Estimation and inference for quantile partially linear varying coefficients models with missing observations*

Francesco Bravo†
University of York

May 2025

Abstract

This paper considers estimation and inference for quantile partially linear varying coefficients models, where some of the observations are missing at random. The unknown parameters are estimated using two different two step procedures, one of which is based on iteration and the other is based on profiling. Both procedures are based on inverse probability weighting, where the weights can be estimated either parametrically or nonparametrically. The paper proposes two computationally simple resampling techniques that can be used to consistently estimate the asymptotic distributions and the asymptotic variances of the unknown finite dimensional parameters estimators. For inference, the paper proposes new test statistics for both the finite and infinite dimensional parameters, including a test for constancy of the varying coefficients part of the model. Monte Carlo simulations show that the proposed estimators and test statistics have good finite sample properties. Finally, the paper contains a real data application.

Keywords: Inverse probability weighting, Distance statistic, Local linear estimation, MAR, MM algorithm, Wilks' phenomenon, Wald statistic.

2010 mathematics classification codes: 62G05 , 62G08

---

*Comments from participants at ESEM2023 and EC^{2}2023 are gratefully acknowledged.

†Address correspondence to: Department of Economics, University of York, York YO10 5DD, UK. E-mail: francesco.bravo@york.ac.uk. Web Page: https://sites.google.com/a/york.ac.uk/francescobravo/

1 Introduction

Since its introduction as a generalization of the linear regression model, *parametric* quantile regressions (Basset & Koenker 1978, Koenker & Bassett 1978) have been widely used in economics, finance, and statistics - see Koenker (2005) for a review of applications. Compared to linear regressions, quantile regressions provide a more complete characterization of the conditional distribution of the responses given a set of covariates, being at the same time more robust to the presence of possible outliers. Despite these appealing features, parametric quantile regressions can be limited due to the potential risk of misspecification and lack of flexibility. For these reasons, various *nonparametric* and *semiparametric* extensions to quantile regression models have been considered in the literature; here we mention a number of contributions (among many others) that are most related to the results of this paper. Chauduri (1991) considered local polynomial estimation of a nonparametric quantile regression model and obtained a (pointwise) Bahadur expansion for the resulting estimator; Chauduri, Doksum & Samarov (1997) built upon the results of Chauduri (1991) and considered an average quantile regression model; Yu & Jones (1998) considered (a possibly double kernel based) local linear estimation of a nonparametric quantile regression; Kim (2007) and Cai & Xu (2008) considered quantile varying coefficients models; Kong, Linton & Xia (2010) and Guerre & Sabbah (2012) extended the results of Chauduri (1991) to obtain Bahadur expansions of their proposed local polynomial estimators of a nonparametric quantile regression model that are uniform in the conditioning variables and also in the bandwidth, respectively. Lee (2003) considered efficient estimation of a quantile partially linear regression model; Kai, Li & Zou (2011), Wang, Zhu & Zhou (2009) and Cai & Xiao (2012) proposed a two step estimation procedure for a quantile partially linear varying coefficients model, and Sherwood (2016) proposed a one step estimation procedure for a partially linear additive quantile regression model with missing covariates.

With the exception of Sherwood (2016), all of the above results assume that the observations are always observable. However, in many situations of empirical relevance some of the observations in the sample are missing; for example, in a survey of empirical research in top economics journals, Abrevaya & Donald (2017) found that missing data occurs in 40% of the publications, and, depending on the missing mechanism, simply ignoring this fact may result in inconsistent and/ or inefficient estimators with possibly great loss of information. The missing mechanism considered in this paper is missing at random (MAR henceforth) (Rubin 1976), which specifies that the probability of missing - often called selection probability- depends on variables that are always observed. MAR has been widely applied in a number of econometric and statistical models, including program evaluation (Imbens 2004), non-classical measurement error (Robins, Hseih & Newey 1995, Chen, Hong & Tamer 2005), missing covariates (Robins, Rotnitzky & Zhao 1994) and attrition in panel data (Robins, Rotnisky & Zhao 1995); see Little & Rubin (2002) for other applications of MAR.

In this paper, we provide a unifying framework for estimating and testing quantile partially linear varying coefficients (QPLVC henceforth) models with MAR observations. As mentioned by Kai et al. (2011) and others, compared to the fully nonparametric approach of Chauduri (1991) and Guerre & Sabbah (2012), and the quantile varying coefficients models of Kim (2007) and Cai & Xu (2008), the QPLVC specification avoids the curse of dimensionality and allows partial information about the linearity of some of the components to be incorporated while retaining the flexibility offered by the nonparametric part of the model. An important feature of this paper is the fact that the MAR observations are allowed to be in both the responses and some of the covariates, or in the responses only, or in the covariates only, making the results of this paper very general and applicable to most situations with missing data problems. To deal with MAR observations we use the inverse probability weighting (IPW henceforth) method (Horvitz & Thompson 1952), which has been used in many semiparametric models with MAR observations, including semiparametric regressions (Wang, Hardle & Linton 2004, Bianco, Boente, Gonzales-Mantiega & Perez-Gonzales 2010) and semiparametric treatment effects (Hirano, Imbens & Ridder 2003), among many others. IPW has been used previously in the context of quantile models with missing data: Firpo (2007) considered efficient estimation of quantile treatment effects, Chen, Wan & Zhou (2015) considered efficient estimation of parametric quantile models with MAR observations, whereas Wang, Tian & Tang (2022) considered estimation of nonparametric quantile models with MAR observations. None of these contributions considered the class of semiparametric quantile regression models considered in this paper. In fact, to the best of our knowledge, this is the first paper that considers IPW-based estimation (and inference) for QPLVC models with the general MAR assumption considered.

We propose two different estimation procedures for the unknown parameters: the first one is based on a two step iterative M-type estimation (often called backfitting), in which the first step is used to estimate locally all the unknown parameters using the local linear estimator of Fan & Gijbels (1996), while the second step is used to re-estimate the finite dimensional unknown parameters, and then iterate between the two steps until convergence. This procedure is similar to the one proposed by Kai et al. (2011) and Cai & Xiao (2012), although neither of these authors considered missing data, and the latter used a different estimation method for the second step estimation. The second procedure is based on a profiled two step Z-type estimation, in which the unknown infinite dimensional parameter is indexed by the finite dimensional parameter, and estimation of the latter is not iterative. Each methods have their own merits: the one based on iteration is simpler to compute but requires undersmoothing and is computationally more intensive. The one based on profiling is not computationally intensive but requires the computation of the derivative of the unknown infinite dimensional parameter, which is difficult given the nonsmoothness of the model. In order to simplify the computation of the proposed estimators, we use the MM algorithm (Hunter & Lange 2000), which replaces the nonsmooth objective function used in the quantile estimation with a certain smooth majorizing function that can be easily minimized by standard iterative methods - see Section 6 for more details. We note that if the unknown infinite dimensional parameters are of direct interest, as for example in 4.4), an additional step can be added, in which the infinite dimensional parameters are re-estimated locally, see Remark 1 in Section 2 below.

For inference, we consider Wald statistics that can be used to test local and global linear hypotheses on, respectively, the infinite and finite dimensional unknown parameters; we also propose a "distance" statistic that can be used to test general hypotheses on the infinite dimensional parameters, including the important one of constancy over its whole support. The proposed distance statistic is in the same spirit as the one proposed by Fan, Zhang & Zhang (2001) for varying coefficients models, and, as we are aware of, has not been proposed for QPLVC models, even without missing variables.

We now discuss in some detail the novel contributions this paper makes to the literature on quantile semiparametric models with missing data:
First, profile estimation for the finite dimensional parameter in QPLVC models is new (even without missing observations). We note that without missing data, a simple modification of the proposed profile estimator achieves the semiparametric efficiency bound, which, in the context of this paper is given by

$$\tau(1-\tau)(E[(f_{\varepsilon|X}(0))^2 X_1^{\otimes 2}] - E[E(f_{\varepsilon|X}(0)^2 X_1 X_2^T|X_3)E(f_{\varepsilon|X}(0)^2 X_2^{\otimes 2}|X_3)^{-1}E(f_{\varepsilon|X}(0)^2 X_2 X_1^T|X_3)). \tag{1.1}$$

We also note that its asymptotic distribution is different from that of corresponding iterative two step estimator because of the presence of missing observations. This result is consistent with that of Hu, Wang & Carroll (2004), who showed that once you move away from the i.i.d. assumption, backfitting and profile estimation in semiparametric models results in estimators with different asymptotic variances.

Second, we consider two different estimators for the probabilities of missing appearing in the IPW, one based on a parametric specification and one based on a nonparametric one. The former has the advantage of being computationally simpler and not depending on the dimension of the missing variables vector, whereas the latter has the advantage of being robust to possible misspecification of the probability of missing mechanism, but it may suffer from the curse of dimensionality. We show that the asymptotic variance of the infinite dimensional parameters estimator is the same, regardless of the choice of the probability of missing estimator, as long as the additional "undersmoothing" condition A2(ii) is satisfied, see the discussion after the assumptions in Section 3.1 and Remark 2 for more details. On the other hand, choosing a parametric or nonparametric estimator for the probabilities of missing has bearings for the asymptotic variance (and hence efficiency) of the finite dimensional parameters estimator, which are very different, see Remarks 3 and 4 for a discussion.

Third, in order to derive the asymptotic distribution of the unknown infinite dimensional parameters estimator, we obtain a Bahadur expansion that is uniform in the conditioning variable regardless as to which estimator is used for the probabilities of missing. The expansion is based on the quadratic approximation lemma of Fan & Gijbels (1996), which avoids stochastic equicontinuity arguments often used in the literature, see the proof of Theorem 1 for more details. For the unknown finite dimensional parameters estimator, we show that using a nonparametric estimator for the probability of missing results in an asymptotic variance that corresponds to that obtained by using the so-called augmented IPW estimating equations originally proposed by Robins et al. (1994), see also Chen et al. (2015), to increase the efficiency of the estimator. On the other hand stochastic equicontinuity arguments are needed to derive the asymptotic distribution of the profile estimator, see the proof of Theorem 5 for more details.

Fourth, we propose a computationally simple resampling method for the estimation of the unknown finite dimensional parameters that is well suited for both estimators with MAR observations, as it preserves the missing structure of the observations in the original sample. The method is based on the so-called multiplier bootstrap (see for example Van der Vaart & Wellner (1996) and Kosorov (2008)) and consists of randomly perturbing the objective functions by a sequence of independent and identically distributed random variables independent of the original sample of observations, and re-estimate the unknown parameters. Bose & Chatterjee (2003), Chen et al. (2015), and Cheng & Huang (2010) showed the consistency of such resampling method for parametric quantile regression and general semiparametric M estimators, respectively. We show the consistency of the proposed multiplier bootstrap, and how it can be used to consistently estimate the asymptotic variances of the proposed estimators., which is a topic often ignored in the multiplier bootstrap literature.

Fifth, we consider inference for both the unknown finite and infinite dimensional parameters. For the former, we propose a Wald statistic for a set of linear restrictions that, under a standard undersmoothing condition, is shown to be asymptotically Chi-squared distributed under the null hypothesis and a sequence of Pitman-type alternatives, as well as consistent under fixed alternative hypotheses. For the latter, we propose a Wald statistic for local linear hypotheses (that is hypotheses evaluated at a single point in the support of the random variate associated to the infinite dimensional parameter) that are asymptotically Chi-squared distributed under the null hypothesis and a sequence of Pitman-type alternatives, as well as consistent under fixed alternative hypotheses. We also consider global hypotheses (that is hypotheses evaluated over the whole support of the random variate associated to the infinite dimensional parameter) and show that a distance statistic based on the IPW-quantile objective function is asymptotically normal when appropriately standardized. The proposed distance statistic can be interpreted as a generalized likelihood ratio as in Fan et al. (2001), however, as opposed to Fan et al. (2001), the so-called Wilks' phenomenon, that is the proposed statistic is asymptotically independent of nuisance parameters and (nearly) Chi-squared distributed, does not hold because of the IPW. On the other hand, without MAR observations the Wilks' phenomenon still holds, see Proposition 10 and the simulation results in Section 6 for more details.

Finally, we use a Monte Carlo study and an empirical application to illustrate the finite sample properties and the applicability of the proposed estimators and test statistics.

The rest of the paper is structured as follows: next section introduces the model and the estimators. Sections 2 and 4 introduce the estimators and test statistics, whereas sections 3 and 5 contain the main asymptotic results; Section 6 first describes some details on the MM algorithm used to compute the proposed estimators, and then reports the results of the Monte Carlo study, whereas Section 7 contains the empirical application. Finally, Section 8 contains some concluding remarks. All proofs are contained in a Supplemental Appendix, which also contains some additional simulations' results.

The following notation is used throughout the paper: "^{T}" indicates transpose, a prime " $'$ " and double prime " $''$ " denote first and second derivatives of the unknown vector of real valued functions $\theta_{0\tau}(\cdot)$ with respect to the argument $\cdot$; finally for any vector v, $v^{\otimes 2} = vv^T$.

2 The model and the estimators

Consider the QPLVC model

$$Y = X_1^T \beta_{0\tau} + X_2^T \theta_{0\tau}(X_3) + \varepsilon, \tag{2.1}$$

where $\beta_{0\tau}$ is a k dimensional vector of unknown parameters, $\theta_{0\tau}(\cdot)$ is a p dimensional vector of unknown real valued functions and the unobservable error ε satisfies the τth conditional quantile restriction $q_\tau(\varepsilon|X)=0$ for $X=\left[X_1^T,X_2^T,X_3\right]^T$. Model (2.1) assumes that for a chosen τth conditional quantile q_τ, X_1 and X_2 are the key covariates while allowing for possible nonlinear interactions between X_2 and X_3 such that a different level of X_3 is associated to a different quantile regression, and it is this feature that makes (2.1) very flexible and useful in practice.

Let $\left(\left[Y_i,X_{1i}^T,X_{2i}^T,X_{3i}\right]^T\right)_{i=1}^n$ denote an (incomplete) random sample, and let $(Z_{oi})_{i=1}^n$ denote the corresponding sample containing all the always observed data. For example, if some of the $(Y_i)_{i=1}^n$ responses and some of the $(X_{1i})_{i=1}^n$ and $(X_{2i})_{i=1}^n$ covariates (could be either of them or both) are missing, then $(Z_{oi})_{i=1}^n=\left(\left[X_{oi}^T,X_{3i}\right]^T\right)_{i=1}^n$, where X_{oi} are the always observed covariates; if some of the observations in all of the $\left(\left[X_{1i}^T,X_{2i}^T\right]^T\right)_{i=1}^n$ covariates are missing, then $(Z_{oi})_{i=1}^n=\left([Y_i,X_{3i}]^T\right)_{i=1}^n$. In what follows, we assume that $Z_{oi}=\left[X_{oi}^T,X_{3i}\right]^T$, noting that the cases of missing covariates only or missing responses only can be easily accommodated by changing the selection probability defined in (2.2) and the related expressions in Sections 3 and 5 below, accordingly. Let δ^Y and δ^{X_m} denote the binary indicators for the missing responses and covariates, where a 0 indicates a missing observation, and, for $\delta=\delta^Y\delta^{X_m}$, let

$$\Pr(\delta=1|Y,X)=\Pr(\delta=1|Z_o):=\pi_0(Z_o)>0 \quad a.s., \tag{2.2}$$

denote the selection probability, which specifies that the probability of missing depends only on the always observed variables.

We first describe the two step iterative estimation procedure, which can be interpreted as an IPW-M estimation process. Let

$$Q_n(\beta_\tau,\theta_\tau,\pi)=\sum_{i=1}^n\frac{\delta_i}{\pi(Z_{oi})}\rho_\tau\left(Y_i-X_{1i}^T\beta_\tau-X_{2i}^T\theta_\tau(X_{3i})\right) \tag{2.3}$$

be the IPW objective function, where $\rho_\tau(\cdot)=\cdot\,(\tau-I(\cdot<0))$ denotes the check function.

Let $\widehat{\pi}(Z_{oi})$ denote an estimator for $\pi_0(Z_{oi})$ and let

$$\theta_{0\tau}(X_3)\approx\theta_{0\tau}(x_3)+\theta_{0\tau}'(x_3)(X_3-x_3):=a_\tau+b_\tau(X_3-x_3) \tag{2.4}$$

denote the local linear approximation of $\theta_{0\tau}(X_3)$ in a neighbourhood of x_3.

The two step iterative estimation procedure for the unknown parameters $\beta_{0\tau}$ and $\theta_{0\tau}(\cdot)$ is based on the following two steps:

Step 1 Estimate $\beta_{0\tau}$ and $\theta_{0\tau}(\cdot)$ locally using (2.4), that is

$$(\widehat{\beta}_\tau^l,\widehat{a}_\tau^l,\widehat{b}_\tau^l)=\arg\min_{a_\tau,b_\tau,\beta_\tau}Q_n(\beta_\tau,a_\tau+b_\tau(X_{3i}-x_3),\widehat{\pi})K_h(X_{3i}-x_3), \tag{2.5}$$

where $K_h(\cdot)=K(\cdot/h)$ is a kernel function and $h:=h(n)$ is the bandwidth.

Step 2 Estimate $\beta_{0\tau}$ using

$$\widehat{\beta}_\tau=\arg\min_{\beta_\tau\in B}Q_n\left(\beta_\tau,\widehat{\theta}_\tau^l,\widehat{\pi}\right). \tag{2.6}$$

where $\widehat{\theta}_\tau^l=\widehat{a}_\tau^l$, obtained in Step 1.

Then iterate between the two steps until convergence of $\widehat{\beta}_\tau$.

Remark 1 *Note that to further improve the efficiency of the estimators $\widehat{a}_\tau^l$ and $\widehat{b}_\tau^l$ obtained in Step 1, an additional third step local estimation can be added, which consists of re-estimating $\theta_{0\tau}(\cdot)$ using*

$$(\widehat{a}_\tau, \widehat{b}_\tau) = \arg\min_{a_\tau, b_\tau} Q_n\left(\widehat{\beta}_\tau, a_\tau + b_\tau (X_{3i} - x_3), \widehat{\pi}\right) K_h (X_{3i} - x_3),$$

where $\widehat{\beta}_\tau$ is defined in Step 2.

For the profile estimation procedure we follow the same approach as that used by Wong & Severini (1991) and Severini & Wong (1992), which is based on the notion of least favourable curve $\theta_{\beta_\tau}(x_3)$, which, in the context of this paper, is defined as the minimizer of

$$E[\rho_\tau(Y_i - X_{1i}^T\beta_\tau - X_{2i}^T\eta)|X_{3i} = x_3] \tag{2.7}$$

satisfying

$$\frac{\partial}{\partial\eta} E[\rho_\tau(Y_i - X_{1i}^T\beta_\tau - X_{2i}^T\eta]|X_{3i} = x_3]|_{\eta=\theta_{\beta_\tau(u)}} = 0.$$

As with the two step estimator we consider the local linear approximation $\theta_{0\tau}(X_{3i}) \approx a_\tau + b_\tau(X_{3i} - x_3)$ so that for a fixed β_τ the least favourable curve minimises $Q_n(\beta_\tau, a_\tau + b_\tau(X_{3i} - x_3), \widehat{\pi}_i)K_b(X_{3i} - x_3)$. Using $\widehat{\theta}_{\beta_\tau} =: a_\tau$ and $\partial\widehat{\theta}_{\beta_\tau}/\partial\beta_\tau^T =: b_\tau$ the profile estimator $\widehat{\beta}_\tau^p$ is defined as

$$\widehat{\beta}_\tau^p = \arg\min_{\beta_\tau \in B} ||M_n(\beta_\tau, \widehat{\theta}_{\beta_\tau}, \partial\widehat{\theta}_{\beta_\tau}/\partial\beta_\tau^T, \widehat{\pi})||, \tag{2.8}$$

where

$$M_n(\beta_\tau, \theta_{\beta_\tau}, \partial\theta_{\beta_\tau}/\partial\beta_\tau^T, \pi) = \frac{1}{n}\sum_{i=1}^n \frac{\delta_i}{\pi_i}(X_{1i} + \left(\frac{\partial\theta_{\beta_\tau}(X_{3i})}{\partial\beta_\tau^T}\right)^T X_{2i})\rho_\tau'(Y_i - X_{1i}^T\beta_\tau - X_{2i}^T\theta_{\beta_\tau}(X_{3i})),$$

that is the subgradient of $Q_n(\beta_\tau, \theta_{\beta_\tau}, \pi)$ with $\rho_\tau'(\cdot) = \tau - I(\cdot < 0)$.

We conclude this section by discussing the form of $\widehat{\pi}(Z_{oi})$, which depends on whether we assume a parametric or a nonparametric specification for $\pi_0(Z_o)$. For the former, we assume that $\pi_0(Z_o) = \pi(Z_o, \alpha)$ is a parametric model (such as a probit or logit model) where $\alpha \in A \subseteq \mathbb{R}^l$ is an unknown parameter. For the latter, the estimator takes the form

$$\widehat{\pi}(z) = \frac{\sum_{i=1}^n \delta_i L_b(Z_{oi} - z)}{\sum_{i=1}^n L_b(Z_{oi} - z)}, \tag{2.9}$$

where $L_b(\cdot) = L(\cdot/b)$ is a product kernel function with another bandwidth $b := b(n)$.

3 Asymptotic results for estimation

3.1 Two step iterative estimation

Let $F_{\varepsilon|X}(\cdot)$, $f_{\varepsilon|X}(\cdot)$ and $f_{X_3}(\cdot)$ denote the conditional distribution and density of ε, and the marginal density of X_3, respectively. Assume that:

A1 (i) $F_{\varepsilon|X=x}(0) = \tau$ and $f_{\varepsilon|X=x}(0)$ are continuous and positive for all $x \in \mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3$, (ii) the marginal density $f_{X_3}(x)$ of X_3 is continuous and positive at $x = x_3$, (iii) X_1, X_2 and X_3 have bounded supports $\mathcal{X}_1$, $\mathcal{X}_2$ and $\mathcal{X}_3$, (iv) the parameter space B is a compact set.

A2 (i) The kernel functions $K(\cdot)$ and $L(\cdot)$ are symmetric with bounded support, with bandwidths satisfying, respectively, $nh \to \infty$ and $nb^{\dim(Z_o)} \to \infty$, (ii) $h = o\left(b^{dim(Z_o)}\right)$ and $nhb^4 \to 0$.

A3 (i) $\theta''_\tau(x)$ is continuous at $x = x_3$, (ii) the matrix $\Sigma(x_3)$ defined in (10.4) in the Supplemental Appendix is nonsingular for all $x_3 \in \mathcal{X}_3$.

Either

A4 (i) $\inf_{Z_o \in \mathcal{Z}_o} \pi(Z_o, \alpha) > 0$ for all $\alpha \in A$, (ii) there exists a $\alpha_0 \in A$ such that $\pi(Z_o, \alpha_0) = \pi_0(Z_o)$, (iii) $E \sup_{\alpha \in A} \|\partial \pi(Z_o, \alpha) / \partial \alpha\|^\delta < \infty$ for some $\delta > 2$, (iv) the maximum likelihood estimator $\widehat{\alpha}$ has the following stochastic expansion:

$$n^{1/2}(\widehat{\alpha} - \alpha_0) = I(\alpha_0)^{-1} \frac{1}{n^{1/2}} \sum_{i=1}^{n} s(Z_{oi}, \alpha_0) + o_p(1),$$

where $E[s(Z_o, \alpha_0)] = 0$, $E\left[\partial^2 \log \pi(Z_o, \alpha_0) / (\partial \alpha)^{\otimes 2}\right] = -I(\alpha_0)$ and

$$n^{1/2}(\widehat{\alpha} - \alpha_0) \xrightarrow{d} N\left(0, I(\alpha_0)^{-1}\right).$$

Or

A5 (i) $\inf_{Z_o \in \mathcal{Z}_o} \pi_0(Z_o) > 0$, (ii) $\pi_0(Z_o)$ is twice continuously differentiable with bounded derivatives.

