Online Appendix:

Testing a parametric transformation model versus a nonparametric alternative

Arkadiusz Szydłowski*

University of Leicester

October 29, 2019

A Additional proofs

A.1 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 $B_n(y) = O_p(n^{-1/2})$ uniformly over $y \in [y_1, y_2]$ 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 $\lim_{n\to\infty} P(T_n > c^*_\kappa) = 1$.

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

A.2 Proof of Theorem 5:

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

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

where $\{\psi_j : j = 1, 2, ...\}$ 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)).

B Additional Monte Carlo simulations

B.1 Kolmogorov-Smirnov statistic

We re-run our Monte Carlo simulations in Table 1 in the main text using the following statistic:

$$T_n^{KS} = n \max_{y \in [y_1, y_2]} |(a_n \Lambda_n(y) - \Lambda(y, \hat{\gamma}))w(y)| \tag{1}$$

instead of the statistic in (2) (in the main text), with uniform weights $w(y) = 1$. In order to calculate the max we perform a grid search with 100 equally-spaced points (same as the size of Halton sequences used in the MC integration). Computation times are very similar to the Cramer-von-Mises test. Results are in Table B.1.

Overall, coverage probabilities under the Null and Alternative 1 are very similar to our C-v-M test. However, the KS test displays much lower power against Alternative 2 for $n = 100$ and $n = 500$ (especially with logistic shocks).

B.2 Alternative testing intervals $[y_1, y_2]$

The designs are the same as in Table 1 in the main text besides that we set $[y_1, y_2] = [-3, 3]$ and $[y_1, y_2] = [-4, 4]$ here. Results are in Table B.2.

B.3 Different censoring rates

We re-run the Monte Carlo simulations in Table 2 with censoring rates changed to 10% and 30%.

Comparing the results in Table B.3 to those in Table 2 we conclude that for moderate sample sizes ($n = 500$ or $n = 1000$) higher rates of censoring lead to poorer size control and lower power of our test. For small sample sizes in fact higher rate of censoring leads to better performance of the test in many cases. This is due to the fact that higher rate of censoring leads to more observations on C_i and, thus, more precise estimates of C_i's survival function: with $n = 100$ we have 10 or 30 observations for estimation of this function with low and high censoring rate, respectively.

C Assumptions for $\sqrt{n}$-consistency of Han's and Ichimura's estimator

We show that our Assumptions 1-4 are not at odds with assumptions in Han (1987) and Ichimura (1993) and thus, after necessary strengthening, imply $\sqrt{n}$-consistency of their estimators of β_0. Let β_n denote either of these estimators.

C.1 Maximum rank correlation, Han (1987)

The $\sqrt{n}$-consistency of β_n follows from Theorem 4 in Sherman (1993). Let:

$$\tau((x,y),b) = E[\mathbb{1}\{y > Y_i\}\mathbb{1}\{x'b > X_i'b\} + \mathbb{1}\{Y_i > y\}\mathbb{1}\{X_i'b > x'b\}]$$

Translated into our setup his assumptions A.1-A.4 are as follows:

A.1 β_0 is an interior point of a compact set $\Theta_\beta \subset \mathbb{R}^q$

A.2 X and U are independent.

A.3 The support of X is not contained in any proper linear subspace of $\mathbb{R}^q$ and X_1 has an everywhere positive Lebesgue density, conditional on other components.

A.4 There exist a neighborhood of β_0, $\mathcal{N}_\beta$, such that:

(i) For each (x, y) all mixed second partial derivatives of $\tau((x, y), \cdot)$ exist on $\mathcal{N}_\beta$.

(ii) There is an integrable function $M(x,y)$ such that for all (x,y) and all $b \in \mathcal{N}_\beta$:

$$\|vec(\partial^2\tau((x,y),b)) - vec(\partial^2\tau((x,y),\beta_0))\| \leq M(x,y)\|b-\beta_0\|$$

(iii) $E\|\partial\tau(\cdot,\beta_0)\|^2 < \infty$

(iv) $E\left[\sum_{i_1,\ldots,i_q}\left|\frac{\partial^q}{\partial\beta_{i_1}\ldots\partial\beta_{i_q}}\tau(\cdot,\beta_0)\right|\right] < \infty$

(v) The matrix $E[\partial^2\tau(\cdot,\beta_0)]$ is negative definite.

