Ann Inst Stat Math (2014) 66:383–412
DOI 10.1007/s10463-013-0420-2

Varying coefficients partially linear models with randomly censored data

Francesco Bravo

Received: 20 July 2012 / Revised: 16 April 2013 / Published online: 31 August 2013

Abstract This paper considers the problem of estimation and inference in semiparametric varying coefficients partially linear models when the response variable is subject to random censoring. The paper proposes an estimator based on combining inverse probability of censoring weighting and profile least squares estimation. The resulting estimator is shown to be asymptotically normal. The paper also proposes a number of test statistics that can be used to test linear restrictions on both the parametric and nonparametric components. Finally, the paper considers the important issue of correct specification and proposes a nonsmoothing test based on a Cramer von Mises type of statistic, which does not suffer from the curse of dimensionality, nor requires multidimensional integration. Monte Carlo simulations illustrate the finite sample properties of the estimator and test statistics.

Keywords Empirical likelihood · Goodness of fit · Kaplan–Meier estimator · Profile least squares · Wilks phenomenon · Wald statistic

1 Introduction

Varying coefficient models (Cleveland et al. 1991; Hastie and Tibshirani 1993) arise in many situations of practical relevance in economics, finance and statistics. They have been used in the context of generalised linear models and quasi-likelihood esti-

Electronic supplementary material The online version of this article (doi:10.1007/s10463-013-0420-2) contains supplementary material, which is available to authorized users.

F. Bravo (✉)
Department of Economics, University of York, York YO10 5DD, UK
e-mail: francesco.bravo@york.ac.uk
https://sites.google.com/a/york.ac.uk/francescobravo

mation (Cai et al. 2000a), time series (Cai et al. 2000b), longitudinal data (Fan and Wu 2008), survival analysis (Cai et al. 2008) to name just a few applications-(see Fan and Jiang 2008) for a recent review containing further applications and a number of examples. Varying coefficient partially linear models assume that some of the varying coefficients are constant, and thus are extension of the popular partially linear model considered by Engle et al. (1986) and Speckman (1988) among many others. Compared to the latter, varying coefficient partially linear models offer additional flexibility because they allow interactions between a vector of covariates and a vector of unknown functions depending on another covariate. Ahmad et al. (2005) and Fan and Huang (2005) suggest two general estimation techniques based, respectively, on nonparametric series and profile least square estimation. Both approaches yield semiparametric efficient estimators for the parametric components under the assumption of conditional homoskedasticity. Fan and Huang (2005) also consider inference and show that the so-called Wilks phenomenon (Fan et al. 2001; Fan and Jiang 2007) holds for the profile likelihood ratio statistic, implying that its distribution does not depend on unknown parameters.

In this paper, we consider varying coefficient partially linear models when the response variable is not directly observed; instead it is subjected to (right) random censoring. Censoring is important in empirical applications, arising naturally in biostatistics and in medical statistics; it can also be used in economics to model for example unemployment spells.

We propose a unifying theory for estimation, inference and specification for varying coefficient partially linear models under random censoring. The theory relies on the same inverse probability of censoring weighting (IPCW henceforth) approach as that used, for example, by Wang and Li (2002) and Lu and Burke (2005), which is based on the transformation of Koul et al. (1981). This transformation yields an estimator that is not as efficient as that based on Buckley and James (1979)'s transformation or the more recent approach of Heuchenne and van Keilegom (2007), but as opposed to the former it does not require complex iterative computations that can lead to unstable solutions, nor it is subjected to the curse of dimensionality as the latter.

In this paper, we make a number of contributions: first, we derive the asymptotic distribution of the estimators for both the parametric and nonparametric components using profile likelihood (least squares) estimation as in Fan and Huang (2005). We show that in both cases random censoring has an effect on the variance of the estimators whereas it has no effect on the bias of the nonparametric estimator. The latter result is consistent with the findings of Fan and Gijbels (1994) and Cai (2003) who both investigated the effect of random censoring in nonparametric regressions.

Second, we consider inference for the parametric and nonparametric components and show that the Wilks phenomenon does hold for the profile least squares ratio (generalised likelihood ratio) statistic for the nonparametric components, but, as opposed to Fan and Huang (2005), not for the parametric components. This different asymptotic behaviour can be explained by the different effect that the estimation of IPCW has on the asymptotic variance of the estimators of the parametric and nonparametric components. First, note that since the IPCW is estimated at the parametric rate, it can be regarded as known for nonparametric inferences. Therefore, as in Fan and Huang (2005), the asymptotic distribution of the profile least squares ratio for the nonparametric components can be obtained directly from that of a standard varying coefficient model for which the Wilks phenomenon holds (Fan et al., 2001). On the other hand, since the estimator of the IPWC is not orthogonal to the profile least squares estimator of the parametric component, it has an effect on the asymptotic variance of the latter, which is the key fact to explain the failure of the Wilks phenomenon. It is important to note that failures of the Wilks phenomenon can also be observed in the context of parametric likelihood ratio inferences with misspecified models (because the information matrix equality—second Bartlett identity—does not hold, (see e.g. White 1982) and empirical likelihood inferences for certain semiparametric models (because the asymptotic variance of the score is not equal to the expectation of its square, (see e.g. Wang and Veraverbeke 2006; Xue and Zhu 2007; Bravo 2009; Chen and Van Keilegom 2009). In both cases, the test statistics are asymptotically equivalent to a quadratic form in normal random vectors with the "wrong" covariance matrix, which is also what happens in the case of parametric inferences for the profile least squares and empirical likelihood ratio statistics of this paper (see (33) in the Appendix), and explains why the Wilks phenomenon does not hold (see Sect. 3.1 for a further discussion). It is also important to note that even if the Wilks phenomenon does not hold for the original statistics, it can still be obtained by applying the same scale correction to both statistics. Furthermore, the distribution of the profile least square ratio can be accurately approximated by the bootstrap as we show in Sects. 3.2 and 5.

Third we consider the important issue of correct specification of the varying coefficient partially linear model. We follow the so-called nonsmoothing (or unconditional) approach and propose a test based on a Cramer von Mises type of statistic. The test statistic is based on the same dimension reduction approach proposed by Escanciano (2006) in the context of parametric regression models. This approach yields a test statistic that does not suffer from the curse of dimensionality, nor requires multidimensional integration. On the other hand, the test statistic is not asymptotic distribution free but its distribution can be easily simulated by a resampling method that is computationally simpler than the bootstrap. The method is motivated by the so-called random symmetrisation technique described for example by Pollard (1984) and by the multiplier central limit theorems of Van der Vaart and Wellner (1996). It has been previously used by Su and Wei (1991), Delgado et al. (2003), Zhu and Ng (2003) and others.

Finally, we use simulations to assess the finite sample properties of the proposed estimator and test statistics for both the parametric and nonparametric components.

The remaining part of the paper is structured as follows: next section introduces the model and the estimator. Section 3 discusses inference for both the parametric and nonparametric components. Section 4 introduces the Cramer von Mises statistic for the correct specification of the model and shows the consistency of the proposed resampling technique. Section 5 presents the results of the Monte Carlo simulations. Section 6 contains some concluding remarks. An Appendix contains sketches of the proofs. Full proofs and some further Monte Carlo simulations can be found in the online supplement to this paper.

The following notation is used throughout the paper: "$a.a.$" , "$a.s.$" and "$\overset{a}{\sim}$" stand for "almost all ", "almost surely" and "asymptotically distributed as"; $\Rightarrow$, $\overset{d}{\to}$

denote weak convergence in $l^{\infty}(\cdot)$—the space of all real-valued uniformly bounded functions (see Van der Vaart and Wellner (1996) for a definition), and convergence in distribution, respectively. Finally, $\wedge$ denotes min, and $\|\cdot\|$ denotes the Euclidean norm.

2 Estimation

Consider a varying coefficient partially linear model

$$Y = X'\alpha_0(U) + W'\beta_0 + \varepsilon, \tag{1}$$

where $\alpha_0(\cdot)$ is a p-dimensional vector of unknown functions, β_0 is a k-dimensional vector of unknown parameters and the unobservable error ε is such that $E(\varepsilon|U, X, W) = 0\, a.s.$ and $E\left(\varepsilon^2|U, X, W\right) = \sigma^2\, a.s.$

Let $(Y_i, U_i, X_i, W_i)_{i=1}^n$ denote an i.i.d. sample from (Y, U, X, W); Fan and Huang (2005) propose to estimate $\alpha_0(\cdot)$ and β_0 using profile least squares estimation: for a given β_0, (1) can be written as

$$Y_i^* = X_i'\alpha_0(U_i) + \varepsilon_i, \tag{2}$$

where $Y_i^* = Y_i - W_i'\beta_0$, and $\alpha_0(u)$ can be estimated by a local regression. Plugging the resulting estimator $\widehat{\alpha}(U_i)$ back in (2), it follows that β_0 can be estimated using least squares on

$$\widetilde{Y}_i = \widetilde{W}'\beta_0 + \varepsilon_i,$$

where a tilde denotes the (empirical) projection on the varying coefficient functional space, that is

$$\widetilde{W}_j = W_j - W_j S_j(U_i)' \text{ with} \tag{3}$$
$$S_j(U_i) = X_j'\left(\sum_{l=1}^n X_l X_l' K_h(U_l - U_i)\right)^{-1} \sum_{l=1}^n X_l K_h(U_l - U_i),$$

where $K_h(\cdot) := K(\cdot/h)/h$ is a kernel function and $h =: h(n)$ is the bandwidth. Fan and Huang (2005) show that the resulting estimator $\widehat{\beta} = \left(\sum_{i=1}^n \widetilde{W}_i\widetilde{W}_i'\right)^{-1} \sum_{i=1}^n \widetilde{W}_i\widetilde{Y}_i$ is asymptotically normal and achieves the semiparametric efficiency lower bound.

Suppose that the Y_is are randomly censored, so that instead of observing $(Y_i)_{i=1}^n$, we observe $(Z_i, \delta_i)_{i=1}^n$ where

$$Z_i = Y_i \wedge C_i, \quad \delta_i = I(Y_i \le C_i),$$

and C_i is an i.i.d. sample from the censoring variable C with unknown distribution function G_0, assumed to be independent of (Y, U, W, X). To deal with the randomly censored responses, we follow Koul et al. (1981)'s approach and consider the following

IPWC transformed responses (synthetic responses) $Z_{G_0} = Z\delta / (1 - G_0(Z))$ and note that

$$E\left(Z_{G_0}|U, X, W\right) = X'\alpha_0(U) + W'\beta_0 \quad a.s., \tag{4}$$

which suggests that an estimator for β_0 can be based on profile least square estimation using the IPCW observations $Z_{iG_0} = Z_i\delta_i / (1 - G_0(Z_i))$ instead of Y_i $(i = 1, \ldots, n)$, that is

$$Z_{iG_0} = X_i'\alpha_0(U_i) + W_i'\beta_0 + \varepsilon_{iG_0}. \tag{5}$$

To estimate the unknown G_0, we use the Kaplan–Meier product-limit estimator (Kaplan and Meier, 1958)

$$1 - \widehat{G}(z) = \prod_{i=1}^{n}\left(\frac{N(Z_i)}{1 + N(Z_i)}\right)^{I(Z_i \le z, \delta_i = 0)}$$

where $N(z) = \sum_{i=1}^{n} I(Z_i > z)$. We note that the assumption of independence between the censoring variable and the covariates is crucial for the Kaplan–Meier estimator. This assumption is reasonable in many applications, such as, for example, those when censoring is caused by the termination of the study. On the other hand, if this assumption is not reasonable one could replace the Kaplan–Meier estimator with Beran (1981)'s local product-limit estimator. This extension is certainly interesting (although in practice it would be limited by the curse of dimensionality of nonparametric estimators), but is beyond the scope of the paper.