The above regularity conditions are fairly standard: A1(i) is standard in the quantile regression literature, see for example Koenker (2005). A1(ii)-A3 are commonly used in nonparametric estimation, see for example Chauduri (1991); A2(ii) can be interpreted as an undersmoothing type condition, where the degree of undersmoothing depends on the dimension of the observable covariates Z_o and the selected bandwidth b; for example, if $b = n^{-1/5}$ and $\dim(Z_o) = 1$, then $h = n^{-1/4}$ would satisfy it. More generally, for $h \propto n^{-a}$ and $b \propto n^{-c}$ A2(ii) requires $a > c \dim(Z_o)$. Finally A4 and A5 are commonly used in the MAR literature, see for example Robins et al. (1994). Note that A4(i) and A5(i) can be indirectly verified by examining the distribution of the estimated selected probabilities.

The following theorem gives the asymptotic distribution of the estimators $\widehat{\beta}^l_\tau$ and $\widehat{\theta}^l_\tau(x_3) = \widehat{a}^l_\tau$ obtained in Step 1; let $\kappa_j = \int t^j K(t)\, dt$ and $v_j = \int t^j K^2(t)\, dt$ for $j = 0, 1, 2$.

Theorem 1 *Under assumptions A1-A5*

$$(nh)^{1/2} \left[\begin{array}{c} \widehat{\beta}^l_\tau - \beta_{0\tau} \\ \widehat{\theta}^l_\tau(x_3) - \theta_{0\tau}(x_3) \end{array} - B(x_3) \right] \xrightarrow{d} N\left(0, \Sigma_1(x_3)^{-1} \Sigma_{1\pi}(x_3) \Sigma_1(x_3)^{-1}\right),$$

where

$$B(x_3) = \frac{h^2}{2} f_{X_3}(x_3) \Sigma_1(x_3)^{-1} E\left\{\kappa_2 f_{\varepsilon|X}(0) \left[\begin{array}{c} X_1 X_2^T \\ X_2^{\otimes 2} \end{array} \right] | X_3 = x_3 \right\} \theta''_{0\tau}(x_3),$$

$$\Sigma_1(x_3) = f_{X_3}(x_3)\, E\left\{ f_{\varepsilon|X}(0) \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}^{\otimes 2} |X_3 = x_3 \right\},$$

$$\Sigma_{1\pi}(x_3) = f_{X_3}(x_3)\, E\left\{ \frac{\tau(1-\tau)\, v_0}{\pi_0(Z_o)} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}^{\otimes 2} |X_3 = x_3 \right\}.$$

The following theorem gives the asymptotic distribution of the estimator $\widehat{\theta}_\tau(\cdot)$ suggested in Remark 1.

Theorem 2 *Under the same assumptions of Theorem 1*

$$(nh)^{1/2}\left(\widehat{\theta}_\tau(x_3) - \theta_{0\tau}(x_3) - \frac{h^2\kappa_2\theta''_{0\tau}(x_3)}{2}\right) \xrightarrow{d} N\left(0, \Sigma_3(x_3)^{-1}\Sigma_{3\pi}(x_3)\Sigma_3(x_3)^{-1}\right),$$

where

$$\Sigma_3(x_3) = f_{X_3}(x_3)\, E\left[f_{\varepsilon|X}(0)\, X_2^{\otimes 2}|X_3 = x_3\right],$$

$$\Sigma_{3\pi}(x_3) = f_{X_3}(x_3)\, E\left[\frac{\tau(1-\tau)\, v_0}{\pi_0(Z_o)} X_2^{\otimes 2}|X_3 = x_3\right].$$

Remark 2 *Theorem 1 shows that the asymptotic variance of the IPW local estimator depends on the unknown selection probabilities and is larger than the corresponding one without missing observations, see for example Kai et al. (2011) and Wang et al. (2009) for a comparison. The asymptotic variance does not depend on the type of estimator used to estimate the selection probabilities $\pi_0(Z_o)$, because of the faster convergence rate of the parametric estimator $\widehat{\pi}(Z_{io}, \widehat{\alpha})$ and A2(ii), which implies that the estimation effect coming from the nonparametric estimation of $\pi_0(Z_o)$ is asymptotically negligible. Theorem 2 shows that the additional estimator suggested in Remark 1 has the same asymptotic bias as that of the quantile varying coefficient model considered for example by Cai & Xu (2008). The explanation of this result is that $\widehat{\beta}_\tau$ converges at a faster rate than that of the estimator of the unknown infinite dimensional parameters, which effectively makes the QPLVC model a quantile varying coefficient model, meaning that the argument of the check function $\rho_\tau(.)$ in (2.3) can be replaced by say $\widetilde{Y}_i - X_{2i}^T\theta_\tau(X_{3i})$, with $\widetilde{Y}_i = Y_i - X_{1i}^T\widehat{\beta}_\tau$.*

Next we obtain the asymptotic distribution of the estimator (2.6) defined in Step 2. We first consider the case of parametric estimation of the selection probabilities, so that the estimator for $\beta_{\tau 0}$ is defined as

$$\widehat{\beta}_\tau = \arg\min_{\beta_\tau \in B} Q_n\left(\beta_\tau, \widehat{\theta}_\tau, \widehat{\pi}(Z_{oi}, \widehat{\alpha})\right).$$

Let

$$\varphi(X_i) = E\left[f_{\varepsilon|X}(0)\, X_1 X_2^T|X_3 = X_{3i}\right] S\Sigma(X_{3i})^{-1} \begin{bmatrix} X_{1i} \\ X_{2i} \\ 0_p \end{bmatrix},$$

where $S = [O_{pk}, I_p, O_{pp}]$ is a selection matrix with O_{pk} a $p \times k$ matrix of zeroes, I_p the identity matrix of order p, O_{pp} a $p \times p$ matrix of zeroes, 0_p a $p \times 1$ vector of zeroes, and $\Sigma(X_{3i})$ is defined in (10.4) in the Supplemental Appendix. Assume that

A6 $E\left(f_{\varepsilon|X}(0)\,X_1^{\otimes 2}\right) := \Sigma_2$ is nonsingular.

Theorem 3 *Under assumptions A1-A4, A6 and* $E\sup_{\alpha\in A}\left\|\left(\partial\pi_0(Z,\alpha)/\partial\alpha\right)/\pi_0(Z_o,\alpha)\right\|^2 < \infty$*, for* $nh^4 \to 0$

$$n^{1/2}\left(\widehat{\beta}_\tau - \beta_{0\tau}\right) \xrightarrow{d} N\left(0, \Sigma_2^{-1}\Sigma_{2p}\Sigma_2^{-1}\right),$$

where

$$\Sigma_{2p} = E\left[\frac{\left((X_1-\varphi(X))\,\rho_\tau'(\varepsilon)\right)^{\otimes 2}}{\pi_0(Z_o,\alpha)}\right] - E\left[\frac{(X_1-\varphi(X))\,\rho_\tau'(\varepsilon)}{\pi_0(Z_o,\alpha)}\frac{\partial\pi_0(Z_o)}{\partial\alpha^T}\right]\times$$
$$I(\alpha_0)^{-1}\,E\left[\frac{(X_1-\varphi(X))\,\rho_\tau'(\varepsilon)}{\pi_0(Z_o,\alpha)}\frac{\partial\pi_0(Z_o,\alpha)}{\partial\alpha^T}\right]^T.$$

In the case of nonparametric estimation of the selection probabilities, the estimator for $\beta_{0\tau}$ is defined as

$$\widehat{\beta}_\tau = \arg\min_{\beta\in B} Q_n\left(\beta_\tau, \widehat{\theta}_\tau, \widehat{\pi}(Z_{oi})\right),$$

where $\widehat{\pi}(Z_{oi})$ is defined in (2.9).

Theorem 4 *Under assumptions A1-A3, A5 and A6 for* $nh^4 \to 0$ *and* $nb^4 \to 0$

$$n^{1/2}\left(\widehat{\beta}_\tau - \beta_{0\tau}\right) \xrightarrow{d} N\left(0, \Sigma_2^{-1}\Sigma_{2np}\Sigma_2^{-1}\right),$$

where

$$\Sigma_{2np} = E\left[\frac{\left((X_1-\varphi(X))\,\rho_\tau'(\varepsilon)\right)^{\otimes 2}}{\pi_0(Z_o)}\right] - E\left(\frac{1-\pi_0(Z_o)}{\pi_0(Z_o)}E\left[\left((X_1-\varphi(X))\,\rho_\tau'(\varepsilon)\right)|Z_o\right]^{\otimes 2}\right).$$

Remark 3 *Note that for* $h \propto n^{-a}$ *and* $b \propto n^{-c}$ *the undersmoothing condition* $nh^4 \to 0$ *requires* $a > 1/4$*, which implies* $\max\{1/4, c\dim(Z_o)\} < a < 1$*, which in turn implies that for a second order kernel (like the one used in this paper) the undersmoothing condition* $nb^4 \to 0$ *is satisfied for* $\dim(Z_o) < 4$*, which represents a limitation (the well known curse of dimensionality) of the proposed nonparametric estimation of* $\pi_0(Z_o)$*. Alternatively, one could use a higher order kernel, say of order* $r > 2$*, which would imply* $\dim(Z_o) < 2r$ *for the resulting undersmoothing condition* $nb^{2r} \to 0$ *to be satisfied. However higher order kernels might result in negative estimates of the selection probabilities, which is clearly something undesirable.*

Remark 4 *It is important to note that the asymptotic variance* Σ_{2np} *corresponds to the asymptotic variance of the augmented IPW estimating equation*

$$0 = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\delta_i}{\pi_0(Z_{oi})}(X_{1i}-\varphi(X_i))\,\rho_\tau'(\varepsilon_i) - \frac{\delta_i-\pi_0(Z_{oi})}{\pi_0(Z_{oi})}E\left[(X_{1i}-\varphi(X_i))\,\rho_\tau'(\varepsilon_i)\,|Z_{oi}\right]\right\}, \tag{3.1}$$

which can be used to obtain a more efficient estimator for $\beta_{\tau 0}$*, see for example Robins et al. (1994) for the case of MAR covariates. Thus, the proposed estimation method results in more efficient estimators without having to estimate the additional conditional expectation in* (3.1).

3.2 Profile estimation

For some $\alpha > 1$, let $C_M^\alpha(\mathbb{R}_X)$ denote the space of continuous functions $\mathbb{R}_X \to \mathbb{R}$ with Holder norm bounded by a finite M. Assume that:

A1' (i) A1(i)-(iv) hold; (ii) $\Theta_B = \{\theta_{\beta_\tau}, \partial\theta_{\beta_\tau}/\partial\beta_\tau^T \in C_M^\alpha(\mathcal{X}_3)\}$, (iii) $ng^8 \to 0$ and $nh^8g^{-4} \to 0$, (iv) $\partial E(X_1 + (\partial\theta_{0\tau}/\partial\beta_\tau^T)^T X_2)\rho_\tau'(Y - X_1^T\beta_\tau - X_2^T\theta_{0\tau})/\partial\beta_\tau^T$ exists, is continuous at β_τ and has full column rank,

A6' (i) $E(f_{\varepsilon|X}(0)(X_1 + (\partial\theta_{\beta_\tau}(X_3)/\partial\beta_\tau^T)^T X_2)^{\otimes 2}) := \Sigma_4$ is nonsingular, (ii) $E||\partial^2\theta_{\beta_\tau}(x_3)/\partial\beta_\tau^T\partial\beta_{\tau j}|| < \infty$ uniformly in $x_3 \in \mathcal{X}_3$ for $j = 1, ..., k$,

and note that A1'(iii) is satisfied for $h \propto n^{-1/5}$ and $g \propto n^{-1/7}$. Let

$$\varphi^p(X_i) = [E(X_2X_2^T|X_3 = X_{3i})^{-1}E(X_2X_1^T|X_3 = X_{3i})]^T X_{2i}$$

Theorem 5 *Under A1', A2-A4, A6' and* $E\sup_{\alpha\in A}\left\|\left(\partial\pi_0\left(Z,\alpha\right)/\partial\alpha\right)/\pi_0\left(Z_o,\alpha\right)\right\|^2 < \infty$

$$n^{1/2}(\widehat{\beta}_\tau^p - \beta_{0\tau}) \xrightarrow{d} N\left(0, \Sigma_4^{-1}\Sigma_{4p}\Sigma_4^{-1}\right),$$

where

$$\Sigma_{4p} = E\left[\frac{\left(\left(X_1 - \varphi^p\left(X\right)\right)\rho_\tau'\left(\varepsilon\right)\right)^{\otimes 2}}{\pi_0\left(Z_o,\alpha\right)}\right] - E\left[\frac{\left(X_1 - \varphi^p\left(X\right)\right)\rho_\tau'\left(\varepsilon\right)}{\pi_0\left(Z_o,\alpha\right)}\frac{\partial\pi_0\left(Z_o\right)}{\partial\alpha^T}\right]\times$$
$$I\left(\alpha_0\right)^{-1}E\left[\frac{\left(X_1 - \varphi^p\left(X\right)\right)\rho_\tau'\left(\varepsilon\right)}{\pi_0\left(Z_o,\alpha\right)}\frac{\partial\pi_0\left(Z_o,\alpha\right)}{\partial\alpha^T}\right]^T.$$

Under A1', A2-A3, A5 and A6'

$$n^{1/2}\left(\widehat{\beta}_\tau^p - \beta_{0\tau}\right) \xrightarrow{d} N\left(0, \Sigma_4^{-1}\Sigma_{4np}\Sigma_4^{-1}\right),$$

where

$$\Sigma_{4np} = E\left[\frac{\left(\left(X_1 - \varphi^p\left(X\right)\right)\rho_\tau'\left(\varepsilon\right)\right)^{\otimes 2}}{\pi_0\left(Z_o\right)}\right] - E\left(\frac{1 - \pi_0\left(Z_o\right)}{\pi_0\left(Z_o\right)}E\left[\left(\left(X_1 - \varphi^p\left(X\right)\right)\rho_\tau'\left(\varepsilon\right)\right)|Z_o\right]^{\otimes 2}\right).$$

Remark 5 *As mentioned in the Introduction the profile estimator does not require undersmoothing, however we still need the same type of undersmoothing condition in A5 for the nonparametric estimation of the selection probabilities, although a wider range of bandwidths can be used. Note that the asymptotic variances have the same structure as that of those given in Theorems 3 and 4, but they are different because of the profiling estimation. We also note that* $\widehat{\beta}_\tau^p$ *can be used in Theorem 2.*

3.3 Resampling

The asymptotic variances of the estimators of Theorems 3, 4 and 5 are rather complicated to estimate, so in this section we suggest a resampling technique that is based on the multiplier bootstrap and has been previously used in quantile regressions by Jin, Ying & Wei (2001), Zhou (2006) and Xie, Wan &

Zhou (2015) among others. We generate B random samples $\{\xi_i\}_{i=1}^n$ from the random variable ξ with $E(\xi)=1$ and $Var(\xi)=1$ and compute

$$\widehat{\beta}_\tau^* = \arg\min_{\beta\in B} Q_{\xi n}\left(\beta_\tau, \widehat{\theta}_\tau, \widehat{\pi}\right)$$

where

$$Q_{\xi n}\left(\beta_\tau, \widehat{\theta}_\tau, \widehat{\pi}\right) = \sum_{i=1}^n \frac{\delta_i \xi_i}{\widehat{\pi}(Z_{oi})} \rho_\tau\left(Y_i - X_{1i}^T\beta_\tau - X_{2i}^T \widehat{\theta}_\tau(X_{3i})\right)$$

for the two step iterative estimator $\widehat{\beta}_\tau$. For the profile estimator $\widehat{\beta}_\tau^p$ we compute

$$\widehat{\beta}_\tau^{p*} = \arg\min_{\beta_\tau\in B} ||M_{\xi n}(\beta_\tau, \widehat{\theta}_{\beta_\tau}, \partial\widehat{\theta}_{\beta_\tau}/\partial\beta_\tau^T)||,$$

where

$$M_{\xi n}(\beta_\tau, \widehat{\theta}_{\beta_\tau}, \partial\widehat{\theta}_{\beta_\tau}/\partial\beta_\tau^T) = \widehat{\Sigma}_4^{-1}\frac{1}{n}\sum_{i=1}^n \frac{\delta_i\xi_i}{\widehat{\pi}(Z_{oi})}(X_{1i}-\widehat{\varphi}^p(X_i))\rho_\tau'(Y_i - X_{1i}^T\beta_\tau - X_{2i}^T\widehat{\theta}_{\beta_\tau}(X_{3i}))(X_{1i}-\widehat{\varphi}^p(X_i)),$$

with

$$\widehat{\Sigma}_4 = \frac{1}{n}\sum_{i=1}^n \frac{\delta_i}{\widehat{\pi}_i}\widehat{f}_{\widehat{\varepsilon}_i|X_i}(0)(X_{1i} + \left(\frac{\partial\widehat{\theta}_{\beta_\tau}(X_{3i})}{\partial\beta_\tau^T}\right)^T X_{2i})^{\otimes 2},$$

$$\widehat{\varphi}^p(X_i) = \left[\left(\frac{1}{ng}\sum_{j\neq i}^n \frac{\delta_i}{\widehat{\pi}_i}X_{2i}^{\otimes 2}H_g(X_{3j}-X_{3i})\right)^{-1} \frac{1}{ng}\sum_{j\neq i}^n \frac{\delta_i}{\widehat{\pi}_i}X_{2i}X_{1i}^T H_g(X_{3j}-X_{3i})\right]^T X_{2i},$$

$\widehat{f}_{\widehat{\varepsilon}_i|X_i}(0)$ is a nonparametric (conditional) density estimator, $\widehat{\varepsilon}_i$ is the QPLVC residual and $H_g(\cdot)$ is another kernel with bandwidth g.

Theorem 6 *Under the same assumptions of Theorems 3-4, conditionally on* $\left(\left[Y_i, \delta_i, X_i^T\right]^T\right)_{i=1}^n$

$$n^{1/2}\left(\widehat{\beta}_\tau^* - \widehat{\beta}_\tau\right) \xrightarrow{d} N\left(0, \Sigma_2^{-1}\Sigma_{2*}\Sigma_2^{-1}\right),$$

where Σ_2 *and* Σ_{2*}*, with* $*$ *corresponding to either* Σ_{2p} *or* Σ_{2np}*, are given in Theorems 3-4.*
Under the assumptions of Theorem 5 and the additional assumption (i) $\sup_{X_i\in\mathcal{X}}\left|\widehat{f}_{\widehat{\varepsilon}_i|X_i}(0) - f_{\varepsilon|X}(0)\right| = o_p(1)$*, conditionally on* $\left(\left[Y_i, \delta_i, X_i^T\right]^T\right)_{i=1}^n$

$$n^{1/2}\left(\widehat{\beta}_\tau^{p*} - \widehat{\beta}_\tau^p\right) \xrightarrow{d} N\left(0, \Sigma_4^{-1}\Sigma_{4*}\Sigma_4^{-1}\right),$$

where Σ_{4*}*, with* $*$ *is either* Σ_{4p} *or* Σ_{4np} *given in Theorem 5.*

Theorem 6 shows that the proposed resampling technique consistently estimate the distributions of the estimators proposed in Sections 3.1 and 3.2. However, it is not sufficient to obtain consistent asymptotic variance estimators. To do so we need the following additional assumptions:

A7 (i) $E\left\|\left[X_1^T, X_2^T\right]^T\right\|^{2+\epsilon} < \infty$, (ii) $E\left\|s\left(Z, \alpha_0\right)\right\|^{2+\epsilon} < \infty$, (iii) $\inf_{z\in\mathcal{Z}}\left|\pi_0\left(Z\right)\right|^{2+\epsilon} > 0$, (iv) $E\left|\delta - \pi_0\left(Z\right)\right|^{2+\epsilon} < \infty$ and (v) $E\left|\xi\right|^{2+\epsilon} < \infty$ for some $\epsilon > 0$.

Let

$$\widehat{V}^* = \frac{1}{B}\sum_{b=1}^{B}\left(\widehat{\beta}_\tau^{*(b)} - \widehat{\beta}_\tau\right)^{\otimes 2}, \quad \widehat{V}^{p*} = \frac{1}{B}\sum_{b=1}^{B}(\widehat{\beta}_\tau^{p*(b)} - \widehat{\beta}_\tau^{p})^{\otimes 2},$$

denote the resampled variances, where $\widehat{\beta}_\tau^{*(b)}$ and $\widehat{\beta}_\tau^{p*(b)}$ denote the estimators from the $b-th$ sample.

Corollary 7 *Under the assumptions of Theorem 6 and A7, conditionally on* $\left(\left[Y_i, \delta_i, X_i^T\right]^T\right)_{i=1}^n$

$$\widehat{V}^* \xrightarrow{p} \Sigma_2^{-1}\Sigma_{2*}\Sigma_2^{-1}, \quad \widehat{V}^{p*} \xrightarrow{p} \Sigma_4^{-1}\Sigma_{4*}\Sigma_4^{-1}.$$

Corollary 7 is important because it can be used to obtain confidence intervals for β_τ using the normal approximation and test statistical hypotheses on β_τ using the χ^2 approximation and the delta method.

4 Some tests of statistical hypotheses

The results of Section 3 can be used to test statistical hypotheses about both the finite and infinite dimensional parameters β_τ and $\theta_\tau(\cdot)$. First, Theorem 2 can be used to construct a Wald statistic to test local hypotheses about $\theta_\tau(\cdot)$, that is hypotheses that are valid at a given point $x_3^* \in \mathcal{X}_3$. To investigate the asymptotic properties of such statistic, we consider the following local hypothesis with a Pitman drift

$$H_n : R\theta_\tau\left(x_3^*\right) = r_\tau\left(x_3^*\right) + \gamma_{n\tau}\left(x_3^*\right), \tag{4.1}$$

where R is an $l \times p$ $(l \le p)$ matrix of constants, $r_\tau\left(x_3^*\right)$ is an l-dimensional vector of known constants and $\gamma_{\tau n}(\cdot)$ is a bounded continuous function that may depend on n. Let

$$W_l\left(x_3^*\right) = nh\left(R\widehat{\theta}_\tau\left(x_3^*\right) - r_\tau\left(x_3^*\right)\right)^T\left(R\widehat{\Sigma}_3\left(x_3^*\right)^{-1}\widehat{\Sigma}_{3\widehat{\pi}}\left(x_3^*\right)\widehat{\Sigma}_3\left(x_3^*\right)^{-1}R^T\right)^{-1}\left(R\widehat{\theta}_\tau\left(x_3^*\right) - r_\tau\left(x_3^*\right)\right)$$

denote the local Wald statistic, where

$$\widehat{\Sigma}_3\left(x_3^*\right) = \widehat{f}_{X_3}\left(x_3^*\right)\frac{1}{nh}\sum_{i=1}^{n}\frac{\delta_i}{\widehat{\pi}\left(Z_{oi}\right)}\widehat{f}_{\widehat{\varepsilon}_i|X_i}\left(0\right)X_{2i}^{\otimes 2}K_h\left(X_{3i} - x_3^*\right), \tag{4.2}$$

$$\widehat{\Sigma}_{3\widehat{\pi}}\left(x_3^*\right) = \frac{\tau\left(1-\tau\right)v_0}{nh}\widehat{f}_{X_3}\left(x_3^*\right)\sum_{i=1}^{n}\frac{\delta_i}{\widehat{\pi}\left(Z_{oi}\right)^2}X_{2i}^{\otimes 2}K_h\left(X_{3i} - x_3^*\right),$$

and $\widehat{\pi}(\cdot)$ is either the parametric or the nonparametric estimator of $\pi_0(\cdot)$ described in Section 2.

Second, Theorem 2 can be used to test the global hypothesis

$$H_0 : \theta_\tau(\cdot) = \theta_{0\tau}(\cdot), \tag{4.3}$$

where $\theta_{0\tau}(\cdot)$ is a p-dimensional vector of known functions, where we use the term global to emphasize the fact that (4.3) is over the entire support $\mathcal{X}_3$ and not just over a given value x_3^* as in (4.1). Note that (4.3) includes the important hypothesis of constancy of the varying coefficients $\theta_\tau(\cdot)$, where $\theta_{0\tau}(\cdot)$ is assumed to be a possibly unknown constant function $\theta^c_{0\tau}$, see Proposition 12 below for more details.

