda_0(y))$ converges to a mean zero Gaussian process. Assumptions 2(b),(c) are not relevant if $\Lambda(y, \gamma)$ does not depend on γ as in our leading example of testing a log-linear model versus a nonparametric alternative, i.e. $\Lambda(y, \gamma) = \log(y)$. Assumption 2(b) is satisfied by GMM estimators and estimator proposed by Foster et al. (2001). Assumption 2(c) is satisfied by a Box-Cox transformation (with $y_1 > 0$) and most hedonic pricing models if the utility (profit) function is sufficiently smooth (e.g. Cobb-Douglas). The asymptotic linear representation in Assumption 2(d) is clearly available for the OLS estimator in the loglinear model but also for the estimator developed by Foster et al. (2001) for the Box-Cox model.

Example 1. ***(log-linear model)*** *We test if the wage regression has a log-linear form. For simplicity assume that there is only one regressor and no censoring. We estimate the model by ordinary least squares. In this case we have $\Lambda(y, \gamma) = \log(y)$, $\frac{\partial \Lambda(y,\gamma)}{\partial \gamma} = 0$ and $\Omega_1(Y_i, X_i; \gamma, \beta) = (X_i - \bar{X})(\log Y_i - \beta X_i)/Var(X_i)$.*

Define:

$$B_n(y) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [(D + (1 - D)\beta_1) J(Y_i, X_i; y, \beta_0) - \frac{\partial \Lambda(y, \gamma)'}{\partial \gamma} \Omega_\gamma(Y_i, X_i; \gamma, \beta)$$
$$+ (1 - D)\Lambda(y, \gamma)\Omega_1(Y_i, X_i; \gamma, \beta)] w(y). \quad (3)$$

The following theorem establishes the asymptotic approximation to the distribution of the test statistics.

Theorem 1. *Under H_0 and Assumptions 1-3:*

$$T_n \to^d \int_{y_1}^{y_2} \mathcal{B}^2(y) dy \quad (4)$$

[5] Klein & Sherman (2002) only show point-wise convergence of their estimator to a normal variable. They do not provide a uniform linear representation as in Assumption 2(a). Also the estimator developed by Ye & Duan (1997) does not have a linear representation.

where $\mathcal{B}$ is a mean zero Gaussian process with covariance function:

$$R(y, y') = E[B_n(y)B_n(y')].$$

Alternatively, we can write:

$$T_n \to^d \sum_{j=1}^{\infty} \omega_j \chi_{j1}^2, \tag{5}$$

where χ_{j1}^2's are independent chi-square random variables with one degree of freedom and ω_j's are eigenvalues of the linear integral operator:

$$(\mathcal{R}g)(y) = \int_{y_1}^{y_2} R(y, z)g(z)dz; \qquad g(\cdot) \in L^2([y_1, y_2]). \tag{6}$$

We can obtain the critical value by simulating the process $\mathcal{B}$ and calculating the integral in (4).[6] However, this would require estimating the covariance function R which both for CS and HJ estimators involves kernel smoothing. Since there are no procedures to choose a bandwidth for these estimators in the finite sample and, as evidenced by our simulation studies (available upon request), the results of the test are very sensitive to this choice, we do not pursue this approach. Instead, we suggest using bootstrap critical value.

2.1 MPH duration model

Before we turn to the bootstrap procedure, we briefly discuss how our test can be used to test the parametric specification of the (integrated) baseline hazard in duration models.

A duration model can be seen as a special case of the transformation model. We consider the single-spell mixed proportional hazard (MPH) model:

$$\alpha \log \tilde{\Lambda}(Y) = X'\beta + V - \xi \tag{7}$$

[6] As an alternative, one can use the characterization in (5) and employ the simulation procedure in Horowitz (2006) and Blundell & Horowitz (2007). This would involve truncating the sum in (5) and estimating the remaining eigenvalues ω_j. This is straightforward in the setting analyzed by Horowitz (2006) and Blundell & Horowitz (2007) because the Fourier representation of the covariance kernel can be calculated analytically without numerical integration. This is not the case here since the covariance function includes an at least three dimensional non-separable function $J(\cdot, \cdot; \cdot, \beta_0)$, which entails the need to perform a triple numerical integration in order to obtain the Fourier coefficients. This makes this method unattractive in our setting.

where $\tilde{\Lambda}(Y)^{\alpha}$ is the integrated baseline hazard, ξ has the standard Gumbel distribution and (ξ, V, X) are mutually independent. For simplicity, there is no censoring. We intentionally factored out the scale of the log of integrated hazard, α, to facilitate discussion below. The difference between this model and the general transformation model discussed before is that here β and α are separately identified and we do not need the normalization $\beta_1 = 1$.

Now observe that, if $\tilde{\Lambda}$ is known, equation (7) pins down the scale α because the scale of ξ is fixed (and ξ is independent of X and V). In other words, if there are two MPH models:

$$\alpha^{(1)} \log \tilde{\Lambda}^{(1)}(Y) = X\beta^{(1)} + V^{(1)} - \xi$$
$$\alpha^{(2)} \log \tilde{\Lambda}^{(2)}(Y) = X\beta^{(2)} + V^{(2)} - \xi$$

with $\tilde{\Lambda}^{(1)} = \tilde{\Lambda}^{(2)}$, then they can generate the same population distribution of Y given X only if $\alpha^{(1)} = \alpha^{(2)}$ (excluding a knife-edge case when $V^{(1)}/\alpha^{(1)}$ and $V^{(2)}/\alpha^{(2)}$ have the same distribution as $-\xi$). As a result, if we want to test if the integrated baseline hazard $\alpha \log \tilde{\Lambda}(\cdot)$ belongs to some parametric class, it is enough to test that the estimate up to scale, $\log \tilde{\Lambda}(\cdot)$, belongs to a conjectured parametric family.

Therefore, the following procedure can be used:

1. Estimate the transformation model in (1) imposing the necessary normalizations.

2. Estimate the null parametric transformation $\Lambda(y, \gamma) = \log \tilde{\Lambda}(y, \gamma)$ with (this would correspond to $D = 0$ above) or without ($D = 1$) imposing the normalization $\beta_1 = 1$, for example by using GMM (see Horowitz (2009), Ch. 6.1) or Foster et al. (2001).

3. Run our bootstrap test (see next section for details). If the test statistic is greater than the critical value, conclude that the integrated baseline hazard is misspecified.

This is convenient since the estimators of α do not converge at the $n^{-1/2}$ rate either in parametric or nonparametric model. For the Weibull MPH model Honoré (1990) shows that under the assumption $E[e^{-V}] < \infty$ his estimator converges at a rate that can be made arbitrarily close to $n^{-1/3}$ and establishes its asymptotic normality. The estimator proposed by Ishwaran (1996*b*) achieves this rate under the same assumption. Moreover, Horowitz's estimator for the nonparametric model (Horowitz (1999)) converges at, at most, $n^{-2/5}$ rate under the assumption that $E[e^{-3V}]$ is finite.

As shown by Ishwaran (1996a), the highest rate at which the estimator α converges to the true value under the assumption $E[e^{-V}] < \infty$ is $n^{-1/3}$, and $n^{-2/5}$ under the assumption $E[e^{-3V}] < \infty$ (cf. Horowitz (2009)). On the other hand, the estimators of $\log \tilde{\Lambda}(\cdot)$ converge at the usual $n^{-1/2}$ rate. Thus, by avoiding the need to estimate the scale α in our test we sustain this fast rate of convergence.

Example 2. (Weibull MPH model) *We test if the integrated baseline hazard has a Weibull shape, i.e. if* $\log \tilde{\Lambda}(y) = \log(y)$. *Now the MPH model becomes:*

$$\log(Y) = X' \frac{\beta}{\alpha} + \frac{V - \xi}{\alpha}$$

and $\tilde{\beta}_1 = \beta_1/\alpha$ *can be estimated* $\sqrt{n}$*-consistently by OLS. We can use* $\hat{\tilde{\beta}}_1$ *as the scaling factor. Since the transformation function does not depend on unknown parameters, the second term in the expression for* B_n *(equation* (3)*) vanishes.*

2.2 Bootstrap critical value

The theory developed so far applies both to HJ and CS estimator. Nevertheless, CS is preferred from the computational point of view. Using HJ to compute the test statistic involves double numerical integration to compute Λ_n on top of the integration involved in computing the L^2 distance. Doing that repetitively to obtain the bootstrap critical value would entail a very large computational cost. It is much easier to bootstrap the CS estimator. Thus, from now on we will assume that Λ_n is the CS estimator. In the case of censored observations this estimator is defined as:

$$\Lambda_n(y) = \arg\max_{\Lambda} \frac{1}{n(n-1)} \sum_{i \neq j} \left(\frac{d_{iy}}{\bar{G}_n(y)} - \frac{d_{jy_0}}{\bar{G}_n(y_0)} \right) \mathbb{1}\{Z_i - Z_j \geq \Lambda\} \tag{8}$$

where $\bar{G}_n(y)$ is the Kaplan-Meier estimator of the survival function of the censoring threshold C and d_{iy}, d_{jy_0}, Z_i are the same as in the definition of the original CS estimator.

Let $w_1 = (x^1, y^1)$ and $w_2 = (x^2, y^2)$. Define:

$$r(w_1, w_2, y, G, \Lambda, b) = \left(\frac{\mathbb{1}\{y^1 \geq y\}}{\bar{G}(y)} - \frac{\mathbb{1}\{y^2 \geq y_0\}}{\bar{G}(y_0)} \right) \left(\mathbb{1}\{x^1 b - x^2 b \geq \Lambda\} - \mathbb{1}\{x^1 b - x^2 b \geq \Lambda_0\} \right)$$

and:

$$\tau(w, y, \Lambda) = E[r(w, W, y, G_0, \Lambda, \beta_0) + r(W, w, y, G_0, \Lambda, \beta_0)]$$

for $W = (X, Y)$. Let:

$$V(y) = E\left[-\frac{\partial^2 \tau(W, y, \Lambda)}{\partial \Lambda^2}\bigg|_{\Lambda=\Lambda_0} \right].$$

Finally, let X_1 be the first component of X and X_{-1} denote the remaining $(q-1)$ components.

Our bootstrap procedure will be valid under assumptions similar to those introduced in Chen (2002) and Jochmans (2012):

Assumption 4. (Chen (2002))

(a) The normalization $\beta_1 = 1$ is imposed on the nonparametric estimator Λ_n.

(b) The distribution of X_1 conditional on $X_{-1} = x_{-1}$ is absolutely continuous with respect to the Lesbegue measure.

(c) The support of X is not contained in any proper linear subspace of $\mathbb{R}^q$.

(d) $\Lambda_0(\cdot)$ is strictly increasing, $\Lambda_0(y_0) = 0$, $[\Lambda_0(y_1 - \varepsilon), \Lambda_0(y_2 + \varepsilon)] \subset \Theta_\Lambda$ for a small positive number ε, where Θ_Λ is a compact interval.

(e) The conditional density of X_1 given $X_{-1} = x_{-1}$ and the density of U are bounded and twice continuously differentiable, the derivatives are uniformly bounded and X_{-1} has finite third-order moments.