Given $\widehat{G}$, we can estimate β_0 using profile least squares as in the uncensored case, which results in

$$\widehat{\beta} = \left(\sum_{i=1}^{n} \widetilde{W}_i \widetilde{W}_i'\right)^{-1} \sum_{i=1}^{n} \widetilde{W}_i \widetilde{Z}_{i\widehat{G}}.$$

Let

$$\eta(Z, \delta; z) = \int_0^{Z \wedge z} \left[\overline{H}(s)\right]^{-2} \mathrm{d}H_0(s) + \frac{I(Z \le z, \delta = 0)}{\overline{H}(Z)}, \tag{6}$$

where $H(z) = \Pr(Z \le z)$, $\overline{H}(z) = 1 - H(z)$ and $H_0(z) = \Pr(Z > z, \delta = 0)$ denote the key term appearing in Lo and Singh (1986)'s asymptotic linear representation of the Kaplan–Meier estimator.

Assume that

A1 The random variable U has bounded support $\mathcal{U}$, and its density $f(\cdot)$ is Lipschitz continuous and bounded away from 0 in $\mathcal{U}$,

A2 The $p \times p$ matrix $E\left(XX'|U\right)$ is nonsingular for each U, and $E\left(WX'|U\right)$, $E\left(XX'|U\right)$ have Lipschitz continuous second derivatives in $U \in \mathcal{U}$, and $E\left(XX'|U\right)^{-1}$ is Lipschitz continuous,
A3 $E\left(\|X\|^4\right) < \infty,\ E\left(\|W\|^4\right) < \infty,\ E\left(\varepsilon_{G_0}^4\right) < \infty$,
A4 The functions $\alpha_j\left(U\right)$ have Lipschitz continuous second derivatives in $U \in \mathcal{U}$,
A5 As $n \to \infty$ $n^{1/2}h^4 \to 0$ and $nh^2/\ln\left(n\right) \to \infty$,
A6 The kernel function $K\left(\cdot\right)$ is a symmetric density with compact support .

The above assumptions are similar to the ones used by Fan and Huang (2005) and are routinely imposed in local regression methods, see e.g. Fan and Gijbels (1996) and Masry (1996).

Theorem 1 *Under A1–A6*

$$n^{1/2}\left(\widehat{\beta} - \beta_0\right) \xrightarrow{d} N\left(0, \Gamma^{-1}\Xi_\beta\Gamma^{-1}\right), \tag{7}$$

where

$$\begin{aligned}
\Gamma &= E\left[WW' - \Omega\left(U\right)\Sigma\left(U\right)^{-1}\Omega\left(U\right)'\right],\\
\Xi_\beta &= \sigma_{G_0}^2 E\left[WW' - \Omega\left(U\right)\Sigma\left(U\right)^{-1}\Omega\left(U\right)'\right]\\
&\quad +2E\left\{\left[W_1 - \Omega\left(U_1\right)\Sigma\left(U_1\right)^{-1}X_1\right]\varepsilon_{1G_0}\right.\\
&\quad \left.\times\ E\left[W_2 - \Omega\left(U_2\right)\Sigma\left(U_2\right)^{-1}X_2\eta\left(Z_1,\delta_1;Z_2\right)Z_{2G_0}|Z_1,\delta_1\right]'\right\}\\
&\quad +E\left\{E\left[W_2 - \Omega\left(U_2\right)\Sigma\left(U_2\right)^{-1}X_2\eta\left(Z_1,\delta_1;Z_2\right)Z_{2G_0}|Z_1,\delta_1\right]\right.\\
&\quad \left.\times\ E\left[W_2 - \Omega\left(U_2\right)\Sigma\left(U_2\right)^{-1}X_2\eta\left(Z_1,\delta_1;Z_2\right)Z_{2G_0}|Z_1,\delta_1\right]'\right\},
\end{aligned}$$

$\sigma_{G_0}^2 = E\left(\varepsilon_{G_0}^2|U,X,W\right)$, *and* $\Omega\left(U\right) = E\left(WX'|U\right)$, $E\left(XX'|U\right) = \Sigma\left(U\right)$.

The following theorem establishes the distribution of the estimators $\widehat{\alpha}\left(\cdot\right)$ of $\alpha_0\left(\cdot\right)$. Let $\kappa_2 = \int K\left(v\right)v^2\mathrm{d}v$, $\alpha^{(j)}\left(u\right) = \partial^j\alpha\left(u\right)/\partial u^j$. Assume that

A5' As $n \to \infty$ $n^{1/2}h^{5/2} = O\left(1\right)$ and $nh^2/\ln\left(n\right) \to \infty$,

Theorem 2 *Under A1–A6 with A5' replacing A5 for any fixed value u with* $f\left(u\right) > 0$

$$\left(nh\right)^{1/2}\left(\widehat{\alpha}\left(u\right) - \alpha\left(u\right) - h^2B\left(u\right)\right) \xrightarrow{d} N\left(0, \Delta^{-1}\left(u\right)\Xi_\alpha\left(u\right)\Delta^{-1}\left(u\right)\right),$$

where

$$B(u) = \kappa_2 \Delta^{-1}(u) E \times \left\{ XX' \left[\alpha^{(1)}(u) f(X,U) / f(X|U=u) + f(U) \alpha^{(2)}(u)/2 \right] |U=u \right\},$$
$$\Delta(u) = f(u) \Sigma(u), \quad \Xi_\alpha(u) = f(u) \sigma^2_{G_0} \int K^2(v) dv E\left[XX'|U=u\right].$$

Note that the asymptotic bias $B(u)$ is the same as that for the kernel-based estimator of a standard varying coefficient model with uncensored data (see for example Li et al. 2002). On the other hand, the asymptotic variance is larger. This is, however, typical of models with IPCW responses and more generally with synthetic type of responses (see for example Fan and Gijbels 1994 and Cai 2003).

3 Inference

In this section, we consider inference for both the parametric and nonparametric components of (5). For the former component, we consider a number of test statistics: the profile least squares ratio of Fan and Huang (2005), a generalised Wald and an empirical likelihood ratio. We note that all of these statistics do not rely on any parametric assumptions. For the nonparametric component, we only consider the profile least squares ratio because of its computational simplicity.

3.1 Parametric component

It is often of interest to test linear hypotheses that can be expressed as

$$H_0 : R\beta_0 = r, \tag{8}$$

where R is an $l \times k$ matrix of constants ($l \leq k$). As in Fan and Huang (2005), we propose the following profile least squares ratio test for (8)

$$\mathrm{PL}_n = \frac{n}{2} \frac{\mathrm{RSS}_0 - \mathrm{RSS}_1}{\mathrm{RSS}_1}, \tag{9}$$

where $\mathrm{RSS}_0 = \sum_{i=1}^n \left(Z_{i\widehat{G}} - X_i'\widehat{\alpha}_0(U_i) - W_i'\widehat{\beta}_0 \right)^2$ and $\mathrm{RSS}_1 = \sum_{i=1}^n \left(Z_{i\widehat{G}} - X_i'\widehat{\alpha}(U_i) - W_i'\widehat{\beta} \right)^2$ denote, respectively, the residual sum of squares from the profile least squares under (8) and the unrestricted one. As mentioned in Sect. 1, the Wilks phenomenon does not hold for (9) because as we show in the proof of Proposition 3 PL_n is asymptotically equivalent to a quadratic form in $n^{1/2}\left(\widehat{\beta} - \beta_0\right)$ with a covariance matrix that is not Ξ_β^{-1} [see (33) in the Appendix]. As a result, the asymptotic distribution of PL_n is not the standard χ^2; instead it is given in the following Proposition.

Proposition 3 *Assume that* $\text{rank}(R) = l$*. Then under A1–A6 and the null hypothesis (8)*

$$PL_n \xrightarrow{d} \sum_{j=1}^{l} \omega_j \chi_j^2(1),$$

where ω_j *are the eigenvalues of* $\left[\sigma_{G_0}^2\left(R\Gamma^{-1}R'\right)\right]^{-1}\left(R\Gamma^{-1}\Xi_\beta\Gamma^{-1}R'\right)$.

To obtain a test statistic with an asymptotic χ^2 calibration, we can use a (generalised) Wald statistic, or alternatively an adjusted profile least squares ratio and empirical likelihood ratio statistic defined in (12) below. Note that all of these statistics require an explicit computation of Ξ_β.

The Wald statistic for (8) is

$$W_n = n\left(R\widehat{\beta} - r\right)'\left[R\left(\widehat{\Gamma}^{-1}\widehat{\Xi}_\beta\widehat{\Gamma}^{-1}\right)R'\right]^{-1}\left(R\widehat{\beta} - r\right),$$

where

$$\widehat{\Gamma} = \frac{1}{n}\sum_{i=1}^{n}\left[W_iW_i' - \widehat{E}\left(W_iX_i'|U_i\right)\left[\widehat{E}\left(X_iX_i'|U_i\right)\right]^{-1}\widehat{E}\left(X_iW_i'|U_i\right)\right], \tag{10}$$

$$\begin{aligned}\widehat{\Xi}_\beta &= \frac{\widehat{\sigma}_{\widehat{G}}^2}{n}\sum_{i=1}^{n}\left\{\left[W_iW_i' - \widehat{E}\left(W_iX_i'|U_i\right)\left[\widehat{E}\left(X_iX_i'|U_i\right)\right]^{-1}\widehat{E}\left(X_iW_i'|U_i\right)\right]\right\}\\ &+\frac{2}{n}\sum_{i=1}^{n}\sum_{j\neq i}\left\{\left[W_i - \widehat{E}\left(W_iX_i'|U_i\right)\left[\widehat{E}\left(X_iX_i'|U_i\right)\right]^{-1}X_i\right]\widehat{\varepsilon}_{i\widehat{G}}\right.\\ &\left.\times\widehat{E}\left[W_j - \widehat{E}\left(W_jX_j'|U_j\right)\left[\widehat{E}\left(X_iX_i'|U_i\right)\right]^{-1}X_j\eta\left(Z_i,\delta_i;Z_j\right)Z_{2\widehat{G}}|Z_i,\delta_i\right]'\right\}\\ &+\frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}\left\{\widehat{E}\left[W_j - \widehat{E}\left(W_jX_j'|U_j\right)\left[\widehat{E}\left(X_iX_i'|U_i\right)\right]^{-1}X_j\eta\left(Z_i,\delta_i;Z_j\right)Z_{j\widehat{G}}|Z_i\right]\right.\\ &\left.\times\widehat{E}\left[W_j - \widehat{E}\left(W_jX_j'|U_j\right)\left[\widehat{E}\left(X_iX_i'|U_i\right)\right]^{-1}X_j\eta\left(Z_i,\delta_i;Z_j\right)Z_{j\widehat{G}}|Z_i\right]'\right\},\end{aligned}$$

$\widehat{E}(\cdot|\cdot)$ is the standard leave-one-out kernel estimator of a conditional expectation, $\widehat{\sigma}_{\widehat{G}}^2 = \sum_{i=1}^{n}\widehat{\varepsilon}_{i\widehat{G}}^2/n$, $\widehat{\varepsilon}_{i\widehat{G}} = Z_{i\widehat{G}} - X_i'\widehat{\alpha}(U_i) + W_i'\widehat{\beta}$ and $\widehat{G}$ is the Kaplan–Meier estimator.