To test for (4.3) we use the following distance statistic

$$D_{\widehat{\pi}}(\theta_{0\tau}) = \sum_{i=1}^{n} \frac{\delta_i}{\widehat{\pi}(Z_{oi})} \rho_\tau\left(Y_i - X_{1i}^T \widehat{\beta}_\tau - X_{2i}^T \widehat{\theta}_{\tau-i}(X_{3i})\right) - \sum_{i=1}^{n} \frac{\delta_i}{\widehat{\pi}(Z_{oi})} \rho_\tau\left(Y_i - X_{1i}^T \widehat{\beta}_\tau - X_{3i}^T \theta_{0\tau}(X_{3i})\right). \tag{4.4}$$

where $\widehat{\theta}_{\tau-i}(\cdot)$ is the leave-one-out version of the estimator considered in Theorem 2 (see (10.21) in the Appendix for a definition), and note that the test statistic (4.4) is in the same spirit as that of the generalized likelihood ratio proposed by Fan et al. (2001) for linear varying coefficients models.

Finally, we consider inference for the finite dimensional parameter β_τ; let

$$H_n : R\beta_\tau = r_\tau + \gamma_{n\tau}, \tag{4.5}$$

where R is an $l \times k$ $(l \leq k)$ matrix of constants and $\gamma_{\tau n}$ is a bounded continuous function that may depend on n. Let

$$\begin{aligned} W &= n\left(R\left(\widehat{\beta}_\tau - r_\tau\right)\right)^T \left(R\widehat{\Sigma}_2^{-1}\widehat{\Sigma}_{2*}\widehat{\Sigma}_2^{-1}R^T\right)^{-1} R\left(\widehat{\beta}_\tau - r_\tau\right) \\ W^p &= n(R(\widehat{\beta}^p_\tau - r_\tau))^T (R\widehat{\Sigma}_4^{-1}\widehat{\Sigma}_{4*}\widehat{\Sigma}_\tau^{-1}R^T)^{-1} R(\widehat{\beta}^p_\tau - r_\tau) \end{aligned} \tag{4.6}$$

denote the Wald statistics for (4.5), where $\widehat{\Sigma}_2$, $\widehat{\Sigma}_{2*}$, $\widehat{\Sigma}_4$ and $\widehat{\Sigma}_{4*}$ are estimators of the matrices of Theorems 3, 4 and 5 such as their sample analogues or those obtained using the resampling technique proposed in Section 3.3.

5 Asymptotic results for the statistical hypotheses tests

The following proposition establishes the asymptotic distribution of the local Wald statistic $W_l(x_3^*)$ under (4.1) as well as its consistency, under some mild high level assumptions, which can however be verified by standard assumptions on the uniform convergence of kernel estimators[1], see for example Masry (1996).

Proposition 8 *Under the assumptions of Theorem 2, if* $\operatorname{rank}(R) = l$, $\sup_{X_i \in \mathcal{X}} \left|\widehat{f}_{\widehat{\varepsilon}_i|X_i}(0) - f_{\varepsilon|X}(0)\right| = o_p(1)$, $\sup_{x_3 \in \mathcal{X}_3} \left|\widehat{f}(x_3) - f(x_3)\right| = o_p(1)$, $\sup_{Z \in \mathcal{Z}} |\widehat{\pi}(Z_i) - \pi_0(Z_i)| = o_p(1)$ *and* $nh^4 \to 0$, *then under* (4.1) *(i) for* $(nh)^{1/2}\gamma_{n\tau}(x_3^*) \to \gamma_\tau(x_3^*) > 0$ *(for some* $\|\gamma_\tau(x_3^*)\| < \infty$*)*

$$W_l(x_3^*) \xrightarrow{d} \chi^2(\kappa, l),$$

[1] For the parametric estimator of $\widehat{\pi}(Z_i) = \widehat{\pi}(Z_i, \widehat{\alpha})$, its uniform consistency follows by assuming that $E\sup_{\alpha \in A} \|\partial \pi(Z, \alpha)/\partial \alpha\|^\delta < \infty$ for $\delta > 2$, as in Assumption A4(iii).

where $\chi^2(\kappa, l)$ *is a noncentral Chi-squared distribution with* l *degrees of freedom and noncentrality parameter*

$$\kappa = f_{X_3}(x_3^*)\,\gamma_\tau(x_3^*)^T \left(R\Sigma_3(x_3^*)^{-1}\Sigma_{3\pi}(x_3^*)\Sigma_3(x_3^*)^{-1}R^T\right)^{-1}\gamma_\tau(x_3^*)\,;$$

(ii) for $(nh)^{1/2}\gamma_{\tau n}(x_3^*) \to \infty$,

$$W_l(x_3^*) \xrightarrow{p} \infty.$$

The following theorem establishes the asymptotic distribution of the distance statistic (4.4); let

$$\mu_\pi = \frac{tr}{2h}E\left[\frac{\tau(1-\tau)}{\pi_0(Z_o)\,f_{X_3}(X_3)}\Sigma_3(X_3)^{-1}X_2^{\otimes 2}\right]\kappa_2, \quad d_\pi = n^{1/2}h^2(T_{1\pi} - T_{3\pi}) - nh^4T_2,$$

$$T_{1\pi} = \frac{1}{n^{1/2}}\sum_{i=1}^n \frac{\delta_i}{\pi_0(Z_{oi})}X_{2i}^T\rho'_\tau(\varepsilon_i)\,\theta''_{0\tau}(X_{3i})\,\kappa_2,$$

$$T_2 = -\frac{1}{8}E\left[f_{\varepsilon|X}(0)\,\theta''_{0\tau}(X_3)^T X_2^{\otimes 2}\theta''_{0\tau}(X_3)\right]\int\int t^2(t+s)^2K(t)\,K(t+s)\,dtds,$$

$$T_{3\pi} = \frac{1}{2n^{1/2}}\sum_{i=1}^n \frac{\delta_i}{\pi_0(Z_{oi})}X_{2i}^T\rho'_\tau(\varepsilon_i)\,\theta''_{0\tau}(X_{3i})\int\int t^2(t+s)^2K(t)\,K(t+s)\,dtds,$$

$$\sigma_\pi^2 = \frac{2}{h}tr\left(E\left(\frac{\tau(1-\tau)}{\pi_0(Z_o)\,f_{X_3}(X_3)}\Sigma_{12}(X_3)^{-1}X_2^{\otimes 2}\right)^2\int(2K_h(t) - K_h * K_h(t))^2\,dt\right).$$

Theorem 9 *Under the assumptions of Theorem 2 and if* $h \to 0$ *and* $nh^{3/2} \to \infty$*, then*

$$\frac{1}{\sigma_\pi}\left(D_{\widehat{\pi}}(\theta_{0\tau}) - \mu_\pi - d_\pi\right) \xrightarrow{d} N(0,1).$$

Furthermore, if $\theta_{0\tau}(\cdot)$ *is linear or* $nh^4 \to 0$*, then*

$$\frac{1}{\sigma_\pi}\left(D_{\widehat{\pi}}(\theta_{0\tau}) - \mu_\pi\right) \xrightarrow{d} N(0,1).$$

Theorem 9 shows that the distance statistic $D_\pi(\theta_{0\tau})$, when appropriately scaled and centred, is asymptotically standard normal. As noted in the Introduction, as opposed to the generalized likelihood ratio statistic proposed by Fan et al. (2001), the Wilks' phenomenon does not hold for $D_\pi(\theta_{0\tau})$, because of the IPW estimation, see for example Bravo (2020). On the other hand, without the MAR observations, the Wilks' phenomenon still holds, as next proposition shows. Note that in this case, as in Fan et al. (2001), we use the full estimator $\widehat{\theta}_\tau(\cdot)$ and not its leave-one version $\widehat{\theta}_{\tau-i}(\cdot)$, hence the appearance of the constant $K(0)$ in Proposition 10. Let

$$D(\theta_{0\tau}) = \sum_{i=1}^n \rho_\tau\left(Y_i - X_{1i}^T\widehat{\beta}_\tau - X_{2i}^T\widehat{\theta}_\tau(X_{3i})\right) - \sum_{i=1}^n \rho_\tau\left(Y_i - X_{1i}^T\widehat{\beta}_\tau - X_{3i}^T\theta_{0\tau}(X_{3i})\right).$$

Proposition 10 *Under the assumptions of Theorem 9, if* $\theta_{0\tau}(\cdot)$ *is linear or* $nh^4 \to 0$ *and there are no MAR observations, then*

$$r_K D(\theta_{0\tau}) \xrightarrow{d} \chi^2(r_K\mu),$$

where

$$r_K = \left(K(0) - \frac{\kappa_2}{2}\right) / \int \left(K_h(t) - \frac{K_h * K_h(t)}{2}\right)^2 dt \text{ and}$$
$$\mu = \frac{p}{h} |\mathcal{X}_3| \tau (1-\tau) \left(K(0) - \frac{\kappa_2}{2}\right).$$

To compute the terms in the statistic $D_\pi(\theta_{0\tau})$, we need consistent estimators of μ_π, d_π and σ_π; let

$$\widehat{\mu}_{\widehat{\pi}} = \frac{1}{2nh} \sum_{i=1}^{n} \left[\frac{\tau(1-\tau)}{\widehat{\pi}(Z_{oi}) \widehat{f}_{X_3}(X_{3i})} \widehat{\Sigma}_3 (X_{3_i})^{-1} X_{2i}^{\otimes 2}\right] \kappa_2,$$
$$\widehat{T}_{1\widehat{\pi}} = \frac{1}{n^{1/2}} \sum_{i=1}^{n} \frac{\delta_i}{\widehat{\pi}(Z_{oi})} X_{2i}^T \rho'_\tau(\widehat{\varepsilon}_i) \widehat{\theta}''_\tau(X_{3i}) \kappa_2,$$
$$\widehat{T}_2 = -\frac{1}{8n} \sum_{i=1}^{n} \left[\widehat{f}_{\widehat{\varepsilon}_i|X_i}(0) \widehat{\theta}''_{0\tau}(X_{3i})^T X_{2i}^{\otimes 2} \widehat{\theta}''_\tau(X_{3i})\right] \int\int t^2 (t+s)^2 K(t) K(t+s)\, dtds,$$
$$\widehat{T}_{3\widehat{\pi}} = \frac{1}{2n^{1/2}} \sum_{i=1}^{n} \frac{\delta_i}{\widehat{\pi}(Z_{oi})} X_{2i}^T \rho'_\tau(\widehat{\varepsilon}_i) \widehat{\theta}''_\tau(X_{3i}) \int\int t^2 (t+s)^2 K(t) K(t+s)\, dtds\, (1 + o_p(1)),$$
$$\widehat{\sigma}^2_{\widehat{\pi}} = \frac{2}{nh} tr \left(\sum_{i=1}^{n} \left(\frac{\tau(1-\tau)}{\widehat{\pi}(Z_{oi}) \widehat{f}_{X_3}(X_{3i})} \widehat{\Sigma}_3 (X_{3i})^{-1} X_{2i}^{\otimes 2}\right)^2 \int (2K_h(t) - K_h * K_h(t))^2 dt\right),$$

where, as in Proposition 8 $\widehat{\pi}(\cdot)$, is either a parametric or nonparametric estimator of $\pi_0(\cdot)$, $\widehat{f}_{X_3}(\cdot)$ is a standard kernel estimator for the unknown density of X_3, $\widehat{\Sigma}_3(\cdot)$ and $\widehat{f}_{\widehat{\varepsilon}_i|X_i}(\cdot)$ are as defined in (4.2), $\widehat{\varepsilon}_i$ is the QPLVC residual and $\widehat{\theta}''_\tau(\cdot)$ is an estimator for the second derivative of the unknown parameter $\theta''_{0\tau}(\cdot)$, which can be computed, for example, using a local quadratic estimator. The following proposition is in the same spirit as Proposition 8 in terms of its regularity conditions.

Proposition 11 *Assume that* $\sup_{Z_{oi}\in\mathcal{Z}} |\widehat{\pi}(Z_{oi}) - \pi_0(Z_{oi})| = o_p(1)$, $\sup_{X_{3i}\in\mathcal{X}_3} \left|\widehat{f}_{X_3}(X_{3i}) - f_{X_3}(X_{3i})\right| = o_p(1)$, $\sup_{X_{3i}\in\mathcal{X}_3} \left|\widehat{\theta}''_{0\tau}(X_{3i}) - \theta''_{0\tau}(X_{3i})\right| = o_p(1)$, $\sup_{X_i\in\mathcal{X}} \left|\widehat{f}_{\widehat{\varepsilon}_i|X_i}(0) - f_{\varepsilon|X}(0)\right| = o_p(1)$*; then*

$$|\widehat{\mu}_{\widehat{\pi}} - \mu_\pi| = o_p(1),$$
$$\left|\widehat{T}_{j\widehat{\pi}} - T_{j\pi}\right| = o_p(1) \quad j = 1 \text{ and } 3,$$
$$\left|\widehat{T}_2 - T_2\right| = o_p(1)$$
$$\left|\widehat{\sigma}^2_{\widehat{\pi}} - \sigma^2_\pi\right| = o_p(1),$$

and

$$\frac{1}{\widehat{\sigma}_{\widehat{\pi}}} \left(D_{\widehat{\pi}}(\theta_{0\tau}) - \widehat{\mu}_{\widehat{\pi}} - \widehat{d}_{\widehat{\pi}}\right) \xrightarrow{d} N(0,1).$$

Theorem 9 and Proposition 11 can be used to test the empirically relevant hypothesis of constancy of the varying coefficients $H_0 : \theta_{0\tau}(\cdot) = \theta^c_\tau$, where θ^c_τ can be a specific value, say $\theta^c_{\tau 0}$, or is unknown, in which case it can be the parametric quantile estimator $\widehat{\theta}_\tau$; let

$$D_{\widehat{\pi}}\left(\theta_\tau^c\right) = \sum_{i=1}^n \frac{\delta_i}{\widehat{\pi}\left(Z_{oi}\right)} \rho_\tau \left(Y_i - X_{1i}^T \widehat{\beta}_\tau - X_{2i}^T \widehat{\theta}_{\tau - i}\left(X_{3i}\right) \right) - \sum_{i=1}^n \frac{\delta_i}{\widehat{\pi}\left(Z_{oi}\right)} \rho_\tau \left(Y_i - X_{1i}^T \widehat{\beta}_\tau - X_{2i}^T \theta_\tau^c \right), \tag{5.1}$$

denote the resulting distance statistic.

Proposition 12 *Under the same assumptions of Theorem 9*

$$\frac{1}{\widehat{\sigma}_{\widehat{\pi}}} \left(D_{\widehat{\pi}}\left(\theta_\tau^c\right) - \widehat{\mu}_{\widehat{\pi}} \right) \xrightarrow{d} N\left(0, 1\right).$$

To investigate the power properties of the statistic $D_\pi\left(\theta_{0\tau}\right)$, we focus on the case where $\theta_{0\tau}$ is linear (or assume that $h = o\left(n^{-1/4}\right)$ so that the term d_π can be ignored asymptotically). We consider local hypotheses of the form

$$H_n : \theta_{n\tau}\left(\cdot\right) = \theta_{0\tau}\left(\cdot\right) + \gamma_{n\tau}\left(\cdot\right), \tag{5.2}$$

where $\gamma_{n\tau}\left(\cdot\right)$ is a bounded function with bounded first and second derivatives, and note that $\gamma_{n\tau}\left(\cdot\right) = \gamma_\tau\left(\cdot\right) / \left(nh\right)^{1/2}$ corresponds to the standard Pitman drift. Let

$$d_n = \frac{n}{2} E\left(f_{\varepsilon|X}\left(0\right) \gamma_{n\tau}\left(X_3\right)^T X_2^{\otimes 2} \gamma_{n\tau}\left(X_3\right) \right) - \frac{nh^4}{8} E\left[f_{\varepsilon|X}\left(0\right) \gamma_{n\tau}''\left(X_3\right)^T X_2^{\otimes 2} \gamma_{n\tau}''\left(X_3\right) \right] \int\int t^2 \left(t+s\right)^2 K\left(t\right) K\left(t+s\right) dtds$$

$$\sigma_{\gamma\pi}^2 = \sigma_\pi^2 + nE\left(\frac{\tau\left(1-\tau\right)}{\pi_0\left(Z_o\right)} X_2^T \gamma_{n\tau}\left(X_3\right)^{\otimes 2} X_2 \right)$$

Theorem 13 *Under the same assumption of Theorem 9 and (5.2), if* $nhE\left(\gamma_{n\tau}\left(X_3\right)^T X_2^{\otimes 2} \gamma_{n\tau}\left(X_3\right) \right) = O\left(1\right)$ *and* $E\left(\gamma_{n\tau}\left(X_3\right)^T X_2^{\otimes 2} \gamma_{n\tau}\left(X_3\right) \rho'\left(\varepsilon\right)^2 \right)^2 = O\left(\left(nh\right)^{-3/2} \right)$*, then*

$$\frac{1}{\widehat{\sigma}_{\gamma\widehat{\pi}}} \left(D_{\widehat{\pi}}\left(\theta_{0\tau}\right) - \widehat{\mu}_{\widehat{\pi}} - \widehat{d}_n \right) \xrightarrow{d} N\left(0, 1\right),$$

where $\widehat{\sigma}_{\gamma\widehat{\pi}}$ *and* $\widehat{d}_{2n}$ *are the sample analogues of* $\sigma_{\gamma\pi}$ *and* d_n*.*

Finally, the following proposition establishes the asymptotic distributions of the Wald statistics W and W^p given in (4.6) under (4.5) as well as their consistency.

Proposition 14 *Under the assumptions of Theorems 3 and 4, if* $\operatorname{rank}\left(R\right) = l, \left\| \widehat{\Sigma}_2 - \Sigma_2 \right\| = o_p\left(1\right)$, $\left\| \widehat{\Sigma}_{2*} - \Sigma_{2*} \right\| = o_p\left(1\right)$ *and* $nh^4 \to 0$*, then under (4.5) (i) for* $n^{1/2}\gamma_{\tau n} \to \gamma_\tau > 0$ *(for some* $\left\|\gamma_\tau\right\| < \infty$*)*

$$W \xrightarrow{d} \chi^2\left(\kappa, l\right),$$

where $\chi^2(\kappa, l)$ *is a noncentral Chi-squared distribution with* l *degrees of freedom and noncentrality parameter* $\kappa = \gamma_\tau^T \left(R\Sigma_2^{-1}\Sigma_{2*}\Sigma_2^{-1}R^T\right)^{-1}\gamma_\tau$*; (ii) for* $n^{1/2}\gamma_{\tau n} \to \infty$,

$$W \xrightarrow{p} \infty.$$

Under the assumptions of Theorem 5, if $rank(R) = l$, $||\widehat{\Sigma}_4 - \Sigma_4|| = o_p(1)$, $||\widehat{\Sigma}_{4*} - \Sigma_{4*}|| = o_p(1)$, *then under (4.5) (i) for* $n^{1/2}\gamma_{\tau n} \to \gamma_\tau > 0$ *(for some* $\|\gamma_\tau\| < \infty$*)*

$$W^p \xrightarrow{p} \infty,$$

where $\chi^2(\kappa, l)$ *is a noncentral Chi-squared distribution with* l *degrees of freedom and noncentrality parameter* $\kappa = \gamma_\tau^T \left(R\Sigma_4^{-1}\Sigma_{4*}\Sigma_4^{-1}R^T\right)^{-1}\gamma_\tau$*; (ii) for* $n^{1/2}\gamma_{\tau n} \to \infty$,

$$W^p \xrightarrow{p} \infty.$$

6 Simulation study

We first discuss some computational aspects of the proposed estimators and describe how to use the MM algorithm to estimate the unknown parameters. We begin with the two step iterative estimator; let $\varepsilon_{i(k)} = Y_i - X_{1i}^T\beta_{\tau(k)} - X_{2i}^T\theta_{\tau(k)}(X_3)$ denote the kth iterate in finding the minimum of the objective function and let

$$\varsigma_\tau\left(\varepsilon_i|\varepsilon_{i(k)}\right) = \frac{1}{4}\left[\frac{\varepsilon_i^2}{\epsilon + \left|\varepsilon_{i(k)}\right|} + (4\tau - 2)\,\varepsilon_i + c_{(k)}\right]$$

denote the so-called surrogate function, where the constant $c_{(k)}$ is such that $\varsigma\left(\varepsilon_{(k)}|\varepsilon_{(k)}\right)$ is equal to $\rho_\tau\left(\varepsilon_{(k)}\right)$ and $0 < \epsilon \leq 1$ is a tuning parameter to be selected. Then, since $\varsigma\left(\varepsilon_i|\varepsilon_{i(k)}\right) \geq \rho_\tau(\varepsilon_i)$ for all ε_i, the unknown parameters can be estimated by minimising both the local and the global majorising objective functions

$$\sum_{i=1}^n \frac{\delta_i}{\widehat{\pi}(Z_{oi})}\varsigma_\tau\left(\varepsilon_i|\varepsilon_{i(k)}\right)K_h(X_{3i} - x_3), \quad \sum_{i=1}^n \frac{\delta_i}{\widehat{\pi}(Z_{oi})}\varsigma_\tau\left(\widehat{\varepsilon}_i|\widehat{\varepsilon}_{i(k)}\right),$$

where $\widehat{\varepsilon}_i = Y_i - X_{1i}^T\beta_\tau - X_{2i}^T\widehat{\theta}_\tau(X_{3i})$. As in Hunter & Lange (2000), we use the Gauss-Newton algorithm with direction

$$\begin{aligned}\Delta_{(k)}(x_3) &= -\left[X(x_3)^T W\left(\delta, \widehat{\pi}(\cdot), \varepsilon_{(k)}, K\right)X(x_3)\right]^{-1} X(x_3)^T d\left(\delta, \widehat{\pi}(\cdot), \varepsilon, K\right),\\ \Delta_{(k)} &= -\left[X_1^T W\left(\delta, \widehat{\pi}(\cdot), \varepsilon_{(k)}\right)X_1\right]^{-1} X_1^T d\left(\delta, \widehat{\pi}(\cdot), \varepsilon\right),\end{aligned}$$

where $X(x_3)$ is an $n \times (k+2p)$ matrix containing the k, p and p covariates X_{1i}^T, X_{2i}^T and $X_{2i}^T(X_{3i}-x_3)$ $(i=1,...,n)$,

$$W\left(\delta,\widehat{\pi}(\cdot),\varepsilon_{(k)},K\right) = diag\left[\frac{\delta_1}{\widehat{\pi}(Z_{o1})}\frac{1}{\epsilon+\varepsilon_{1(k)}}K_h(X_{31}-x_3),...,\right.$$
$$\left.\frac{\delta_n}{\widehat{\pi}(Z_{on})}\frac{1}{\epsilon+\varepsilon_{n(k)}}K_h(X_{3n}-x_3)\right]^T,$$
$$d(\delta,\widehat{\pi}(\cdot),\varepsilon,K) = \left[\frac{\delta_1}{\widehat{\pi}(Z_{o1})}\left(1-2\tau-\frac{\varepsilon_1}{\epsilon+\varepsilon_1}\right)K_h(X_{31}-x_3),...,\right.$$
$$\left.\frac{\delta_n}{\widehat{\pi}_(Z_{on})}\left(1-2\tau-\frac{\varepsilon_n}{\epsilon+\varepsilon_n}\right)K_h(X_{3n}-x_3)\right]^T,$$

with $W\left(\delta,\widehat{\pi}(\cdot),\varepsilon_{(k)}\right)$ and $d(\delta,\widehat{\pi}(\cdot),\varepsilon)$ defined similarly.