(f) $V(y)$ is positive for each $y \in [y_1, y_2]$ and uniformly bounded away from zero.

(g) The first step estimator of β_0 from the nonparametric model, b_n, has the following asymptotic representation:[7]

$$\sqrt{n}(b_n - \beta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Omega^{NP}(Y_i, X_i, \delta_i; \beta_0) + o_p(1).$$

[7] Recall that the estimator obtained from the model with parametric Λ is denoted by $\hat{\beta}$. We can have $\hat{\beta}_1 \neq 1$ whereas $b_{n1} = 1$ by assumption.

where Ω^{NP} is a mean zero vector valued function with finite variance-covariance matrix.

We will employ the following bootstrap procedure to obtain a critical value for our test:

1. Draw a random sample $\{(Y_i^*, X_i^*, \delta_i^*) : i = 1, \ldots, n\}$ with replacement from $\{(Y_i, X_i, \delta_i) : i = 1, \ldots, n\}$ or use a parametric bootstrap:
   - Estimate $(\hat{\beta}, \hat{\gamma})$ using $\{(Y_i, X_i, \delta_i) : i = 1, \ldots, n\}$.
   - Generate $\hat{U}_i = \Lambda(Y_i, \hat{\gamma}) - X_i'\hat{\beta}$.
   - Draw a random sample $\{U_i^* : i = 1, \ldots, n\}$ with replacement from $\{\hat{U}_i : i = 1, \ldots, n\}$ and calculate $Y_i^* = \Lambda^{-1}(X_i'\hat{\beta} + U_i^*, \hat{\gamma})$.
2. Using the bootstrap sample calculate $(\hat{\beta}_1, \hat{\gamma})$ from the parametric model and (Λ_n, b_n) from the nonparametric model. Let the resulting estimates be denoted by $(\hat{\beta}_1^*, \hat{\gamma}^*)$ and (Λ_n^*, b_n^*).
3. Calculate the bootstrap statistic:
$$T_n^* = n \int_{y_1}^{y_2} \left[(a_n^* \Lambda_n^*(y) - a_n \Lambda_n(y) - (\Lambda(y, \hat{\gamma}^*) - \Lambda(y, \hat{\gamma})))w(y)\right]^2 dy.$$
   if nonparametric bootstrap has been used, or:
$$T_n^* = n \int_{y_1}^{y_2} \left[(a_n^* \Lambda_n^*(y) - \Lambda(y, \hat{\gamma}^*))w(y)\right]^2 dy.$$
   for parametric bootstrap, where $a_n^* = D + (1 - D)\hat{\beta}_1^*$.
4. Obtain the empirical distribution of T_n^* by repeating steps 1-3 many times. Calculate the $1 - \kappa$ quantile of this empirical distribution. Denote it by c_κ^*.

If data is not censored, then we recommend to use the parametric bootstrap as it usually leads to more precise results. On the other hand, applying parametric bootstrap is complicated with censored data so we prefer nonparametric bootstrap in this case. Finally, note that the statistic corresponding to parametric bootstrap does not require recentering as the parametric bootstrap imposes the null hypothesis contrary to nonparametric resampling.

On top of the assumptions above we will need an asymptotic linear approximation in the bootstrap sample:

Assumption 5. (Bootstrap asymptotic linearity) *We have:*

$$E\left|\hat{\gamma}^* - \gamma - \frac{1}{n}\sum_{i=1}^{n}\Omega_\gamma(Y_i^*, X_i^*, \delta_i^*; \gamma, \beta)\right| = o(n^{-1/2}) \tag{9}$$

$$E\left|\hat{\beta}_1^* - \beta_1 - \frac{1}{n}\sum_{i=1}^{n}\Omega_1(Y_i^*, X_i^*, \delta_i^*; \gamma, \beta)\right| = o(n^{-1/2}) \tag{10}$$

$$E\left|b_n^* - \beta_0 - \frac{1}{n}\sum_{i=1}^{n}\Omega^{NP}(Y_i^*, X_i^*, \delta_i^*; \beta_0)\right| = o(n^{-1/2}) \tag{11}$$

where $\Omega_\gamma, \Omega_1, \Omega^{NP}$ have mean zero and finite variance-covariance matrix.

In the leading case when the parametric model in the null hypothesis does not depend on any free parameters (e.g. testing the Weibull model in duration analysis), condition (9) is redundant. In the next section, using arguments in Subbotin (2007), we show that this condition is satisfied for the estimator in Foster et al. (2001). Condition (10) will be satisfied for the OLS estimator. An asymptotic bootstrap linear representation for the rank estimators b_n introduced in Han (1987), Cavanagh & Sherman (1998) and Abrevaya (2003) follows from Subbotin (2007).

The following theorem states that the bootstrap critical value gives the correct approximation to the asymptotic critical value:

Theorem 2. *Under H_0 and Assumptions 1, 2*(b)-(d)*,3-5:*

$$\lim_{n\to\infty} P(T_n \le c_\kappa^*) = 1 - \kappa.$$

The proof of this theorem relies on the results in Subbotin (2007) who proves bootstrap validity for rank estimators. A slight complication in the proof compared to his work comes from the fact that the rank objective function in (8) contains estimators b_n and $\bar{G}_n$, which will contribute to the asymptotic distribution of Λ_n and Λ_n^*. The argument leading to Theorem 2 implies also a following useful corollary:

Corollary 1. *Nonparametric bootstrap approximates consistently the asymptotic distribution of the CS estimator.*

This result is important because it provides an operational method for obtaining standard errors for the CS estimator. Previous approaches based on numerical derivatives or kernel smoothing relied on arbitrary choices of the approximation step or bandwidth with the results being very sensitive to inappropriate choices of these tuning parameters.

2.3 Bootstrap asymptotic linear approximation for semiparametric Box-Cox model

We will verify that Assumption 5 holds for the estimators of γ and β in the Box-Cox transformation model proposed by Foster et al. (2001). The Box-Cox transformation is given by:

$$\Lambda(y,\gamma) = \begin{cases} \frac{y^\gamma - 1}{\gamma} & \text{if } \gamma \neq 0 \\ \log y & \text{otherwise} \end{cases}$$

Foster et al. (2001) suggest to estimate (γ_0, β_0) by minimizing: [8]

$$S_n(\gamma,\beta) = \int_0^\infty \frac{1}{n(n-1)(n-2)} \sum_{i,\,j,\,k \text{ distinct}} \left(\mathbb{1}\{Y_i \leq y\} - \mathbb{1}\{\Lambda(Y_j,\gamma) - X_j'\beta \leq \Lambda(y,\gamma) - X_i'\beta\}\right)$$
$$\times \left(\mathbb{1}\{Y_i \leq y\} - \mathbb{1}\{\Lambda(Y_k,\gamma) - X_k'\beta \leq \Lambda(y,\gamma) - X_i'\beta\}\right) d\Psi(y)$$

where $\Psi(y)$ is a differentiable, strictly increasing, deterministic and bounded weight function, subject to the constraint:

$$\frac{1}{n}\sum_{i=1}^n X_i \left(\Lambda(Y_i,\gamma) - X_i'\beta\right) = 0$$

This problem is equivalent to minimizing:

$$L_n(\theta) = S_n(\gamma,\beta) + \mu' \frac{1}{n}\sum_{i=1}^n X_i \left(\Lambda(Y_i,\gamma) - X_i'\beta\right)$$

over $\theta = (\gamma, \beta, \mu) \in \Theta$ where μ is the Lagrange multiplier. Let θ^* be the corresponding estimators calculated on the bootstrap sample.

[8] In fact, Foster et al. (2001) state S_n in a form of V-statistic. However, throughout their proofs they use the U-statistic formulation given here. It follows from Lemma 5.7.3 in Serfling (1980) (p.206) that these two formulations are asymptotically equivalent.

Let $w_l = (x^l, y^l), l = 1, 2, 3$. Define:

$$h_{\theta,y}^{BC}(w_1, w_2, w_3) = \left(\mathbb{1}\{y^1 \leq y\} - \mathbb{1}\{\Lambda(y^2,\gamma) - x^{2\prime}\beta \leq \Lambda(y,\gamma) - x^{1\prime}\beta\}\right)$$
$$\times \left(\mathbb{1}\{y^1 \leq y\} - \mathbb{1}\{\Lambda(y^3,\gamma) - x^{3\prime}\beta \leq \Lambda(y,\gamma) - x^{1\prime}\beta\}\right)$$

and:

$$\tau^{BC}(w, y, \theta) = E\left[h_{\theta,y}^{BC}(w, W_1, W_2) + h_{\theta,y}^{BC}(W_1, w, W_2) + h_{\theta,y}^{BC}(W_1, W_2, w)\right]$$

where the expectation is taken with respect to $W_1 = (X_1, Y_1)$ and $W_2 = (X_2, Y_2)$. It will be convenient to define $R(W_i, \theta) = \mu' X_i \left(\Lambda(Y_i, \gamma) - X_i'\beta\right)$. Now:

$$V_{BC} = E\left[\int_0^\infty \partial^2 \tau_{BC}(W, y, \theta_0) d\Psi(y) - \partial^2 R(W, \theta_0)\right].$$

with $\partial^2 \tau_{BC}(W, y, \theta)$ and $\partial^2 R(W, \theta)$ denoting the matrices of second derivatives of $\tau_{BC}(w, y, \theta)$ and $R(W, \theta)$ with respect to θ.

Theorem 3. *Let Assumptions 4*(b),(c),(e) *hold. Furthermore, assume:*

(a) $\Psi(y)$ *is supported on a compact interval* $\mathcal{Y} \subset (0, \infty)$, $\Theta = \Theta_\gamma \times \Theta_\beta \times \Theta_\mu$ *is compact,*

(b) $E\left[\sup_{\gamma \in \Theta_\gamma} \left|\frac{Y^\gamma \log Y - \Lambda(Y,\gamma)}{\gamma}\right|\right]^2 < \infty$,

(c) the elements of the matrix $\partial^2 R(W, \theta_0)$ *have finite variance,*

(d) V_{BC} *is non-singular,*

then Assumption 5 is satisfied for the estimators of (γ, β_1) *introduced in Foster et al. (2001).*

The proof follows lines similar to the proof of Theorem 2 and is given in the Appendix. The requirement that the support of the weight function is compact and does not contain zero is of technical nature and implies that the derivatives of the Box-Cox transformation are bounded. In practice, if the weight function has full support on $[0, \infty]$, it can always be truncated above and below such that the value of the objective function S_n is not affected. Similarly, Assumption (b) ensures that the derivatives needed for a Taylor expansion have bounded moments. Further, although $\partial^2 \tau_{BC}(W, y, \theta)$ is singular for every y, $E[\partial^2 R(W, \theta)]$ is non-singular in most of the cases, which implies invertibility of V_{BC}. For example, when X is one-dimensional and $\gamma \neq 0$:

$$E[\partial^2 R(W,\theta)] = \begin{bmatrix} \frac{\mu}{\gamma^2} E[2(\Lambda(Y,\gamma) - Y^\gamma \log Y) + Y^\gamma \log^2 Y] & 0 & \frac{1}{\gamma} E[X(Y^\gamma \log Y - \Lambda(Y,\gamma))] \\ 0 & 0 & -E[X^2] \\ \frac{1}{\gamma} E[X(Y^\gamma \log Y - \Lambda(Y,\gamma))] & -E[X^2] & 0 \end{bmatrix}$$

2.4 Consistency and behaviour under local alternatives

We conclude this section with an analysis of power and local behaviour of our bootstrap test. Assume that the null hypothesis is false, i.e. there is no $\gamma \in \Theta_\gamma$ such that $\Lambda_0(\cdot) = \Lambda(\cdot, \gamma)$ a.e. Define:

$$q(y) = \Lambda_0(y) - \Lambda(y, \gamma)$$

where γ is a probability limit of $\hat{\gamma}$. The following theorem establishes consistency of the test under a fixed alternative:

Theorem 4. *Let Assumptions 1, 2*(b)-(d)*,3-5 hold. Additionally, let H_0 be false and*

$$\int_{y_1}^{y_2} [q(y)w(y)]^2 dy > 0.$$

Then, for $\kappa \in (0,1)$ we have:

$$\lim_{n\to\infty} P(T_n > c^*_\kappa) = 1.$$

Here and in the next theorem the values γ and β_1 (from the parametric model) described in Assumptions 2(b),(d) are interpreted as pseudo true values because the parametric model is misspecified.