Theorem 4 *Assume that* $\text{rank}(R) = l$*. Then under A1–A6 and the null hypothesis (8)*

$$W_n \xrightarrow{d} \chi^2(l).$$

An alternative method to test the null hypothesis (8) is to consider empirical likelihood, which is introduced by Owen (1988, 1990, 2001) as a nonparametric likelihood alternative to traditional likelihood-based methods for inference. To introduce empirical likelihood in the context of the model considered in this paper, let

$$\rho_i(\beta) = \max_{\pi_i} \prod_{i=1}^{n} n\pi_i \quad s.t.\ \pi_i \geq 0, \sum_{i=1}^{n} \pi_i = 1, \sum_{i=1}^{n} \pi_i \widetilde{W}_i \left(Z_{iG_0} - \widetilde{W}_i'\beta\right)$$

denote a profile empirical likelihood ratio function. Then, a profile empirical likelihood ratio statistic for (8) can be based on the following test statistic

$$\max_{\beta} -2\sum_{i=1}^{n} \log\left(\rho_i(\beta)\right) \quad \text{s.t.} \;\; R\beta = r. \tag{11}$$

The computation of (11) can be numerically difficult because it involves the computation of a saddlepoint. We suggest an alternative asymptotically equivalent but computationally simpler test statistic based on the constrained profile least squares estimator $\widehat{\beta}_0$. To be specific, we construct a profile empirical likelihood ratio statistic for the restricted profile residuals $Z_{i\widehat{G}} - \widetilde{W}_i'\widehat{\beta}_0$ $(i = 1, \ldots, n)$, which is in the same spirit as that of a standard score statistic in ordinary likelihood inference, but instead of using the information matrix as the metric we use the empirical likelihood function. The resulting test statistic is

$$\text{EL}_n = 2\sum_{i=1}^{n} \log\left(1 + \widehat{\lambda}'\widetilde{W}_i\left(Z_{i\widehat{G}} - \widetilde{W}_i'\widehat{\beta}_0\right)\right), \tag{12}$$

where $\widehat{\lambda}$ is the solution to

$$0 = \sum_{i=1}^{n} \frac{\widetilde{W}_i\left(Z_{i\widehat{G}} - \widetilde{W}_i'\widehat{\beta}_0\right)}{1 + \widehat{\lambda}'\widetilde{W}_i\left(Z_{i\widehat{G}} - \widetilde{W}_i'\widehat{\beta}_0\right)}.$$

Theorem 5 *Assume that rank $(R) = l$. Then, under A1–A6 and under the null hypothesis (8)*

$$\text{EL}_n = \text{PL}_n + o_p(1).$$

An immediate consequence of the theorem is that the profile empirical likelihood ratio has the same nonstandard nonpivotal distribution as that of the profile least squares ratio. To obtain a standard distribution, we propose to adjust both EL_n and PL_n by a scale correction as in Rao and Scott (1981). Let S_n denote either EL_n or PL_n.

Proposition 6 *Assume that rank $(R) = l$. Then, under A1–A6 and the null hypothesis (8)*

$$\widehat{\varrho} S_n \xrightarrow{d} \chi^2(l),$$

where $\widehat{\varrho} = l/trace\left(\left[\widehat{\sigma}_G^2\left(R\widehat{\Gamma}^{-1}R'\right)\right]^{-1}\left(R\widehat{\Gamma}^{-1}\widehat{\Xi}_\beta\widehat{\Gamma}^{-1}R'\right)\right).$

3.2 Nonparametric component

It is also of interest to test whether the unknown varying-coefficients $\alpha_0(U)$'s can be modelled as parametric functions, that is

$$H_0 : \alpha_{0j}(U) = \alpha_j(U, \varsigma_0) \quad j = 1, \ldots, p \tag{13}$$

for some unknown parameter vector ς_0. Examples of (13) include the hypothesis of no significance of the covariate X_j $(j = 1, \ldots, p)$, e.g. $\alpha_{0j}(U, \varsigma_0) = 0$, or the hypothesis of homogeneity (constancy) of one or more of the varying coefficients, e.g. $\alpha_{0j}(U, \varsigma_0) = \alpha_j$.

Fan and Huang (2005) showed that the Wilks phenomenon holds for the profile least squares ratio statistic PL_n for (13). The key observation is that because the unknown parametric components can be estimated at the $n^{1/2}$ rate, they can be assumed known for nonparametric inferences. The same observation applies also to the Kaplan–Meier estimator for G_0, which, therefore, implies that inferences on $\alpha_0(U)$ can be based on the nonparametric varying coefficient model

$$Z^*_{iG_0} = X_i'\alpha_0(U_i) + \varepsilon_{iG_0} \tag{14}$$

where $Z^*_{iG_0} = Z_{iG_0} - W_i'\beta_0$. The general results of Fan et al. (2001) imply that the Wilks phenomenon holds for the resulting profile least squares ratio. As an illustration, we consider the same test of homogeneity as considered by Fan and Huang (2005), that is

$$H_0 : \alpha_{0j}(U, \varsigma_0) = \alpha_j \quad j = 1, \ldots, p. \tag{15}$$

As in Sect. 3.1 let RSS_0 and RSS_1 denote the restricted and unrestricted residual sum of squares and let

$$\mathrm{PL}_n^\alpha = \frac{n}{2}\frac{\mathrm{RSS}_0 - \mathrm{RSS}_1}{\mathrm{RSS}_1}$$

denote the profile least squares ratio statistic.

Theorem 7 *Under A1–A6 with $nh^{3/2} \to \infty$ and under the null hypothesis (15)*

$$r_K PL_n^\alpha \overset{a}{\sim} \chi^2(p_K),$$

where

$$r_K = \frac{K(0) - \frac{1}{2}\int K^2(v)\,\mathrm{d}v}{\int \left(K(v) - \frac{1}{2}K * K(v)\right)^2 \mathrm{d}v},$$
$$p_K = r_K \frac{p\,|\mathcal{U}|}{h}\left(K(0) - \frac{1}{2}\int K^2(v)\,\mathrm{d}v\right).$$

The distribution of $r_K \mathrm{PL}_n^\alpha$ can be approximated either by

$$\mathrm{PL}_n^{\alpha,S} := \frac{\left(r_K \mathrm{PL}_n^\alpha - p_K\right)}{\left(2p_K\right)^{1/2}} \xrightarrow{d} N\left(0,1\right) \tag{16}$$

or by the bootstrap. The bootstrap approximation can be based on the following steps:

1. generate $Y_i^* = X_i'\widehat{\alpha}\left(U_i\right) + W_i'\widehat{\beta} + \varepsilon_i^*$ where ε_i^* are randomly sampled from the centred unrestricted residuals $\widehat{\varepsilon}_{i\widehat{G}}$,
2. generate the bootstrap censoring indicator δ_i^* as a Bernoulli random variate with $\Pr\left(\delta_i^* = 1\right) = 1 - \widehat{G}\left(Y_i^*\right)$ and $\widehat{G}$ is the Kaplan–Meier estimator of G_0,
3. generate the bootstrap censoring variable C_i^* as follows: if Y_i^* and $\delta_i^* = 1$, C_i^* is taken from $\widehat{G}$ restricted to $\left[Y_i^*, \infty\right)$, otherwise C_i^* is taken from $\widehat{G}$ restricted to $\left[0, Y_i^*\right)$,
4. re-estimate the parameters of the model using the bootstrap sample $\left(Z_i^*, \delta_i^*, X_i, W_i, U_i\right)_{i=1}^n$ with the same bandwidth as that used in step 1 and with the parameters under the null hypothesis replaced by their original estimates,
5. compute RSS_1^* and RSS_0^* and hence the bootstrap analogue $\mathrm{PL}_n^{\alpha *}$ of PL_n^α,
6. repeat steps 1–5 B times.

Let $\Pr^*$ denote probability conditional on the original sample; the following theorem establishes the consistency of the proposed bootstrap procedure.

Theorem 8 *Under A1–A6 with $nh^{3/2} \to \infty$ and under the null hypothesis* (15)

$$\sup_{c \in \mathbb{R}_+} \left| \overset{*}{\Pr}\left(\mathrm{PL}_n^{\alpha *} \geq c\right) - \Pr\left(\mathrm{PL}_n^\alpha \geq c\right) \right| \xrightarrow{p} 0.$$

Note that it is also possible to construct an empirical likelihood ratio statistic for (15) using for example the sieve empirical likelihood approach of Fan and Zhang (2004). As with the profile least squares ratio, the key observation is that both β_0 and G_0 can be considered as if they were known for nonparametric inferences. However, the calculation of the resulting sieve empirical likelihood ratio statistic is demanding and potentially unstable because it involves the computation of n saddlepoints. Thus, for practical purposes, the profile least squares ratio test statistic PL_n^α seems a more viable option.

4 Specification analysis

In this section, we consider the important issue of testing whether the varying coefficient partially linear specification of (1) is correct. Note that under (4)

$$H_0 : E\left(\varepsilon_{G_0} | U, X, W\right) = E\left(Z_{G_0} - X'\alpha_0\left(U\right) - W'\beta_0 | U, X, W\right) = 0 \quad a.s., \tag{17}$$

or equivalently

$$H_0 : E\left(\overline{\varepsilon}_{G_0} | U, X, W\right) = E\left(\overline{Z}_{G_0} - \overline{W}'\beta_0 | U, X, W\right) = 0 \quad a.s. \tag{18}$$

where $\overline{Z}_{G_0} = Z_{G_0} - X'\Sigma(U)^{-1} E\left(XZ_{G_0}|U\right)$ and $\overline{W} = W - \Omega(U)\Sigma(U)^{-1}X$.

It is well-known (Bierens 1982) that (17) and (18) are equivalent to

$$H_0 : E\left[\varepsilon_{G_0}\Psi(U, X, W; u, x, w)\right] = 0, \tag{19}$$

provided the linear span of the weight function $\Psi(U, X, W; u, x, w)$ is dense in the space of bounded measurable functions on the support of U, X, W. Two popular choices of $\Psi(\cdot)$ are the indicator function

$$I\left(\left[U, X', W'\right] \leq \left[u, x', w'\right]\right) = I(U \leq u)\prod_{j=1}^{d_x} I\left(X^{(j)} \leq x^{(j)}\right)\prod_{j=1}^{d_w} I\left(W^{(j)} \leq w^{(j)}\right)$$

(see e.g. Stute 1997) and the exponential function $\exp\left(\iota\left[U, X', W'\right]\left[u, x', w'\right]'\right)$, where $\iota = (-1)^{1/2}$ (see e.g. Bierens 1982). Other possible choices are discussed in Stinchcombe and White (1998).

In this paper, we use the same projection-based approach proposed by Escanciano (2006), which avoids the deficiencies associated with using the indicator function and exponential functions, while preserving their merits. To be specific, we consider

$$E\left[\varepsilon_{G_0} I\left(U \leq u, \theta'\left[X', W'\right]' \leq s\right)\right] = 0 \quad a.a.u, s, \theta \in \Pi, \tag{20}$$

where $\Pi = [-\infty, \infty]^2 \times \mathbb{S}^{k+p-1}$ and $\mathbb{S}^{k+p-1}$ is the unit sphere in $\mathbb{R}^{k+p}$, that is $\mathbb{S}^{k+p-1} = \left\{\theta \in \mathbb{R}^{k+p} : \|\theta\| = 1\right\}$.

The advantage of specifying $\Psi(U, X, W; u, x, w)$ as in (20) over the standard indicator function is apparent in its dimension reduction character, which implies that potentially high dimensional covariates X, W are not a problem as they would be with the standard indicator function, i.e. too many zeroes negatively affecting both size and power properties of any test statistic based on it. At the same time, it also avoids the choice of the integrating measure as well as potentially high-dimensional numerical integration required for example to compute Cramer von Mises type statistic based on the exponential function (see e.g. Bierens 1982). A more detailed discussion of the merits of (20), including also a comparison with the related approach of Stute and Zhu (2002) in the context of generalised linear models, can be found in Escanciano (2006).

A test statistic for the null hypothesis (18) can be constructed by considering a functional of the so-called projected marked empirical process

$$n^{1/2}\widehat{v}(u, s, \theta) = \frac{1}{n^{1/2}}\sum_{i=1}^{n}\widehat{\varepsilon}_{i\widehat{G}} I\left(U_i \leq u, \theta'\left[X_i', W_i'\right]' \leq s\right),$$

where $\widehat{\varepsilon}_{i\widehat{G}} = \widetilde{Z}_{i\widehat{G}} - \widetilde{W}_i'\widehat{\beta}$, $\widetilde{Z}_{i\widehat{G}} = Z_{i\widehat{G}} - S_i(U_i) Z_{i\widehat{G}}$, $\widetilde{W}_i = W_i - X_i S_i(U_i)'$ and $\widehat{\beta}$ is the profile least squares estimator from the previous section.

Let $F_\theta(u,s)$ and $F_{n,\theta}(u,s)$ denote, respectively, the distribution and empirical distribution of U and $\theta'\left[X', W'\right]'$, and let $\mathrm{d}\theta$ denote the uniform distribution on the sphere $\mathbb{S}^{k+p-1}$.

Assume that

A7 The product measure $F_{n,\theta}(u,s)\,\mathrm{d}\theta$ is absolutely continuous with respect to the Lebesgue measure in Π.