The implementation of the MM algorithm for the two step iterative estimator involves the following steps:

1. Set $k=0$, choose *either* the initial values $\left[\beta_\tau^{0T},a_\tau^{0T},b_\tau^{0T}\right]^T$ *or* β_τ^0 and set $\epsilon n\,|\ln\epsilon| = \delta$, with $\delta = 10^{-6}$,
2. Define *either* $\left[\beta_\tau^{k+1T},a_\tau^{k+1T},b_\tau^{k+1T}\right]^T = \left[\beta_\tau^{kT},a_\tau^{kT},b_\tau^kT\right]^T + \Delta_{(k)}(x_2)/2^k$ *or* $\beta_\tau^k = \beta_\tau^k + \Delta_{(k)}/2^k$,
3. Iterate until *either* $\left\|\left[\beta_\tau^{k+1T},a_\tau^{k+1T},b_\tau^{k+1T}\right]^T - \left[\beta_\tau^{kT},a_\tau^{kT},b_\tau^{kT}\right]^T\right\| < \delta$ *or* $\left\|\beta_\tau^{k+1}-\beta_\tau^k\right\| < \delta$.

For the profile estimator we solve directly the first order conditions

$$\sum_{i=1}^n \frac{\delta_i}{\widehat{\pi}(Z_{oi})}\frac{\partial\varsigma_\tau(\widehat{\varepsilon}_i|\widehat{\varepsilon}_{i(k)})}{\partial\beta_\tau^T} = 0, \tag{6.1}$$

for β_τ,, where $\widehat{\varepsilon}_i = Y_i - X_{1i}^T\beta_\tau - X_{2i}^T\widehat{\theta}_{\beta_\tau}(X_{3i})$.
Given an initial value β_τ^0, the computation of the estimator can be carried out with few iterations (typically one or two) until $||\beta_\tau^{p,k+1}-\beta_\tau^{p,k}|| < \delta$ with $\delta = 10^{-6}$.
Next, we discuss how to choose the bandwidths b, h and g. For the profile estimator we use standard cross-validation for b and h, whereas we use $g = s(X_{3i})n^{-1/7}$ with $s(X_{3i})$ the sample standard deviation of X_{3i}. For the two step iterative estimator we still use cross-validation for b, but because of the assumed undersmoothing, the choice of h is more delicate because of the nonparametric nature of the estimation in Step 1, for which, as noted by El Gouch & van Keilegom (2009), the problem of optimally choosing the bandwidth is still an open one. However, given the plug-in nature of the estimation in Step 2, as long as the selected bandwidth does not result in a large bias for the infinite dimensional parameter estimator, the finite dimensional parameter estimator should not be very sensitive to the bandwidth choice, see Bickel & Kwon (2002) for a thorough discussion on this important point. In this paper, we propose a two-fold method, which consists of computing for a random subset of the sample - the

training set - S_t with $0 < t < 1$

$$\left[\beta_\tau^{-tT}, a_\tau^{-tT}, b_\tau^{-tT}\right]^T (h) = \arg\min_{\beta_\tau, a_\tau, b_\tau} \sum_{i \in S_t} \frac{\delta_i}{\widehat{\pi}(Z_{oi})} \varsigma_\tau \left(\varepsilon_i | \varepsilon_{i(k)}\right) K_h (X_{3i} - x_3),$$

$$\widehat{\beta}_\tau^{-t}(h) = \arg\min_{\beta_\tau} \sum_{i \in S_t} \frac{\delta_i}{\widehat{\pi}(Z_{oi})} \varsigma_\tau \left(\widehat{\varepsilon}_i^{-t} | \widehat{\varepsilon}_{i(k)}^{-t}\right),$$

where $\widehat{\varepsilon}_i^{-t} = Y_i - X_{1i}^T \beta_\tau^{-t} - X_{2i}^T \widehat{\theta}_\tau^{-t}(X_{3i})$ and then using the remaining part of the sample S_{1-t} - the validation set- to select h as

$$\widehat{h} = \arg\min_{h} \sum_{i \in S_{1-t}} \frac{\delta_i}{\widehat{\pi}(Z_{oi})} \varsigma_\tau \left(\widehat{\varepsilon}_i^{-t}(h) | \widehat{\varepsilon}_{i(k)}^{-t}(h)\right). \tag{6.2}$$

In the simulations, 80% of the sample is used as the training set and the remaining 20% is used as the validation set.

We consider the following QPLVC model

$$Y_i = X_{1i}^T \beta_{0\tau} + X_{2i}^T \left[\cos(\pi X_{3i}), X_{3i}^2\right]^T + \varepsilon_{i\tau} \quad i = 1, ..., n, \tag{6.3}$$

where $\beta_{0\tau} = [\beta_{10\tau}, \beta_{20\tau}]^T = [1, 1/4]^T$, $X_{1i} = [1, X_{11i}]^T$, X_{11i} is $N(0, 0.2)$, $X_{2i} = [X_{21i}, X_{22i}]^T$ is a bivariate normal with unit variance and correlation coefficient $\rho = 0.1$, X_{3i} is $U(0, 2)$ and the unobservable (zero τ quantile) error term $\varepsilon_{i\tau}$ generated independently from the X_i covariates as either a standard normal or a t distribution with 5 degrees of freedom ($t(5)$) or a (centred) Chi-squared distribution with 4 degrees of freedom ($\chi^2(4) - 4$); the selection probabilities (2.2) are specified as either

$$\pi_0(Z_{oi}) = \frac{\exp(\alpha_{10} + \alpha_{20} X_{21i} + \alpha_{30} X_{3i})}{1 + \exp(\alpha_{10} + \alpha_{20} X_{21i} + \alpha_{30} X_{3i})}, \tag{6.4}$$

or

$$\pi_0(Z_{oi}) = \frac{\exp(\alpha_{10} + \alpha_{20} Y_i + \alpha_{30} X_{3i})}{1 + \exp(\alpha_{10} + \alpha_{20} Y_i + \alpha_{30} X_{3i})}, \tag{6.5}$$

corresponding, respectively, to the cases where some of the responses Y_i and of the covariates X_{11i} and X_{22i} are MAR (6.4), and some of the covariates in X_{11i} and X_{2i} are MAR (6.5) with $\alpha_0 = [\alpha_{10}, \alpha_{20}, \alpha_{30}]^T$ chosen so that the average percentage of missing at the τ quantile are approximately 10% and 40%.

In the simulations, we use the Epanechnikov kernel for $K(\cdot)$, $L(\cdot)$, and $H(\cdot)$ with bandwidth $h = n^{-2/9}\widehat{h}$ with $\widehat{h}$ defined in (6.2) for $K(\cdot)$ for the two step iterative estimator. We consider three quantiles $\tau = [0.25, 0.5, 0.75]^T$, two sample sizes: $n = 100$ and $n = 400$ and six different estimators for $[\beta_{10\tau}, \beta_{20\tau}]^T$, namely the complete case $\left[\widehat{\beta}_{1\tau c}, \widehat{\beta}_{2\tau c}\right]^T$, $\left[\widehat{\beta}_{1\tau c}^p, \widehat{\beta}_{2\tau c}^p\right]^T$ and the IPW based $\left[\widehat{\beta}_{1\tau p}, \widehat{\beta}_{2\tau p}\right]^T$, $\left[\widehat{\beta}_{1\tau np}, \widehat{\beta}_{2\tau np}\right]^T$, $\left[\widehat{\beta}_{1\tau p}^p, \widehat{\beta}_{2\tau np}^p\right]^T$ and $\left[\widehat{\beta}_{1\tau np}^p, \widehat{\beta}_{2\tau np}^p\right]^T$ estimators. Tables 1a-3c report the absolute bias (*bias*), standard error (*se*), average length (*length*) and coverage (*cov*) of nominal 95% confidence intervals for the six proposed estimators based on 1000 replications, with standard errors calculated using the resampling technique of Section 3.3 with the number of replications B set to 500 and the random variables ξ_i generated from an Exponential distribution with mean 1.

Tables 1a-3c approximately here

The first two rows of each tables report the finite sample properties of the estimators $\left[\widehat{\beta}_{1\tau},\widehat{\beta}_{2\tau}\right]^T$ and $\left[\widehat{\beta}^p_{1\tau},\widehat{\beta}^p_{2\tau}\right]^T$ for the case without missing observations, and are used as benchmark for the missing observations cases. We note that across the three different quantiles and distributions of the unobservable errors the finite sample biases are statistically insignificant, the standard errors and average lengths of the confidence intervals are decreasing by a factor of two as the sample size is increased fourfold, as implied by the asymptotic theory developed in the previous section, whereas the confidence intervals are characterized by some undercoverage, which is however diminishing as the sample size increases. With missing observations, a number of clear patterns emerge: first, as the percentage of MAR observations increases the bias of the complete estimators increases (albeit it is still statistically insignificant), whereas that of the IPW estimators is comparable to that of the estimators without missing observations for both sample sizes. The profile estimator seems to have slightly better standard errors, average lengths and coverage of the confidence intervals. Second, as expected, the standard errors of the IPW estimators are typically larger than those based on the complete case, and this is reflected in the average length of the corresponding confidence intervals, which are slightly longer than those based on the complete case. Third, the coverage of the confidence intervals for the complete case show considerable undercoverage compared to those based on the IPW estimators.

Figure 1 shows the nonparametric quantiles estimates at $\tau=[0.25,0.5,0.75]^T$ of the two unknown infinite dimensional parameters for the case of no missing observations and two different distributions of the unobservable errors $\varepsilon_{i\tau}$. Figure 2 shows the nonparametric quantile estimates with 40% missing observations under the (6.5) MAR mechanism and IPW based on the nonparametric estimator (2.9) for the selection probabilities. Figure 2 clearly shows that despite the missing observations the IPW based estimates fit well the original unknown infinite dimensional parameters.

Figures 1-2 approximately here

In the remaining part of this section we only consider the MAR mechanism (6.5), as the results based on (6.4) are similar or slightly better, especially for the IPW based estimators. We first consider the finite sample properties of the distance statistic (4.4). Tables 4a-4b and Figure 3 report the finite sample size and power of (4.4) for the hypothesis

$$H_n:\theta_{1\tau}\left(X_3\right)=\left(1+\gamma\right)\cos\left(\pi X_3\right);\ \theta_{2\tau}\left(X_3\right)=\left(1+\gamma\right)X_3^2 \tag{6.6}$$

for $\gamma=[-1,-0.9,...,0.9,1]$ with $\gamma=0$ corresponding to the null hypothesis. The results are based on 1000 replications with the same bandwidth as that chosen in the previous simulation. The tables show that with 10% MAR observations the finite sample sizes of the $D_\pi\left(\theta_{0\tau}\right)$ statistic based on the complete case and IPW based estimators are broadly comparable, whereas with 40% MAR observations the $D_\pi\left(\theta_{0\tau}\right)$ statistic based on the complete case estimator is characterized by a considerably bigger size distortion compared to that based on both the IPW estimators. Figure 3 clearly shows that the size adjusted finite sample power of the statistic $D_\pi\left(\theta_{0\tau}\right)$ based on the IPW estimators is higher compared to the one based on the complete estimator.

Tables 4a-4b approx. here

Figure 3 approx. here

Figure 4 demonstrates the Wilks' phenomenon for the scaled statistic $D(\theta_{0\tau})$ defined in Proposition (10). The figure is based on a kernel estimate of the distribution of the statistic based on 1000 simulations under the null hypothesis $H_n : \theta_{1\tau}(X_3) = \cos(\pi X_3)$ with no missing observations and three bandwidths, namely the same one used in the previous simulations b and two alternative ones based on 1/2 and 3/2 of b. As expected, the simulated distribution looks like a Chi-squared regardless of the bandwidth choice.

Figure 4 approx. here

Next, we consider the finite sample properties of the statistic $D_\pi(\theta_\tau^c)$ defined in (5.1) for the constancy of the functional parameters. Tables 5a-5c show the finite sample power of $D_\pi(\theta_\tau^c)$ for the hypothesis

$$H_n = \theta_{1\tau}(X_3) = \gamma\cos(\pi X_3)\,;\;\; \theta_{2\tau}(X_3) = \gamma X_3^2$$

with $\gamma = [-1, -0.8, ..., 0, ..., 0.8, 1]$ with $\gamma = 0$ corresponding to the null hypothesis. The results are based on 1000 replications with 40% MAR observations using the same undersmoothed bandwidth as that used in the previous simulations, and they show that the $D_\pi(\theta_\tau^c)$ statistic finite sample performance in terms of both size and (size adjusted) power is clearly better for $D_\pi(\theta_\tau^c)$ based on the IPW estimators.

Finally, we consider the finite sample properties of the Wald statistics W and W^p (4.6) for the finite dimensional parameter $\beta_{0\tau} = [\beta_{10\tau}, \beta_{20\tau}]^T$ in (6.3). The null hypothesis is specified as $H_0 = [\beta_{10\tau}, \beta_{20\tau}]^T = [1, 1/4]^T$ with the alternative hypothesis specified as the grid $\gamma = [\gamma_1, \gamma_2] = [-1, -0.8, ..., 0.8, 1] \times [-1, -0.8, ..., 0.8, 1]$. Tables 6a-6b report the finite sample sizes of W and W^p, using 1000 replications and the asymptotic variances $\widehat{\Sigma}_2^{-1}\widehat{\Sigma}_{2*}\widehat{\Sigma}_2^{-1}$ and $\widehat{\Sigma}_4^{-1}\Sigma_{4*}\widehat{\Sigma}_4^{-1}$ estimated by the same resampling technique of Section 3.3 used to compute the standard errors of Tables 1a-3c.

Tables 6a-6b approx. here

As with Tables 4a-4b, Tables 6a-6b show that with 10% MAR observations the finite sample sizes of the W andW^p statistics based on the complete case and IPW based estimators are broadly comparable, whereas with 40% MAR observations the W and W^p statistic based on the complete case estimator are characterized by larger size distortions compared to those based on both the IPW estimators. We also note that W^p has slightly better finite sample properties than those of the W statistic. Figure 5 shows the contour plots at the level 0.40 of the size adjusted finite sample powers of W^p with 40% MAR observations, $N(0,1)$ unobservable errors and $n = 400$. Note that smaller contour plots indicate higher finite sample power.

Figure 5 approx. here

7 Empirical application

We illustrate the applicability of the proposed estimation and inference methods by considering the New York air quality measurements data (from May to September 1973, available in the R package `datasets` which consists of 153 daily observations of mean ozone parts (per billion) (O), solar radiations

(S), wind speed (in mph) (W) and temperature (in degrees F) (T) and contains 37 missing ozone parts observations (missing rate of around 24%) and 7 missing solar radiations observations (missing rate of around 4.5%). As some of the missing responses and the covariates are missing at the same time, the overall missing rate is around 27.3%. After some preliminary data analysis, the following quantile partial linear regression specification

$$q_\tau\left(O|S,W,T\right) = \beta_{1\tau} + \beta_{2\tau}S + \beta_{3\tau}T + \theta_\tau\left(W\right), \tag{7.1}$$

is chosen; we consider the complete case $\widehat{\beta}_{j\tau c}$, IPW parametric $\widehat{\beta}_{j\tau p}$ and IPW nonparametric $\widehat{\beta}_{j\tau np}$ estimators ($j = 1, 2, 3$) for the three quantiles $\tau = [0.25, 0.50, 0.75]^T$, with the selection probabilities $\pi\left(T, W\right)$, which seems plausible given the well-known results of the effects of the temperature and the wind on the ozone level and solar radiation, estimated either with a standard logit model or a bivariate product Epanechnikov kernel $L\left(T, W\right)$. Tables 7a-7c report the estimates, standard errors, length of 95% confidence intervals and p-values of the three different sets of estimators, with the standard errors calculated using the same resampling technique of Section 3.3.

Tables 7a-7c approx. here

Tables 7a-7c show that, across the three estimators, at the 0.25 quantile there is a positive relationship between solar radiations and the mean ozone parts, but the same relation becomes statistically insignificant at the higher quantiles. Temperature is also positively related with the mean ozone parts, but as opposed to the solar radiations, the relationship is statistically significant at the three quantiles, which confirms the widely accepted view among climate and environmental scientists that there is a positive relationship between ozone (hence pollution) and temperature. Figure 6 shows the nonparametric quantile estimates for $\theta_\tau\left(W\right)$; interestingly, as opposed to the finite dimensional parameters case, there is a notable difference between the complete case estimator and the IPW based ones, as the former shows a pattern that is counter-intuitive in that the wind speed negatively affects the mean ozone parts up to a certain speed and then the relationship becomes positive. On the other hand, both IPW estimators show a negative relationship between the ozone level and the wind speed, which seems to be more in line with current empirical evidence, see for example Jammalamadoka & Lund (2006).

Figure 6 approx. here

To this end, we tested the constancy of the infinite dimensional parameter $\theta_\tau\left(W\right)$ using the statistic (5.1) with the quantile parametric estimate as $\theta^c_\tau\left(W\right)$; Table 8 reports the corresponding sample values and corresponding p-values, which clearly supports the quantile partially linear specification 7.1.

Table 8 approx. here

To further support the chosen semiparametric specification, we compare the local goodness of fit measures $R^1_{\tau *}$ proposed by Koenker & Machado (1999a)[2], where $*$ indicates the complete case, IPW

[2]The local, as it depends on the chosen quantile τ, goodness of fit $R^1_{\tau *}$ is defined as $1 - \widehat{V}_{\tau *}/\widetilde{V}_{\tau *}$, where $\widehat{V}_{\tau *} = \sum_{i=1}^n \rho_\tau\left(\widehat{\varepsilon}_{i*}\right)$, $\widetilde{V}_{\tau *} = \sum_{i=1}^n \rho_\tau\left(\widetilde{\varepsilon}_{i*}\right)$, and $\widehat{\varepsilon}_{i*}$, $\widetilde{\varepsilon}_{i*}$ are the residuals of the unrestricted and the restricted quantile regressions, respectively. Here the restricted quantile regression model consists only of the intercept.

parametric and IPW nonparametric estimators, between (7.1) and the restricted parametric model $q_\tau(O|S,W,T) = \beta_{1\tau} + \beta_{2\tau}S + \beta_{3\tau}T + \beta_{4\tau}W$.

Table 9 approx. here

Table 9 clearly shows that the chosen quantile semiparametric specification has a higher $R^1_{\tau*}$ compared to that of the parametric one, across the three estimators and three chosen quantiles.

8 Conclusions

In this paper we propose a general method to estimate and test statistical hypotheses of the unknown parameters in QPLVC models when some of the observations are missing at random. The proposed estimators are based on the IPW method and can be efficiently computed using the MM algorithm. For inference, we consider Wald statistics that can be used to test local linear hypotheses for the infinite dimensional parameter and linear hypotheses for the finite dimensional parameter; we also consider a distance statistic that can be used to test global hypotheses on the infinite dimensional parameter, including the important one of constancy over the whole support of the underlying conditioning random variate. Monte Carlo simulations show that the proposed IPW based estimators perform well (compared to those based on the complete case) in finite samples, especially when the percentage of MAR observations is higher, and similarly for both the Wald and distance statistics. Finally, an empirical application illustrates the applicability and usefulness of the proposed estimation and inference methods.

References

Abrevaya, J. & Donald, S. (2017), 'A GMM approach for dealing with missing data on regressors', *Review of Economics and Statistics* **99**, 657–662.

Basset, G. & Koenker, R. (1978), 'Asymptotic theory of least absolute error regression', *Journal of the American Statistical Association* **73**, 618–622.

Bianco, A., Boente, G., Gonzales-Mantiega, W. & Perez-Gonzales, A. (2010), 'Estimation of the marginal location under a partially linear model with missing responses', *Computational Statistics and Data Analysis* **54**, 546–564.

Bickel, P. & Kwon, J. (2002), 'Inference for semiparametric models: Some current frontiers and an answer (with discussion)', *Statistica Sinica* **11**, 863–960.

Bose, A. & Chatterjee, S. (2003), 'Generalized bootstrap for estimators of minimizers of convex functions', *Journal of Statistical Planning and Inference* **117**, 1238–1249.

Bravo, F. (2020), 'Robust estimation and inference for general varying coefficient models with missing observations', *TEST* **29**, 966–988.

Cai, Z. & Xiao, Z. (2012), 'Semiparametric quantile regression estimation in dynamic models with partially varying coefficients', *Journal of Eonometrics* **167**, 413–425.

Cai, Z. & Xu, X. (2008), 'Nonparametric quantile estimations for dynamic smooth coefficient models', *Journal of the American Statistical Association* **103**, 1595–1608.

Chauduri, P. (1991), 'Nonparametric estimates of regression quantiles and their local Bahadur representation', *Annals of Statistics* **19**, 760–777.

Chauduri, P., Doksum, K. & Samarov, A. (1997), 'On average derivative quantile regression', *Annals of Statistics* **25**, 715–744.

Chen, X., Hong, H. & Tamer, E. (2005), 'Measurement error models with auxiliary data', *Review of Economic Studies* **72**, 343–366.

Chen, X., Wan, A. & Zhou, Y. (2015), 'Efficient quantile regression analysis with missing observations', *Journal of the American Statistical Association* **110**, 723–741.

Cheng, G. & Huang, J. (2010), 'Bootstrap consistency for general semiparametric m-estimation', *Annals of Statistics* **38**, 2884–2915.

de Jong, P. (1987), 'A central limit theorem for quadratic forms', *Probability Theory and Related Fields* **75**, 261–277.

El Gouch, A. & van Keilegom, I. (2009), 'Local linear quantile regression with dependent censored data', *Statistica Sinica* **19**, 1621–1640.

Fan, J. & Gijbels, I. (1996), *Local polynomial modeling and its applications*, Chapman Hall, London, UK.

Fan, J., Zhang, C. & Zhang, J. (2001), 'Generalized likelihood ratio statistics and Wilks' phenomenon', *Annals of Statistics* **29**, 153–193.

Firpo, S. (2007), 'Efficient semiparametric estimation of quantile treatment effects', *Econometrica* **75**, 259–276.

Guerre, E. & Sabbah, C. (2012), 'Uniform bias study and bahadur representation for local polynomial estimators of the conditional quantile function', *Econometric Theory* **28**, 87–129.

Hirano, K., Imbens, G. & Ridder, G. (2003), 'Efficient estimation of average treatment effects using the estimated propensity score', *Econometrica* **71**, 1161–1189.

Horvitz, D. & Thompson, D. (1952), 'A generalization of sampling without replacement from a finite universe', *Journal of the American Statistical Association* **47**, 663–685.

Hu, Z., Wang, N. & Carroll, R. (2004), 'Profile kernel vs backfitting in the partially linear model for longitudinal/clustered data', *Biometrika* **91**, 251–262.

Hunter, D. & Lange, K. (2000), 'Quantile regression via an mm algorhitm', *Journal of Compu- tational and Graphical Statistics* **9**, 60–77.

Imbens, G. (2004), 'Nonparametric estimation of average treatment effects and exogeneity: A review', *Review of Economics and Statistics* **86**, 4–29.

Jammalamadoka, S. & Lund, E. (2006), 'The effect of wind direction on ozone level:A case study', *Enviromental and Ecological Statistics* **13**, 287–298.

Jin, Z., Ying, Z. & Wei, L. (2001), 'A simple resampling method by perturbing the minimand', *Biometrika* **88**, 381–390.

Kai, B., Li, R. & Zou, H. (2011), 'New efficient estimation and variable selection methods for semi-parametric varying coefficients partially linear models', *Annals of Statistics* **39**, 305–332.

Kim, M. (2007), 'Quantile regression with varying coefficients', *Annals of Statistics* **35**, 92–108.

Knight, K. (1999), 'Limiting distributions for L1 regression estimators under general conditions', *Annals of Statistics* **26**, 755–770.

Koenker, R. (2005), *Quantile regression*, Cambridge Universty Press, Cambridge, UK.

Koenker, R. & Bassett, G. (1978), 'Regression quantiles', *Econometrica* **46**, 33–50.

Koenker, R. & Machado, J. (1999*a*), 'Goodness of fit and related inference processes for quantile regression', *Journal of the American Statistical Association* **94**, 1296–1310.

Koenker, R. & Machado, J. (1999*b*), 'Goodness of fit and related inference processes for quantile regression', *Journal of the American Statistical Association* **94**, 1296–1310.

Kong, E., Linton, O. & Xia, Y. (2010), 'Uniform bahadur representation for local polynomial estimates of m-regression and its application to the additive model', *Econometric Theory* **26**, 1529–1564.

Kosorov, M. (2008), *An introduction to empirical processes and semiparametric inference*, Springer, New York.

Lee, S. (2003), 'Efficient semiparametric estimation of a partially linear quantile regression model', *Econometric Theory* **19**, 1–31.