Now consider local alternatives of the form:

$$\Lambda(y) = \Lambda(y, \gamma) + \frac{1}{\sqrt{n}} \Lambda^{loc}(y), \tag{12}$$

where $\Lambda(y, \gamma) = \Lambda_0(y)$ and $\Lambda^{loc}(\cdot) \in L^2([y_1, y_2])$. Let the sequence of functions $\{\phi_j\}_{j=0}^{\infty}$ form an orthonormal basis of $L^2([y_1, y_2])$. The following theorem provides local asymptotics:

Theorem 5. *Let Assumptions 1, 2*(b)-(d)*,3-5 hold. Under the sequence of local alternatives described in* (12)*:*

$$T_n \to^d \sum_{j=1}^{\infty} \omega_j \chi^2_{j1}\left(\frac{\vartheta_j^2}{\omega_j}\right),$$

where:

$$\vartheta_j = \int_{y_1}^{y_2} \Lambda^{loc}(y) w(y) \phi_j(y) dy$$

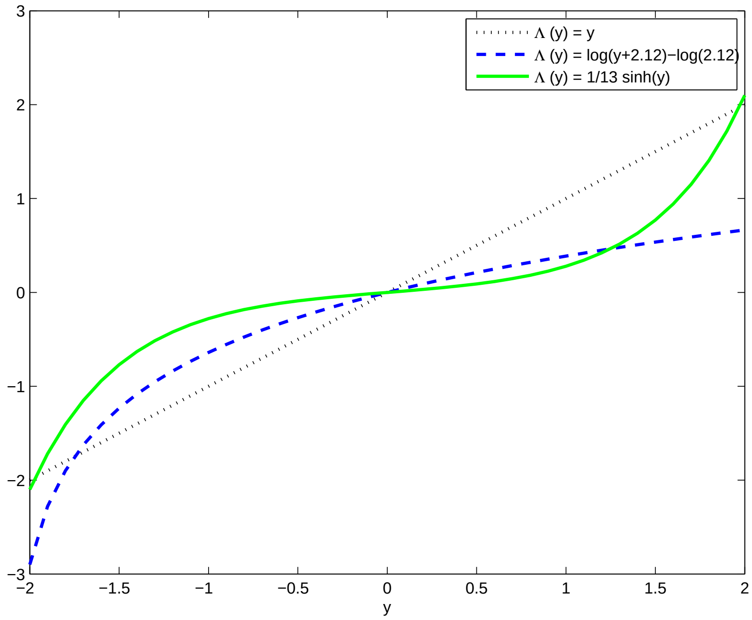
and $\chi^2_{j1}\left(\frac{\vartheta_j^2}{\omega_j}\right)$ *denotes a noncentral chi-square random variable with 1 degree of freedom and noncentrality parameter* $\frac{\vartheta_j^2}{\omega_j}$*.*

Theorem 5 implies that the test has local power against local alternatives that are $n^{-1/2}$ away from the null hypothesis. In principle, different choices of the nonparametric estimator Λ_n will yield different eigenvalues ω_j and thus different local power. However, it is difficult to compare them theoretically since the kernels of the operator $\mathcal{R}$ in (6) for different estimators (HJ, CS and Ye & Duan (1997)) are complicated functions of y. The eigenvalues are usually computed as solutions to differential equations that involve derivatives of the kernels. Hence, the general expressions are hard to get.

3 Monte Carlo simulations

We investigate finite sample performance of the aforementioned testing procedures using several simple designs. We consider both the case when the model in the null hypothesis does not (linear transformation) and does (Box-Cox transformation) depend on the unknown parameter.

Figure 1: Monte Carlo design I



3.1 Linear transformation

The data is generated from the following three models:

$$Y = X + U \tag{Null}$$

$$\log(Y + 2.12) - \log(2.12) = X + U \tag{Alternative 1}$$

$$\frac{1}{13}\sinh(2Y) = X + U \tag{Alternative 2}$$

where X is drawn from the standard normal distribution and U is drawn either from the standard normal, the standard Gumbel or from the logistic distribution. We shifted the logarithmic function by 2.12 in order to minimize L^2 distance of the logarithmic transformation in Alternative 1 to the linear function in the null. We set $[y_1, y_2] = [-2, 2]$. The transformation functions under the null and under the alternatives are normalized at the same point $y_0 = 0$ (though, we do not use this information for running our test i.e. $D = 0$). This design is similar to the one used in Horowitz (1996). Figure 1 shows the shape of the transformation functions.

The model with logistic U can be interpreted as a MPH model with V having the standard Gumbel distribution. Under this interpretation the null model assumes an increasing baseline hazard $\lambda(y) = e^y$, Alternative 1 implies that this hazard is constant and in Alternative 2 the baseline hazard equals $\frac{2}{13}\cosh(2y)e^{\frac{1}{13}\sinh(2y)}$ and is non-monotonic.

We consider both the case when Y is fully observed as well as the case when Y is randomly censored. In the former case we use parametric bootstrap. In the latter case the censoring threshold C is drawn from $N(\mu, 1)$ and μ is chosen such that the probability of being censored is roughly equal to 20%. The coefficient vector β is either estimated by OLS or RCLAD estimator of Honoré et al. (2002).

We run 2000 Monte Carlo replications. We calculate the integral in the test statistic using Halton sequences of size 100. Optimization needed to compute the nonparametric estimator Λ_n was performed using MATLAB's *fminsearch* function with default parameter values. The starting values for the optimization were taken from the null model whether the data was generated by this model or the alternative. The number of bootstrap replications used to calculate the critical value is 500. One Monte Carlo replication in the case with no censoring takes 2.1, 3.2 and 6.2 minutes on average for $n = 100, 500$ and 1000 respectively. For the censored case the respective computing times are 2.1, 8.3 and 12.6 minutes.

Table 1: Rejection probabilities, no censoring

	$U \sim Normal$			$U \sim Gumbel$			$U \sim Logistic$		
				$n = 100$					
	10%	5%	1%	10%	5%	1%	10%	5%	1%
Null	7.6	3.7	0.6	6.1	3.6	1.2	5.9	2.3	0.2
Alternative 1	99.8	99.6	98.7	98.7	97.0	89.6	90.5	88.6	82.9
Alternative 2	98.7	94.4	69.0	95.8	87.2	44.3	71.5	45.6	10.6
				$n = 500$					
	10%	5%	1%	10%	5%	1%	10%	5%	1%
Null	10.6	5.7	1.0	9.0	4.4	1.0	8.5	4.3	0.7
Alternative 1	100.0	100.0	100.0	100.0	100.0	100.0	96.5	96.1	95.7
Alternative 2	100.0	100.0	100.0	100.0	100.0	100.0	99.8	97.3	74.7
				$n = 1000$					
	10%	5%	1%	10%	5%	1%	10%	5%	1%
Null	9.7	4.5	0.9	10.4	5.0	1.1	9.4	4.9	1.1
Alternative 1	100.0	100.0	100.0	100.0	100.0	100.0	97.4	97.2	96.4
Alternative 2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.1

Note: 2000 Monte Carlo simulations, 500 bootstrap replications (parametric bootstrap).

The results for the model without censoring (Table 1) show that our bootstrap test performs very well when $n \geq 500$ with some underrejection for smaller sample size. The test is consistent against both alternatives. Already with a sample size of 500 the test rejects the log-linear and hyperbolic *sin* model almost with certainty.

Table 2: Rejection probabilities, random censoring

	$U \sim Normal$			$U \sim Gumbel$			$U \sim Logistic$		
				$n = 100$					
Null	5.2	3.1	0.9	4.3	2.3	0.5	1.9	0.4	0.0
Alternative 1	57.7	36.0	8.5	56.5	35.6	8.8	27.8	12.8	1.5
Alternative 2	26.9	11.5	1.2	16.8	7.7	0.9	7.7	2.6	0.6
				$n = 500$					
	10%	5%	1%	10%	5%	1%	10%	5%	1%
Null	3.4	0.9	0.0	4.4	1.5	0.4	3.3	1.1	0.0
Alternative 1	100.0	99.9	98.7	99.6	97.9	89.1	98.6	96.1	81.0
Alternative 2	99.9	99.9	99.3	99.9	99.7	97.1	98.9	96.2	82.1
				$n = 1000$					
	10%	5%	1%	10%	5%	1%	10%	5%	1%
Null	5.8	2.5	0.4	5.6	2.2	0.1	5.3	2.1	0.4
Alternative 1	100.0	100.0	99.9	99.9	99.7	98.6	99.9	99.8	97.9
Alternative 2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.1

Note: 2000 Monte Carlo simulations, 500 bootstrap replications (nonparametric bootstrap).

As we can see from Table 2 the finite sample performance of our bootstrap test deteriorates when the dependent variable Y is censored. This is expected because compared to the model with no censoring the rank estimation in the censored model involves additional estimation of the survival function of the censoring threshold C. For example, with a sample of size 100 and censoring rate of 20% we have only about 20 censored observations to estimate this function so the resulting estimator will be quite imprecise. This manifests itself with low power of the test (especially for Alternative 2). However, the power increases fast with the sample size and already with $n = 500$ we reject the alternative models with probability close to one. When it comes to controlling size, even for $n = 1000$ the null rejection probabilities are significantly below the nominal levels which suggests that our test may be conservative in small to medium sized samples. Similar finding was obtained by Subbotin (2007) in his Monte Carlo simulations for the maximum rank correlation estimator of the β coefficients in the transformation model.

Overall, our bootstrap test performs reasonably well in small to moderate samples with a tendency to be on the conservative side.