Theorem 9 *Under A1–A7*

$$n^{1/2}\widehat{v}(u,s,\theta) \Longrightarrow v_\infty(u,s,\theta) \quad in\ l^\infty(\Pi),$$

where $v_\infty(u,s,\theta)$ *is a centred Gaussian process with covariance function*

$$E\left[\sigma(u_1,s_1,\theta_1)\,\sigma(u_2,s_2,\theta_2)\right], \tag{21}$$

and

$$\begin{aligned}
\sigma(u,s,\theta) &= \varepsilon_{G_0} I\left(U \le u, \theta'\left[X', W'\right]' \le s\right) \\
&+E\left[\frac{X_2'\Sigma(U_2)^{-1}\eta(Z_1,\delta_1;Z_2)X_2 Z_{2G_0}\delta_2 I\left(U_2 \le u, \theta'\left[X_2', W_2'\right]' \le s\right)}{1-G_0(Z_2)}|U_1,Z_1,\delta_1,X_1\right] \\
&-E\left[X_2'\Sigma(U_2)^{-1}E\left[\frac{\eta(Z_1,\delta_1;Z)X_2Z_{2G_0}\delta_2}{1-G_0(Z_2)}|U_1\right]I\left(U_2\le u,\theta'\left[X_2',W_2'\right]'\le s\right)|U_1,Z_1,\delta_1,X_1\right] \\
&-\left(W_1-\Omega(U_1)\Sigma(U_1)^{-1}X_1\right)'\left\{\varepsilon_{1G_0}+E\left[\frac{\eta(Z_1,\delta_1;Z_2)}{1-G_0(Z_2)}Z_{2G_0}\delta_2|U_1,Z_1,X_1,W_1\right]\right\}\Gamma^{-1} \\
&\times E\left[W_1-\Omega(U_1)\Sigma(U_1)^{-1}X_1 I\left(U_1\le u,\theta'\left[X_1',W_1'\right]'\le s\right)\right] \\
&-X_1'\Sigma(U_1)^{-1}X_1\left(Z_{1G_0}-W_1'\beta_0\right)F_\theta\left(\theta'\left[X',W'\right]'|U_1\right)I(U_1\le u) \\
&+E\left[X_1'\alpha(U_1)I\left(U_1\le u,\theta'\left[X_1',W_1'\right]'\le s\right)\right].
\end{aligned} \tag{22}$$

Given the result of Theorem 9, we can use a Cramer von Mises type of functional to construct a test statistic for the null hypothesis H_0 that is

$$\mathrm{CM}_n = n\int_\Pi \widehat{v}(u,s,\theta)^2\,\mathrm{d}F_{n,\theta}(u,s)\,\mathrm{d}\theta. \tag{23}$$

A straightforward application of the continuous mapping theorem gives the following:

Corollary 10 *Under A1–A7 and under the null hypothesis* (18)

$$\mathrm{CM}_n \xrightarrow{d} \int_\Pi v_\infty(u,s,\theta)^2\,\mathrm{d}F_\theta(u,s)\,\mathrm{d}\theta. \tag{24}$$

It is easy to see that under the global alternative hypothesis

$$
\begin{aligned}
&H_{1g}:\Pr\left(E\left(Z_{G_0}-X'\alpha\left(U\right)-W'\beta|U,X,W\right)\neq 0\right)>0\quad\forall\beta,\alpha\left(U\right)\\
&\frac{\mathrm{CM}_n}{n}\overset{d}{\to}\int_{\Pi}E\left[\left(Z_{G_0}-X'\alpha\left(U\right)-W'\beta\right)\right.\\
&\quad\left.\times I\left(U\leq u,\theta'\left[X',W'\right]'\leq s\right)\right]^2\mathrm{d}F_\theta\left(u,s\right)\mathrm{d}\theta>0,
\end{aligned}
$$

which implies the consistency of the test based on (23). Also local alternatives of the form

$$
H_{1a}:E\left(\varepsilon_{G_0}|U,X,W\right)=\frac{\gamma\left(U,X,W\right)}{n^{1/2}} \tag{25}
$$

for some known real-valued function can be detected by (23) as the following theorem shows. Let

$$
\overline{\gamma}=\Gamma^{-1}E\left[\left(W-\Omega\left(U\right)\Sigma^{-1}\left(U\right)X\right)\left(\gamma\left(U,X,W\right)-E\left[\gamma\left(U,X,W\right)|U\right]\right)\right],
$$

and assume that

A8 The function $E\left[\gamma\left(U,X,W\right)|U\right]$ is Lipschitz and $E\left[\gamma\left(U,X,W\right)\right]=0$.

Theorem 11 *Under A1–A8 and under the alternative hypothesis* (25)

$$
\mathrm{CM}_n\overset{d}{\to}\int_{\Pi}\left[v_\infty\left(u,s,\theta\right)+\gamma^*\left(u,s,\theta\right)\right]^2\mathrm{d}F_\theta\left(u,s\right)\mathrm{d}\theta,
$$

where

$$
\begin{aligned}
\gamma^*\left(u,s,\theta\right)=&E\left[\left(\gamma\left(U,X,W\right)-E\left[\gamma\left(U,X,W\right)|U\right]\right)I\left(U\leq u,\theta'\left[X',W'\right]'\leq s\right)\right]\\
&-E\left[\left(W-\Omega\left(U\right)\Sigma^{-1}\left(U\right)X\right)'I\left(U\leq u,\theta'\left[X',W'\right]'\right)\right]\overline{\gamma}.
\end{aligned}
$$

It is important to note that the asymptotic distribution (24) is non standard and, more importantly, nonpivotal. This is a well-known problem for the nonsmooth approach to specification testing and it is a direct consequence of the estimation of unknown parameters, which affects the covariance kernel of the limiting Gaussian process—see e.g. Durbin (1973)'s seminal contribution on the effect of estimated parameters on the distribution of the classical parametric empirical process. However, this distribution can be easily simulated using the same resampling technique used for example by Su and Wei (1991), Delgado et al. (2003) and Zhu and Ng (2003).

Let $\{\xi_i\}_{i=1}^n$ denote a random sample from the distribution of the random variable ξ with zero mean and unit variance that is independent from U,X,W, and let

$$
\sigma_i^*\left(u,s,\theta\right)=\widehat{\sigma}_i\left(u,s,\theta\right)\xi_i \tag{26}
$$

where $\widehat{\sigma}(\cdot)$ is as in (22) with $\varepsilon_{G_0 i}$, β_0, α_0 and $E(\cdot)$ replaced by consistent estimators, while $\int_{\mathbb{S}^{k+p-1}} \mathrm{d}\theta$ can be easily simulated by drawing say M samples from $k+p$ standard normal variates scaled by their norm, see e.g. Marsaglia (1972). For example, the last term in (22) is

$$\frac{1}{nM}\sum_{i=1}^{n}\sum_{j=1}^{M}\left[X_{1i}'\widehat{\alpha}(U_{1i})\, I\left(U_{1i} \le u, \theta_j'\left[X_{1i}', W_{1i}'\right]' \le s\right)\right].$$

Note that conditionally on η_i $\sum_{i=1}^{n} \sigma_i^*(u, s, \theta)/n^{1/2}$ has 0 mean and covariance structure

$$\sum_{i=1}^{n} \widehat{\sigma}_i(u_1, s_1, \theta_1)\,\widehat{\sigma}_i(u_2, s_2, \theta_2)/n;$$

by randomly sampling from ξ we can obtain the simulated version of CM_n, that is

$$\mathrm{CM}_n^* = n\int_{\Pi} \widehat{\nu}^*(u, s, \theta)^2\, \mathrm{d}F_{n,\theta}(u, s)\, \mathrm{d}\theta. \tag{27}$$

The following theorem shows that the proposed resampling method consistently estimates the distribution CM_n. Assume that

A9 $|\xi| \le C < \infty$ $a.s.$

Theorem 12 *Under A1–A7 and A9*

$$\sup_{c\in\mathbb{R}_+}\left|\overset{*}{\Pr}\left(\mathrm{CM}_n^* \ge c\right) - \Pr\left(\mathrm{CM}_n \ge c\right)\right| \overset{p}{\to} 0.$$

5 Monte Carlo evidence

In this section, we use simulations to assess the finite sample properties of the estimators and test statistics introduced in the previous sections. The model we consider is similar to the one considered in Fan and Huang (2005), namely

$$Y_i = \cos(\pi U_i) + X_i'\left[\sin(6\pi U_i), U_i^2\right]' + W_i'\beta_0 + \varepsilon_i, \tag{28}$$

where $X_i = [X_{1i}, X_{2i}]' \sim N(0, \Sigma)$, Σ is a symmetric matrix with 1 on the main diagonal and 1/3 on the off-diagonal, $U_i \sim U[0, 1]$ where $U[0, 1]$ is the uniform distribution on $[0, 1]$, $W_i = [W_{1i}, W_{2i}]'$ and ε_i are independent $N(0, 1)$, $Z_{3i} \sim B(0.4)$ where $B(0.4)$ is the binomial distribution with 0.4 probability, $\beta_0 = [2, 1/2, 0]'$, the censoring variable $C_i \sim 5U[-1, 1] + c$ where c is a constant chosen so as to satisfy an (approximate) fixed censoring proportion.

One important practical aspect of the results of this paper concerns the choice and dependence of the methods proposed in this paper on the bandwidth h. Bearing in

Table 1 Finite sample bias, standard error and RMSE

	$h^{cv}/2$			$h^{cv\,a}$			$3h^{cv}$		
	Bias	Std.err.	RMSE	Bias	Std.err.	RMSE	Bias	Std.err.	RMSE
$n = 100$									
$\widehat{\beta}_1^N$	−0.327	0.143	0.356	−0.336	0.137	0.362	−0.346	0.130	0.369
$\widehat{\beta}_1$	−0.122	0.191	0.226	−0.141	0.186	0.241	−0.146	0.178	0.230
$\widehat{\beta}_2^N$	−0.120	0.155	0.196	−0.125	0.150	0.195	−0.132	0.142	0.193
$\widehat{\beta}_2$	−0.035	0.193	0.196	−0.041	0.190	0.194	−0.048	0.180	0.186
$n = 400$									
$\widehat{\beta}_1^N$	−0.320	0.070	0.327	−0.326	0.068	0.333	−0.337	0.062	0.342
$\widehat{\beta}_1$	−0.070	0.102	0.123	−0.073	0.099	0.123	−0.079	0.089	0.119
$\widehat{\beta}_2^N$	−0.104	0.089	0.136	−0.110	0.085	0.139	−0.119	0.079	0.142
$\widehat{\beta}_2$	−0.020	0.110	0.111	−0.022	0.108	0.110	−0.025	0.105	0.107

[a] h^{cv} bandwidth chosen with least squares crossvalidation

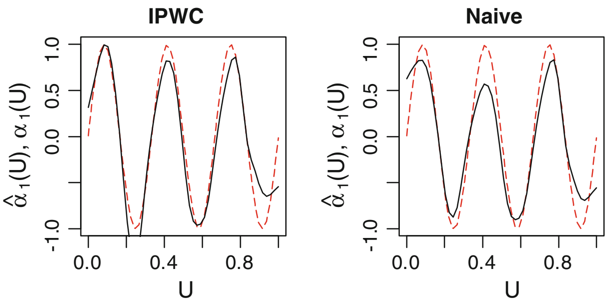

Fig. 1 Estimated (*solid line*) and true (*dashed line*) second varying coefficient using the estimated IPWC method (*left*) and the naive one (*right*) for $n = 400$

mind that no optimal bandwidth selection theory is available in the context of testing, we investigate this problem by first choosing h by standard crossvalidation and then considering alternative bandwidths. In the simulations, we consider two sample sizes, namely $n = 100$ and $n = 400$ and we fix the censoring level at approximately 30 per cent.

Table 1 reports the finite sample bias (bias), standard error (std. err.) and root mean squared error (RMSE) for two estimators for $\beta_0 = [\beta_{10}, \beta_{20}]'$: the naive estimator $\widehat{\beta}^N$ that ignores the censoring, and the proposed one $\widehat{\beta}$. The results are based on 1000 replications.

Figures 1 and 2 show the estimated varying coefficients $\widehat{\alpha}_1(U)$ and $\widehat{\alpha}_2(U)$, together with their true counterparts $\alpha_{10}(U) = \sin(6\pi U)$ and $\alpha_{20}(U) = U^2$ with IPWC and without it (i.e. the naive method) for $n = 400$ with a crossvalidated bandwidth. Figure 2 also shows the estimated varying coefficient with fixed bandwidth equal to 0.2 the length of the support of U.