A.1-A.3 follow from assumptions of the model (see p. 2) and our Assumption 4. According to the discussion in Section 8 of Sherman (1993) a sufficient condition for the first four assumptions in A.4 is that the conditional density of X_1 given $X_{-1} = x_{-1}$ and $Y = y$ has bounded derivatives up to order three for each (x_{-1}, y). Assuming that $[y_1, y_2]$ is contained in the support of Y and using Bayes law, the latter would follow from strengthening of our Assumption 4(e) to:

Assumption 4(e)'. *The conditional density of X_1 given $X_{-1} = x_{-1}$ and the density of U are bounded and thrice continuously differentiable, the derivatives are uniformly bounded and X_{-1} has finite third-order moments.*

Also, as discussed in Section 6 of Sherman (1993), if Y is continuously distributed conditional on $X'b$ (which is implied by our Assumption 4), condition A.4(v) follows from A.3.

C.2 Semiparametric least squares, Ichimura (1993)

Let $Z = X'b$ and f_z denote its density. Assumptions in Ichimura (1993) translated into our model are:

A.4.1 Λ is differentiable and non-degenerate.

A.4.2 The support of X is not contained in any proper linear subspace of $\mathbb{R}^q$ and X_1 has an everywhere positive Lebesgue density, conditional on other components.

A.5.1 $\{X_i, Y_i, \delta_i : i = 1, \ldots, n\}$ is a random sample.

A.5.2 β_0 is an interior point of a compact set $\Theta_\beta \subset \mathbb{R}^q$.

A.5.3 There exist a compact subset of the support of X, $\mathcal{X}^{comp}$, such that:

(i) $\sup_{x \in \mathcal{X}^{comp}} f_z(x'b) > 0$

(ii) f_z and $E[Y|Z=z]$ are three times continuously differentiable with respect to z, and the third derivatives satisfy Lipschitz conditions for all $\{z : z = x'b, x \in \mathcal{X}^{comp}\}$ uniformly in $b \in \Theta_\beta$.

A.5.4 $E|Y|^3 < \infty$ and $Var(Y|X=x)$ is uniformly bounded and bounded away from 0 on $\mathcal{X}^{comp}$.

A.4.1-A.5.4 follow from our Assumptions 1 and 4, if in the latter assumption we strengthen part (e) to:

Assumption 4(e)". *The conditional density of X_1 given $X_{-1} = x_{-1}$ and the density of U are bounded and thrice differentiable with third derivatives being Lipschitz continuous, the derivatives are uniformly bounded and X_{-1} has finite third-order moments.*

Additionally, in order to satisfy A.4.1 and A.5.3(ii) we need to assume that Λ is differentiable with uniformly bounded first derivative.

D Sufficient conditions for Assumption 2(b) in Box-Cox and Bickel-Doksum models

In this section we list assumptions needed for an asymptotic linear representation of Han's estimator in the Box-Cox and Bickel-Doksum transformation models. These assumptions are listed in Asparouhova et al. (2002) where they also show that the Euclidean property in their A7 is satisfied for these models, thus we skip it.

For $w_l = (x^l, y^l), l = 1, \ldots, 4$ define:

$$h(w_1, w_2, w_3, w_4, b, g) = \mathbb{1}\{\Lambda(y^1, g) - \Lambda(y^2, g) > \Lambda(y^3, g) - \Lambda(y^4, g)\}\mathbb{1}\{(x^1 - x^2)'b > (x^3 - x^4)'b\}$$
$$- \mathbb{1}\{\Lambda(y^1, \gamma) - \Lambda(y^2, \gamma) > \Lambda(y^3, \gamma) - \Lambda(y^4, \gamma)\}\mathbb{1}\{(x^1 - x^2)'\beta > (x^3 - x^4)'\beta\}$$

and:

$$\tau(w, b, g) = E[h(w, W_2, W_3, W_4, g, b) + h(W_1, w, W_3, W_4, g, b) + h(W_1, W_2, w, W_4, g, b) + h(W_1, W_2, W_3, w, b, g)]$$

Let $\partial_\gamma \tau$ and $\partial^2_{\gamma\gamma}\tau$ denote the first and second derivative of τ with respect to the last argument.

A1 $\{U_i\}_{i=1}^n$ are i.i.d.

A2 The X_i's are i.i.d. and independent of the U_i's.

A3 $X'\beta$ is continuously distributed and U has a non-degenerate distribution.

A4 (β, γ) is an interior point of a compact set $\Theta_\beta \times \Theta_\gamma$.

A5 For each $g \in \Theta_\gamma$, $\Lambda(\cdot, g)$ is continuous and strictly increasing.

A6 With positive probability, $\Lambda(\cdot, g)$ is differentiable and nonlinear in $\Lambda(\cdot, \gamma)$ for $g \neq \gamma$.