3.2 Box-Cox transformation

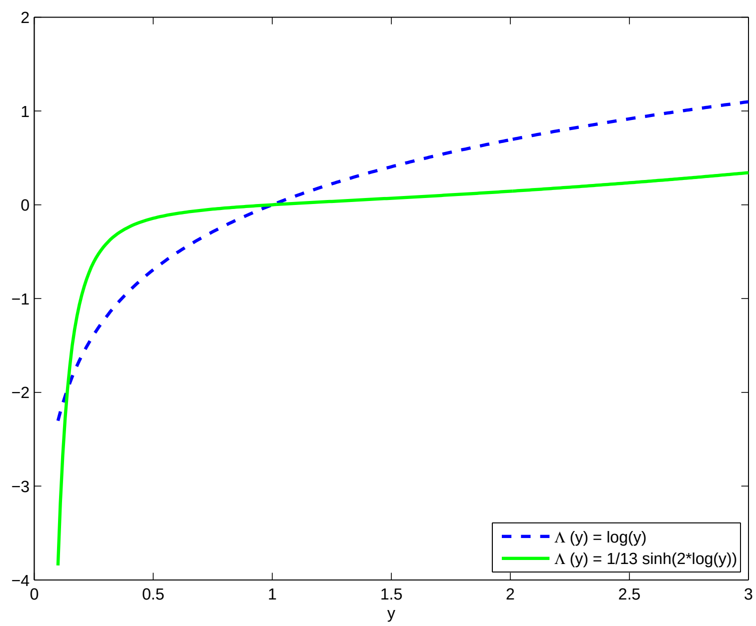
Due to the high computational burden of implementing the test for the Box-Cox model (note that the estimator in Foster et al. (2001) requires minimizing a third order U statistic), we only run a small scale simulation study. We generate data from the log-linear and hyperbolic *sin* model:

$$\log Y = X + U \qquad \text{(Null)}$$

$$\frac{1}{13}\sinh(2\log(Y)) = X + U \qquad \text{(Alternative)}$$

where both X and U are drawn from the standard normal distribution (see Figure 2).

Figure 2: Monte Carlo design II



Following the recommendation in Foster et al. (2001) we use standard normal distribution with mean and variance equal to sample mean and variance of Y as a weighting function Ψ. We set $[y_1, y_2] = [0.1, 3.1]$. Note that both functions are normalized at $y_0 = 1$. We consider only the case without censoring (modifying the semiparametric estimator of the Box-Cox model to accommodate censoring is still an open question).

The results in Table 3 confirm the conclusions from the previous section. The test performs well even in small samples with a tendency to be slightly conservative. Moreover, the results suggest that the test is consistent.

Table 3: Box-Cox model, rejection probabilities

	$U \sim Normal$		
	$n = 100$		
	10%	5%	1%
Null	8.2	3.5	1.0
Alternative	98.2	94.6	73.3
	$n = 200$		
	10%	5%	1%
Null	9.2	4.7	0.8
Alternative	100	100	100
	$n = 300$		
	10%	5%	1%
Null	9.2	4.5	0.3
Alternative	100	100	100

Note: 1000 Monte Carlo simulations, 500 bootstrap replications (parametric bootstrap).

4 Application to Kennan's strike duration data

In this section we apply our testing procedure in the study of the relation between strike durations and the level of economic activity. Kennan (1985) was the first to empirically investigate this relation using data on strikes involving 1000 or more workers in US manufacturing during 1968-1976. He measured the level of economic activity by an index of industrial production in manufacturing (INDP). Table 4 presents summary statistics.

Table 4: Summary statistics, $n = 566$

	Mean	Std. Dev.	Min	Max
strike duration	43.624	44.666	1	235
INDP	.00604	.04991	-.13996	.08554

Note: Strike durations are recorded in days.

Horowitz (2009) re-investigates this question using three models that differ with the parametric assumptions on the transformation function and the distribution F:

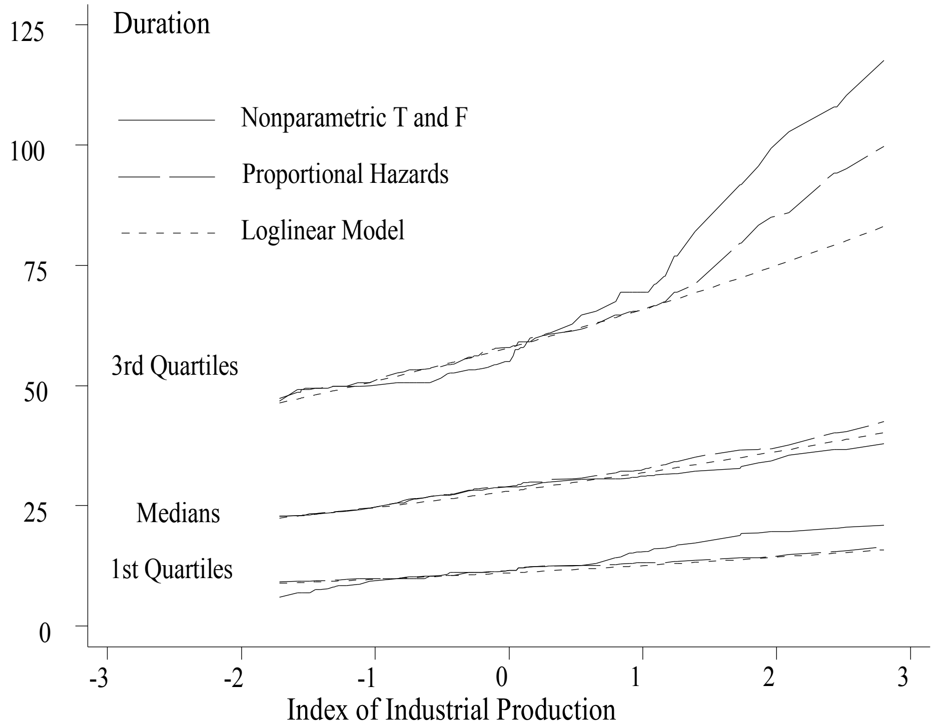
1. proportional hazards model (nonparametric Λ, parametric F)

2. loglinear model (parametric Λ, nonparametric F)

3. nonparametric model (both Λ and F nonparametric)

The results of estimating these three models are summarized in Figure 3, which shows estimates

Figure 3: Results of estimating three models of strike duration



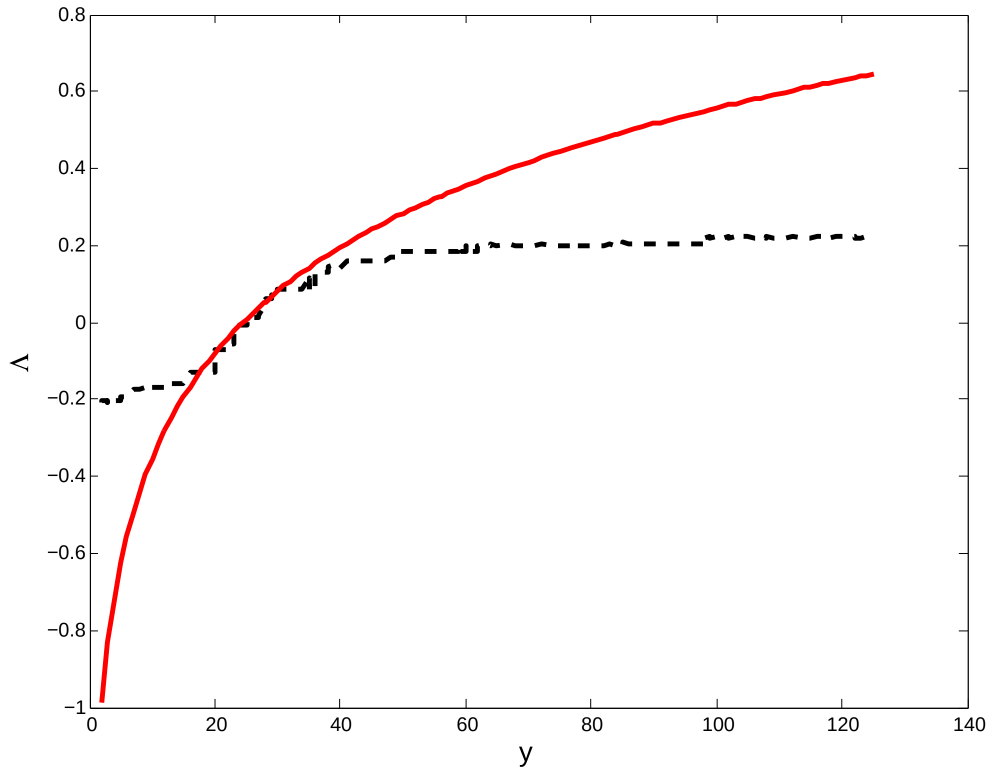
Note: This figure comes from Horowitz (2009), Section 6.5. Higher values of the index correspond to lower levels of economic activity.

of the conditional first quartile, median, and third quartile of the distribution of strike durations given INDP obtained from each of these models.

For our purpose, it is interesting to compare the loglinear and nonparametric model. These two models differ only with respect to the assumptions on the transformation function, which is exactly the setting that we analyzed above. We notice that the loglinear model and the nonparametric model deliver quite similar predictions for the median strike duration but the results diverge for the first and the third quartile, especially for high values of INDP (i.e. periods of low economic activity). In particular, nonparametric estimates suggest that the distribution of strike durations is more highly skewed to the right than the distribution resulting from the estimation of the loglinear model.

The differences in the estimated parametric and nonparametric transformations are also evident from Figure 4. In particular, the nonparametric curve agrees with log specification around the center of the data (median duration is equal to 28) but diverges further from the median. It is interesting to

Figure 4: Nonparametric and parametric estimates of the transformation function



Note: Solid line corresponds to the log transformation and dashed line to the nonparametric estimator obtained using the MRC estimator in Chen (2002).

formally verify if this discrepancies are due merely to the imprecision of the nonparametric estimate in the tails of the data or they signify misspecification of the loglinear model.

For the purpose of our test we set $y_1 = 2$ and $y_2 = 125$ (around 90% of observations on strike durations fall in this range) and use Halton sequence of length 100 to evaluate the integral in (2). We run 500 bootstrap replications to obtain the critical value. The test statistic is equal to 43.24 with the bootstrap critical value of 20.35 at the 1% level. We also run a test for $y_1 = 2$ and $y_2 = 61$ (75% of the sample falls in this range) and obtained $T_n = 11.63$ and $c^*_{0.01} = 5.87$. Thus, we reject the loglinear specification and conclude that the differences between the nonparametric and parametric functions in Figure 4 are caused by misspecification of the transformation function rather than being merely a consequence of the estimation error.

5 Discussion

Our test can be embedded into a formalized specification search procedure using ideas in Romano & Wolf (2005). In other words, one can consider multiple parametric null models and run a stepwise multiple testing procedure to chose the correct specification, controlling family-wise error rate at the desired level.