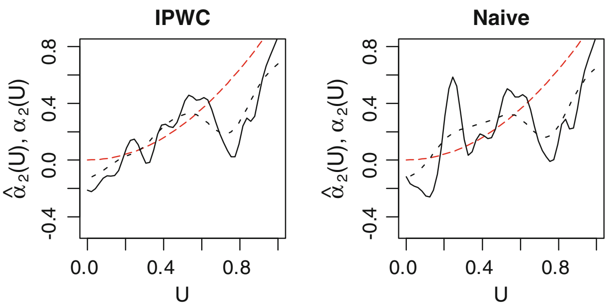

Fig. 2 Estimated with crossvalidated (*solid line*) and fixed (*dashed-dotted line*) bandwidths, and true (*dashed*) third varying coefficient using the estimated IPWC method (*left*) and the naive one (*right*) for $n = 400$

Table 2 Finite sample size and power for profile least squares ratio PL_n, profile empirical likelihood ratio EL_n and Wald statistic W_n

δ	$h^{cv}/2$				h^{cv} [a]				$3h^{cv}$			
	PL_n	EL_n	W_n	PL_n^*	PL_n	EL_n	W_n	PL_n^*	PL_n	EL_n	W_n	PL_n^*
$n = 100$												
0.0	0.068	0.063	0.058	0.053	0.079	0.066	0.060	0.055	0.084	0.071	0.064	0.58
0.2	0.122	0.149	0.108	0.125	0.141	0.209	0.122	0.136	0.152	0.224	0.137	0.150
0.4	0.214	0.323	0.233	0.300	0.341	0.434	0.324	0.355	0.371	0.440	0.368	0.380
0.8	0.621	0.715	0.616	0.644	0.676	0.762	0.644	0.650	0.700	0.781	0.700	0.702
1.2	0.917	0.940	0.876	0.930	0.985	1.00	0.906	0.976	1.00	1.00	0.995	1.00
$n = 400$												
0.0	0.061	0.056	0.057	0.052	0.071	0.060	0.056	0.053	0.078	0.069	0.060	0.055
0.2	0.165	0.200	0.140	0.154	0.186	0.231	0.147	0.165	0.203	0.242	0.187	0.193
0.4	0.335	0.412	0.400	0.412	0.435	0.499	0.418	0.424	0.490	0.521	0.480	0.399
0.8	0.876	0.947	0.819	0.997	0.929	0.999	0.932	0.905	0.942	1.00	0.951	0.964
1.2	0.995	1.00	0.934	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

[a] h^{cv} bandwidth chosen with least squares crossvalidation

Table 2 reports finite sample size and (size corrected) power for the three test statistics proposed in Sect. 3.1 as well as for a bootstrap version of the profile least squares ratio statistic based on

$$\mathrm{PL}_n^* = \frac{n}{2}\frac{\left(R\widehat{\beta}^* - R\widehat{\beta}\,\right)'\left(R\widehat{\Gamma}^{-1}R'\right)^{-1}\left(R\widehat{\beta}^* - R\widehat{\beta}\,\right)}{\mathrm{RSS}_1^*}, \tag{29}$$

where $\widehat{\Gamma}$ is defined in (10) $\mathrm{RSS}_1^* = \sum_{i=1}^n \left(Z_{i\widehat{G}}^* - X_i'\widehat{\alpha}^*(U_i) - W_i'\widehat{\beta}^*\right)^2$ and $Z_{i\widehat{G}}^*$ is generated using the same unrestricted bootstrap procedure as that described in Sect. 3.2. The following proposition shows that the proposed bootstrap procedure is consistent.

Table 3 Bootstrap p values for PL_n^α for constancy of coefficients

	$H_0^{(1)}$	$H_0^{(2)}$	$H_0^{(3)}$	$H_0^{(4)}$
$h^{0.25}$				
PL_n^α	8.077	21.24	17.43	1.051
p-value	0.442	0.002	0.006	0.831
h^{cv}				
PL_n^α	16.024	24.87	29.87	1.602
p-value	0.432	0.002	0.001	0.663

$H_0^{(4)}: \alpha_4(U) = \alpha_4 H_0^{(1)}: \alpha_4(U) = \beta_1, \alpha_5(U) = \beta_2$, $H_0^{(2)}: \alpha_2(U) = \alpha_2$, $H_0^{(3)}: \alpha_3(U) = \alpha_3$

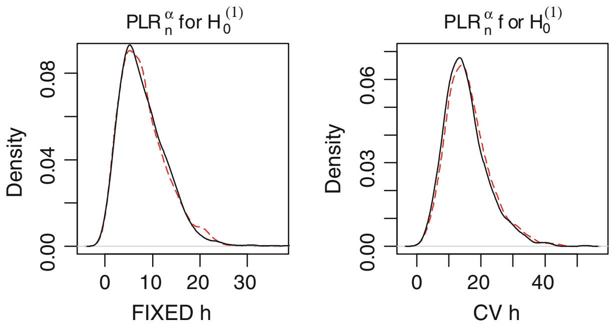

Fig. 3 Simulated distributions of the PLR (*solid line*) and its bootstrapped version (*dotted line*) for $H_0^{(1)}$ using fixed bandwidth (*left*) and crossvalidated bandwidth (*right*)

Proposition 13 *Under the same assumptions as those given in Proposition 3*

$$\sup_{c\in\mathbb{R}_+} \left| \overset{*}{\Pr}\left(\mathrm{PL}_n^* \geq c\right) - \Pr\left(\mathrm{PL}_n \geq c\right) \right| \xrightarrow{p} 0.$$

The null hypothesis in Table 2 is $H_0 : [\beta_2, \beta_3]' = [0.5, 0]'$ while the set of alternative hypotheses is $H_{1a} : [\beta_2, \beta_3]' = [0.5, 0]' + \delta$ where $\delta = [0.2, 0.4, 0.8, 1.0]$; note that the size and power are calculated at the 0.05 nominal significant level using 1000 replications with 500 bootstrap replications for the size, and 500 replications for the power calculations.

Table 3 reports the bootstrap p values for the profile least squares ratio for the null four hypotheses $H_0^{(1)} : \alpha_4(U) = \alpha_4 =: \beta_1, \alpha_5(U) = \beta_2$, and $H_0^{(j)} : \alpha_j(U) = \alpha_j (j = 2, 3, 4)$, that is the first hypothesis tests whether the last two coefficients in 28 are constant, while the second, third and fourth hypotheses test, respectively, whether the second, third and fourth coefficients are constant. The results are based on 1000 replications with p values calculated from 500 bootstrap replications with sample size $n = 100$.

Figures 3 and 4 show the distribution of the profile least squares ratio and its bootstrapped version for, respectively, $H_0^{(1)}$ and $H_0^{(2)}$ using two different bandwidths: one chosen by crossvalidation and one chosen as 0.25 of the length of the support of U.

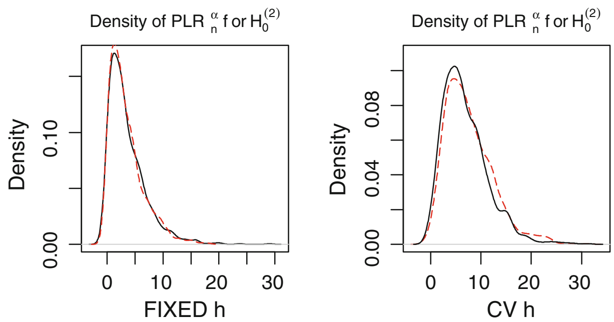

Fig. 4 Simulated distributions of the PLR (*solid line*) and its bootstrapped version (*dotted line*) for $H_0^{(2)}$ using fixed bandwidth (*left*) and crossvalidated bandwidth (*right*)

Table 4 Finite sample size and power for Cramer von Mises statistics CM_n and CM_n^o

δ	$h^{cv}/2$		h^{cv} [a]		$3h^{cv}$	
	CM_n	CM_n^o	CM_n	CM_n^o	CM_n	CM_n^o
$n = 100$						
0.00	0.055	0.040	0.059	0.040	0.066	0.044
0.40	0.270	0.145	0.271	0.148	0.278	0.150
0.80	0.534	0.205	0.545	0.206	0.553	0.208
1.20	0.780	0.299	0.781	0.296	0.798	0.300
1.60	0.800	0.360	0.799	0.359	0.802	0.368
$n = 400$						
0.00	0.054	0.043	0.054	0.042	0.058	0.045
0.40	0.335	0.221	0.340	0.218	0.344	0.210
0.80	0.612	0.287	0.621	0.286	0.628	0.282
1.20	0.791	0.404	0.802	0.410	0.806	0.432
1.60	0.910	0.495	0.912	0.490	0.911	0.493

[a] h^{cv} bandwidth chosen with least squares crossvalidation

Finally, Table 4 reports finite sample size and size corrected power for the Cramer von Mises type of statistic based on the projection approach [i.e. (23)] and of the same statistic without the projection on θ, that is

$$\mathrm{CM}_n^o = n \int_{\Pi} \widehat{\nu}\,(u, s, t)^2 \, \mathrm{d}F_n\,(u, s, t)\,, \tag{30}$$

where $\Pi = [-\infty, \infty]^{k+p+1}$ and $n^{1/2}\widehat{\nu}\,(u, s, t) = \sum_{i=1}^{n} \widehat{\varepsilon}_{i\widehat{G}} I\,(U_i \le u, X_i \le s, W_i \le t)/n^{1/2}$ is the standard marked empirical process. As in Table 2, the finite sample size and power are calculated at a 0.05 nominal significance level with the size calculated from 1000 replications and the power from 500. The critical values are computed from 1000 replications of (27) using the same two points distribution attaching probability masses $\left(5^{1/2}+1\right)/2\left(5^{1/2}\right)$ and $\left(5^{1/2}-1\right)/2\left(5^{1/2}\right)$ to the points $-\left(5^{1/2}-1\right)/2$ and $\left(5^{1/2}+1\right)/2$ as that used for example by Stute et al. (1998). The alternative hypothesis is parameterised as $\gamma\,(U, X, W) = \delta W^2$ for $\delta = [0.4, 0.8, 1.2, 1.6]$

The results of the simulations can be summarised as follows: first, Table 1 indicates that for the parametric components, the proposed estimator outperforms the naive one in terms of bias. It is less precise than the naive one, which is to be expected because of the IPCW, but still dominates in terms of RMSE. As the sample size increases, the bias of the proposed estimator decreases as opposed to that of the naive estimator which is still very substantial; at the same time, both estimators become more precise. More importantly, both bias and variance do not seem to significantly depend on the choice of the bandwidth, which confirms the simulation findings of Fan and Huang (2005) in the case of uncensored responses. On the other hand, the bandwidth choice is important for the estimation of the nonparametric components, as clearly illustrated by Fig. 2, where a smaller bandwidth than the one suggested by crossvalidation results in a better fit. Note, however, that as with the parametric estimators, both nonparametric estimators based on IPWC perform better than those based on the naive method, as indicated by the partial R^2 and mean squared error criterion that are, respectively, 0.543 and 1.021 against 0.451 and 1.122.

Second, the three test statistics for the parametric component based on the asymptotic χ^2 calibration have reasonably good finite sample size and power properties. Table 2 indicates that the Wald statistic has the smallest size distortion, whereas the empirical likelihood statistic has the largest power. As expected as the sample size increases, the three test statistics become more accurate and more powerful. Table 2 shows that the size distortion can be partially reduced by choosing a smaller bandwidth than the one chosen by crossvalidation, but the resulting test statistics seem to lose some power for alternatives closer to the null hypothesis. Table 2 also shows that the bootstrap approximation (29) yields a profile least squares ratio statistic with smaller size distortions. Further simulations results (available in the supplement) with a smaller bandwidth confirm these findings, while they also indicate that larger values of the bandwidth can have a negative effect on the size of the test statistics.