A8 $\sqrt{n}(\hat{\beta} - \beta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \Omega_\beta(Y_i, X_i, \beta) + o_p(1)$ as $n \to \infty$ and $\frac{1}{\sqrt{n}} \sum_{i=1}^n \Omega_\beta(Y_i, X_i, \beta) \to^d N(0, E[\Omega_\beta(\cdot, \cdot, \beta)\Omega_\beta(\cdot, \cdot, \beta)'])$

A9 There exist a nondegenerate convex neighborhood of (β, γ), $\mathcal{N}_{\beta\gamma}$, such that:

(i) For each w, $\tau(w, \cdot, \cdot)$ has continuous mixed third partial derivatives on $\mathcal{N}_{\beta\gamma}$.

(ii) There is an integrable function $M(w)$ such that for each w and $(b, g) \in \mathcal{N}_{\beta\gamma}$:

$$|\partial^2_{\gamma\gamma}\tau(w, b, g) - \partial^2_{\gamma\gamma}\tau(w, b, \gamma)| \leq M(w)|g - \gamma|$$

(iii) $E|\partial_\gamma \tau(\cdot, \beta, \gamma)|^2 < \infty$

(iv) $E\|\partial^2_{\gamma\gamma}\tau(\cdot, \beta, \gamma)\| < \infty$

(v) The matrix $E[\partial^2_{\gamma\gamma}\tau(\cdot, \beta, \gamma)]$ is negative definite.

A1-A9 essentially follow from assumptions of our model, Assumptions 2(c), 4(b)-(e) and additionally assuming $E\left[\sup_{g \in \mathcal{N}_\gamma} \left|\frac{\partial^2 \Lambda(Y,g)}{\partial g^2}\right|\right]^2 < \infty$. Linear representation in A8 is consistent with our Assumption 2(d) and is satisfied by Han's and Ichimura's estimators under mild strengthening of Assumption 4(e) (see Section C). Condition A9(v) follows from identification, established by Han (1987).

E 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: [1]

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

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.

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

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

[1] In fact, Foster et al. (2001) state S_n in a form of a 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.

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 E.1. *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 and (γ, β) is an interior point of Θ,

(b) $E\left[\sup_{g \in \Theta_g} \left|\frac{\partial^2 \Lambda(Y, g)}{\partial g^2}\right|\right]^2 < \infty$

(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).

Proof. The proof follows lines similar to the proof of Theorem 2 in the main text. 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 3 to show $P\sup_{\theta,y}|U_n^{*(3)}h_{\theta,y}^{BC} - P^3 h_{\theta,y}^{BC}| \to 0$ and:

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

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 3. 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 with $m = 3$ condition (d) follows from continuity of the distribution of U and X_1.

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

$$
L_n^*(\theta) = \int_0^\infty P^3 h_{\theta,y}^{BC}(W_1, W_2, W_3)d\Psi(y) + PR(W,\theta) + o_p(1)
$$

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 $\theta \to \theta_0$:

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

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. □

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 $\gamma \neq 0$:

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

E.1 Monte Carlo simulations

We verify finite sample performance of the test for the Box-Cox model using Monte Carlo simulations. 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 1).

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 the weighting function Ψ. We set $[y_1, y_2] = [0.1, 3.1]$. Note that both functions are normalized at $y_0 = 1$.

The results in Table E.1 confirm the conclusions from Section 3. The test performs well even in small samples with a tendency to be slightly conservative. Moreover, the results suggest that the test is consistent.

References

Asparouhova, E., Golanski, R., Kasprzyk, K., Sherman, R. P. & Asparouhov, T. (2002), 'Rank estimators for a transformation model', *Econometric Theory* **18**(5), 1099–1120.

Durbin, J. & Knott, M. (1972), 'Components of Cramér-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 Cramér-von Mises statistics. II', *Journal of the Royal Statistical Society. Series B (Methodological)* **37**(2), 216–237.

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.

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

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

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

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.

Table B.1: Rejection probabilities, Kolmogorov-Smirnov test, no censoring

	$U \sim Normal$			$U \sim Gumbel$			$U \sim Logistic$		
	$n = 100$								
	10%	5%	1%	10%	5%	1%	10%	5%	1%
Null	9.2	3.6	0.5	9.8	5.5	1.5	7.0	3.2	0.6
Alternative 1	100.0	100.0	99.9	99.9	99.9	99.2	92.8	91.5	88.4
Alternative 2	79.6	52.4	8.4	55.2	26.6	2.6	17.4	5.7	0.2
	$n = 500$								
	10%	5%	1%	10%	5%	1%	10%	5%	1%
Null	10.5	5.4	1.0	8.8	5.1	1.7	9.0	4.6	1.1
Alternative 1	100.0	100.0	100.0	100.0	100.0	100.0	97.5	97.1	96.7
Alternative 2	100.0	100.0	100.0	100.0	100.0	99.4	93.9	80.2	31.1
	$n = 1000$								
	10%	5%	1%	10%	5%	1%	10%	5%	1%
Null	8.9	4.4	0.8	10.1	5.3	1.4	9.5	4.8	1.1
Alternative 1	100.0	100.0	100.0	100.0	100.0	100.0	97.9	97.7	97.5
Alternative 2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.6	91.6