Similarly to testing the form of Λ, one may test the form of the distribution of U using the estimator for F proposed in Ye & Duan (1997) or Horowitz (1996). One can also apply a procedure used in Horowitz (1996) to derive an estimator for F based on CS. Since the estimators of F usually satisfy conditions equivalent to Assumptions 1-5, the same reasoning may be used to derive a CvM test. Such test may be used to test the form of unobserved heterogeneity (i.e. distribution of V in (7), F_V) in the MPH model. As pointed out by Heckman & Singer (1984), the estimates of the parameters of the MPH model can be very sensitive to the choice of the parametric form of F_V. Therefore, it may be interesting to see if some parametric specifications are at odds with the nonparametric estimate. Specifically, one may want to test for the presence of unobserved heterogeneity, i.e. if $V = 0$ in (7). The tests available so far require X to be discrete (usually X distinguishes separate samples), whereas the procedure applied here allows continuously distributed explanatory variables.

In the context of specification testing for F, the nonparametric estimator in Ye & Duan (1997) seems especially attractive since it does not involve estimation of the transformation function in the first step in order to obtain the estimator of F. Also, it does not require numerical optimization (as CS) or multiple numerical integration (as HJ), which makes bootstrap an attractive route for obtaining the critical value (plug-in asymptotic approach would not be practical in this case since it would rely on arbitrarily chosen smoothing parameters). However, given a different nature of the estimator in Ye & Duan (1997) proving bootstrap validity would require a separate treatment from the one employed above. We plan to investigate this topic in our future research.

A Proofs

Let:

$$\mathcal{H} = \{h_{\theta,y}(w_1, w_2, \ldots, w_m) : \theta \in \Theta \subset \mathbb{R}^d, y \in \mathcal{Y} \subset \mathbb{R}_+\}$$

be a family of real-valued functions defined on $\mathcal{W}^m$. We will use the operator notation common in the U-statistics literature. For example, for the case of $m = 2$ we will have $P^0 h = h$, $P^2 h = Eh(W_1, W_2)$, $P_n h(w_1) = 1/n \sum_{i=1}^n h(w_1, W_i)$ and $P_n^* h(w_1) = 1/n \sum_{i=1}^n h(w_1, W_i^*)$ etc. Additionally, let:

$$h_{\theta,y}^{[m-2]}(w_1, w_2, \ldots, w_{m-2}) = \int h_{\theta,y}(w_1, w_2, \ldots w_{m-2}, W, W) dP(W)$$

Define an U-process:

$$U_n^{(m)} h_{\theta,y} = \frac{(n-m)!}{n!} \sum_{i_1, i_2, \ldots, i_m \text{ distinct}} h_{\theta,y}(W_1, W_2, \ldots, W_m)$$

and denote the same process evaluated on a bootstrap sample as $U_n^{*(m)} h_{\theta,y}$.

We will only discuss the model with censoring (i.e. we focus on nonparametric bootstrap) so Y is the censored observation on the dependent variable. Define $\pi(y) = P(Y \geq y)$ and:

$$M(y) = \mathbb{1}\{Y \leq y, \delta = 0\} - \int_0^y \mathbb{1}\{Y \geq u\} d\Lambda_C(u)$$

where Λ_C is the integrated hazard of the censoring variable C. Proofs for the uncensored case (including proofs for parametric bootstrap) follow similar and, in fact, simpler arguments and therefore are omitted.

We will frequently use the following stochastic order arithmetic, for a sequence a_n:

$$o_p^*(a_n) + o_p(a_n) = o_p(a_n), \quad O_p^*(a_n) + O_p(a_n) = O_p(a_n)$$

which follows from Law of Iterated Expectations.[9]

[9]Cheng & Huang (2010) derive such arithmetic for convergence in outer probability.

A.1 Useful lemmas

Lemma 1. (Lo & Singh (1986)) *Let $\bar{G}_0$ be a continuous survival function of the censoring variable and $\bar{G}_n$ and $\bar{G}_n^*$ be Kaplan-Meier estimators of $\bar{G}_0$ on the original and the bootstrap sample, respectively. Then:*

$$\begin{aligned}\frac{\bar{G}_0(y) - \bar{G}_n(y)}{\bar{G}_0(y)} &= P_n \int_0^y \frac{1}{\pi(s)} dM(s) + o_p(n^{-1/2}) \\ \frac{\bar{G}_0(y) - \bar{G}_n^*(y)}{\bar{G}_0(y)} &= P_n^* \int_0^y \frac{1}{\pi(s)} dM(s) + o_p(n^{-1/2})\end{aligned}$$

uniformly over $\{y : \pi(y) > c\}$ for some $c > 0$.

Proof. This lemma follows from Theorem 1 in Lo & Singh (1986). They show that uniformly over $\{y : \pi(y) > c\}$:

$$\begin{aligned}\frac{G_n(y) - G_0(y)}{\bar{G}_0(y)} &= P_n \xi(y) + o_p(n^{-1/2}) \\ \frac{G_n^*(y) - G_n(y)}{\bar{G}_0(y)} &= (P_n^* - P_n)\xi(y) + o_p^*(n^{-1/2})\end{aligned}$$

where:

$$\xi(y) = \frac{1}{\pi(Y)} \mathbb{1}\{Y \leq y, \delta = 0\} + \int_0^{\min\{Y,y\}} \frac{1}{\pi(s)^2} d\pi_1(s)$$

and

$$\pi_1(s) = 1 - P(Y \leq s, \delta = 0).$$

But we have $\frac{d\pi_1(s)}{\pi(s)} = d \log \bar{G}_0(s)$ (see equation (7) in Lo & Singh (1986)). Now using the fact that the integrated hazard can be expressed as $\Lambda_C(s) = -\log \bar{G}_0(s)$ and $\frac{1}{\pi(Y)} \mathbb{1}\{Y \leq y, \delta = 0\} = \int_0^y \frac{1}{\pi(s)} d\mathbb{1}\{Y \leq y, \delta = 0\}$ we obtain $\xi(y) = \int_0^y \frac{1}{\pi(s)} dM(y)$. Finally:

$$\frac{\bar{G}_0(y) - \bar{G}_n^*(y)}{\bar{G}_0(y)} = \frac{\bar{G}_0(y) - \bar{G}_n(y)}{\bar{G}_0(y)} + \frac{\bar{G}_n(y) - \bar{G}_n^*(y)}{\bar{G}_0(y)} = P_n^* \int_0^y \frac{1}{\pi(s)} dM(s) + o_p(n^{-1/2})$$

□

Lemma 2. (Subbotin (2007)) *Define* $\tau_{\theta,y}(w) = P^{m-1}h_{\theta,y}(w)$ *and let* $h_{\theta_0,y} \equiv 0$*. If* $|h_{\theta,y}(w_1,\ldots,w_m)| < H$ *for some* $0 < H < \infty$ *and:*

(a) Θ *and* $\mathcal{Y}$ *are compact sets,* $P^m h_{\theta,y}$ *is continuous on* Θ *for every* $y \in \mathcal{Y}$*,*

(b) $\mathcal{H}$ *is an Euclidean class of symmetric functions,*

(c) there is an open neighborhood $\mathcal{N} \subset \Theta$ *of* θ_0 *such that:*

  (i) all mixed partial derivatives of $\tau_{\theta,y}(w)$ *with respect to* θ *of orders 1 and 2 exist on* $\mathcal{N}$ *for all* $y \in \mathcal{Y}$*,*

  (ii) there is a square P-integrable function $K(w)$ *such that for all* w*,* $y, y' \in \mathcal{Y}$ *and all* θ *in* $\mathcal{N}$*:*

$$||vec(\partial^2\tau_{\theta,y}(w)) - vec(\partial^2\tau_{\theta_0,y'}(w))|| \leq K(w)\sqrt{||\theta-\theta_0||^2 + (y-y')^2}$$

  where $\partial^2\tau$ *is the Hessian matrix of* τ *with respect to* θ*,*

  (iii) the gradient of $\tau_{\theta,y}$ *with respect to* θ *at* θ_0*,* $\partial\tau_{\theta_0,y}(w)$*, has finite variance relative to* P *for all* $y \in \mathcal{Y}$ *and* $P\partial\tau_{\theta_0,y} = 0$*,*

  (iv) the elements of the matrix $A(y) = -P[\partial^2\tau_{\theta_0,y}]$ *are finite for all* $y \in \mathcal{Y}$

(d) as $\theta \to \theta_0$*,* $P^2\left[P^{m-2}h_{\theta,y} - P^{m-2}h_{\theta_0,y}\right]^2 \to 0$ *for all* $y \in \mathcal{Y}$*,*

(e) as $\theta \to \theta_0$*,* $P^{m-2}h^{[m-2]}_{\theta,y} - P^{m-2}h^{[m-2]}_{\theta_0,y} \to 0$ *for all* $y \in \mathcal{Y}$*,*

then

$$P \sup_{\theta\in\Theta, y\in\mathcal{Y}} |U_n^{*(m)}h_{\theta,y} - P^m h_{\theta,y}| \to 0 \tag{13}$$

and uniformly over $y \in \mathcal{Y}$*:*

$$U_n^{(m)}h_{\theta,y} = (\theta-\theta_0)'mP_n\partial\tau_{\theta_0,y} - \frac{1}{2}(\theta-\theta_0)'A(y)(\theta-\theta_0) + o_p(\|\theta-\theta_0\|^2) + o_p(n^{-1}) \tag{14}$$

$$U_n^{*(m)}h_{\theta,y} = (\theta-\theta_0)'mP_n^*\partial\tau_{\theta_0,y} - \frac{1}{2}(\theta-\theta_0)'A(y)(\theta-\theta_0) + o_p(\|\theta-\theta_0\|^2) + o_p(n^{-1}) \tag{15}$$

as $\theta \to \theta_0$*.*

Proof. This result follows from Lemma 8 and arguments leading to Theorem 2 in Subbotin (2007). The only difference is that the function h is indexed by y in addition to θ. Also note that we do not need invertibility of A here (his Assumption 3(iv)). For completeness we give details of the proof of (15) ((14) follows by similar arguments).

Use the following Hoeffding decomposition for the bootstrapped U-statistic (see Subbotin (2007) for details):

$$U_n^{*(m)} h_{\theta,y} = (\theta-\theta_0)' m P_n^* \partial \tau_{\theta_0,y} - \frac{1}{2}(\theta-\theta_0)' A(y)(\theta-\theta_0) + \hat{\zeta}_{\theta,y}$$

where

$$\begin{aligned}
\hat{\zeta}_{\theta,y} &= P_n^* R_{\theta,y} + \sum_{k=2}^{m} \binom{m}{k} U_n^{*(k)}((\delta_{w_1}-P)\dots(\delta_{w_k}-P)P^{m-k} h_{\theta,y}) \\
\delta_{w_k} h_{\theta,y}(\cdot) &= h_{\theta,y}(\cdot, w_k, \cdot) \\
R_{\theta,y}(w) &= [P^m h_{\theta,y} + m(\delta_{w_1}-P)\tau_{\theta,y}](w) - m(\theta-\theta_0)'\partial\tau_{\theta_0,y}(w) + \frac{1}{2}(\theta-\theta_0)'A(y)(\theta-\theta_0).
\end{aligned}$$

Condition (c) and second order Taylor expansion around θ_0 imply:

$$|P_n^* R_{\theta,y}| \le m\|(P_n^*-P)\partial^2\tau_{\theta_0,y}\|\|\theta-\theta_0\|^2 + m(PK + P_n^* K)\|\theta-\theta_0\|^3$$

in the neighborhood of θ_0.