Third the profile least squares ratio for testing hypotheses on the nonparametric component performs well and correctly detect the constancy of the linear part of the model while strongly rejecting the constancy of the varying coefficients (Table 3). The same conclusions hold for the standardised version $\mathrm{PL}_n^{\alpha,S}$ defined in (16) using the normal approximation (see Table 3b in the supplement). Figures 3 and 4 show that the bootstrap provides a very good approximation to the distribution of the profile least squares ratio using both a fixed and a crossvalidated bandwidth. The normal approximation also works well (see the Q–Q plots reported in the supplement), but it appears to be more dependent on the choice of bandwidth, which could have a negative effect particularly for larger values of the bandwidth.

Finally, the Cramer von Mises type of statistic based on a projected marked empirical process has very good size and power properties. Compared to the standard Cramer von Mises statistic based on a marked empirical process, Table 4 indicates that the one based on the projection is less size distorted and has considerably higher power.

Taken together, these results seem to indicate that the proposed estimator and test statistics are characterised by good finite sample properties which compare favourably to standard alternatives. They also indicate that the choice of bandwidth is less crucial for inferences on the parametric components as long as it does not create excessive bias. In this respect, the bandwidth chosen by least squares cross validation provides a good initial value from which it is possible to obtain test statistics with improved accuracy by carefully choosing a smaller bandwidth. On the other hand, the choice of bandwidth is more important for estimation and inferences on the nonparametric components. However, thanks to the Wilks phenomenon and the bootstrap it is still possible to obtain reliable and accurate inferences.

6 Conclusion

In this paper, we consider the problem of estimation and inference of varying coefficient partially linear models when the responses are subject to random censoring. We propose to use profile least squares with IPCW to estimate the unknown parameters and suggest a number of statistics for testing linear hypotheses about both the parametric and nonparametric parameters. We show that the proposed estimation method yields an asymptotically normal estimator. We also show that as opposed to the uncensored case, the Wilks phenomenon does not hold for either the profile least squares ratio or the empirical likelihood ratio for testing hypotheses about the parametric component; however, it is still possible to obtain test statistics with an asymptotic Chi-squared calibration by adjusting the test statistics with an appropriate scale factor that can be consistently estimated. We consider the important issue of correct specification of the assumed varying coefficient partially linear structure and propose a Cramer von Mises type of statistic that does not suffer from the curse of dimensionality, nor requires multidimensional integration. We investigate the finite sample properties of the proposed estimator and test statistics with simulations. The results of the simulations are encouraging and suggest that profile least squares estimation combined with IPCW is a useful method to deal with semiparametric varying partially linear models with random censoring.

7 Appendix

Let C denote a generic constant, $c_n = (\log n/(nh))^{1/2} + h^2$ respectively. "CLT", "CMT", "LLN" stand, respectively, for "central limit theorem", "continuous mapping theorem" and (possibly uniform) "law of large numbers". Recall also that $\Sigma(U) = E\left(XX'|U\right)$ and $\Omega(U) = E\left(WX'|U\right)$.

7.1 Auxiliary Lemmas

Lemma 14 *Let $A(W, X, U)$ denote a generic possibly matrix-valued function of W and X such that $E[\|A(W, X, U)\|] < \infty$ and $E[A(W, X, u)|U = u]$ have Lipschitz continuous second derivatives in u, and let ϵ be a possibly vector-valued random variable such that $E[\|A(W, X, U)\epsilon\|] < \infty$. Then, under A1, A5 and A6*

$$\sup_{u\in\mathcal{U}}\left\|\frac{1}{n}\sum_{i=1}^{n}K_h\left(U_i-u\right)A\left(W_i,X_i,U_i\right)-f\left(u\right)E\left[A\left(W,X,u\right)|U=u\right]\right\|=O_p\left(c_n\right),$$

$$\sup_{u\in\mathcal{U}}\left\|\frac{1}{n}\sum_{i=1}^{n}K_h\left(U_i-u\right)A\left(W_i,X_i,U_i\right)\epsilon_i-f\left(u\right)E\left[A\left(W,X,u\right)\epsilon|U=u\right]\right\|=O_p\left(c_n\right).$$

Lemma 15 *Let $A\left(W,X,U\right)$ be as in Lemma 14, and assume that $E[\overline{A}(W,X,U)\overline{A}(W,X,U)']$ is nonsingular where $\overline{A}\left(W,X,U\right)=A\left(W,X,U\right)-E[A(W,X,U)X'|U]\Sigma(U)X$. Then, for $\widetilde{A}\left(W_i,X_i,U\right)$ as defined in 3*

$$\sup_{U_i\in\mathcal{U}}\left\|\left[\frac{1}{n}\sum_{i=1}^{n}\widetilde{A}\left(W_i,X_i,U_i\right)\widetilde{A}\left(W_i,X_i,U_i\right)'\right]^{-1}-\left\{E\left[\overline{A}\left(W,X,U\right)\overline{A}\left(W,X,U\right)'\right]\right\}^{-1}\right\|=o_p\left(1\right).$$

Lemma 16 *Let $A\left(W,X,U\right)$ be as in Lemma 14 and let ξ denote a possibly vector valued random variable such that either (i) $E\left(\xi|W,X,U\right)=0$ a.s. or (ii) $\xi=\xi_1\xi_2$ and $|\xi_1|=O_p\left(n^{-1/2}\right)$ uniformly in its support $\mathcal{S}_{\xi_1}$. Then,*

$$\sup_{U_i\in\mathcal{U}}\left\|\frac{1}{n^{1/2}}\sum_{i=1}^{n}\widetilde{A}\left(W_i,X_i,U_i\right)\widetilde{\xi}_i-\frac{1}{n^{1/2}}\sum_{i=1}^{n}\widetilde{A}\left(W_i,X_i,U_i\right)\xi_i\right\|=O_p\left(c_n\right).$$

Lemma 17 *Let $A\left(W,X,U\right)$ be as in Lemma 14 and let ξ_1 and ξ_2 denote two possibly vector-valued random variables such that (ξ_2) $\|\xi_2\|=O_p\left(n^{-1/2}\right)$. Then,*

$$\sup_{U_i\in\mathcal{U}}\left\|\frac{1}{n}\sum_{i=1}^{n}\xi_2'\widetilde{A}\left(W_i,X_i,U_i\right)\widetilde{A}\left(W_i,X_i,U_i\right)'\xi_2\right\|=O_p\left(n^{-1}\right),$$

$$\sup_{U_{ii}\in\mathcal{U}}\left\|\frac{1}{n}\sum_{i=1}^{n}\xi_2'\widetilde{A}\left(W_i,X_i,U_i\right)\widetilde{\xi}_{1i}\right\|=O_p\left(n^{-1/2}\right),$$

$$\sup_{U_i\in\mathcal{U}}\left\|\frac{1}{n}\sum_{i=1}^{n}\xi_2'\widetilde{A}\left(W_i,X_i,U_i\right)\widetilde{X_i}\xi_{1i}\right\|=O_p\left(n^{-1/2}c_n\right).$$

7.2 Proof of the main results

Proof of Theorem 1 By Lemma 15

$$n^{1/2}\left(\widehat{\beta}-\beta_0\right)=\Gamma^{-1}\frac{1}{n^{1/2}}\sum_{i=1}^{n}\widetilde{W}_i\left(\widetilde{\varepsilon}_{iG_0}+\frac{\widehat{G}\left(Z_i\right)-G_0\left(Z_i\right)}{\left(1-G_0\left(\left(Z_i\right)\right)\right)}\widetilde{Z}_{iG_0}\delta_i\right)$$
$$-\frac{1}{n^{1/2}}\sum_{i=1}^{n}\widetilde{W}_i\widetilde{X}_i'\alpha_0\left(U_i\right)+o_p\left(1\right):=T_{1n}+T_{2n},$$

uniformly in $\mathcal{U}$. We first consider T_{1n} and note that by Lemma 16 and

$$\sup_{0\le z\le \max_i Z_i}\left|\frac{\widehat{G}(z)-G_0(z)}{1-G_0(z)}\right| = O_p\left(n^{-1/2}\right) \tag{31}$$

Zhou (1991)

$$\sum_{i=1}^n \widetilde{W}_i\widetilde{\varepsilon}_{iG_0} = \sum_{i=1}^n \widetilde{W}_i\varepsilon_{iG_0} + O_p(c_n),$$
$$\frac{1}{n^{1/2}}\sum_{i=1}^n \widetilde{W}_i\frac{\widehat{G}(Z_i)-G_0(Z_i)}{(1-G_0(Z_i))}\widetilde{Z}_{iG_0}\delta_i = \frac{1}{n^{1/2}}\sum_{i=1}^n \widetilde{W}_i\frac{\widehat{G}(Z_i)-G_0(Z_i)}{(1-G_0(Z_i))}Z_{iG_0}\delta_i + o_p(1),$$

uniformly in $\mathcal{U}$. Next by Lemma 14 and LLN $\|T_{2n}\| = O_p\left(n^{1/2}c_n^2\right)$ uniformly in $\mathcal{U}$, and thus

$$n^{1/2}\left(\widehat{\beta}-\beta_0\right) = \frac{1}{n^{1/2}}\left(\sum_{i=1}^n \widetilde{W}_i\widetilde{W}_i'/n\right)^{-1}\sum_{i=1}^n \widetilde{W}_i\left(\varepsilon_{iG_0} + \frac{\widehat{G}(Z_i)-G_0(Z_i)}{(1-G_0(Z_i))}Z_{iG_0}\delta_i\right). \tag{32}$$

By Lo and Singh (1986)'s representation of the Kaplan–Meier estimator, we have that

$$T_{1n} = T_{1n}^U + R_n$$

where

$$T_{1n}^U = \frac{1}{n^{3/2}}\sum_{i<j} h\left(U_i, X_i, W_i, Z_i, \delta_i, U_j, X_j, W_j, Z_j, \delta_j\right),$$

with

$$\begin{aligned} &h\left(U_i, X_i, W_i, Z_i, \delta_i, U_j, X_j, W_j, Z_j, \delta_j\right)\\ &= \left(W_i - \Omega(U_i)\Sigma(U_i)^{-1}X_i\right)\left(\varepsilon_{iG_0} + \frac{\eta\left(Z_j,\delta_j;Z_i\right)}{(1-G_0(Z_i))^2}Z_i\delta_i\right)\\ &\quad + \left(W_j - \Omega\left(U_j\right)\Sigma\left(U_j\right)^{-1}X_j\right)\left(\varepsilon_{jG_0} + \frac{\eta\left(Z_i,\delta_i;Z_j\right)}{(1-G_0(Z_i))^2}Z_j\delta_j\right), \end{aligned}$$

and $\|R_n\| = o_p(1)$ (see the supplement). Clearly $E[h(\cdot)] = 0$, $E[\eta(Z_1, \delta_1; Z_2)|Z_2] = 0$, $E[\eta(Z_2, \delta_2; Z_1)|Z_1] = 0$; furthermore for $j, k = 1, 2$ and $j \neq k$

$$E\left[\eta^2\left(Z_j, \delta_j; Z_k\right)|Z_k\right] = -\left(1 - G_0(Z_k)\right)^2 \int_0^{Z_k} \frac{d\overline{H}_0(s)}{[H_0(s)]^2},$$

hence, for any $k \times 1$ vector ξ, A3 implies that $E[\xi' h(U_1, X_1, W_1, Z_1, \delta_1, U_2, X_2, W_2, Z_2, \delta_2)]^2 < \infty$.