Little, R. & Rubin, D. (2002), *Statistical analysis with missing data*, John Wiley and Sons, New Jersey.

Masry, E. (1996), 'Multivariate local polynomial regression for time series: uniform consistency and rates', *Journal of Time Series Analysis* **17**, 571–599.

Pollard, D. (1991), 'Asymptotics for least absolute deviation regression estimators', *Econometric Theory* **7**, 186–199.

Robins, J., Hseih, F. & Newey, W. (1995), ‘Semiparametric efficient estimation of a conditional density function with missing or mismeasured covariates’, *Journal of the Royal Statistical Society, B* **57**, 409–424.

Robins, J., Rotnisky, A. & Zhao, L. (1995), ‘Analysis of semiparametric regression models for repeated outcomes in the presence of missing data’, *Journal of the American Statistical Association* **90**, 106–121.

Robins, J., Rotnitzky, A. & Zhao, L. (1994), ‘Estimation of regression coefficients when some regressors are not always observed’, *Journal of the American Statistical Association* **89**, 846–866.

Rubin, D. (1976), ‘Inference and missing data’, *Biometrika* **63**, 581–591.

Severini, T. & Wong, W. (1992), ‘Profile likelihood and conditionally parametric models’, *Annals of Statistics* **20**, 1768–1802.

Sherwood, B. (2016), ‘Variable selection for additive partial linear quantile regression with missing covariates’, *Journal of Multivariate Analysis* **152**, 206–223.

Van der Vaart, A. & Wellner, J. (1996), *Weak Convergence and Empirical Processes*, Springer, New York.

Wang, C., Tian, M. & Tang, M. (2022), ‘Nonparametric quantile regression with missing data using local estimating equations’, *Journal of Nonparametric Statistics* **34**, 164–186.

Wang, H., Zhu, Z. & Zhou, J. (2009), ‘Quantile regression in partially linear varying coefficient models’, *Annals of Statistics* **37**, 3841–3866.

Wang, Q., Hardle, W. & Linton, O. (2004), ‘Semiparametric regression analysis with missing response at random’, *Journal of the American Statistical Association* **99**, 334–345.

Wong, W. & Severini, T. (1991), ‘On maximum likelihood estimation in infinite dimensional parameter spaces’, *Annals of Statistics* **19**, 663–632.

Xie, S., Wan, A. & Zhou, Y. (2015), ‘Quantile regression methods with varying-coefficients models for censored data’, *Computational Statistics and Data Analysis* **88**, 154–172.

Yu, K. & Jones, M. (1998), ‘Local linear quantile regression’, *Journal of the American Statistical Association* **93**, 228–237.

Zhou, L. (2006), ‘A simple censored median estimator’, *Statistica Sinica* **16**, 1043–1058.

9 Tables and figures

Table 1a $\varepsilon_\tau \sim N(0,1)$, $\tau = 0.25$

n	100				400			
	bias	*se*	*length*	*cov*	*bias*	*se*	*length*	*cov*
$\widehat{\beta}_{1\tau}$	.031	.183	.412	.943	.021	.096	.251	.946
$\widehat{\beta}_{2\tau}$	.071	.812	.888	.944	.056	.432	.459	.946
$\widehat{\beta}^p_{1\tau}$	.034	.175	.400	.951	.024	.086	.244	.951
$\widehat{\beta}^p_{2\tau}$	.075	.786	.864	.953	.057	.401	.415	.953
	MAR (6.4) 10%				MAR (6.4) 10%			
$\widehat{\beta}_{1\tau c}$	.088	.189	.422	.893	.078	.098	.258	.902
$\widehat{\beta}_{2\tau c}$	.108	.831	.898	.896	.085	.440	.486	.904
$\widehat{\beta}^p_{1\tau c}$	.090	.194	.432	.884	.081	.102	.255	.898
$\widehat{\beta}^p_{2\tau c}$	.112	.829	.901	.890	.088	.448	.492	.900
$\widehat{\beta}_{1\tau p}$	.032	.193	.431	.940	.018	.098	.223	.942
$\widehat{\beta}_{2\tau p}$	.073	.829	.905	.942	.036	.451	.493	.943
$\widehat{\beta}^p_{1\tau p}$	.030	.192	.910	.945	.021	.453	.490	.947
$\widehat{\beta}^p_{2\tau p}$	.078	.199	.441	.946	.041	.445	.491	.946
$\widehat{\beta}_{1\tau np}$	.033	.196	.435	.940	.019	.101	.231	.942
$\widehat{\beta}_{2\tau np}$	.074	.834	.910	.941	.037	.496	.496	.943
$\widehat{\beta}^p_{1\tau np}$	.038	.201	.438	.945	.021	.099	.229	.945
$\widehat{\beta}^p_{2\tau np}$	.080	.845	.921	.952	.041	.475	.481	.947
	MAR (6.4) 40%				MAR (6.4) 40%			
$\widehat{\beta}_{1\tau c}$	.112	.199	.441	.880	.099	.119	.262	.890
$\widehat{\beta}_{2\tau c}$	.124	.895	.913	.878	.109	.495	.499	.899
$\widehat{\beta}^p_{1\tau c}$	.132	.212	.453	.874	.104	.121	.265	.884
$\widehat{\beta}^p_{2\tau c}$	.136	.899	.907	.872	.112	.501	.509	.942
$\widehat{\beta}_{1\tau p}$	.037	.198	.438	.941	.023	.101	.258	.943
$\widehat{\beta}_{2\tau p}$	.076	.835	.814	.940	.040	.483	.499	.944
$\widehat{\beta}^p_{1\tau p}$	.041	.202	.441	.938	.020	.099	.255	.941
$\widehat{\beta}^p_{2\tau p}$	.081	.826	.818	.938	.041	.478	.491	.951
$\widehat{\beta}_{1\tau np}$	.038	.201	.441	.941	.029	.105	.260	.941
$\widehat{\beta}_{2\tau np}$	.078	.841	.916	.940	.041	.491	.501	.942
$\widehat{\beta}^p_{1\tau np}$	.034	.197	.445	.946	.031	.108	.255	.943
$\widehat{\beta}^p_{2\tau np}$	.084	.838	.924	.943	.045	.485	.494	.948

Table 1a Continued

n	100				400			
	bias	*se*	*length*	*cov*	*bias*	*se*	*length*	*cov*
	MAR (6.4) 10%				MAR (6.4) 10%			
$\widehat{\beta}_{1\tau c}$	.095	.193	.423	.893	.081	.099	.251	.900
$\widehat{\beta}_{2\tau c}$	.111	.873	.908	.888	.093	.455	.490	.896
$\widehat{\beta}^p_{1\tau c}$	.101	.199	.442	.899	.083	.102	.254	.897
$\widehat{\beta}^p_{2\tau c}$	.112	.877	.451	.890	.095	.459	.493	.899
$\widehat{\beta}_{1\tau p}$	.033	.198	.434	.939	.020	.101	.230	.941
$\widehat{\beta}_{2\tau p}$	.076	.881	.903	.941	.035	.459	.499	.944
$\widehat{\beta}^p_{1\tau p}$	.034	.201	.431	.942	.022	.110	.231	.942
$\widehat{\beta}^p_{2\tau p}$	.081	.889	.908	.943	.038	109	.236	.941
$\widehat{\beta}_{1\tau np}$	.037	.197	.438	.940	.021	.103	.233	.940
$\widehat{\beta}_{2\tau np}$	.078	.836	.912	.941	.039	.468	.507	.941
$\widehat{\beta}^p_{1\tau np}$	.039	.201	.441	.942	.023	.109	.246	.943
$\widehat{\beta}^p_{2\tau np}$	.075	.823	.913	.943	.029	.456	.494	.940
	MAR (6.5) 40%				MAR (6.5) 40%			
$\widehat{\beta}_{1\tau c}$	.116	.210	.456	.878	.098	.109	.268	.891
$\widehat{\beta}_{2\tau c}$	.132	.958	.988	.881	.105	.489	.512	.889
$\widehat{\beta}^p_{1\tau c}$	.124	.221	.476	.873	.102	.116	.278	.890
$\widehat{\beta}^p_{2\tau c}$	.138	.949	.985	.870	.112	.492	.514	.887
$\widehat{\beta}_{1\tau p}$	.039	.200	.440	.940	.018	.099	.260	.942
$\widehat{\beta}_{2\tau p}$	.078	.842	.816	.941	.038	.421	.501	.943
$\widehat{\beta}^p_{1\tau p}$	.041	.197	.443	.938	.019	.101	.254	.946
$\widehat{\beta}^p_{2\tau p}$	.081	.837	.812	.937	.041	.451	.503	.944
$\widehat{\beta}_{1\tau np}$	.040	.202	.444	.940	.025	.095	.260	.944
$\widehat{\beta}_{2\tau np}$	.079	.848	.920	.941	.023	.101	.262	.943
$\widehat{\beta}^p_{1\tau np}$	.042	.199	.431	.945	.041	.451	.503	.944
$\widehat{\beta}^p_{2\tau np}$	.081	.845	.918	.943	.043	.444	.495	.947

Table 1b $\varepsilon_\tau \sim N(0,1)$, $\tau = 0.5$

n	100				400			
	bias	*se*	*length*	*cov*	*bias*	*se*	*length*	*cov*
$\widehat{\beta}_{1\tau}$	.041	.163	.381	.941	.033	.087	.193	.943
$\widehat{\beta}_{2\tau}$	.031	.721	.831	.942	.023	.402	.488	.943
$\widehat{\beta}^p_{1\tau}$	.043	.154	.373	.942	.035	.073	.186	.952
$\widehat{\beta}^p_{2\tau}$	.032	.702	.811	.944	.025	.388	.476	.955
	MAR (6.4) 10%				MAR (6.4) 10%			
$\widehat{\beta}_{1\tau c}$	.098	.171	.393	.893	.077	.103	.172	.901
$\widehat{\beta}_{2\tau c}$	.099	.789	.849	.894	.084	.532	.458	.902
$\widehat{\beta}^p_{1\tau c}$	.092	.165	.388	.895	.081	.100	.174	.897
$\widehat{\beta}^p_{2\tau c}$	.095	.771	.833	.892	.088	.527	.455	.903
$\widehat{\beta}_{1\tau p}$	.040	.174	.403	.942	.028	.088	.219	.942
$\widehat{\beta}_{2\tau p}$	.034	.791	.889	.943	.030	.362	.471	.946
$\widehat{\beta}^p_{1\tau p}$	.041	.170	.835	.944	.027	.090	.218	.942
$\widehat{\beta}^p_{2\tau p}$	.042	.169	.936	.946	.033	.366	.213	.942
$\widehat{\beta}_{1\tau np}$	.043	.179	.402	.942	.036	.090	.213	.942
$\widehat{\beta}_{2\tau np}$	.035	.789	.891	.941	.028	.371	.478	.942
$\widehat{\beta}_{1\tau np}$	.046	.175	.405	.943	.037	.095	.481	.941
$\widehat{\beta}^p_{2\tau np}$	.038	.781	.407	.944	.027	.369	.212	.942
	MAR (6.4) 40%				MAR (6.4) 40%			
$\widehat{\beta}_{1\tau c}$	.121	.185	.399	.881	.112	.129	.199	.894
$\widehat{\beta}_{2\tau c}$	.128	.805	.836	.883	.109	.596	.503	.898
$\widehat{\beta}^p_{1\tau c}$	.125	.189	.403	.884	.110	.131	.202	.893
$\widehat{\beta}^p_{2\tau c}$	.130	.832	.841	.889	.112	.584	.509	.896
$\widehat{\beta}_{1\tau p}$	.045	.183	.407	.941	.030	.091	.224	.942
$\widehat{\beta}_{2\tau p}$	.038	.792	.841	.940	.028	.378	.593	.943
$\widehat{\beta}^p_{1\tau p}$	.044	.181	.400	.942	.031	.087	.221	.942
$\widehat{\beta}^p_{2\tau p}$	.036	.799	.843	.942	.027	.375	.590	.943
$\widehat{\beta}_{1\tau np}$	.047	.184	.403	.942	.032	.094	.219	.944
$\widehat{\beta}_{2\tau np}$	.040	.801	.883	.942	.031	.396	.496	.943
$\widehat{\beta}^p_{1\tau np}$	.048	.187	.405	.941	.033	.090	.212	.948
$\widehat{\beta}^p_{2\tau np}$	.041	.803	.875	.943	.030	.399	.491	.944

Table 1b Continued

n	100				400			
	bias	*se*	*length*	*cov*	*bias*	*se*	*length*	*cov*
	MAR (6.4) 10%				MAR (6.4) 10%			
$\widehat{\beta}_{1\tau c}$	.094	.176	.394	.882	.081	.116	.188	.901
$\widehat{\beta}_{2\tau c}$	.105	.759	.880	.885	.091	.511	.491	.904
$\widehat{\beta}^{p}_{1\tau c}$	.096	.172	.389	.888	.084	.119	.191	.900
$\widehat{\beta}^{p}_{2\tau c}$	.104	.761	.881	.883	.093	.516	.495	.905
$\widehat{\beta}_{1\tau p}$	.036	.176	.405	.940	.031	.092	.213	.948
$\widehat{\beta}_{2\tau p}$	.033	.789	.846	.942	.036	.361	.476	.943
$\widehat{\beta}^{p}_{1\tau p}$	.039	.173	.408	.941	.033	.089	.212	.947
$\widehat{\beta}^{p}_{2\tau p}$	.035	.170	.405	.943	.037	.360	.473	.944
$\widehat{\beta}_{1\tau np}$	.045	.181	.402	.939	.037	.099	.216	.942
$\widehat{\beta}_{2\tau np}$	.036	.731	.890	.940	.030	.363	.484	.941
$\widehat{\beta}^{p}_{1\tau np}$	.046	.179	.404	.941	.036	.096	.213	.943
$\widehat{\beta}^{p}_{2\tau np}$	.039	.177	.884	.942	.029	.094	.210	.942
	MAR (6.5) 40%				MAR (6.5) 40%			
$\widehat{\beta}_{1\tau c}$	.131	.199	.398	.884	.107	.131	.193	.893
$\widehat{\beta}_{2\tau c}$	.138	.810	.823	.885	.122	.592	.541	.896
$\widehat{\beta}^{p}_{1\tau c}$	.134	.203	.402	.886	.110	.136	.196	.890
$\widehat{\beta}^{p}_{2\tau c}$	.134	.813	.827	.882	.126	.590	.546	.892
$\widehat{\beta}_{1\tau p}$	.041	.179	.410	.942	.031	.090	.218	.943
$\widehat{\beta}_{2\tau p}$	.040	.795	.848	.943	.032	.376	.481	.946
$\widehat{\beta}^{p}_{1\tau p}$	.042	.182	.408	.944	.033	.087	.211	.942
$\widehat{\beta}^{p}_{2\tau p}$	.039	.793	.850	.944	.034	.370	.474	.948
$\widehat{\beta}_{1\tau np}$	.042	.183	.407	.940	.032	.096	.211	.943
$\widehat{\beta}_{2\tau np}$	.040	.763	.892	.942	.033	.389	.485	.944
$\widehat{\beta}^{p}_{1\tau np}$	.041	.180	.410	.941	.036	.094	.209	.942
$\widehat{\beta}^{p}_{2\tau np}$	.039	.178	.890	.942	.033	.386	.482	.943

Table 1c $\varepsilon_\tau \sim N(0,1)$, $\tau = 0.75$

n	100				400			
	bias	*se*	*length*	*cov*	*bias*	*se*	*length*	*cov*
$\widehat{\beta}_{1\tau}$	.052	.193	.413	.940	.036	.093	.225	.942
$\widehat{\beta}_{2\tau}$	.068	.751	.798	.942	.038	.381	.368	.944
$\widehat{\beta}^p_{1\tau}$	.054	.186	.410	.942	.037	.090	.223	.943
$\widehat{\beta}^p_{2\tau}$	.070	.744	.800	.943	.039	.375	.360	.943
	MAR (6.4) 10%				MAR (6.4) 10%			
$\widehat{\beta}_{1\tau c}$	.108	.201	.444	.898	.089	.110	.281	.901
$\widehat{\beta}_{2\tau c}$	.101	.799	.828	.899	.092	.399	.438	.900
$\widehat{\beta}^p_{1\tau c}$	.110	.205	.449	.896	.090	.108	.279	.900
$\widehat{\beta}^p_{2\tau c}$	.103	.803	.832	.897	.095	.395	.436	.899
$\widehat{\beta}_{1\tau p}$	.054	.205	.449	.942	.030	.101	.231	.944
$\widehat{\beta}_{2\tau p}$	.069	.803	.838	.941	.031	.418	.431	.943
$\widehat{\beta}^p_{1\tau p}$	.056	.201	.447	.943	.032	.099	.229	.943
$\widehat{\beta}^p_{2\tau p}$	.071	.796	.835	.942	.033	.404	.429	.942
$\widehat{\beta}_{1\tau np}$	.058	.197	.436	.944	.028	.104	.238	.944
$\widehat{\beta}_{2\tau np}$	.056	.204	.467	.943	.030	.416	.430	.945
$\widehat{\beta}^p_{1\tau np}$	.068	.809	.835	.944	.030	.100	.231	.943
$\widehat{\beta}^p_{2\tau np}$	.057	.200	.465	.942	.027	.421	.427	.944
	MAR (6.4) 40%				MAR (6.4) 40%			
$\widehat{\beta}_{1\tau c}$	.119	.216	.494	.883	.103	.185	.299	.889
$\widehat{\beta}_{2\tau c}$	.106	.812	.888	.887	.099	.508	.486	.894
$\widehat{\beta}^p_{1\tau c}$	.121	.214	.492	.884	.100	.180	.294	.886
$\widehat{\beta}^p_{2\tau c}$	.109	.814	.891	.890	.102	.505	.485	.895
$\widehat{\beta}_{1\tau p}$	.056	.210	.420	.944	.032	.115	.260	.946
$\widehat{\beta}_{2\tau p}$	.071	.822	.848	.943	.038	.425	.440	.942
$\widehat{\beta}^p_{1\tau p}$	.057	.205	.415	.943	.034	.110	.258	.944
$\widehat{\beta}^p_{2\tau p}$	.073	.819	.850	.945	.040	,421	.443	.943
$\widehat{\beta}_{1\tau np}$	.057	.209	.407	.943	.033	.118	.243	.944
$\widehat{\beta}_{2\tau np}$	.070	.818	.840	.942	.034	.421	.438	.943
$\widehat{\beta}^p_{1\tau np}$	.059	.205	.409	.945	.035	.113	.239	.946
$\widehat{\beta}^p_{2\tau np}$	.072	.812	.835	.948	.032	.418	.440	.945

Table 1c Continued

n	100				400			
	bias	*se*	*length*	*cov*	*bias*	*se*	*length*	*cov*
	MAR (6.5) 10%				MAR (6.5) 10%			
$\widehat{\beta}_{1\tau c}$	.110	.209	.542	.888	.088	.138	.284	.898
$\widehat{\beta}_{2\tau c}$	.105	.803	.833	.892	.085	.501	.449	.896
$\widehat{\beta}^{p}_{1\tau c}$	.112	.214	.548	.886	.091	.132	.280	.895
$\widehat{\beta}^{p}_{2\tau c}$	.105	.798	.836	.895	.085	.494	.445	.893
$\widehat{\beta}_{1\tau p}$	.055	.208	.418	.943	.030	.110	.238	.946
$\widehat{\beta}_{2\tau p}$	.068	.808	.836	.942	.036	.406	.439	.944
$\widehat{\beta}^{p}_{1\tau p}$	.056	.210	.421	.946	.031	.105	.227	.943
$\widehat{\beta}^{p}_{2\tau p}$	.070	.203	.832	.945	.034	.102	.431	.943
$\widehat{\beta}_{1\tau np}$	.050	.210	.460	.942	.029	.112	.241	.944
$\widehat{\beta}_{2\tau np}$	.069	.811	.836	.944	.035	.403	.433	.945
$\widehat{\beta}^{p}_{1\tau np}$	.051	.205	.454	.941	.030	.109	.243	.946
$\widehat{\beta}^{p}_{2\tau np}$	.065	.809	.832	.947	.034	.399	.436	.947
	MAR (6.5) 40%				MAR (6.5) 40%			
$\widehat{\beta}_{1\tau c}$	.118	.228	.501	.880	.102	.181	.303	.889
$\widehat{\beta}_{2\tau c}$	.110	.822	.890	.885	.099	.452	.491	.895
$\widehat{\beta}^{p}_{1\tau c}$	.121	.225	.497	.878	.110	.179	.312	.887
$\widehat{\beta}^{p}_{2\tau c}$	.111	.819	.887	.880	.102	.447	.490	.894
$\widehat{\beta}_{1\tau p}$	.058	.210	.421	.943	.029	.118	.263	.941
$\widehat{\beta}_{2\tau p}$	.072	.818	.842	.945	.034	.429	.442	.946
$\widehat{\beta}^{p}_{1\tau p}$	.060	.208	.419	.942	.031	.110	.256	.943
$\widehat{\beta}^{p}_{2\tau p}$	.070	.809	.840	.943	.033	.420	.440	.945
$\widehat{\beta}_{1\tau np}$	.059	.213	.456	.945	.031	.113	.244	.946
$\widehat{\beta}_{2\tau np}$	.071	.817	.819	.943	.035	.407	.436	.945
$\widehat{\beta}^{p}_{1\tau np}$	.061	.210	.449	.947	.030	.110	.232	.944
$\widehat{\beta}^{p}_{2\tau np}$	.073	.815	.817	.942	.036	.401	.435	.943

Table 2a $\varepsilon_\tau \sim t\,(5)\,,\ \tau = 0.25$

n	100				400			
	bias	*se*	*length*	*cov*	*bias*	*se*	*length*	*cov*
$\widehat{\beta}_{1\tau c}$	.046	.173	.392	.942	.025	.096	.219	.947
$\widehat{\beta}_{2\tau c}$	.038	.819	.848	.944	.022	.424	.459	.943
$\widehat{\beta}^{p}_{1\tau c}$	.048	.170	.390	.945	.027	.092	.217	.945
$\widehat{\beta}^{p}_{2\tau c}$	.039	.810	.840	.947	.024	.420	.450	.948
	MAR (6.4) 10%				MAR (6.4) 10%			
$\widehat{\beta}_{1\tau c}$	.095	.196	.420	.900	.075	.110	.218	.901
$\widehat{\beta}_{2\tau c}$	.093	.835	.868	.897	.071	.499	.496	.901
$\widehat{\beta}^{p}_{1\tau c}$	.097	.199	.419	.898	.077	.105	.212	.900
$\widehat{\beta}^{p}_{2\tau c}$	,095	.833	.867	.896	.073	.496	.493	.901
$\widehat{\beta}_{1\tau p}$	.047	.200	.412	.940	.027	.107	.223	.942
$\widehat{\beta}_{2\tau p}$	.040	.820	.871	.941	.024	.437	.483	.943
$\widehat{\beta}^{p}_{1\tau p}$	.045	.196	.414	.942	.028	.103	.220	.943
$\widehat{\beta}^{p}_{2\tau p}$	.046	.205	.415	.939	.030	.430	.485	.944
$\widehat{\beta}_{1\tau np}$	.041	.821	.870	.941	.025	.110	.209	.940
$\widehat{\beta}_{2\tau np}$	.047	.200	.413	.941	.021	.505	.476	.943
$\widehat{\beta}^{p}_{1\tau np}$	.042	.818	.867	.943	.026	.107	.205	.942
$\widehat{\beta}^{p}_{2\tau np}$					.020	.500	.471	.943
	MAR (6.4) 40%				MAR (6.4) 40%			
$\widehat{\beta}_{1\tau c}$	.125	.210	.461	.882	.120	.118	.222	.900
$\widehat{\beta}_{2\tau c}$	.121	.841	.883	.878	.116	.514	.439	.901
$\widehat{\beta}^{p}_{1\tau c}$	.128	.204	.458	.883	.123	.115	.224	.902
$\widehat{\beta}^{p}_{2\tau c}$	.120	.839	.881	.879	.118	.510	.442	.900
$\widehat{\beta}_{1\tau p}$	.048	.210	.415	.941	.028	.115	.208	.945
$\widehat{\beta}_{2\tau p}$	.041	.828	.874	.940	.024	.438	.429	.942
$\widehat{\beta}^{p}_{1\tau p}$	.049	.203	.410	.942	.030	.110	.205	.946
$\widehat{\beta}^{p}_{2\tau p}$	.043	.820	.403	.942	.026	.430	.425	.946
$\widehat{\beta}_{1\tau np}$	.047	.211	.421	.939	.023	.119	.210	.945
$\widehat{\beta}_{2\tau np}$	.043	.825	.876	.941	.022	.510	.434	.943
$\widehat{\beta}^{p}_{1\tau np}$	.048	.207	.423	.942	.024	.110	.207	.946
$\widehat{\beta}^{p}_{2\tau np}$	.045	.820	.874	.943	.023	.501	.433	.945