First we will show that

$$\sup_{y, \|\theta-\theta_0\|\le\delta_n} \frac{|P_n^* R_{\theta,y}|}{\|\theta-\theta_0\|^2} = o_p(1) \tag{16}$$

for $\delta_n \to 0$. Note that condition (c) implies that $PK + P_n^* K = O_p(1)$ (by Theorem 2.1 in Bickel & Freedman (1981)) and that $\partial^2\mathcal{T}_0 = \{vec(\partial^2\tau_{\theta_0,y}(w)) : y \in \mathcal{Y} \subset \mathbb{R}_+\}$ is an Euclidean class of functions (by Lemma 2.13 in Pakes & Pollard (1989)). Thus, by uniform law of large numbers and bootstrap uniform law of large numbers (Theorem 3.5 in Gine & Zinn (1990)) we have $\|(P_n-P)\partial^2\tau_{\theta_0,y}\| = o_p(1)$ and $\|(P_n^*-P_n)\partial^2\tau_{\theta_0,y}\| = o_p^*(1)$ uniformly over y, which implies $\sup_y \|(P_n^*-P)\partial^2\tau_{\theta_0,y}\| = o_p(1)$ and (16) follows.

Using conditions (b), (d), (e) and Lemma 8 in Subbotin (2007) we get:

$$P \sup_{y,\|\theta-\theta_0\|\le\delta_n} \left| \sum_{k=2}^{m} \binom{m}{k} U_n^{*(k)}((\delta_{w_1}-P)\dots(\delta_{w_k}-P)P^{m-k}h_{\theta,y}) \right| = o(n^{-1})$$

which, together with (16), implies

$$\sup_{y,\|\theta-\theta_0\|\le\delta_n} |\hat{\zeta}_{\theta,y}| = o_p(\|\theta-\theta_0\|^2) + o_p(n^{-1}).$$

This concludes the proof of (15). □

A.2 Proof of Theorem 1

Assumptions 1-2 imply that $\hat{\gamma}-\gamma = O_p(n^{-1/2})$, $\hat{\beta}_1-\beta_1 = O_p(n^{-1/2})$ and $\Lambda_n(y)-\Lambda_0(y) = O_p(n^{-1/2})$ uniformly over $y \in \mathcal{Y} = [y_1, y_2]$. Thus:

$$T_n = \int_{y_1}^{y_2} [S_{n1}(y) + S_{n2}(y) + S_{n3}(y) + S_{n4}(y)]^2 dy + o_p(1)$$

uniformly over y, where:

$$\begin{aligned}
S_{n1}(y) &= \sqrt{n}\beta_1(\Lambda_n(y) - \Lambda_0(y))w(y) \\
S_{n2}(y) &= -\sqrt{n}(\Lambda(y,\hat{\gamma}) - \Lambda(y,\gamma))w(y) \\
S_{n3}(y) &= \sqrt{n}\Lambda_0(y)(\hat{\beta}_1 - \beta_1)w(y) \\
S_{n4}(y) &= \sqrt{n}(\beta_1\Lambda_0(y) - \Lambda(y,\gamma))w(y).
\end{aligned}$$

Under the null we have $S_{n4}(y) = 0$ and by Assumption 2(c):

$$\sqrt{n}(\Lambda(y,\hat{\gamma}) - \Lambda(y,\gamma)) = -\sqrt{n}\frac{\partial\Lambda(y,\gamma)}{\partial\gamma}' P_n\Omega_\gamma + o_p(1)$$

uniformly over y. Hence, using Assumptions 1-3 we get:

$$T_n = \int_{y_1}^{y_2} B_n(y)^2 dy + o_p(1)$$

and the statement of the theorem follows from extended continuous mapping theorem (Theorem 1.11.1 in Van der Vaart & Wellner (1996)) and the results in Durbin & Knott (1972), Durbin et al. (1975).

A.3 Proof of Theorem 2

We have:

$$T_n^* = \int_{y_1}^{y_2} [S_{n1}^*(y) + S_{n2}^*(y) + S_{n3}^*(y)]^2 dy$$

uniformly over y, where:

$$\begin{aligned}
S_{n1}^*(y) &= \sqrt{n}\hat{\beta}_1(\Lambda_n^*(y) - \Lambda_n(y))w(y) \\
S_{n2}^*(y) &= -\sqrt{n}(\Lambda(y, \hat{\gamma}^*) - \Lambda(y, \hat{\gamma}))w(y) \\
S_{n3}^*(y) &= \sqrt{n}\Lambda_n^*(y)(\hat{\beta}_1^* - \hat{\beta}_1)w(y).
\end{aligned}$$

We need to obtain a bootstrap linear approximation to $\sqrt{n}(\Lambda_n^*(y) - \Lambda_n(y))$. Let $\theta = (b, \Lambda)$ where $b \in \Theta_\beta$ and $\Lambda \in \Theta_\Lambda$. Let $\Gamma^*(y, G, \Lambda, b) = U_n^*[r(w_1, w_2, G, y, \Lambda, b) + r(w_2, w_1, G, y, \Lambda, b)]$ denote the symmetrized bootstrap rank objective function recentered at the true value Λ_0 and note that Λ_n^* is its arg max. Similarly, let $\Gamma(y, G, \Lambda, b) = P^2[r(W_1, W_2, G, y, \Lambda, b) + r(W_2, W_1, G, y, \Lambda, b)]$. Define:

$$\begin{aligned}
h_{\theta,y}^1(w_1, w_2) = \mathbb{1}\{y^1 \geq y\}&(\mathbb{1}\{x^1 b - x^2 b \geq \Lambda\} - \mathbb{1}\{x^1 b - x^2 b \geq \Lambda_0\}) \\
&+ \mathbb{1}\{y^2 \geq y\}(\mathbb{1}\{x^2 b - x^1 b \geq \Lambda\} - \mathbb{1}\{x^2 b - x^1 b \geq \Lambda_0\}) \\
h_{\theta,y}^2(w_1, w_2) = \mathbb{1}\{y^1 \geq y_0\}&(\mathbb{1}\{x^1 b - x^2 b \geq \Lambda\} - \mathbb{1}\{x^1 b - x^2 b \geq \Lambda_0\}) \\
&+ \mathbb{1}\{y^2 \geq y_0\}(\mathbb{1}\{x^2 b - x^1 b \geq \Lambda\} - \mathbb{1}\{x^2 b - x^1 b \geq \Lambda_0\}).
\end{aligned}$$

We have:

$$\begin{aligned}
\Gamma^*(y, G^*, \Lambda, b) &= \Gamma^*(y, G_0, \Lambda, b) + \Gamma^*(y, G^*, \Lambda, b) - \Gamma^*(y, G_0, \Lambda, b) \\
&= \frac{1}{\bar{G}_0(y)} U_n^* h_{\theta,y}^1 - \frac{1}{\bar{G}_0(y_0)} U_n^* h_{\theta,y}^2 + \frac{\bar{G}_0(y) - \bar{G}^*(y)}{\bar{G}^*(y)\bar{G}_0(y)} U_n^* h_{\theta,y}^1 - \frac{\bar{G}_0(y_0) - \bar{G}^*(y_0)}{\bar{G}^*(y_0)\bar{G}_0(y_0)} U_n^* h_{\theta,y}^2. \qquad (17)
\end{aligned}$$

Define $\tau^l_{\theta,y}(w) = Ph^l_{\theta,y}(w)$ and $A^l(y) = -P[\partial^2 \tau^l_{\theta,y}]$ for $l = 1, 2$. We will use Lemma 2 to show that

$$P \sup_{\theta\in\Theta, y\in\mathcal{Y}} |U^*_n h^l_{\theta,y} - P^2 h^l_{\theta,y}| \to 0 \tag{18}$$

and uniformly over $y \in [y_1, y_2]$:

$$U_n h^l_{\theta,y} = (\theta - \theta_0)' 2P_n \partial\tau^l_{\theta_0,y} - \frac{1}{2}(\theta - \theta_0)' A^l(y)(\theta - \theta_0) + o_p(\|\theta - \theta_0\|^2) + o_p(n^{-1}) \tag{19}$$

$$U^*_n h^l_{\theta,y} = (\theta - \theta_0)' 2P^*_n \partial\tau^l_{\theta_0,y} - \frac{1}{2}(\theta - \theta_0)' A^l(y)(\theta - \theta_0) + o_p(\|\theta - \theta_0\|^2) + o_p(n^{-1}) \tag{20}$$

for $l = 1, 2$ as $\theta \to \theta_0$.

Let us verify conditions of Lemma 2. Condition (a) is implied by Assumptions 4(b),(d),(e). Chen (2002) showed that the classes of functions

$$\mathcal{H}^l = \{h^l_{\theta,y}(w_1, w_2) : \theta \in \Theta \subset \mathbb{R}^d, y \in \mathcal{Y} \subset \mathbb{R}_+\} \qquad l = 1, 2$$

are Euclidean for the envelope $H = 2$, thus condition (b) is satisfied. Condition (c) is implied by Assumption 4(e). Finally, continuity of the distribution of U and X_1 imply condition (d) and condition (e) is satisfied vacuously since $h^{l[m-2]}_{\theta_0,y} \equiv 0$.

Now note that Lemma 1 and Assumption 5 imply that $\frac{\bar{G}_0(y) - \bar{G}^*_n(y)}{\bar{G}_0(y)} = o_p(1)$ and $b^*_n \to^p \beta_0$. Combining this, equation (17), the result in (18) and using Assumption 4(e) we obtain:

$$\Gamma^*(y, G^*, \Lambda, b^*_n) = \Gamma(y, G_0, \Lambda, \beta_0) + o_p(1)$$

uniformly over $y \in \mathcal{Y}$ and $\Lambda \in \Theta_\Lambda$. Chen (2002) showed that Λ_0 is the unique maximizer of the expression on the right, which implies consistency of $\Lambda^*_n(y)$ for $\Lambda_0(y)$. Now monotonicity of $\Lambda^*_n(y)$ implies uniform consistency, i.e. $\sup_y |\Lambda^*_n(y) - \Lambda_0(y)| = o_p(1)$, by the same argument as in the proof of Theorem 1 in Chen (2002).

Note that $\frac{\partial\tau^l_{\theta,y}}{\partial b}\big|_{\Lambda=\Lambda_0} = 0$ and $P\frac{\partial^2\tau^l_{\theta,y}}{\partial b^2}\big|_{\Lambda=\Lambda_0} = 0$. Let $V^l_{\Lambda b}(y) = -P\frac{\partial^2\tau^l_{\theta,y}}{\partial\Lambda\partial b}\big|_{\theta=\theta_0}$ and $V^l(y) =$

$-P\frac{\partial^2 \tau^l_{\theta,y}}{\partial \Lambda^2}\big|_{\theta=\theta_0}$. Then (20) becomes:

$$U_n^* h^l_{\theta,y} = (\Lambda-\Lambda_0)2P_n^*\frac{\partial \tau^l_{\theta_0,y}}{\partial \Lambda} - (\Lambda-\Lambda_0)V^l_{\Lambda b}(y)'(b-\beta_0) - \frac{1}{2}(\Lambda-\Lambda_0)^2 V^l(y) + o_p((\Lambda-\Lambda_0)^2) + o_p(n^{-1}) \tag{21}$$

as $\Lambda \to \Lambda_0$ and $b \to \beta_0$.