Let

$$\begin{aligned}
&h_1(U_1, X_1, W_1, Z_1, \delta_1) := E[(U_1, X_1, W_1, Z_1, \delta_1, U_2, X_2, W_2, Z_2, \delta_2)|U_1, X_1, W_1, Z_1, \delta_1]\\
&= \left[W_1 - E\left(W_1 X_1'|U_1\right)\left[E\left(W_1 W_1'|U_1\right)\right]^{-1}\right] X_1 \varepsilon_{1G_0}\\
&\quad + E\left[W_2 - E\left(W_2 X_2'|U_2\right)\left[E\left(W_2 W_2'|U_2\right)\right]^{-1} X_2 \eta(Z_1, \delta_1; Z_2) Z_{2G_0}|Z_1, \delta_1\right],
\end{aligned}$$

and note that

$$\begin{aligned}
\operatorname{Var}\left(\xi' h(U_1, X_1, W_1, Z_1, \delta_1, U_2, X_2, W_2, Z_2, \delta_2)\right) &= E\left(\xi' h_1(U_1, X_1, W_1, Z_1, \delta_1)\right)\\
&= \xi' \Xi \xi.
\end{aligned}$$

The conclusion follows by a standard CLT for second-order U-statistics (Serfling 1980) and the Cramer–Wold device. □

Proof of Theorem 2 See the supplement. □

Proof of Proposition 3 We first show that $\left|\mathrm{RSS}_1/n - \sigma^2_{G_0}\right| = o_p(1)$. By the same arguments as those used in the proof of Theorem 5.1 of Fan and Huang (2005)

$$\begin{aligned}
\frac{\mathrm{RSS}_1}{n} &= \frac{1}{n}\sum_{i=1}^n \left(\widetilde{\varepsilon}_{iG_0}\right)^2 + \frac{1}{n}\sum_{i=1}^n \left[\frac{\widehat{G}(Z_i) - G_0(Z_i)}{(1 - G_0((Z_i)))}\widetilde{Z}_{iG_0}\delta_i\right]^2\\
&\quad + \frac{2}{n}\sum_{i=1}^n \frac{\widehat{G}(Z_i) - G_0(Z_i)}{(1 - G_0((Z_i)))}\widetilde{Z}_{iG_0}\delta_i \widetilde{\varepsilon}_{iG_0}\\
&\quad + \frac{2}{n}\sum_{i=1}^n \frac{\widehat{G}(Z_i) - G_0(Z_i)}{(1 - G_0((Z_i)))}\widetilde{X}_i' \alpha_0(U_i)\\
&\quad + \frac{2}{n}\sum_{i=1}^n \frac{\widehat{G}(Z_i) - G_0(Z_i)}{(1 - G_0((Z_i)))}\widetilde{Z}_{iG_0}\delta_i \widetilde{W}_i'\left(\widehat{\beta} - \beta_0\right) + o_p(1):\\
&= \sum_{j=1}^5 T_{2jn}.
\end{aligned}$$

By LLN $\left|T_{21n} - \sigma^2_{G_0}\right| = o_p(1)$, while by Lemma 17 $|T_{22n}| = O_p\left(n^{-1}\right)$, $|T_{23n}| = O_p\left(n^{-1/2}\right)$, $|T_{24n}| = O_p(c_n)$ and $|T_{25n}| = O_p\left(n^{-1/2}c_n\right)$. Note that $\widehat{\varepsilon}_{0i\widehat{G}} = \widehat{\varepsilon}_{i\widehat{G}} + \widetilde{W}_i'\left(\widehat{\beta}_0 - \widehat{\beta}\right)$ hence by LLN and simple algebra it follows that

$$\left| \frac{\mathrm{RSS}_0}{n} - \sigma^2_{G_0} - \frac{\left(\widehat{\beta}_0 - \widehat{\beta}\right)'}{n} \sum_{i=1}^{n} \widetilde{W}_i \widetilde{W}_i' \left(\widehat{\beta}_0 - \widehat{\beta}\right) \right| = o_p(1),$$

so that the profile least squares ratio

$$\begin{aligned} \mathrm{PL}_n &= n \frac{\left(\widehat{\beta}_0 - \widehat{\beta}\right)'}{\sigma^2_{G_0}} \Gamma \left(\widehat{\beta}_0 - \widehat{\beta}\right) + o_p(1) \\ &= \left(R\widehat{\beta} - r\right)' \left(\sigma^2_{G_0} R \Gamma^{-1} R'\right)^{-1} \left(R\widehat{\beta} - r\right) + o_p(1), \end{aligned} \tag{33}$$

and the conclusion follows by the result of Kent (1982). □

Proof of Theorem 4 By LLN, the consistency of $\widehat{\alpha}(U_i)$ and $\widehat{\beta}$, 31 and iterated expectations it can be shown that

$$\left\| \frac{\widehat{\sigma}^2_{\widehat{G}}}{n} \sum_{i=1}^{n} W_i W_i' - \sigma^2_{G_0} E\left(W_i W_i'\right) \right\| = o_p(1).$$

Lemma 14, LLN and triangle inequality imply that

$$\begin{aligned} &\left\| \frac{\widehat{\sigma}^2_{\widehat{G}}}{n} \sum_{i=1}^{n} \widehat{E}\left(W_i X_i' | U_i\right) \left[\widehat{E}\left(X_i X_i' | U_i\right)\right]^{-1} \widehat{E}\left(X_i W_i' | U_i\right) \right. \\ &\left. -\sigma^2_{G_0} E\left[\Omega(U_i) \Sigma^{-1}(U_i) \Omega(U_i)'\right] \right\| = o_p(1) \end{aligned}$$

uniformly in $\mathcal{U}$, whereas by repeated applications of Lemmas 14 and 15

$$\begin{aligned} &\left\| \widehat{E}\left[W_j - \widehat{E}\left(W_j X_j' | U_j\right) \left[\widehat{E}\left(X_i X_i' | U_i\right)\right]^{-1} X_j \eta\left(Z_i, \delta_i; Z_j\right) Z_{j\widehat{G}} | Z_i\right] \right. \\ &\left. - E\left[W_2 - \Omega(U_2) \Sigma(U_2)^{-1} X_2 \eta(Z_1, \delta_1; Z_2) Z_{2G_0} | Z_1, \delta_1\right] \right\| = o_p(1), \end{aligned}$$

and similarly for all the other terms appearing in $\widehat{\Xi}_\beta$; hence by CMT $\left\| \widehat{\Xi}_\beta - \Xi_\beta \right\| = o_p(1)$ and the conclusion follows by CLT and CMT. □

Proof of Theorem 5 By the triangle inequality, Borel Cantelli lemma, CMT and results of Masry (1996), it is possible to show that $\max_i \left\| \widetilde{W}_i \widehat{\widetilde{\varepsilon}}_{i\widehat{G}} \right\| = o_p\left(n^{1/2}\right)$. Note that the constrained estimator $\widehat{\beta}_0$ can be expressed as

$$\widehat{\beta}_0 - \beta_0 = \Gamma^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} \widetilde{W}_i \varepsilon_{0i\widehat{G}} \right) - \Gamma^{-1} R' \left(R \Gamma^{-1} R'\right)^{-1} R\left(\widehat{\beta} - \beta_0\right) + o_p\left(n^{-1/2}\right),$$

and by Lemma 16

$$\frac{1}{n} \sum_{i=1}^{n} \widetilde{W}_i \widehat{\widetilde{\varepsilon}}_{0i\widehat{G}} = R' \left(R \Gamma^{-1} R'\right)^{-1} R\left(\widehat{\beta} - \beta_0\right) + o_p\left(n^{-1/2}\right). \tag{34}$$

Then (31), LLN, Lemmas 14, 16, 17 and the consistency of $\widehat{\beta}_0$ can be used to show that

$$\left\| \sum_{i=1}^{n} \widetilde{\varepsilon}^2_{0iG_0} \widetilde{W}_i \widetilde{W}'_i / n - \sigma^2_{G_0} \Gamma \right\| = o_p(1).$$

Using Owen (1990)'s arguments, it is possible to show that $\widehat{\lambda} = \Gamma^{-1} \sum_{i=1}^{n} \widetilde{W}_i \widehat{\widetilde{\varepsilon}}_{0i\widehat{G}} / n + o_p\left(n^{-1/2}\right)$. The conclusion follows by a Taylor expansion, (34) and CMT. □

Proof of Proposition 6 See the supplement. □

Proof of Theorem 7 Following Fan and Huang (2005), it suffices to show that $\left|\mathrm{RSS}_j - \mathrm{RSS}_j^0\right| = o_p(n)$ $(j = 0, 1)$ where RSS_j^0 is as RSS_j but with G_0 and β_0 assumed known. Note that

$$\begin{aligned}\left| \frac{\mathrm{RSS}_0}{n} - \frac{\mathrm{RSS}_0^0}{n} \right| &\le \frac{2}{n} \sum_{i=1}^{n} \left(\widetilde{Z}_i \delta_i \frac{\widehat{G}(Z_i) - G_0(Z_i)}{(1 - G_0(Z_i))^2} \right)^2 \\ &\quad + \frac{2}{n} \left(\widehat{\beta}_0 - \beta_0\right)' \sum_{i=1}^{n} \widetilde{W}_i \widetilde{W}'_i \left(\widehat{\beta}_0 - \beta_0\right) \\ &\quad + 2 \left[\sum_{i=1}^{n} \left(\frac{\widehat{\varepsilon}_{i\widehat{G}}}{n} \right)^2 \right]^{1/2} \left[\sum_{i=1}^{n} \left(\frac{\widehat{\varepsilon}_{i\widehat{G}} - \varepsilon_{iG_0}}{n} \right)^2 \right]^{1/2} \\ &\quad + O_p\left(c_n^2\right) = O_p\left(\frac{1}{n}\right) + O_p\left(c_n^2\right),\end{aligned}$$

and similarly for RSS_1. Hence

$$\mathrm{PL}_n^{\alpha} = \frac{n}{2} \frac{\mathrm{RSS}_0^0 - \mathrm{RSS}_1^0}{\mathrm{RSS}_1^0} + o_p(1),$$

and the result follows by the same arguments used by Fan et al. (2001). □

Proof of Theorem 8 By the results of Lo and Singh (1986) (see also Akritas 1986), the bootstrap Kaplan–Meier estimator $\widehat{G}^*$ is $n^{1/2}$ consistent. This together with the bootstrap LLN and CLT of Bickel and Freedman (1981) can be used to show that $n^{1/2}\left(\widehat{\beta}^* - \widehat{\beta}\right) = O_{p^*}(1)$ and $(nh)^{1/2}\left(\widehat{\alpha}^*(u) - \widehat{\alpha}(u)\right) = O_{p^*}(1)$ except if the original sample is on a set with probability tending to 0 as $n \to \infty$. As in the proof of Theorem 7, let $\mathrm{RSS}_1^{*0} = \mathrm{RSS}_1^*$ with $\widehat{G}^*$ and $\widehat{\beta}^*$ assumed to be known. Then, it is possible to show that $\left|\mathrm{RSS}_1^* - \mathrm{RSS}_1^{*0}\right| = o_{p^*}(n)$, $\left|\mathrm{RSS}_1^{*0} - \sigma^2_{G_0}\right| = o_{p^*}(1)$ and $\left|\mathrm{RSS}_0^* - \mathrm{RSS}_0^{*0}\right| = o_{p^*}(1)$ except if the original sample is on a set with probability tending to 0 as $n \to \infty$. The conclusion follows by the bootstrap CLT using the same arguments as those used by Fan et al. (2001). □

Proof of Theorem 9 Note that

$$n^{1/2}\widehat{v}(u,s,\theta)=\frac{1}{n^{1/2}}\sum_{i=1}^{n}\varepsilon_{iG_0}I\left(U_i\le u,\theta'\left[X_i',W_i'\right]'\le s\right)$$
$$+\frac{1}{n^{1/2}}\sum_{i=1}^{n}\left(\widehat{\varepsilon}_{i\widehat{G}}-\varepsilon_{iG_0}\right)I\left(U_i\le u,\theta'\left[X_i',W_i'\right]'\le s\right):=T_{6n}+T_{7n},$$

the class of functions

$$\left\{(r,q,t)\to(r-d(q)\,t)\,I\left(r\le u,\theta't\le s\right),\quad u,s,\theta\in\Pi\right\}\tag{35}$$

is Vapnik–Chervonenkis, and

$$\begin{aligned}T_{7n}&=\frac{1}{n^{1/2}}\sum_{i=1}^{n}\left(Z_{i\widehat{G}}-Z_{iG_0}\right)I\left(U_i\le u,\theta'\left[X_i',W_i'\right]'\le s\right)\\&-\frac{1}{n^{1/2}}\sum_{i=1}^{n}\left(\sum_{j=1}^{n}S_j(U_i)\right)\left(Z_{i\widehat{G}}-Z_{iG_0}\right)I\left(U_i\le u,\theta'\left[X_i',W_i'\right]'\le s\right)\\&-\frac{1}{n^{1/2}}\sum_{i=1}^{n}\left(\sum_{j=1}^{n}S_j(U_i)\left(Z_{iG_0}-W_i'\beta_0\right)-X_i'\Sigma(U_i)^{-1}E\left(X\left(Z_{G_0}-W'\beta_0\right)|U_i\right)\right)\\&\times I\left(U_i\le u,\theta'\left[X_i',W_i'\right]'\le s\right)\\&+\frac{(\widehat{\beta}-\beta_0)'}{n^{1/2}}\sum_{i=1}^{n}\left(W_i-\Omega(U_i)\Sigma(U_i)^{-1}X_i\right)I\left(U_i\le u,\theta'\left[X_i',W_i'\right]'\le s\right)\\&+\frac{(\widehat{\beta}-\beta_0)'}{n^{1/2}}\sum_{i=1}^{n}\left(\Omega(U_i)\Sigma(U_i)^{-1}X_i-\sum_{j=1}^{n}W_jS_j(U_i)'\right)I\left(U_i\le u,\theta'\left[X_i',W_i'\right]'\le s\right).\end{aligned}$$