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

Table B.2: Rejection probabilities, no censoring

	$[y_1, y_2] = [-3, 3]$								
	$U \sim Normal$			$U \sim Gumbel$			$U \sim Logistic$		
	$n = 100$								
	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
	$[y_1, y_2] = [-4, 4]$								
	$n = 100$								
	10%	5%	1%	10%	5%	1%	10%	5%	1%
Null	5.9	3.1	0.6	5.8	2.8	0.8	6.0	3.2	0.6
Alternative 1	99.8	98.9	96.5	98.9	97.6	91.0	91.4	89.8	84.3
Alternative 2	93.3	85.8	53.0	74.1	61.3	32.0	38.1	24.9	7.9
	$n = 500$								
	10%	5%	1%	10%	5%	1%	10%	5%	1%
Null	10.7	5.2	1.1	10.6	5.3	1.0	10.2	5.6	1.5
Alternative 1	100.0	100.0	100.0	100.0	100.0	100.0	96.7	96.4	95.8
Alternative 2	100.0	100.0	100.0	100.0	100.0	99.3	96.6	88.2	53.2
	$n = 1000$								
	10%	5%	1%	10%	5%	1%	10%	5%	1%
Null	10.0	5.7	1.3	11.2	5.6	1.0	11.2	5.1	1.1
Alternative 1	100.0	100.0	100.0	100.0	100.0	100.0	97.4	97.3	96.7
Alternative 2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.7	88.0

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

Table B.3: Rejection probabilities, different censoring rates

10% censoring rate

	$U \sim Normal$			$U \sim Gumbel$			$U \sim Logistic$		
	$n = 100$								
	10%	5%	1%	10%	5%	1%	10%	5%	1%
Null	2.3	1.1	0.1	2.2	0.4	0.2	1.5	0.2	0.0
Alternative 1	44.8	24.1	4.3	44.9	24.9	4.4	27.1	13.3	1.7
Alternative 2	30.5	15.0	1.7	18.9	7.9	1.0	7.7	2.8	0.4
	$n = 500$								
	10%	5%	1%	10%	5%	1%	10%	5%	1%
Null	4.7	1.8	0.0	4.2	1.7	0.2	4.3	1.6	0.2
Alternative 1	100.0	100.0	97.8	99.6	97.9	88.2	98.6	96.1	81.0
Alternative 2	100.0	100.0	100.0	100.0	99.9	98.7	99.6	98.5	90.7
	$n = 1000$								
	10%	5%	1%	10%	5%	1%	10%	5%	1%
Null	6.7	3.2	0.4	5.5	2.6	0.4	4.5	1.9	0.3
Alternative 1	100.0	100.0	99.9	99.9	99.8	98.4	99.9	99.8	97.9
Alternative 2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9
30% censoring rate									
	$n = 100$								
	10%	5%	1%	10%	5%	1%	10%	5%	1%
Null	13.5	8.5	3.0	13.7	8.0	2.2	3.5	1.5	0.4
Alternative 1	72.1	54.5	21.0	66.6	47.7	15.2	37.4	20.9	3.1
Alternative 2	17.3	7.5	2.7	12.5	6.2	2.1	8.4	3.7	1.3
	$n = 500$								
	10%	5%	1%	10%	5%	1%	10%	5%	1%
Null	3.2	1.4	0.4	3.3	1.4	0.3	3.5	1.8	0.2
Alternative 1	100.0	99.8	98.7	99.7	98.5	90.9	98.8	96.2	82.1
Alternative 2	99.3	97.6	86.5	99.0	97.0	82.1	93.9	85.8	56.8
	$n = 1000$								
	10%	5%	1%	10%	5%	1%	10%	5%	1%
Null	4.7	1.8	0.3	4.2	0.9	0.2	4.0	1.7	0.2
Alternative 1	100.0	100.0	99.8	99.9	99.8	99.3	99.9	99.8	98.0
Alternative 2	100.0	100.0	99.6	100.0	100.0	99.5	99.6	98.9	91.4

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

Table E.1: 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).