Chen (2002) shows that under our assumptions the class of functions $\partial\mathcal{T}_0 = \{\partial\tau_{\theta_0,y}(w) : y \in \mathcal{Y} \subset \mathbb{R}_+\}$ is Euclidean with a square integrable envelope. Similar argument shows that the same property holds for $\partial^2\mathcal{T}_0 = \{vec(\partial^2\tau_{\theta_0,y}(w)) : y \in \mathcal{Y} \subset \mathbb{R}_+\}$ (see also proof of Lemma 2). Thus, Theorem 3.5 in Gine & Zinn (1990) gives: $\sup_y \|(P_n^* - P_n)\partial\tau_{\theta_0,y}\| = O_p^*(n^{-1/2})$ and $\sup_y \|(P_n^* - P_n)\partial^2\tau_{\theta_0,y}\| = O_p^*(n^{-1/2})$ and similarly $\sup_y \|(P_n - P)\partial\tau_{\theta_0,y}\| = O_p(n^{-1/2})$ and $\sup_y \|(P_n - P)\partial^2\tau_{\theta_0,y}\| = O_p(n^{-1/2})$.

This and Lemma 1 imply that the third and the fourth term in (17) can be written as:

$$(\Lambda-\Lambda_0)\left(\frac{2}{\bar{G}_0(y)}P\frac{\partial\tau^1_{\theta_0,y}}{\partial\Lambda}P_n^*\int_0^y \frac{1}{\pi}dM - \frac{2}{\bar{G}_0(y_0)}P\frac{\partial\tau^2_{\theta_0,y}}{\partial\Lambda}P_n^*\int_0^{y_0}\frac{1}{\pi}dM\right)$$
$$+ o_p((\Lambda-\Lambda_0)^2) + o_p((\Lambda-\Lambda_0)/\sqrt{n}) + o_p(n^{-1}) \tag{22}$$

uniformly over y.

Note that $V(y) = \frac{V^1(y)}{\bar{G}_0(y)} - \frac{V^2(y)}{\bar{G}_0(y_0)}$ and $\frac{\partial\tau(W,y,\Lambda_0)}{\partial\Lambda} = \frac{1}{\bar{G}_0(y)}\frac{\partial\tau^1_{\theta_0,y}}{\partial\Lambda} - \frac{1}{\bar{G}_0(y_0)}\frac{\partial\tau^2_{\theta_0,y}}{\partial\Lambda}$. Define $V_{\Lambda b} = \frac{V^1_{\Lambda b}}{\bar{G}_0(y)} - \frac{V^2_{\Lambda b}}{\bar{G}_0(y_0)}$. Thus, substituting (21) and (22) into (17) and using Assumption 5 we obtain:

$$\Gamma^*(y, G^*, \Lambda, b_n^*) = (\Lambda-\Lambda_0)P_n^*\Omega_{\Lambda,y} - \frac{1}{2}(\Lambda-\Lambda_0)^2 V(y)$$
$$+ o_p((\Lambda-\Lambda_0)^2) + o_p((\Lambda-\Lambda_0)/\sqrt{n}) + o_p(n^{-1})$$

where

$$\Omega_{\Lambda,y} = 2\frac{\partial\tau(W,y,\Lambda_0)}{\partial\Lambda} + \frac{2}{\bar{G}_0(y)}\int_0^y \frac{1}{\pi}dM\left(P\frac{\partial\tau^1_{\theta_0,y}}{\partial\Lambda}\right) - \frac{2}{\bar{G}_0(y_0)}\int_0^{y_0}\frac{1}{\pi}dM\left(P\frac{\partial\tau^2_{\theta_0,y}}{\partial\Lambda}\right) - V_{\Lambda b}(y)'\Omega^{NP}$$

uniformly over y. Now using $\sup_y |\Lambda_n^*(y) \to \Lambda_0(y)| \to 0$ one can proceed as in Sherman (1993) to show that:

$$\sqrt{n}(\Lambda_n^*(y) - \Lambda_0(y)) = V(y)^{-1} P_n^* \Omega_{\Lambda,y} + o_p(1)$$

uniformly over y. From Chen (2002) and Jochmans (2012):

$$\sqrt{n}(\Lambda_n(y) - \Lambda_0(y)) = V(y)^{-1} P_n \Omega_{\Lambda,y} + o_p(1)$$

and the class of functions $\mathcal{J} = \{J(\cdot, y) = V(y)^{-1}\Omega_{\Lambda,y}(\cdot)\}$ is Euclidean with square integrable envelope. Now:

$$\sqrt{n}(\Lambda_n^*(y) - \Lambda_n(y)) = V(y)^{-1}(P_n^* - P_n)\Omega_{\Lambda,y} + o_p(1)$$

and by Theorem 3.5 in Gine & Zinn (1990) we have $\sup_y |\Lambda_n^*(y) - \Lambda_n(y)| = O_p(n^{-1/2})$, which together with Assumption 2(d) yields:

$$S_{n1}^*(y) = \sqrt{n}\beta_1 V(y)^{-1}(P_n^* - P_n)\Omega_{\Lambda,y} w(y) + o_p(1).$$

uniformly over y. Further, Assumptions 2(b)-(d), Assumption 5 and Theorem 2.2 in Bickel & Freedman (1981) imply:

$$S_{n2}^*(y) = -\sqrt{n}\frac{\partial \Lambda(y,\gamma)}{\partial \gamma}'(P_n^* - P_n)\Omega_\gamma w(y) + o_p(1)$$
$$S_{n3}^*(y) = \sqrt{n}\Lambda_0(y)(P_n^* - P_n)\Omega_1 w(y) + o_p(1)$$

uniformly over y. Denote $\mathcal{B}_n^*(y) = \sqrt{n}(P_n^* - P_n)[\beta_1 V(y)^{-1}\Omega_{\Lambda,y} - \frac{\partial\Lambda(y,\gamma)}{\partial\gamma}\Omega_\gamma + \Lambda_0(y)\Omega_1]w(y)$. Now note that functions Ω_γ and Ω_1 are not indexed by y and $w(y)$, $\frac{\partial\Lambda(y,\gamma)}{\partial\gamma}w(y)$ and $\Lambda_0(y)w(y)$ are constant for fixed y. Thus, by Lemma 2.14 in Pakes & Pollard (1989), Theorem 3.5 in Gine & Zinn (1990) and the extended continuous mapping theorem (Theorem 1.11.1 in Van der Vaart & Wellner (1996)) we have that $\int_{y_1}^{y_2}\mathcal{B}_n^{*2}(y)dy$ converges weakly to $\int_{y_1}^{y_2}\mathcal{B}^2(y)dy$ in conditional probability. Additionally, by continuity of the distribution of $\int_{y_1}^{y_2} \mathcal{B}^2(y)dy$ and monotonicity of CDFs this implies:

$$\sup_{t\in[0,1]} \left| P_n^* \mathbb{1}\left\{ \int_{y_1}^{y_2} \mathcal{B}_n^{*2}(y)dy \le t \right\} - P\mathbb{1}\left\{ \int_{y_1}^{y_2} \mathcal{B}^2(y)dy \le t \right\} \right| = o_p(1) \tag{23}$$

By Theorem 1:

$$\sup_{t\in[0,1]} \left| P\mathbb{1}\{T_n \le t\} - P\mathbb{1}\left\{ \int_{y_1}^{y_2} \mathcal{B}^2(y)dy \le t \right\} \right| = o(1) \tag{24}$$

By our derivation above $T_n^* = \int_{y_1}^{y_2} \mathcal{B}_n^{*2}(y)dy + o_p(1)$ which implies:

$$\sup_{t\in[0,1]} \left| P\mathbb{1}\{T_n^* \le t\} - P\mathbb{1}\left\{ \int_{y_1}^{y_2} \mathcal{B}_n^{*2}(y)dy \le t \right\} \right| = o(1) \tag{25}$$

Putting (23), (24) and (25) together and using Law of Iterated Expectations we obtain

$$\sup_{t\in[0,1]} |P\mathbb{1}\{T_n^* \le t\} - P\mathbb{1}\{T_n \le t\}| = o_p(1)$$

Now taking $t = c_\kappa^*$ concludes the proof.

A.4 Proof of Theorem 3

We can write

$$L_n(\theta) = \int_0^\infty U_n^{(3)} h_{\theta,y}^{BC} d\Psi(y) + P_n R(W,\theta)$$

and

$$L_n^*(\theta) = \int_0^\infty U_n^{*(3)} h_{\theta,y}^{BC} d\Psi(y) + P_n^* R(W,\theta)$$

Note that minimizing $L_n(\theta) = S_n(\gamma,\beta) + P_n R(W,\theta)$ is the same as minimizing $\tilde{L}_n(\theta) = S_n(\gamma,\beta) - S_n(\gamma_0,\beta_0) + P_n[R(W,\theta) - R(W,\theta_0)]$ and similarly for the bootstrap problem. Thus, without loss of generality we take $S_n(\gamma_0,\beta_0) = 0$ and $R(W,\theta_0) = 0$.

We can use Lemma 2 to show $P\sup_{\theta,y}|U_n^{*(3)}h_{\theta,y}^{BC} - P^3 h_{\theta,y}^{BC}| \to 0$ and:

$$U_n^{(3)}h_{\theta,y}^{BC} = (\theta-\theta_0)'3P_n\partial\tau_{\theta_0,y}^{BC} - \frac{1}{2}(\theta-\theta_0)'A(y)(\theta-\theta_0) + o_p(\|(\theta-\theta_0)\|^2) + o_p(n^{-1})$$
$$U_n^{*(3)}h_{\theta,y}^{BC} = (\theta-\theta_0)'3P_n^*\partial\tau_{\theta_0,y}^{BC} - \frac{1}{2}(\theta-\theta_0)'A(y)(\theta-\theta_0) + o_p(\|(\theta-\theta_0)\|^2) + o_p(n^{-1})$$

as $\theta \to \theta_0$, uniformly over y, where $A(y) = -P\partial^2\tau_{BC}(W, y, \theta_0)$.

Let us verify the conditions of Lemma 2. Clearly, $h_{\theta,y}^{BC}$ is uniformly bounded. Assumption (a) is satisfied with $\mathcal{Y} = \{y : \frac{d\Psi(y)}{dy} > 0\}$ and follows from Assumptions 4(b),(e). Part (b) has been shown by Foster et al. (2001). Assumption 4(e), boundedness of $\mathcal{Y}$ and $E\left[\sup_{g\in\Theta_\gamma}\left|\frac{Y^g\log Y - \Lambda(Y,g)}{g}\right|\right]^2 < \infty$ imply condition (c). Now note that $P^2[(Ph_{\theta,y}^{BC} - Ph_{\theta_0,y}^{BC})^2] \leq 2P^2[Ph_{\theta,y}^{BC} - Ph_{\theta_0,y}^{BC}]$ and condition (d) follows from continuity of the distribution of U and X_1. Condition (e) follows from a similar argument.