Table 2a Continued

n	100				400			
	bias	*se*	*length*	*cov*	*bias*	*se*	*length*	*cov*
	MAR (6.5) 10%				MAR (6.5) 10%			
$\widehat{\beta}_{1\tau c}$	.094	.209	.425	.891	.081	.114	.220	.894
$\widehat{\beta}_{2\tau c}$	.096	.836	.878	.897	.075	.505	.494	.890
$\widehat{\beta}^p_{1\tau c}$	.095	.205	.422	.890	.082	.110	.222	.890
$\widehat{\beta}^p_{2\tau c}$	.097	.830	.877	.894	.077	.500	.490	.891
$\widehat{\beta}_{1\tau p}$	.046	.202	.428	.940	.025	.105	.215	.947
$\widehat{\beta}_{2\tau p}$	.043	.816	.880	.941	.028	.425	.435	.945
$\widehat{\beta}^p_{1\tau p}$	.048	.199	.424	.942	.024	.103	.212	.944
$\widehat{\beta}^p_{2\tau p}$	.045	.810	.883	.942	.029	.420	.430	.946
$\widehat{\beta}_{1\tau np}$	.045	.207	.428	.940	.023	.110	.213	.943
$\widehat{\beta}_{2\tau np}$	.046	.822	.882	.939	.024	.507	.437	.943
$\widehat{\beta}^p_{1\tau np}$	.044	.201	.430	.941	.024	.104	.210	.944
$\widehat{\beta}^p_{2\tau np}$	.047	.818	.880	.942	.025	.501	.435	.945
	MAR (6.5) 40%				MAR (6.5) 40%			
$\widehat{\beta}_{1\tau c}$	.128	.221	.458	.880	.083	.118	.230	.890
$\widehat{\beta}_{2\tau c}$	.129	.854	.888	.878	.082	.508	.419	.886
$\widehat{\beta}^p_{1\tau c}$	.131	.219	.454	.881	.088	.116	.234	.888
$\widehat{\beta}^p_{2\tau c}$	.130	.849	.884	.876	.085	.512	.415	.883
$\widehat{\beta}_{1\tau p}$	.046	.218	.430	.940	.029	.120	.235	.943
$\widehat{\beta}_{2\tau p}$	.048	.821	.886	.942	.031	.443	.451	.941
$\widehat{\beta}^p_{1\tau p}$	.047	.210	.431	.942	.030	.115	.232	.945
$\widehat{\beta}^p_{2\tau p}$	.048	.818	.882	.943	.032	,438	.437	.943
$\widehat{\beta}_{1\tau np}$	.047	.210	.436	.941	.026	.114	.232	.942
$\widehat{\beta}_{2\tau np}$	.046	.828	.890	.940	.028	.512	.453	.944
$\widehat{\beta}^p_{1\tau np}$	.045	.205	.433	.942	,025	.112	.230	.944
$\widehat{\beta}^p_{2\tau np}$	.048	.825	.893	.943	.030	.509	.450	.943

Table 2b $\varepsilon_\tau \sim t(5)$, $\tau = 0.5$

n	100				400			
	bias	*se*	*length*	*cov*	*bias*	*se*	*length*	*cov*
$\widehat{\beta}_{1\tau}$	.051	.170	.391	.941	.030	.098	.183	.945
$\widehat{\beta}_{2\tau}$	.042	.764	.841	.943	.019	.364	.368	.946
$\widehat{\beta}^p_{1\tau}$	.053	.165	.386	.946	.031	.094	.178	.946
$\widehat{\beta}^p_{2\tau}$	.044	.759	.837	.947	.018	.360	.365	.947
	MAR (6.4) 10%				MAR (6.4) 10%			
$\widehat{\beta}_{1\tau c}$	.105	.185	.413	.900	.078	.105	.215	.901
$\widehat{\beta}_{2\tau c}$	.095	.816	.882	.894	.067	.498	.418	.900
$\widehat{\beta}^p_{1\tau c}$	.110	.180	.414	.901	.079	.101	.214	.902
$\widehat{\beta}^p_{2\tau c}$	.099	.812	.880	.892	.069	.496	.415	.901
$\widehat{\beta}_{1\tau p}$	.052	.196	.418	.942	.032	.096	.219	.943
$\widehat{\beta}_{2\tau p}$	.046	.781	.865	.941	.034	.396	.431	.945
$\widehat{\beta}^p_{1\tau p}$	.055	.192	.414	.945	.033	.091	.214	.944
$\widehat{\beta}^p_{2\tau p}$	,047	.775	.860	.943	.035	.389	.429	.946
$\widehat{\beta}_{1\tau np}$	.054	.204	.420	.943	.034	.099	.213	.942
$\widehat{\beta}_{2\tau np}$	.045	.797	.875	.941	.026	.386	.418	.944
$\widehat{\beta}^p_{1\tau np}$	.055	.201	.417	.945	.033	.093	.210	.944
$\widehat{\beta}^p_{2\tau np}$	.046	.790	.871	.944	.027	.380	.414	.946
	MAR (6.4) 40%				MAR (6.4) 40%			
$\widehat{\beta}_{1\tau c}$	.119	.222	.423	.883	.110	.128	.223	.894
$\widehat{\beta}_{2\tau c}$	.127	.875	.895	.887	.118	.484	.458	.895
$\widehat{\beta}^p_{1\tau c}$	.120	.218	.420	.882	.112	,126	.222	.895
$\widehat{\beta}^p_{2\tau c}$	.128	.872	.893	.888	.120	.482	.455	.896
$\widehat{\beta}_{1\tau p}$	.055	.198	.421	.946	.035	.101	.219	.940
$\widehat{\beta}_{2\tau p}$	.051	.793	.878	.945	.037	.412	.461	.944
$\widehat{\beta}^p_{1\tau p}$	.056	.195	.419	.945	.036	.099	.215	.942
$\widehat{\beta}^p_{2\tau p}$	.053	.790	.875	.944	.038	.410	.456	.945
$\widehat{\beta}_{1\tau np}$	.056	.206	.423	.944	.037	.103	.220	.940
$\widehat{\beta}_{2\tau np}$	.048	.804	.890	.948	.038	.421	.431	.945
$\widehat{\beta}^p_{1\tau np}$	.058	.202	.419	.945	.038	.099	.216	.942
$\widehat{\beta}^p_{2\tau np}$	.050	.800	.887	.947	.039	.418	.427	.943

Table 2b Continued

n	100				400			
	bias	*se*	*length*	*cov*	*bias*	*se*	*length*	*cov*
	MAR (6.5) 10%				MAR (6.5) 10%			
$\widehat{\beta}_{1\tau c}$	.109	.196	.410	.894	.080	.104	.219	.900
$\widehat{\beta}_{2\tau c}$	.099	.820	.878	.898	.071	.438	.431	.901
$\widehat{\beta}^{p}_{1\tau c}$	.111	.194	.407	.892	.081	.101	.216	.901
$\widehat{\beta}^{p}_{2\tau c}$	.101	.818	.876	.899	.073	.433	.429	.902
$\widehat{\beta}_{1\tau p}$	.053	.198	.411	.948	.033	.099	.218	.945
$\widehat{\beta}_{2\tau p}$	.048	.797	.870	.945	.035	.405	.438	.944
$\widehat{\beta}^{p}_{1\tau p}$	.055	.194	.408	.949	.035	,095	.215	.946
$\widehat{\beta}^{p}_{2\tau p}$	.050	.793	.868	.947	.036	.402	.434	.945
$\widehat{\beta}_{1\tau np}$	.053	.203	.412	.946	.036	.098	.210	.945
$\widehat{\beta}_{2\tau np}$	.046	.787	.880	.948	.033	.408	.414	.946
$\widehat{\beta}^{p}_{1\tau np}$	,052	.200	.410	.945	.037	.095	.206	.946
$\widehat{\beta}^{p}_{2\tau np}$	.048	.784	.879	.947	.034	.405	.410	.947
	MAR (6.5) 40%				MAR (6.5) 40%			
$\widehat{\beta}_{1\tau c}$	.120	.220	.423	.887	.111	.126	.225	.890
$\widehat{\beta}_{2\tau c}$	.124	.883	.888	.885	.108	.491	.435	.889
$\widehat{\beta}^{p}_{1\tau c}$	.122	.217	.419	.884	.115	.123	.222	.887
$\widehat{\beta}^{p}_{2\tau c}$	.125	.886	.890	.884	.109	.487	.433	.888
$\widehat{\beta}_{1\tau p}$	.055	.204	.414	.946	.036	.104	.220	.941
$\widehat{\beta}_{2\tau p}$	.050	.799	.872	.944	.038	.415	.445	.945
$\widehat{\beta}^{p}_{1\tau p}$	.056	.200	.410	.946	.037	.100	.215	.943
$\widehat{\beta}^{p}_{2\tau p}$	.051	.199	.869	.947	.040	.411	.440	.946
$\widehat{\beta}_{1\tau np}$	.055	.206	.416	.944	.039	.106	.218	.949
$\widehat{\beta}_{2\tau np}$	.048	.790	.883	.946	.037	.412	.426	.948
$\widehat{\beta}^{p}_{1\tau np}$	.054	.205	.417	.945	.038	.102	.214	.948
$\widehat{\beta}^{p}_{2\tau np}$	.049	.788	.881	.946	.038	.409	.420	.948

Table 2c $\varepsilon_\tau \sim t\,(5)\,,\,\tau = 0.75$

n	100				400			
	bias	*se*	*length*	*cov*	*bias*	*se*	*length*	*cov*
$\widehat{\beta}_{1\tau}$	.055	.202	.383	.944	.034	.118	.195	.943
$\widehat{\beta}_{2\tau}$	.069	.806	.812	.942	.024	.407	.408	.940
$\widehat{\beta}^p_{1\tau}$	.056	.200	.380	.945	.035	.113	.192	.945
$\widehat{\beta}^p_{2\tau}$	.068	.801	.810	.943	.025	.401	.404	.942
	MAR (6.4) 10%				MAR (6.4) 10%			
$\widehat{\beta}_{1\tau c}$	.112	.211	.406	.892	.086	.107	.208	.901
$\widehat{\beta}_{2\tau c}$	.108	.828	.872	.893	.071	.446	.421	.900
$\widehat{\beta}^p_{1\tau c}$	.113	.209	.403	.890	.888	.103	.203	.902
$\widehat{\beta}^p_{2\tau c}$	.106	.825	.870	.892	.072	.445	.418	.899
$\widehat{\beta}_{1\tau p}$	.056	.215	.412	.942	.036	.111	.211	.945
$\widehat{\beta}_{2\tau p}$	.073	.838	.878	.940	.028	.452	.431	.948
$\widehat{\beta}^p_{1\tau p}$	.057	.210	.410	.943	.037	.107	.209	.947
$\widehat{\beta}^p_{2\tau p}$	.074	.830	.873	.942	.030	.448	.428	.945
$\widehat{\beta}_{1\tau np}$	.058	.214	.414	.941	.038	.114	.214	.944
$\widehat{\beta}_{2\tau np}$	.072	.830	.881	.941	.028	.455	.498	.947
$\widehat{\beta}^p_{1\tau np}$	.060	.210	.412	.942	.039	.112	.212	.945
$\widehat{\beta}^p_{2\tau np}$	.071	.825	.874	.943	.029	.450	.495	.946
	MAR (6.4) 40%				MAR (6.4) 40%			
$\widehat{\beta}_{1\tau c}$	.122	.230	.410	.883	.103	.119	.229	.891
$\widehat{\beta}_{2\tau c}$	.116	.896	.883	.884	.089	.452	.431	.889
$\widehat{\beta}^p_{1\tau c}$	.123	.225	.406	.882	.104	.115	.227	.892
$\widehat{\beta}^p_{2\tau c}$	.117	.890	.880	.881	.091	.447	.428	.887
$\widehat{\beta}_{1\tau p}$	.059	.217	.416	.940	.038	.121	.218	.943
$\widehat{\beta}_{2\tau p}$	.076	.837	.889	.939	.030	.455	.439	.945
$\widehat{\beta}^p_{1\tau p}$	.060	.212	.413	.942	.040	.118	.215	.944
$\widehat{\beta}^p_{2\tau p}$	.075	.833	.885	.941	.031	.452	.434	.943
$\widehat{\beta}_{1\tau np}$	.050	.218	.420	.943	.039	.123	.218	.942
$\widehat{\beta}_{2\tau np}$	.075	.838	.883	.940	.033	.458	.439	.943
$\widehat{\beta}^p_{1\tau np}$	.051	.215	.417	.944	.040	.120	.216	.943
$\widehat{\beta}^p_{2\tau np}$	.074	.832	.880	.942	.034	.450	.436	.944

Table 2c Continued

n	100				400			
	bias	*se*	*length*	*cov*	*bias*	*se*	*length*	*cov*
	MAR (6.5) 10%				MAR (6.5) 10%			
$\widehat{\beta}_{1\tau c}$	.118	.212	.410	.890	.088	.108	.204	.900
$\widehat{\beta}_{2\tau c}$	.106	.829	.881	.901	.076	.447	.417	.901
$\widehat{\beta}^{p}_{1\tau c}$	.120	.210	.408	.888	.089	.107	.200	.901
$\widehat{\beta}^{p}_{2\tau c}$	.107	.824	.880	.902	.077	.445	.410	.902
$\widehat{\beta}_{1\tau p}$	.058	.219	.416	.940	.037	.112	.216	.943
$\widehat{\beta}_{2\tau p}$	.074	.828	.869	.941	.032	.459	.436	.946
$\widehat{\beta}^{p}_{1\tau p}$	.060	.216	.413	.941	.038	.110	.213	.944
$\widehat{\beta}^{p}_{2\tau p}$	.075	.827	.865	.942	.033	.453	.431	.947
$\widehat{\beta}_{1\tau np}$	.060	.216	.418	.941	.040	.118	.217	.943
$\widehat{\beta}_{2\tau np}$	.073	.832	.873	.940	.030	.463	.436	.945
$\widehat{\beta}^{p}_{1\tau np}$	.061	.214	.416	.942	.041	.114	.213	.944
$\widehat{\beta}^{p}_{2\tau np}$	.074	.830	.872	.943	.031	.461	.433	.946
	MAR (6.5) 40%				MAR (6.5) 40%			
$\widehat{\beta}_{1\tau c}$	.128	.232	.414	.883	.104	.120	.226	.893
$\widehat{\beta}_{2\tau c}$	.119	.840	.891	.881	.091	.456	.436	.886
$\widehat{\beta}^{p}_{1\tau c}$	.130	.230	.409	.885	.040	.122	.225	.942
$\widehat{\beta}^{p}_{2\tau c}$	.116	.835	.892	.882	.032	.461	.436	.943
$\widehat{\beta}_{1\tau p}$	.060	.221	.418	.941	.041	.121	.227	.941
$\widehat{\beta}_{2\tau p}$	.079	.824	.891	.943	.036	.467	.438	.941
$\widehat{\beta}^{p}_{1\tau p}$	.061	.217	.415	.942	.110	.123	.230	.895
$\widehat{\beta}^{p}_{2\tau p}$	.080	.819	.885	.945	.093	453	.435	.890
$\widehat{\beta}_{1\tau np}$	.060	.220	.423	.945	.042	.124	.230	944
$\widehat{\beta}_{2\tau np}$	.076	.836	.881	.948	.033	.446	.433	.945
$\widehat{\beta}^{p}_{1\tau np}$	.061	.217	.420	.945	.043	.124	.225	.942
$\widehat{\beta}^{p}_{2\tau np}$	.077	.830	.415	.946	.037	.465	.435	.943

Table 3a $\varepsilon_\tau \sim \chi^2(4) - 4$, $\tau = 0.25$

n	100				400			
	bias	*se*	*length*	*cov*	*bias*	*se*	*length*	*cov*
$\widehat{\beta}_{1\tau}$	.047	.181	.402	.942	.029	.097	.221	.943
$\widehat{\beta}_{2\tau}$	.037	.826	.841	.944	.028	.421	.464	.943
$\widehat{\beta}^p_{1\tau}$	.045	.175	.400	.944	.030	.094	.218	.945
$\widehat{\beta}^p_{2\tau}$	.038	.820	.837	.945	.031	.416	.460	.946
	MAR (6.4) 10%				MAR (6.4) 10%			
$\widehat{\beta}_{1\tau c}$	.097	.199	.422	.900	.085	.108	.238	.902
$\widehat{\beta}_{2\tau c}$	.098	.831	.898	.897	.094	.444	.479	.901
$\widehat{\beta}^p_{1\tau c}$	.098	.195	.416	.901	.084	.109	.239	.900
$\widehat{\beta}^p_{2\tau c}$	.099	.830	.899	.899	.095	.440	.476	.902
$\widehat{\beta}_{1\tau p}$	.048	.202	.428	.942	.031	.107	.223	.948
$\widehat{\beta}_{2\tau p}$	.041	.802	.857	.941	.031	.450	.473	.943
$\widehat{\beta}^p_{1\tau p}$	.049	.199	.425	.943	.033	.104	.220	.947
$\widehat{\beta}^p_{2\tau p}$	.042	.799	.853	.943	.032	.446	.468	.946
$\widehat{\beta}_{1\tau np}$	.047	.236	.430	.940	.033	.109	.236	.941
$\widehat{\beta}_{2\tau np}$	.049	.810	.852	.941	.036	.451	.476	.942
$\widehat{\beta}^p_{1\tau np}$	.048	.233	.425	.942	.034	.105	.230	.942
$\widehat{\beta}^p_{2\tau np}$	.050	.803	.850	.943	.037	.445	.471	.944
	MAR (6.4) 40%				MAR (6.4) 40%			
$\widehat{\beta}_{1\tau c}$	.119	.216	.428	.889	.108	.120	.256	.893
$\widehat{\beta}_{2\tau c}$	.126	.852	.889	.894	.110	.468	.496	.896
$\widehat{\beta}^p_{1\tau c}$	.120	.210	.423	.891	.110	.115	.252	.894
$\widehat{\beta}^p_{2\tau c}$	.125	.853	.888	.890	.109	.462	.493	.895
$\widehat{\beta}_{1\tau p}$	.049	.220	.433	.942	.036	.122	.258	.943
$\widehat{\beta}_{2\tau p}$	.043	.851	.859	.942	.037	.464	.491	.943
$\widehat{\beta}^p_{1\tau p}$	.050	.215	.430	.943	.037	.117	.250	.945
$\widehat{\beta}^p_{2\tau p}$	.044	.847	.855	.944	.038	.467	.482	.945
$\widehat{\beta}_{1\tau np}$	.048	.223	.436	.942	.037	.118	.262	.945
$\widehat{\beta}_{2\tau np}$	.050	.866	.858	.944	.036	.473	.499	.948
$\widehat{\beta}^p_{1\tau np}$	.047	.220	.430	.943	.039	.115	.260	.946
$\widehat{\beta}^p_{2\tau np}$	.051	.860	.855	.946	.037	,478	.494	.947

Table 3a Continued

n	100				400			
	bias	*se*	*length*	*cov*	*bias*	*se*	*length*	*cov*
	MAR (6.5) 10%				MAR (6.5) 10%			
$\widehat{\beta}_{1\tau c}$	.099	.201	.424	.894	.088	.110	.240	.898
$\widehat{\beta}_{2\tau c}$	.102	.836	.882	.890	.093	.457	.483	.893
$\widehat{\beta}^p_{1\tau c}$	.100	.196	.420	.895	.090	.105	.241	.896
$\widehat{\beta}^p_{2\tau c}$	.103	.832	.879	.891	.092	.455	.480	.894
$\widehat{\beta}_{1\tau p}$	.049	.206	.430	.942	.033	.113	.226	.947
$\widehat{\beta}_{2\tau p}$	.043	.839	.853	.941	.032	.455	.489	.944
$\widehat{\beta}^p_{1\tau p}$	.050	.200	.426	.943	.034	.109	.220	.946
$\widehat{\beta}^p_{2\tau p}$	.045	.827	.849	.944	.033	.452	.483	.945
$\widehat{\beta}_{1\tau np}$	.048	.208	.436	.941	.036	.114	.228	.944
$\widehat{\beta}_{2\tau np}$	.051	.840	.854	.942	.037	.453	.480	.943
$\widehat{\beta}^p_{1\tau np}$	.049	.204	.430	.942	.037	.110	.223	.945
$\widehat{\beta}^p_{2\tau np}$	.050	.839	.852	.943	.038	.450	.476	.946
	MAR (6.5) 40%				MAR (6.5) 40%			
$\widehat{\beta}_{1\tau c}$	.126	.214	.430	.884	.110	.122	.250	.889
$\widehat{\beta}_{2\tau c}$	.121	.859	.890	.885	.112	.470	.492	.893
$\widehat{\beta}^p_{1\tau c}$	.128	.211	.428	.885	.111	.120	.246	.890
$\widehat{\beta}^p_{2\tau c}$	.120	.855	.887	.886	.111	.471	.493	.892
$\widehat{\beta}_{1\tau p}$	.051	.222	.438	.943	.032	.129	.263	.944
$\widehat{\beta}_{2\tau p}$	.048	.860	.860	.942	.036	.460	.493	.945
$\widehat{\beta}^p_{1\tau p}$	.053	.218	.433	.945	.033	.126	.260	.945
$\widehat{\beta}^p_{2\tau p}$	.049	.853	.853	.943	.037	.455	.490	.947
$\widehat{\beta}_{1\tau np}$	.050	.218	.444	.941	.038	.126	.259	.943
$\widehat{\beta}_{2\tau np}$	.052	.859	.860	.942	.039	.470	.504	.943
$\widehat{\beta}^p_{1\tau np}$	.051	.214	.440	.943	.039	.123	.255	.945
$\widehat{\beta}^p_{2\tau np}$	.053	.855	.856	.944	.037	.471	.505	.944