Using 6, Lemma 16, (31) and some kernel calculations, it is possible to show that

$$\begin{aligned}T_{7n}&=\frac{1}{n^{1/2}}\sum_{i=1}^{n}\left\{\frac{1}{n}\left[\sum_{j=1}^{n}\frac{\eta(Z_j,\delta_j;Z_i)}{1-G_0(Z_i)}Z_{iG_0}\delta_i-X_i'\Sigma(U_i)^{-1}E\left[\frac{\eta(Z_j,\delta_j;Z)}{1-G_0(Z)}Z_{G_0}\delta|U_i\right]\right]\right\}\\&\times I\left(U_i\le u,X_i\le x,W_i\le w\right)\\&+\frac{1}{n^{1/2}}\left(\sum_{i=1}X_i'\Sigma(U_i)^{-1}X_i\left(Z_{iG_0}-W_i'\beta_0\right)F_\theta(s|U_i)\,f(U_i)\,I(U_i\le u)\right.\\&-\left.E\left[X'\alpha(U)\,I\left(U\le u,\theta'\left[X_i',W_i'\right]'\le s\right)\right]\right)\\&+\frac{1}{n^{1/2}}\sum_{i=1}^{n}\left(W_i-\Omega(U_i)\Sigma(U_i)^{-1}X_i\right)'\left(\varepsilon_{iG_0}+\frac{\eta(Z_j,\delta_j;Z_i)}{(1-G_0(Z_i))^2}Z_i\delta_i\right)\\&\times\Gamma^{-1}E\left[W-\Omega(U)\Sigma(U)^{-1}XI\left(U\le u,\theta'\left[X_i',W_i'\right]'\le s\right)\right]+o_p(1).\end{aligned}\tag{36}$$

The fidi convergence of (36) follows by CLT, whereas the asymptotic equicontinuity follows by a direct application of Theorem 2.5.2 of Van der Vaart and Wellner (1996), which implies the weak convergence of (36) to a Gaussian process with the same covariance as that given in (21). □

Proof of Theorem 11 See the supplement. □

Proof of Theorem 12 Let $\sigma_i^*(u,s,\theta)=\sigma_i^*\left(\widehat{\beta},\widehat{\alpha},\widehat{G};u,s,\theta\right)$ to emphasise the dependence on β, α and G. It is possible to show that

$$\sup_{u,s,\theta\in\Pi}\left|\frac{1}{n^{1/2}}\sum_{i=1}^{n}\sigma_i^*\left(\widehat{\beta},\widehat{\alpha},\widehat{G};u,s,\theta\right)-\frac{1}{n^{1/2}}\sum_{i=1}^{n}\sigma_i^*\left(\beta_0,\alpha_0,G_0;u,s,\theta\right)\right|=o_p(1),$$

and by LLN

$$\left|\frac{1}{n}\sum_{i=1}^{n}\sigma_i^*\left(\beta_0,\alpha_0,G_0;u_1,s_1,\theta_1\right)\sigma_i^*\left(\beta_0,\alpha_0,G_0;u_2,s_2,\theta_2\right)-E\left[\sigma\left(u_1,s_1,\theta_1\right)\sigma\left(u_2,s_2,\theta_2\right)\right]\right|=o_p(1).$$

The same arguments used in the proof of Theorem 1 combined with the multiplier CLT Van der Vaart and Wellner (1996, Lemma 2.9.5) imply the fidi convergence of $n^{1/2}\widehat{\sigma}^*(\beta_0,\alpha_0,G_0;u,s,\theta)$ for any fixed u, s and θ, whereas the asymptotic equicontinuity of the process $n^{1/2}\widehat{\sigma}^*(\beta_0,\alpha_0,G_0;u,s,\theta)$ follows as in the proof of Theorem 9. Thus $n^{1/2}\widehat{\sigma}^*(\beta_0,\alpha_0,G_0;u,s,\theta)$ converges weakly to a Gaussian process with the same covariance structure as that given in (21), and the result follows by CMT. □

Proof of Proposition 13 See the supplement. □

Acknowledgments I am grateful to an associate editor and two referees for very useful comments and suggestions that improved considerably the original version.

References

Ahmad, I., Leelahanon, S., Li, Q. (2005). Efficient estimation of a semiparametric partially linear varying coefficient model. *Annals of Statistics*, *33*, 258–283.

Akritas, M. (1986). Bootstrapping the Kaplan-Meier estimator. *Journal of the American Statistical Association*, *81*, 1032–1038.

Beran, R. (1981). *Nonparametric regression with randomly censored data*. Technical report: University of California, Berkeley.

Bickel, P., Freedman, D. (1981). Some asymptotic theory for the bootstrap. *Annals of Statistics*, *9*, 1196–1297.

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

Bravo, F. (2009). Two-step generalised empirical likelihood inference for semiparametric models. *Journal of Multivariate Analysis*, *100*, 1412–1431.

Buckley, J., James, I. (1979). Linear regression with censored data. *Biometrika*, *66*, 429–436.

Cai, J., Fan, J., Jiang, J., Zhou, H. (2008). Partially linear hazard regression with varying coefficients for multivariate survival data. *Journal of the Royal Statistical Society B*, pp. 141–158.

Cai, Z. (2003). Weighted local linear approach to censored nonparametric regression. In M. Akritas & D. Politis (Eds.), *Recent advances and trends in nonparametric statistics* (pp. 217–231). San Diego, CA: Elsevier.

Cai, Z., Fan, J., Li, R. (2000a). Efficient estimation and inference for varying-coefficient models. *Journal of the American Statistical Association*, *95*, 888–902.

Cai, Z., Fan, J., Yao, Q. (2000b). Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association*, *95*, 941–956.

Chen, S., Van Keilegom, I. (2009). A goodness-of-fit test for parametric and semiparametric models in multiresponse regression. *Bernoulli*, *15*, 955–976.
Cleveland, W., Grosse, E., Shyu, W. (1991). *Local regression models*, Wadsworth: Brooks-Cole, pp. 309–376.
Delgado, M., Dominguez, M., Lavergne, P. (2003). Consistent tests of conditional moment restrictions. *Annales de Economie et de Statistique*, *81*, 33–67.
Durbin, J. (1973). Weak convergence of the sample distribution function when parameters are estimated. *Annals of Statistics*, *1*, 279–290.
Engle, R., Granger, C., Rice, J., Weiss, A. (1986). Nonparametric estimation of the relation between weather and electricity sales. *Journal of the American Statistical Association*, *81*, 310–320.
Escanciano, J. (2006). A consistent diagnostic test for regression model using projections. *Econometric Theory*, *22*, 1030–1051.
Fan, J., Gijbels, I. (1994). Censored regression: Local linear approximation and their applications. *Journal of the American Statistical Association*, *89*, 560–570.
Fan, J., Gijbels, I. (1996). *Local polynomial modeling and its applications*. London: Chapman and Hall.
Fan, J., Huang, T. (2005). Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli*, *11*, 1031–1057.
Fan, J., Wu, Y. (2008). Semiparametric estimation of covariance matrices for longitudinal data. *Journal of the American Statistical Association*, *103*, 1520–1533.
Fan, J., Zhang, J. (2004). Sieve empirical likelihood ratio tests for nonparametric functions. *Annals of Statistics*, *32*, 1858–1907.
Fan, J., Zhang, W. (2008). Statistical methods with varying coefficients models. *Statistics and its interface*, *1*, 179–195.
Fan, J., Zhang, C., Zhang, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *Annals of Statistics*, *29*, 153–193.
Fan, Y., Jiang, J. (2007). Nonparametric inference with generalized likelihood ratio tests. *Test*, *16*, 409–444.
Hastie, T., Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society B*, *55*, 757–796.
Heuchenne, C., van Keilegom, I. (2007). Polynomial regression with censored data based on preliminary nonparametric estimation. *Annals of the Institute of Statistical Mathematics*, *59*, 273–279.
Kaplan, E., Meier, P. (1958). Nonparametric estimation from incomplete data. *Journal of the American Statistical Association*, *53*, 457–481.
Kent, J. (1982). Robust properties of likelihood ratio tests. *Biometrika*, *69*, 19–27.
Koul, H., Susarla, V., Ryzin, J. (1981). Regression analysis with randomly right-censored data. *Annals of Statistics*, *9*, 244–248.
Li, Q., Huang, C., Li, D., Fu, T. (2002). Semiparametric smooth coefficient models. *Journal of Business and Economic Statistics*, *20*, 412–422.
Lo, S., Singh, K. (1986). The product-limit estimator and the bootstrap: Some asymptotic representations. *Probability Theory and Related Fields*, *71*, 455–465.
Lu, X., Burke, M. (2005). Censored multiple regression by the method of average derivatives. *Journal of Multivariate Analysis*, *95*(1), 182–205.
Marsaglia, C. (1972). Choosing a point from the surface of a sphere. *Annals of Mathematical Statistics*, *43*, 645–646.
Masry, E. (1996). Multivariate regression estimation: Local polynomial fitting for time series. *Stochastic Processess and their Applications*, *65*, 81–101.
Owen, A. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, *36*, 237–249.
Owen, A. (1990). Empirical likelihood ratio confidence regions. *Annals of Statistics*, *18*, 90–120.
Owen, A. (2001). *Empirical likelihood*. New York, NY: Chapman and Hall.
Pollard, D. (1984). *Convergence of stochastic processes*. New York, NY: Springer.
Rao, J., Scott, A. (1981). The analysis of categorical data from complex sampling surveys: Chi-squared tests for goodness of fit and independence in two-way tables. *Journal of the American Statistical Association*, *76*, 221–230.
Serfling, R. (1980). *Approximation theorems of mathematical statistics*. New York, NY: Wiley.
Speckman, H. (1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society B*, *50*, 413–436.

Stinchcombe, M., White, H. (1998). Consistent specification testing with nuisance parameters present only under the alternative. *Econometric theory*, *14*, 295–325.

Stute, W. (1997). Nonparametric model checks for regressions. *Annals of Statistics*, *25*, 613–641.

Stute, W., Zhu, L. (2002). Model checks for generalized linear models. *Scandinavian Journal of Statistics*, *29*, 535–545.

Stute, W., Gonzalez-Mantiega, W., Quindimil, M. P. (1998). Bootstrap approximations in model checks for regression. *Journal of the American Statistical Association*, *93*, 141–149.

Su, J., Wei, L. (1991). A lack of fit test for the mean function in a generalized linear model. *Journal of the American Statistical Association*, *86*, 420–426.

Van der Vaart, A., Wellner, J. (1996). *Weak convergence and empirical processes*. New York, NY: Springer.

Wang, L., Veraverbeke, N. (2006). Empirical likelihood in a semiparametric model for missing response data. *Communications in Statistics: Theory and Methods*, *35*, 625–639.

Wang, Q., Li, G. (2002). Empirical likelihood semiparametric regression under random censorship. *Journal of Multivariate Analysis*, *83*, 469–486.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, *50*, 1–25.

Xue, L., Zhu, L. (2007). Empirical likelihood semiparametric regression analysis for longitudinal data. *Biometrka*, *94*, 921–937.

Zhou, M. (1991). Some properties of the Kaplan-Meier estimator for independent, nonidentically distributed random variables. *Annals of Statistics*, *19*, 2266–2274.

Zhu, L., Ng, K. (2003). Checking the adequacy of a partial linear model. *Statistica Sinica*, *13*, 763–781.