Now using $E\left[\sup_{g\in\Theta_\gamma}\left|\frac{Y^g\log Y - \Lambda(Y,g)}{g}\right|\right]^2 < \infty$ and Lemma 2.13 in Pakes & Pollard (1989) we find that the class of functions $\mathcal{R} = \{R(\cdot,\theta) : \theta \in \Theta\}$ is Euclidean with square integrable envelope. Hence, $\sup_\theta |P_n^* R(W,\theta) - PR(W,\theta)| = o_p(1)$, which together with the previous derivation implies that

$$L_n^*(\theta) = \int_0^\infty P^3 h_{\theta,y}^{BC}(W)d\Psi(y) + PR(W,\theta) + o_p(1)$$

holds uniformly over $\theta \in \Theta$. Foster et al. (2001) show that the expression on the right is uniquely maximized at θ_0. It follows that θ^* is consistent for θ_0.

Next, we have as $\theta \to \theta_0$:

$$P_n^* R(W,\theta) = (\theta-\theta_0)'P_n^*\partial R(W,\theta_0) + (\theta-\theta_0)'P_n^*\partial^2 R(W,\theta_0)(\theta-\theta_0) + o_p(\|(\theta-\theta_0)\|^2)$$

where $\partial R(W,\theta)$ denotes the gradient of R with respect to θ.

Putting the above linear representations for $U_n^{*(3)}h_{\theta,y}^{BC}$ and $P_n^* R(W,\theta)$ together and noting that

$|P_n^* \partial^2 R(W,\theta_0) - P\partial^2 R(W,\theta_0)| = O_p(n^{-1/2})$ under condition (c) of the theorem we obtain, as $\theta \to \theta_0$:

$$L_n^*(\theta) = (\theta-\theta_0)' P_n^* \left[3\int_0^\infty \partial\tau_{\theta_0,y}^{BC} d\Psi(y) + \partial R(W,\theta_0) \right] - \frac{1}{2}(\theta-\theta_0)' V_{BC}(\theta-\theta_0) + o_p(\|(\theta-\theta_0)\|^2) + o_p(n^{-1})$$

Now using the fact that V_{BC} is invertible and proceeding as in the proof of Theorem 2 we get:

$$\theta^* - \theta_0 = V_{BC}^{-1} P_n^* \left[3\int_0^\infty \partial\tau_{\theta_0,y}^{BC} d\Psi(y) + \partial R(W,\theta_0) \right] + o_p(n^{-1/2})$$

which implies:

$$P\left| \theta^* - \theta_0 - V_{BC}^{-1} P_n^* \left[3\int_0^\infty \partial\tau_{\theta_0,y}^{BC} d\Psi(y) + \partial R(W,\theta_0) \right] \right| = o(n^{-1/2})$$

(see e.g. Theorem 1.4.C in Serfling (1980) where the uniform integrability follows from assumptions of the theorem). Finally, $P\left[3\int_0^\infty \partial\tau_{\theta_0,y}^{BC} d\Psi(y) + \partial R(W,\theta_0)\right] = 0$ by first order condition of the population maximization problem and $Var\left[3\int_0^\infty \partial\tau_{\theta_0,y}^{BC} d\Psi(y) + \partial R(W,\theta_0)\right]$ has finite elements by Assumption 4(e) and condition (c) in the statement of the theorem.

A.5 Proof of Theorem 4

First, note that, due to centering at the sample estimators Λ_n and $\Lambda(y,\hat{\gamma})$, bootstrap gives a valid estimate of the asymptotic distribution of T_n under the null both when the data is generated from the null model and the alternative model (in fact, the same argument as in the proof of Theorem 2 applies with redefining γ and β_1 as pseudo true values). Now, by Assumption 2 we have $B_n = O_p(n^{-1/2})$ which implies:

$$\frac{T_n}{n} = \int_{y_1}^{y_2} \left[\frac{1}{\sqrt{n}} B_n(y) + q(y)w(y) \right]^2 dy + o_p(1) = \int_{y_1}^{y_2} [q(y)w(y)]^2 dy + o_p(1),$$

thus $T_n \to \infty$ and $\lim_{n\to\infty} P(T_n > c_\kappa^*) = 1$.

A.6 Proof of Theorem 5:

Using the spectral decomposition, under the sequence of local alternatives we get:

$$T_n = \int_{y_1}^{y_2} [B_n(y) + \Lambda^{loc}(y)w(y)]^2 dy + o_p(1) = \int_{y_1}^{y_2} \left[\sum_{j=1}^{\infty} (b_j + \vartheta_j)\phi_j(x) \right]^2 dy + o_p(1) =$$

$$= \sum_{j=1}^{\infty} (b_j + \vartheta_j)^2 + o_p(1),$$

where $\{\phi_j : j = 1, 2, ...\}$ form complete orthonormal basis of $L^2([y_1, y_2])$ and b_j's are asymptotically $N(0, \omega_j)$. Therefore, $T_n \to \sum_{j=1}^{\infty} \omega_j \chi_{1j}(\vartheta_j^2/\omega_j)$ (cf. Durbin & Knott (1972), Durbin et al. (1975)).

References

Abrevaya, J. (2003), 'Pairwise-Difference Rank Estimation of the Transformation Model', *Journal of Business & Economic Statistics* **21**(3), 437–47.

Bickel, P. J. & Freedman, D. A. (1981), 'Some asymptotic theory for the bootstrap', *The Annals of Statistics* **9**(6), pp. 1196–1217.

Blundell, R. & Horowitz, J. L. (2007), 'A non-parametric test of exogeneity', *The Review of Economic Studies* **74**(4), pp. 1035–1058.

Cavanagh, C. & Sherman, R. P. (1998), 'Rank estimators for monotonic index models', *Journal of Econometrics* **84**(2), 351 – 381.

Chen, S. (2002), 'Rank estimation of transformation models', *Econometrica* **70**(4), pp. 1683–1697.

Cheng, G. & Huang, J. Z. (2010), 'Bootstrap consistency for general semiparametric m-estimation', *Ann. Statist.* **38**(5), 2884–2915.

Durbin, J. & Knott, M. (1972), 'Components of cramer-von mises statistics. i', *Journal of the Royal Statistical Society. Series B (Methodological)* **34**(2), 290–307.

Durbin, J., Knott, M. & Taylor, C. C. (1975), 'Components of cramer-von mises statistics. ii', *Journal of the Royal Statistical Society. Series B (Methodological)* **37**(2), 216–237.

Ekeland, I., Heckman, J. J. & Nesheim, L. (2004), 'Identification and estimation of hedonic models', *Journal of Political Economy* **112**(S1), S60–S109.

Fernandes, M. & Grammig, J. (2005), 'Nonparametric specification tests for conditional duration models', *Journal of Econometrics* **127**(1), 35 – 68.

Foster, A. M., Tian, L. & Wei, L. J. (2001), 'Estimation for the Box-Cox transformation model without assuming parametric error distribution', *Journal of the American Statistical Association* **96**, 1097–1101.

Gine, E. & Zinn, J. (1990), 'Bootstrapping general empirical measures', *Ann. Probab.* **18**(2), 851–869.

Gørgens, T. & Horowitz, J. L. (1999), 'Semiparametric estimation of a censored regression model with an unknown transformation of the dependent variable', *Journal of Econometrics* **90**(2), 155 – 191.

Han, A. K. (1987), 'A non-parametric analysis of transformations', *Journal of Econometrics* **35**(2-3), 191–209.

Heckman, J. & Singer, B. (1984), 'A method for minimizing the impact of distributional assumptions in econometric models for duration data', *Econometrica* **52**(2), 271–320.

Honoré, B. E. (1990), 'Simple estimation of a duration model with unobserved heterogeneity', *Econometrica* **58**(2), 453–473.

Honoré, B., Khan, S. & Powell, J. L. (2002), 'Quantile regression under random censoring', *Journal of Econometrics* **109**(1), 67 – 105.

Horowitz, J. (2009), *Semiparametric and Nonparametric Methods in Econometrics*, Springer.

Horowitz, J. L. (1996), 'Semiparametric estimation of a regression model with an unknown transformation of the dependent variable', *Econometrica* **64**(1), 103–37.

Horowitz, J. L. (1999), 'Semiparametric estimation of a proportional hazard model with unobserved heterogeneity', *Econometrica* **67**(5), 1001–1028.

Horowitz, J. L. (2006), 'Testing a parametric model against a nonparametric alternative with identification through instrumental variables', *Econometrica* **74**(2), 521–538.

Ichimura, H. (1993), 'Semiparametric least squares (SLS) and weighted least squares estimation of single index models', *Journal of Econometrics* **58**, 71–120.

Ishwaran, H. (1996*a*), 'Identifiability and rates of estimation for scale parameters in location mixture models', *The Annals of Statistics* **24**(4), 1560–1571.

Ishwaran, H. (1996*b*), 'Uniform rates of estimation in the semiparametric weibull mixture model', *The Annals of Statistics* **24**(4), 1572–1585.

Jochmans, K. (2012), 'The variance of a rank estimator of transformation models', *Economics Letters* **117**(1), 168–169.

Kennan, J. (1985), 'The Duration of Contract Strikes in U.S. Manufacturing', *Journal of Econometrics* **28**(1), 5–28.

Klein, R. W. & Sherman, R. P. (2002), 'Shift restrictions and semiparametric estimation in ordered response models', *Econometrica* **70**(2), 663–691.

Lo, S.-H. & Singh, K. (1986), 'The product-limit estimator and the bootstrap: Some asymptotic representations', *Probability Theory and Related Fields* **71**(3), 455–465.

Meyer, B. D. (1990), 'Unemployment insurance and unemployment spells', *Econometrica* **58**(4), 757–782.

Pakes, A. & Pollard, D. (1989), 'Simulation and the asymptotics of optimization estimators', *Econometrica* **57**(5), 1027–57.

Romano, J. P. & Wolf, M. (2005), 'Stepwise multiple testing as formalized data snooping', *Econometrica* **73**(4), 1237–1282.

Serfling, R. J. (1980), *Approximation theorems of mathematical statistics*, Wiley, Chichester; New York.

Sherman, R. P. (1993), 'The limiting distribution of the maximum rank correlation estimator', *Econometrica* **61**(1), pp. 123–137.

Subbotin, V. (2007), Asymptotic and Bootstrap Properties of Rank Regressions. MPRA Working Paper 9030.

Van der Vaart, A. W. & Wellner, J. A. (1996), *Weak Convergence and Empirical Processes: with Applications to Statistics*, Springer-Verlag, New York.

Ye, J. & Duan, N. (1997), 'Nonparametric $\sqrt{n}$-consistent estimation for the general transformation models', *Annals of Statistics* **25**, 2682–2717.