Table 3b $\varepsilon_\tau \sim \chi^2(4) - 4$, $\tau = 0.5$

n	100				400			
	bias	*se*	*length*	*cov*	*bias*	*se*	*length*	*cov*
$\widehat{\beta}_{1\tau}$	.053	.182	.396	.941	.031	.101	.203	.943
$\widehat{\beta}_{2\tau}$	.040	.839	.840	.946	.026	.424	.478	.947
$\widehat{\beta}^p_{1\tau}$	.055	.180	.395	.945	.032	.099	.198	.946
$\widehat{\beta}^p_{2\tau}$	.042	.832	.834	.947	.027	.420	.473	.948
	MAR (6.4) 10%				MAR (6.4) 10%			
$\widehat{\beta}_{1\tau c}$	.099	.201	.443	.901	.086	.112	.226	.902
$\widehat{\beta}_{2\tau c}$	.096	.833	.862	.899	.074	.441	.489	.900
$\widehat{\beta}^p_{1\tau c}$	.101	.196	.438	.903	.088	.110	.223	.903
$\widehat{\beta}^p_{2\tau c}$	.099	.830	.860	.900	.075	.437	.486	.901
$\widehat{\beta}_{1\tau p}$	.054	.203	.416	.941	.033	.108	.232	.946
$\widehat{\beta}_{2\tau p}$	.042	.837	.865	.942	.038	.444	.491	.943
$\widehat{\beta}^p_{1\tau p}$	.055	.197	.410	.943	.034	.104	.230	.945
$\widehat{\beta}^p_{2\tau p}$	.043	.834	.862	.943	.039	.440	.485	.944
$\widehat{\beta}_{1\tau np}$	.057	.205	.420	.940	.033	.110	.236	.944
$\widehat{\beta}_{2\tau np}$	.043	.836	.868	.941	.040	.436	.495	.942
$\widehat{\beta}^p_{1\tau np}$	.059	.201	.418	.942	.034	.105	.233	.945
$\widehat{\beta}^p_{2\tau np}$	.044	.832	.862	.943	.041	.432	.491	.944
	MAR (6.4) 40%				MAR (6.4) 40%			
$\widehat{\beta}_{1\tau c}$	.126	.206	.431	.883	.102	.130	.254	.895
$\widehat{\beta}_{2\tau c}$	.125	.840	.869	.887	.106	.498	.499	.896
$\widehat{\beta}^p_{1\tau c}$	.128	.200	.432	.885	.103	.127	.253	.893
$\widehat{\beta}^p_{2\tau c}$	.127	.843	.871	.888	.107	.495	.495	.899
$\widehat{\beta}_{1\tau p}$	.056	.208	.433	.942	.037	.119	.260	.942
$\widehat{\beta}_{2\tau p}$	.047	.841	.870	.941	.038	.469	.501	.944
$\widehat{\beta}^p_{1\tau p}$	.057	.206	.430	.944	.038	.112	.257	.943
$\widehat{\beta}^p_{2\tau p}$	.046	.840	.865	.946	.039	.464	.497	.945
$\widehat{\beta}_{1\tau np}$	.059	.209	.426	.943	.079	.122	.269	.943
$\widehat{\beta}_{2\tau np}$	.048	.838	.871	.940	.043	.474	.503	.945
$\widehat{\beta}^p_{1\tau np}$	.061	.205	.423	.944	.080	.119	.265	.945
$\widehat{\beta}^p_{2\tau np}$	.050	.836	.868	.942	.044	.470	.500	.946

Table 3b Continued

n	100				400			
	bias	*se*	*length*	*cov*	*bias*	*se*	*length*	*cov*
	MAR (6.5) 10%				MAR (6.5) 10%			
$\widehat{\beta}_{1\tau c}$	.103	.202	.416	.900	.088	.115	.223	.900
$\widehat{\beta}_{2\tau c}$	.099	.835	.870	.892	.077	.461	.482	.896
$\widehat{\beta}^p_{1\tau c}$	.105	.200	.412	.901	.090	.110	.219	.902
$\widehat{\beta}^p_{2\tau c}$	.101	.832	.865	.893	.078	.456	.478	.898
$\widehat{\beta}_{1\tau p}$	.056	.204	.420	.942	.036	.116	.230	.944
$\widehat{\beta}_{2\tau p}$	.047	.836	.868	.943	.035	.474	.490	.946
$\widehat{\beta}^p_{1\tau p}$	.057	.200	.416	.943	.039	.110	.227	.946
$\widehat{\beta}^p_{2\tau p}$	.048	.832	.862	,944	.037	.470	.485	.947
$\widehat{\beta}_{1\tau np}$	.059	.206	.421	.941	.039	.120	.237	.944
$\widehat{\beta}_{2\tau np}$	.045	.838	.872	.942	.042	.486	.498	.944
$\widehat{\beta}^p_{1\tau np}$	.060	.201	.418	.945	.040	.117	.230	.945
$\widehat{\beta}^p_{2\tau np}$	.047	.832	.868	.944	.045	.484	.494	.946
	MAR (6.5) 40%				MAR (6.5) 40%			
$\widehat{\beta}_{1\tau c}$	.129	.205	.432	.892	.104	.136	.250	.891
$\widehat{\beta}_{2\tau c}$	.127	.839	.871	.889	.103	.501	.495	.895
$\widehat{\beta}^p_{1\tau c}$	.130	.203	.430	.893	.105	.130	.246	.892
$\widehat{\beta}^p_{2\tau c}$	.128	.840	.869	.890	.101	.500	.497	.896
$\widehat{\beta}_{1\tau p}$	.128	.200	.427	.894	.038	.124	.261	.942
$\widehat{\beta}_{2\tau p}$	.057	.209	.435	.942	.040	.489	.499	.946
$\widehat{\beta}^p_{1\tau p}$	.050	.840	.872	.941	.039	.122	.257	.941
$\widehat{\beta}^p_{2\tau p}$	.059	.205	.430	.943	.041	.483	.494	.947
$\widehat{\beta}_{1\tau np}$	.049	.836	.870	.942	.077	.126	.239	.945
$\widehat{\beta}_{2\tau np}$	.101	.211	.428	.942	.044	.490	.500	.944
$\widehat{\beta}^p_{1\tau np}$	.090	.840	.872	.941	.078	.124	.235	.945
$\widehat{\beta}^p_{2\tau np}$	.103	.208	.426	.943	.045	.485	.495	.945

Table 3c $\varepsilon_\tau \sim \chi^2(4) - 4$, $\tau = 0.75$

n	100				400			
	bias	*se*	*length*	*cov*	*bias*	*se*	*length*	*cov*
$\widehat{\beta}_{1\tau}$	.056	.190	.399	.942	.032	.162	.205	.946
$\widehat{\beta}_{2\tau}$	.042	.842	.843	.944	.028	.428	.482	.947
$\widehat{\beta}_{1\tau}^p$	.057	.185	.394	.944	.033	.158	.200	.947
$\widehat{\beta}_{2\tau}^p$	.043	.838	.840	.945	.029	.424	.478	.948
	MAR (6.4) 10%				MAR (6.4) 10%			
$\widehat{\beta}_{1\tau c}$	.101	.206	.416	.895	.088	.110	.210	.900
$\widehat{\beta}_{2\tau c}$	.098	.858	.868	.899	.078	.466	.481	.901
$\widehat{\beta}_{1\tau c}^p$	.103	.202	.412	.896	.089	.105	.205	.903
$\widehat{\beta}_{2\tau c}^p$	.099	.855	.864	.900	.080	.460	.474	.902
$\widehat{\beta}_{1\tau p}$	.057	.207	.418	.941	.036	.112	.213	.943
$\widehat{\beta}_{2\tau p}$	.045	.861	.868	.942	.032	.450	.419	.945
$\widehat{\beta}_{1\tau p}^p$	.058	.203	.410	.943	.038	.110	.210	.944
$\widehat{\beta}_{2\tau p}^p$	.060	.856	.861	.944	.034	.446	.410	.946
$\widehat{\beta}_{1\tau np}$	.058	.210	.420	.942	.035	.124	.203	.942
$\widehat{\beta}_{2\tau np}$	.048	.863	.873	.941	.038	.454	.486	.943
$\widehat{\beta}_{1\tau np}^p$	.060	.206	.414	.947	.036	.120	.205	.944
$\widehat{\beta}_{2\tau np}^p$	.049	.860	.873	.943	.037	.450	.479	.945
	MAR (6.4) 40%				MAR (6.4) 40%			
$\widehat{\beta}_{1\tau c}$	.130	.212	.421	.891	.103	.136	.239	.896
$\widehat{\beta}_{2\tau c}$	.128	.871	.874	.894	.104	.473	.499	.901
$\widehat{\beta}_{1\tau c}^p$	.131	.210	.418	.889	.105	.130	.232	.893
$\widehat{\beta}_{2\tau c}^p$	.129	.867	.870	.895	.106	.470	.495	.902
$\widehat{\beta}_{1\tau p}$	.062	.213	.423	.942	.038	.114	.241	.943
$\widehat{\beta}_{2\tau p}$	.049	.874	.878	.943	.036	.467	.501	.945
$\widehat{\beta}_{1\tau p}^p$	.063	.210	.420	.943	.039	.110	.235	.945
$\widehat{\beta}_{2\tau p}^p$	.050	.870	.874	.944	.038	.461	.497	.946
$\widehat{\beta}_{1\tau np}$	.060	.216	.427	.941	.037	.118	.243	.943
$\widehat{\beta}_{2\tau np}$	.050	.869	.880	.941	.040	.465	.499	.942
$\widehat{\beta}_{1\tau np}^p$	.061	.210	.423	.943	.038	.115	.238	.945
$\widehat{\beta}_{2\tau np}^p$	.051	.857	.875	.942	.039	.463	.497	.948

Table 3c Continued

n	100				400			
	bias	*se*	*length*	*cov*	*bias*	*se*	*length*	*cov*
	MAR (6.5) 10%				MAR (6.5) 10%			
$\widehat{\beta}_{1\tau c}$	.102	.208	.418	.894	.090	.112	.213	.898
$\widehat{\beta}_{2\tau c}$	.102	.860	.873	.892	.081	.451	.482	.894
$\widehat{\beta}^p_{1\tau c}$	.103	.206	.415	.989	.091	.113	.210	.899
$\widehat{\beta}^p_{2\tau c}$	.104	.856	.870	.990	.082	.450	.480	.895
$\widehat{\beta}_{1\tau p}$	.059	.208	.420	.939	.038	.113	.216	.942
$\widehat{\beta}_{2\tau p}$	.048	.868	.870	.942	.034	.454	.499	.944
$\widehat{\beta}^p_{1\tau p}$	.060	.205	.417	.941	.036	.110	.214	.943
$\widehat{\beta}^p_{2\tau p}$	.049	.865	.964	.943	.036	.450	.495	.945
$\widehat{\beta}_{1\tau np}$	.061	.203	.413	.942	.036	.116	.218	.942
$\widehat{\beta}_{2\tau np}$	.057	.213	.483	.938	.040	.457	.501	.943
$\widehat{\beta}^p_{1\tau np}$	.050	.862	.871	.941	.037	.113	.210	.944
$\widehat{\beta}^p_{2\tau np}$	.058	.210	.480	.943	.042	.455	.499	.945
	MAR (6.5) 40%				MAR (6.5) 40%			
$\widehat{\beta}_{1\tau c}$	.129	.214	.422	.889	.104	.238	.240	.894
$\widehat{\beta}_{2\tau c}$	.125	.873	.879	.891	.108	.556	.494	.893
$\widehat{\beta}^p_{1\tau c}$	.130	.212	.420	.890	.105	.235	.241	.895
$\widehat{\beta}^p_{2\tau c}$	.126	.870	.877	.892	.107	,553	.492	.984
$\widehat{\beta}_{1\tau p}$	.060	.216	.424	.941	.040	.236	.243	.945
$\widehat{\beta}_{2\tau p}$	.050	.876	.880	.943	.038	.509	.500	.945
$\widehat{\beta}^p_{1\tau p}$	.061	.214	.427	.943	.041	.233	.240	.945
$\widehat{\beta}^p_{2\tau p}$	.052	.874	.874	.944	.042	.233	.242	.943
$\widehat{\beta}_{1\tau np}$	.061	.218	.423	.940	.039	.510	.501	.946
$\widehat{\beta}_{2\tau np}$	.051	.870	.881	.939	.043	.235	.244	.945
$\widehat{\beta}^p_{1\tau np}$	.062	.215	.420	.942	.038	.503	.500	.945
$\widehat{\beta}^p_{2\tau np}$	.053	.871	.882	938	.043	.491	.503	.942

Table 4a: Finite sample size of (4.4) with 10% MAR

τ	0.25		0.50		0.75	
$n = 100$						
$N(0,1)$	0.066^a	0.119^a	0.060^a	0.117^a	0.064^a	0.116^a
	0.066^b	0.113^b	0.063^b	0.112^b	0.065^b	0.113^b
	0.069^c	0.114^c	0.064^c	0.114^c	0.068^c	0.113^c
$\chi^2(4) - 4$	0.069^a	0.121^a	0.062^a	0.120^a	0.067^a	0.120^a
	0.071^b	0.115^b	0.063^b	0.114^b	0.069^b	0.115^b
	0.071^c	0.116^c	0.066^c	0.115^c	0.070^c	0.115^c
$t(5)$	0.065^a	0.120^a	0.062^a	0.118^a	0.066^a	0.119^a
	0.065^b	0.114^b	0.062^b	0.112^b	0.069^b	0.112^b
	0.067^c	0.115^c	0.064^c	0.114^c	0.069^c	0.114^c
$n = 400$						
$N(0,1)$	0.059^a	0.114^a	0.055^a	0.112^a	0.060^a	0.113^a
	0.060^b	0.109^b	0.057^b	0.108^b	0.062^b	0.109^b
	0.062^c	0.111^c	0.058^c	0.110^c	0.062^c	0.110^c
$\chi^2(4) - 4$	0.061^a	0.118^a	0.058^a	0.115^a	0.062^a	0.116^a
	0.061^b	0.111^b	0.060^b	0.109^b	0.064^b	0.110^b
	0.063^c	0.112^c	0.061^c	0.110^c	0.064^c	0.111^c
$t(5)$	0.058^a	0.116^a	0.057^a	0.116^a	0.063^a	0.115^a
	0.060^b	0.110^b	0.059^b	0.110^b	0.065^b	0.111^b
	0.063^c	0.111^c	0.061^c	0.111^c	0.065^c	0.112^c

a complete case, b IPW parametric, c IPW nonparametric

Table 4b: Finite sample size of (4.4) with 40% MAR

τ	0.25		0.50		0.75	
			$n = 100$			
$N(0,1)$	0.071^a	0.125^a	0.069^a	0.123^a	0.069^a	0.124^a
	0.064^b	0.113^b	0.063^b	0.113^b	0.064^b	0.113^b
	0.065^c	0.115^c	0.064^c	0.114^c	0.064^c	0.114^c
$\chi^2(4) - 4$	0.073^a	0.127^a	0.072^a	0.126^a	0.074^a	0.126^a
	0.065^b	0.116^b	0.064^b	0.115^b	0.064^b	0.116^b
	0.065^c	0.117^c	0.064^c	0.115^c	0.065^c	0.115^c
$t(5)$	0.074^a	0.127^a	0.073^a	0.125^a	0.073^a	0.125^a
	0.064^b	0.115^b	0.065^b	0.114^b	0.065^b	0.114^b
	0.065^c	0.116^c	0.065^c	0.115^c	0.066^c	0.115^c
			$n = 400$			
$N(0,1)$	0.068^a	0.121^a	0.067^a	0.119^a	0.067^a	0.120^a
	0.060^b	0.111^b	0.061^b	0.109^b	0.062^b	0.110^b
	0.062^c	0.113^c	0.061^c	0.110^c	0.062^c	0.111^c
$\chi^2(4) - 4$	0.069^a	0.123^a	0.068^a	0.121^a	0.068^a	0.122^a
	0.063^b	0.111^b	0.062^b	0.110^b	0.062^b	0.112^b
	0.063^c	0.112^c	0.062^c	0.112^c	0.063^c	0.113^c
$t(5)$	0.068^a	0.121^a	0.066^a	0.120^a	0.069^a	0.122^a
	0.063^b	0.112^b	0.063^b	0.110^b	0.064^b	0.111^b
	0.064^c	0.113^c	0.063^c	0.112^c	0.064^c	0.112^c

a complete case, *b* IPW parametric, *c* IPW nonparametric

Table 5a Finite sample power of the statistic $D_\pi\left(\theta_\tau^c\right)$ for $\tau = 0.25$

γ	$N(0,1)$	$\chi^2(4)-4$	$t(5)$	$N(0,1)$	$\chi^2(4)-4$	$t(5)$
		$n=100$			$n=400$	
	$.871^a$	$.854^a$	$.861^a$	$.932^a$	$.927^a$	$.931^a$
-1	$.910^b$	$.902^b$	$.908^b$	$.990^b$	$.987^b$	$.990^b$
	$.902^c$	$.900^c$	$.902^c$	$.987^c$	$.985^c$	$.984^c$
	$.655^a$	$.643^a$	$.650^a$	$.712^a$	$.719^a$	$.713^a$
-0.8	$.734^b$	$.723^b$	$.731^b$	$.801^b$	$.795^b$	$.792^b$
	$.710^c$	$.709^c$	$.711^c$	$.790^c$	$.789^c$	$.790^c$
	$.423^a$	$.416^a$	$.421^a$	$.503^a$	$.501^a$	$.503^a$
-0.6	$.512^b$	$.504^b$	$.510^b$	$.589^b$	$.580^b$	$.587^b$
	$.497^c$	$.501^c$	$.499^c$	$.584^c$	$.583^c$	$.582^c$
	$.218^a$	$.206^a$	$.210^a$	$.310^a$	$.303^a$	$.301^a$
-0.4	$.296^b$	$.290^b$	$.290^b$	$.399^b$	$.393^b$	$.391^b$
	$.284^c$	$.287^c$	$.285^c$	$.391^c$	$.390^c$	$.389^c$
	$.106^a$	$.105^a$	$.104^a$	$.142^a$	$.140^a$	$.143^a$
-0.2	$.157^b$	$.153^b$	$.150^b$	$.205^b$	$.201^b$	$.204^b$
	$.146^c$	$.149^c$	$.149^c$	$.201^c$	$.200^c$	$.202^c$
	$.065^a$	$.063^a$	$.064^a$	$.059^a$	$.060^a$	$.058^a$
0	$.058^b$	$.058^b$	$.054^b$	$.055^b$	$.054^b$	$.055^b$
	$.059^c$	$.057^c$	$.056^c$	$.056^c$	$.055^c$	$.056^c$
	$.110^a$	$.112^a$	$.109^a$	$.156^a$	$.153^a$	$.153^a$
0.2	$.165^b$	$.161^b$	$.158^b$	$.210^b$	$.205^b$	$.203^b$
	$.160^c$	$.156^c$	$.157^c$	$.207^c$	$.203^c$	$.202^c$
	$.226^a$	$.220^a$	$.221^a$	$.312^a$	$.310^a$	$.310^a$
0.4	$.305^b$	$.300^b$	$.300^b$	$.403^b$	$.400^b$	$.399^b$
	$.301^c$	$.296^c$	$.297^c$	$.399^c$	$.398^c$	$.397^c$
	$.434^a$	$.421^a$	$.423^a$	$.513^a$	$.512^a$	$.510^a$
0.6	$.521^b$	$.510^b$	$.515^b$	$.599^b$	$.595^b$	$.597^b$
	$.510^c$	$.511^c$	$.510^c$	$.595^c$	$.593^c$	$.595^c$
	$.663^a$	$.660^a$	$.661^a$	$.720^a$	$.716^a$	$.719^a$
0.8	$.742^b$	$.737^b$	$.740^b$	$.812^b$	$.810^b$	$.810^b$
	$.737^c$	$.730^c$	$.737^c$	$.806^c$	$.809^c$	$.805^c$
	$.894^a$	$.890^a$	$.893^a$	$.935^a$	$.933^a$	$.936^a$
1	$.924^b$	$.923^b$	$.921^b$	$.995^b$	$.992^b$	$.994^b$
	$.918^c$	$.920^c$	$.920^c$	$.991^c$	$.990^c$	$.991^c$

a complete case, b IPW parametric, c IPW nonparametric

Table 5b Finite sample power of the statistic $D_\pi\left(\theta_\tau^c\right)$ for $\tau = 0.50$

γ	$N(0,1)$	$\chi^2(4)-4$	$t(5)$	$N(0,1)$	$\chi^2(4)-4$	$t(5)$
		$n=100$			$n=400$	
	$.891^a$	$.864^a$	$.861^a$	$.964^a$	$.957^a$	$.961^a$
-1	$.919^b$	$.913^b$	$.918^b$	$.100^b$	$.100^b$	$.100^b$
	$.912^c$	$.910^c$	$.910^c$	$.998^c$	$.999^c$	$.100^c$
	$.676^a$	$.664^a$	$.670^a$	$.730^a$	$.725^a$	$.727^a$
-0.8	$.754^b$	$.743^b$	$.750^b$	$.822^b$	$.815^b$	$.820^b$
	$.738^c$	$.739^c$	$.734^c$	$.809^c$	$.805^c$	$.804^c$
	$.435^a$	$.432^a$	$.431^a$	$.512^a$	$.507^a$	$.509^a$
-0.6	$.521^b$	$.519^b$	$.520^b$	$.596^b$	$.590^b$	$.594^b$
	$.513^c$	$.513^c$	$.510^c$	$.590^c$	$.587^c$	$.589^c$
	$.225^a$	$.218^a$	$.221^a$	$.318^a$	$.313^a$	$.315^a$
-0.4	$.302^b$	$.299^b$	$.299^b$	$.404^b$	$.400^b$	$.399^b$
	$.299^c$	$.297^c$	$.298^c$	$.399^c$	$.399^c$	$.397^c$
	$.113^a$	$.110^a$	$.111^a$	$.154^a$	$.152^a$	$.153^a$
-0.2	$.164^b$	$.163^b$	$.162^b$	$.211^b$	$.210^b$	$.209^b$
	$.160^c$	$.159^c$	$.160^c$	$.210^c$	$.209^c$	$.208^c$
	$.063^a$	$.061^a$	$.062^a$	$.054^a$	$.056^a$	$.056^a$
0	$.056^b$	$.057^b$	$.055^b$	$.054^b$	$.055^b$	$.055^b$
	$.056^c$	$.056^c$	$.055^c$	$.055^c$	$.054^c$	$.055^c$
	$.118^a$	$.117^a$	$.118^a$	$.160^a$	$.158^a$	$.159^a$
0.2	$.169^b$	$.167^b$	$.168^b$	$.218^b$	$.215^b$	$.216^b$
	$.167^c$	$.166^c$	$.167^c$	$.216^c$	$.216^c$	$.214^c$
	$.234^a$	$.232^a$	$.231^a$	$.321^a$	$.318^a$	$.319^a$
0.4	$.317^b$	$.318^b$	$.316^b$	$.415^b$	$.413^b$	$.412^b$
	$.312^c$	$.312^c$	$.315^c$	$.414^c$	$.413^c$	$.412^c$
	$.451^a$	$.446^a$	$.450^a$	$.523^a$	$.522^a$	$.521^a$
0.6	$.534^b$	$.531^b$	$.532^b$	$.602^b$	$.599^b$	$.600^b$
	$.531^c$	$.530^c$	$.529^c$	$.599^c$	$.600^c$	$.599^c$
	$.679^a$	$.676^a$	$.675^a$	$.731^a$	$.732^a$	$.730^a$
0.8	$.753^b$	$.755^b$	$.754^b$	$.824^b$	$.821^b$	$.821^b$
	$.751^c$	$.754^c$	$.752^c$	$.826^c$	$.822^c$	$.824^c$
	$.899^a$	$.897^a$	$.895^a$	$.938^a$	$.937^a$	$.938^a$
1	$.924^b$	$.924^b$	$.924^b$	$.999^b$	$.997^b$	$.999^b$
	$.921^c$	$.920^c$	$.921^c$	$.995^c$	$.996^c$	$.998^c$

a complete case, b IPW parametric, c IPW nonparametric

Table 5c Finite sample power of the statistic $D_\pi\left(\theta_\tau^c\right)$ for $\tau = 0.75$

γ	$N(0,1)$	$\chi^2(4)-4$	$t(5)$	$N(0,1)$	$\chi^2(4)-4$	$t(5)$
		$n=100$			$n=400$	
	$.862^a$	$.860^a$	$.860^a$	$.930^a$	$.929^a$	$.932^a$
-1	$.904^b$	$.901^b$	$.905^b$	$.989^b$	$.988^b$	$.990^b$
	$.901^c$	$.902^c$	$.903^c$	$.988^c$	$.984^c$	$.985^c$
	$.658^a$	$.653^a$	$.654^a$	$.719^a$	$.715^a$	$.713^a$
-0.8	$.744^b$	$.743^b$	$.741^b$	$.811^b$	$.808^b$	$.805^b$
	$.736^c$	$.733^c$	$.734^c$	$.807^c$	$.806^c$	$.807^c$
	$.431^a$	$.425^a$	$.428^a$	$.513^a$	$.511^a$	$.510^a$
-0.6	$.510^b$	$.508^b$	$.511^b$	$.592^b$	$.589^b$	$.589^b$
	$.502^c$	$.505^c$	$.507^c$	$.594^c$	$.587^c$	$.586^c$
	$.215^a$	$.210^a$	$.211^a$	$.312^a$	$.310^a$	$.312^a$
-0.4	$.292^b$	$.291^b$	$.293^b$	$.395^b$	$.392^b$	$.394^b$
	$.288^c$	$.286^c$	$.290^c$	$.395^c$	$.393^c$	$.394^c$
	$.109^a$	$.107^a$	$.105^a$	$.147^a$	$.145^a$	$.145^a$
-0.2	$.154^b$	$.152^b$	$.150^b$	$.208^b$	$.207^b$	$.208^b$
	$.149^c$	$.148^c$	$.146^c$	$.206^c$	$.205^c$	$.207^c$
	$.063^a$	$.061^a$	$.062^a$	$.057^a$	$.059^a$	$.058^a$
0	$.055^b$	$.054^b$	$.053^b$	$.054^b$	$.053^b$	$.053^b$
	$.056^c$	$.055^c$	$.055^c$	$.055^c$	$.054^c$	$.054^c$
	$.113^a$	$.111^a$	$.110^a$	$.160^a$	$.159^a$	$.157^a$
0.2	$.170^b$	$.168^b$	$.168^b$	$.214^b$	$.210^b$	$.208^b$
	$.168^c$	$.166^c$	$.167^c$	$.217^c$	$.212^c$	$.210^c$
	$.232^a$	$.228^a$	$.227^a$	$.312^a$	$.311^a$	$.313^a$
0.4	$.303^b$	$.301^b$	$.302^b$	$.401^b$	$.402^b$	$.403^b$
	$.300^c$	$.299^c$	$.298^c$	$.398^c$	$.397^c$	$.398^c$
	$.439^a$	$.421^a$	$.423^a$	$.513^a$	$.512^a$	$.510^a$
0.6	$.529^b$	$.510^b$	$.515^b$	$.599^b$	$.595^b$	$.597^b$
	$.528^c$	$.511^c$	$.510^c$	$.595^c$	$.593^c$	$.595^c$
	$.663^a$	$.660^a$	$.661^a$	$.720^a$	$.716^a$	$.719^a$
0.8	$.742^b$	$.737^b$	$.740^b$	$.812^b$	$.810^b$	$.810^b$
	$.737^c$	$.730^c$	$.737^c$	$.806^c$	$.809^c$	$.805^c$
	$.894^a$	$.890^a$	$.893^a$	$.935^a$	$.933^a$	$.936^a$
1	$.924^b$	$.923^b$	$.921^b$	$.995^b$	$.992^b$	$.994^b$
	$.918^c$	$.920^c$	$.920^c$	$.991^c$	$.990^c$	$.991^c$

a complete case, *b* IPW parametric, *c* IPW nonparametric

Table 6a Finite sample size for the Wald statistics (4.6) with 10% MAR

τ	0.25		0.50		0.75	
			$n = 100$			
$N(0,1)$	$.064^{a}$	$.115^{a}$	$.063^{a}$	$.113^{a}$	$.065^{a}$	$.114^{a}$
	$.061^{b}$	$.109^{b}$	$.060^{b}$	$.108^{b}$	$.061^{b}$	$.110^{b}$
	$.062^{c}$	$.110^{c}$	$.061^{c}$	$.110^{c}$	$.062^{c}$	$.111^{c}$
	$.059^{\dagger a}$	$.110^{\dagger a}$	$.058^{\dagger a}$	$.110^{\dagger a}$	$,060^{\dagger a}$	$.110^{\dagger a}$
	$.057^{\dagger b}$	$.106^{\dagger b}$	$.057^{\dagger b}$	$.105^{\dagger b}$	$.055^{\dagger b}$	$.107^{\dagger b}$
	$.057^{\dagger c}$	$.105^{\dagger c}$	$.057^{\dagger c}$	$.106^{\dagger c}$	$.056^{\dagger c}$	$.107^{\dagger c}$
$\chi^2(4) - 4$	$.066^{a}$	$.119^{a}$	$.065^{a}$	$.117^{a}$	$.067^{a}$	$.118^{a}$
	$.063^{b}$	$.111^{b}$	$.062^{b}$	$.109^{b}$	$.062^{b}$	$.110^{b}$
	$.063^{c}$	$.112^{c}$	$.061^{c}$	$.110^{c}$	$.064^{c}$	$.112^{c}$
	$.056^{\dagger a}$	$.110^{\dagger a}$	$.057^{\dagger a}$	$.108^{\dagger a}$	$.057^{\dagger a}$	$.110^{\dagger a}$
	$.059^{\dagger b}$	$.107^{\dagger b}$	$.058^{\dagger b}$	$.107^{\dagger b}$	$.058^{\dagger b}$	$.108^{\dagger b}$
	$.057^{\dagger c}$	$.106^{\dagger c}$	$.059^{\dagger c}$	$.106^{\dagger c}$	$.057^{\dagger c}$	$.109^{\dagger c}$
$t(5)$	$.065^{a}$	$.118^{a}$	$.066^{a}$	$.116^{a}$	$.068^{a}$	$.117^{a}$
	$.062^{b}$	$.112^{b}$	$.063^{b}$	$.110^{b}$	$.064^{b}$	$.111^{b}$
	$.063^{c}$	$.113^{c}$	$.062^{c}$	$.111^{c}$	$.065^{c}$	$.112^{c}$
	$.057^{\dagger a}$	$.106^{\dagger a}$	$.056^{\dagger a}$	$.109^{\dagger a}$	$.058^{\dagger a}$	$.110^{\dagger a}$
	$.058^{\dagger b}$	$.107^{\dagger b}$	$.057^{\dagger b}$	$.106^{\dagger b}$	$.059^{\dagger b}$	$.106^{\dagger b}$
	$.059^{\dagger c}$	$.106^{\dagger c}$	$.058^{\dagger c}$	$.106^{\dagger c}$	$.058^{\dagger c}$	$.107^{\dagger c}$
			$n = 400$			
$N(0,1)$	$.058^{a}$	$.111^{a}$	$.057^{a}$	$.110^{a}$	$.058^{a}$	$.112^{a}$
	$.055^{b}$	$.106^{b}$	$.054^{b}$	$.106^{b}$	$.055^{b}$	$.117^{b}$
	$.056^{c}$	$.107^{c}$	$.055^{c}$	$.108^{c}$	$.056^{c}$	$.108^{c}$
	$.054^{\dagger a}$	$.106^{\dagger a}$	$.054^{\dagger a}$	$.106^{\dagger a}$	$.055^{\dagger a}$	$.107^{\dagger a}$
	$.053^{\dagger b}$	$.104^{\dagger b}$	$.053^{\dagger b}$	$.104^{\dagger b}$	$.054^{\dagger b}$	$.103^{\dagger b}$
	$.053^{\dagger c}$	$.103^{\dagger c}$	$.54^{\dagger c}$	$.105^{\dagger c}$	$.0553^{\dagger c}$	$.109^{\dagger c}$
$\chi^2(4) - 4$	$.060^{a}$	$.117^{a}$	$.058^{a}$	$.115^{a}$	$.059^{a}$	$.117^{a}$
	$.057^{b}$	$.107^{b}$	$.055^{b}$	$.107^{b}$	$.056^{b}$	$.108^{b}$
	$.057^{c}$	$.108^{c}$	$.056^{c}$	$.107^{c}$	$.057^{c}$	$.108^{c}$
	$.056^{\dagger a}$	$.110^{\dagger a}$	$.054^{\dagger a}$	$.109^{\dagger a}$	$056^{\dagger a}$	$.110^{\dagger a}$
	$.055^{\dagger b}$	$.104^{\dagger b}$	$.053^{\dagger b}$	$.105^{\dagger b}$	$.053^{\dagger b}$	$..106^{\dagger b}$
	$.055^{\dagger c}$	$.106^{\dagger c}$	$.053^{\dagger c}$	$.104^{\dagger c}$	$.055^{\dagger c}$	$.104^{\dagger c}$
$t(5)$	$.061^{a}$	$.114^{a}$	$.058^{a}$	$.113^{a}$	$.060^{a}$	$.115^{a}$
	$.056^{b}$	$.109^{b}$	$.056^{b}$	$.107^{b}$	$.055^{b}$	$.108^{b}$
	$.057^{c}$	$.109^{c}$	$.056^{c}$	$.108^{c}$	$.056^{c}$	$.107^{c}$
	$.056^{\dagger a}$	$.108^{\dagger a}$	$.054^{\dagger a}$	$.108^{\dagger a}$	$.056^{\dagger a}$	$.110^{\dagger a}$
	$.053^{\dagger b}$	$.106^{\dagger b}$	$.055^{\dagger b}$	$.106^{\dagger b}$	$.053^{\dagger b}$	$.105^{\dagger b}$
	$.055^{\dagger c}$	$.109^{\dagger c}$	$.054^{\dagger c}$	$.105^{\dagger c}$	$.052^{\dagger c}$	$.105^{\dagger c}$

Table 6b Finite sample size for the Wald statistic (4.6) with 40% MAR

τ	0.25		0.50		0.75	
			$n = 100$			
$N(0,1)$	$.071^{a}$	$.122^{a}$	$.070^{a}$	$.121^{a}$	$.071^{a}$	$.122^{a}$
	$.064^{b}$	$.112^{b}$	$.062^{b}$	$.110^{b}$	$.063^{b}$	$.111^{b}$
	$.064^{c}$	$.112^{c}$	$.063^{c}$	$.111^{c}$	$.063^{c}$	$.112^{c}$
	$.059^{\dagger a}$	$.114^{\dagger a}$	$.065^{\dagger a}$	$.114^{\dagger a}$	$.065^{\dagger a}$	$.113^{\dagger a}$
	$0.57^{\dagger b}$	$.109^{\dagger b}$	$.059^{\dagger b}$	$.105^{\dagger b}$	$.060^{\dagger b}$	$.106^{\dagger b}$
	$.060^{\dagger c}$	$.108^{\dagger c}$	$.058^{\dagger c}$	$.106^{\dagger c}$	$.059^{\dagger c}$	$.107^{\dagger c}$
$\chi^2(4) - 4$	$.074^{a}$	$.125^{a}$	$.073^{a}$	$.124^{a}$	$.073^{a}$	$.123^{a}$
	$.066^{b}$	$.114^{b}$	$.065^{b}$	$.113^{b}$	$.062^{b}$	$.113^{b}$
	$.066^{c}$	$.115^{c}$	$.065^{c}$	$.113^{c}$	$.064^{c}$	$.114^{c}$
	$.070^{\dagger a}$	$.112^{\dagger a}$	$.065^{\dagger a}$	$.117^{\dagger a}$	$.065^{\dagger a}$	$.116^{\dagger a}$
	$.061^{\dagger b}$	$.107^{\dagger b}$	$.062^{\dagger b}$	$.109^{\dagger b}$	$.057^{\dagger b}$	$.110^{\dagger b}$
	$.061^{\dagger c}$	$.110^{\dagger c}$	$.061^{\dagger c}$	$.110^{\dagger c}$	$.058^{\dagger c}$	$.109^{\dagger c}$
$t(5)$	$.074^{a}$	$.123^{a}$	$.072^{a}$	$.124^{a}$	$.075^{a}$	$.125^{a}$
	$.065^{b}$	$.114^{b}$	$.063^{b}$	$.114^{b}$	$.064^{b}$	$.114^{b}$
	$.064^{c}$	$.115^{c}$	$.062^{c}$	$.114^{c}$	$.065^{c}$	$.115^{c}$
	$.065^{\dagger a}$	$.118^{\dagger a}$	$.065^{\dagger a}$	$.114^{\dagger a}$	$.067^{\dagger a}$	$.118^{\dagger a}$
	$.062^{\dagger b}$	$.110^{\dagger b}$	$0.57^{\dagger b}$	$.109^{\dagger b}$	$.059^{\dagger b}$	$.109^{\dagger b}$
	$.060^{\dagger c}$	$.110^{\dagger c}$	$.058^{\dagger c}$	$.108^{\dagger c}$	$.058^{\dagger c}$	$.108^{\dagger c}$
			$n = 400$			
$N(0,1)$	$.068^{a}$	$.120^{a}$	$.067^{a}$	$.119^{a}$	$.068^{a}$	$.119^{a}$
	$.060^{b}$	$.108^{b}$	$.059^{b}$	$.108^{b}$	$.060^{b}$	$.109^{b}$
	$.061^{c}$	$.108^{c}$	$.059^{c}$	$.109^{c}$	$.059^{c}$	$.110^{c}$
	$.056^{\dagger a}$	$.110^{\dagger a}$	$.060^{\dagger a}$	$.110^{\dagger a}$	$.058^{\dagger a}$	$.110^{\dagger a}$
	$.056^{\dagger b}$	$.106^{\dagger b}$	$.056^{\dagger b}$	$.105^{\dagger b}$	$.055^{\dagger b}$	$.105^{\dagger b}$
	$.056^{\dagger c}$	$.104^{\dagger c}$	$0.56^{\dagger c}$	$.104^{\dagger c}$	$.055^{\dagger c}$	$.106^{\dagger c}$
$\chi^2(4) - 4$	$.070^{a}$	$.122^{a}$	$.068^{a}$	$.121^{a}$	$.069^{a}$	$.121^{a}$
	$.061^{b}$	$.109^{b}$	$.059^{b}$	$.109^{b}$	$.060^{b}$	$.109^{b}$
	$.060^{c}$	$.110^{c}$	$.060^{c}$	$.110^{c}$	$.061^{c}$	$.109^{c}$
	$.060^{\dagger a}$	$.112^{\dagger a}$	$.057^{\dagger a}$	$.110^{\dagger a}$	$,055^{\dagger a}$	$.112^{\dagger a}$
	$.061^{\dagger b}$	$.106^{\dagger b}$	$.055^{\dagger b}$	$.105^{\dagger b}$	$.056^{\dagger b}$	$.106^{\dagger b}$
	$.056^{\dagger c}$	$.105^{\dagger c}$	$.055^{\dagger c}$	$.106^{\dagger c}$	$.056^{\dagger c}$	$.104^{\dagger c}$
$t(5)$	$.071^{a}$	$.121^{a}$	$.069^{a}$	$.120^{a}$	$.070^{a}$	$.120^{a}$
	$.060^{b}$	$.110^{b}$	$.059^{b}$	$.109^{b}$	$.060^{b}$	$.109^{b}$
	$.061^{c}$	$.111^{c}$	$.060^{c}$	$.109^{c}$	$.061^{c}$	$.110^{c}$
	$.055^{\dagger a}$	$.112^{\dagger a}$	$.060^{\dagger a}$	$.110^{\dagger a}$	$.064^{\dagger a}$	$.110^{\dagger a}$
	$.056^{\dagger b}$	$.106^{\dagger b}$	$.055^{\dagger b}$	$.105^{\dagger b}$	$0.56^{\dagger b}$	$.106^{\dagger b}$
	$.054^{\dagger c}$	$.105^{\dagger c}$	$.0056^{\dagger c}$	$.104^{\dagger c}$	$.056^{\dagger c}$	$.105^{\dagger c}$

Table 7a Estimates, standard errors, length
of confidence interval and p-values for $\tau = 0.25$

	β_1	β_2	β_3
		complete	
$\widehat{\beta}_j$	-69.928	0.062	1.435
se	34.811	0.029	0.400
$length$	97.950	0.047	1.044
$p-value$	0.047	0.047	0.000
		IPW par	
$\widehat{\beta}_j$	-68.505	0.061	1.410
se	35.328	0.029	0.400
$length$	94.025	0.043	1.105
$p-value$	0.055	0.08	0.000
		IPW nopar	
$\widehat{\beta}_j$	-70.861	0.048	1.481
se	33.902	0.028	0.380
$length$	97.50	0.04	1.122
$p-value$	0.038	0.091	0.000

Table 7b Estimates, standard errors, length
of confidence intervals and p-values for $\tau = 0.50$

	β_1	β_2	β_3
		complete	
$\widehat{\beta}_j$	-75.603	0.033	1.782
se	33.141	0.033	0.324
$length$	40.143	0.077	0.487
$p-value$	0.024	0.324	0.000
		IPW par	
$\widehat{\beta}_j$	-74.618	0.032	1.762
se	34.376	0.034	0.369
$length$	38.727	0.078	0.453
$p-value$	0.032	0.347	0.000
		IPW nonpar	
$\widehat{\beta}_j$	-76.822	0.038	1.758
se	33.606	0.033	0.352
$length$	41.121	0.081	0.441
$p-value$	0.024	0.261	0.000

Table 7c Estimates, standard errors, length of confidence intervals and p-values for $\tau = 0.75$

	β_1	β_2	β_3
		complete	
$\widehat{\beta}_j$	-91.565	0.039	2.116
se	22.671	0.033	0.298
$length$	107.832	0.141	1.086
$p-value$	0.001	0.239	0.000
		IPW par	
$\widehat{\beta}_j$	-89.730	0.039	2.803
se	28.715	0.033	0.303
$length$	93.223	0.132	1.025
$p-value$	0.002	0.244	0.000
		IPW nonpar	
$\widehat{\beta}_j$	-96.633	0.040	2.077
se	29.267	0.033	0.297
$length$	107.902	0.133	0.928
$p-value$	0.001	0.229	0.000

,

Table 8 Sample and p values of the statistic $D_\pi(\theta_\tau^c)$

	Complete		IPW par		IPW nonpar	
$\tau = 0.25$	2.43	0.007	2.12	0.017	2.10	0.017
$\tau = 0.50$	2.51	0.006	2.18	0.014	2.17	0.015
$\tau = 0.75$	2.46	0.007	2.15	0.016	2.14	0.015

Table 9 Comparisons of $R^1_{\tau *}$

	unrestricted	restricted
$R^1_{0.25c}$	0.423	0.321
$R^1_{0.25p}$	0.441	0.323
$R^1_{0.25np}$	0.432	0.322
$R^1_{0.50c}$	0.487	0.397
$R^1_{0.05p}$	0.496	0.403
$R^1_{0.50np}$	0.494	0.402
$R^1_{0.75c}$	0.542	0.445
$R^1_{0.75p}$	0.559	0.452
$R^1_{0.75np}$	0.551	0.450

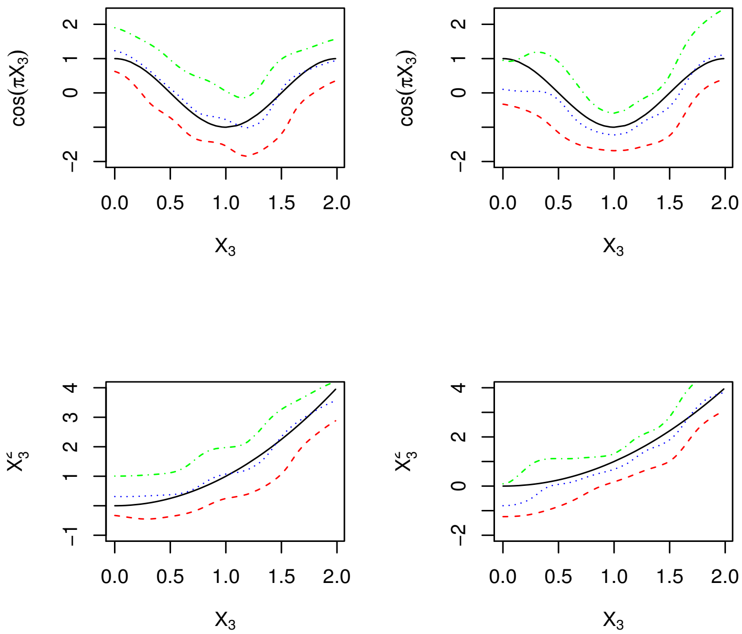

Figure 1: Nonparametric quantile $(\tau = 0.25, 0.50, 0.75)$ estimates of the varying coefficients $\cos(\pi X_3)$ and X_3^2 with no missing observations, $n = 100$, $\varepsilon \sim N(0,1)$ (left column) and $\varepsilon \sim \chi_4^2 - 4$ (right column)

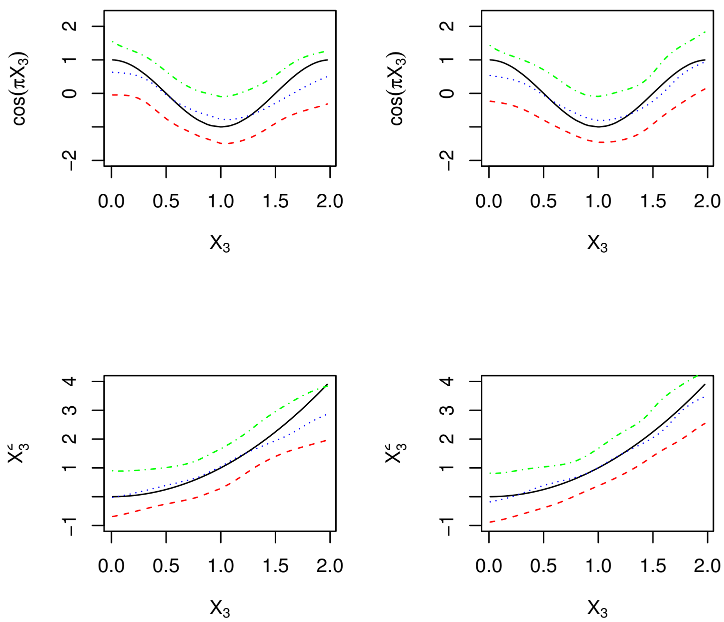

Figure 2: Nonparametric quantile $(\tau = 0.25, 0.50, 0.75)$ estimates of the varying coefficients $\cos(\pi X_3)$ and X_3^2 with 40% MAR observations, $n = 100$ and $\varepsilon \sim N(0,1)$

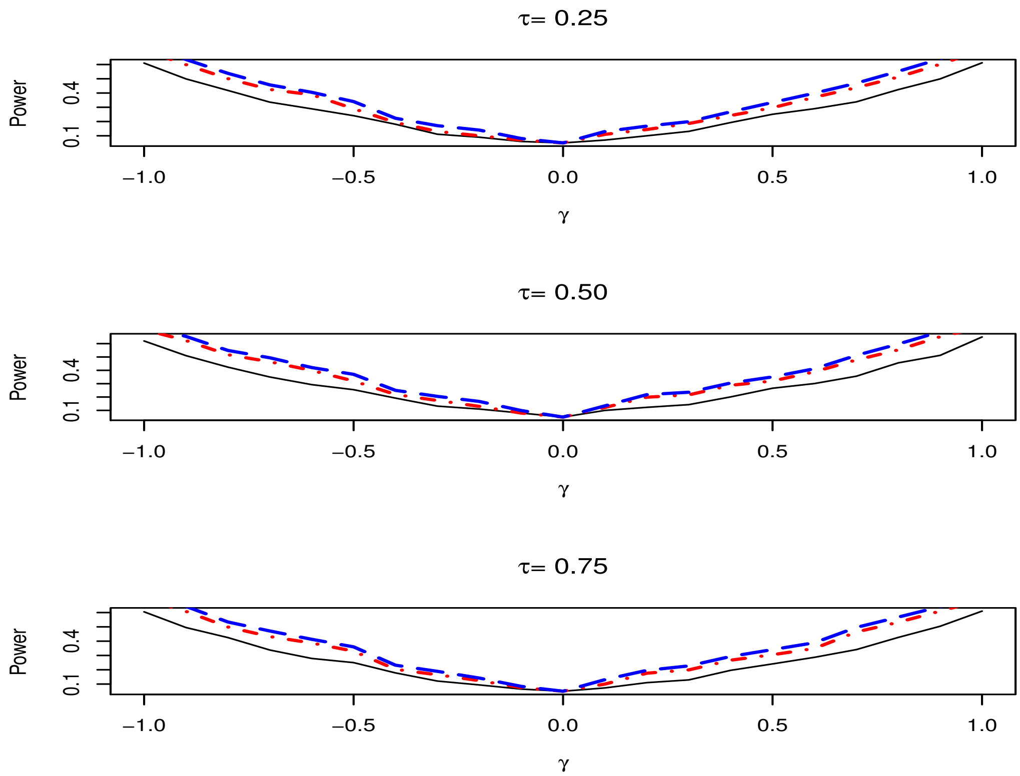

Figure 3: Size adjusted power of (4.4) for the 3 nonparametric quantile estimators based on the complete case (solid line), the parametric IPW estimator (dash line) and the nonparametric IPW estimator (dash dot line) for $n = 100$.

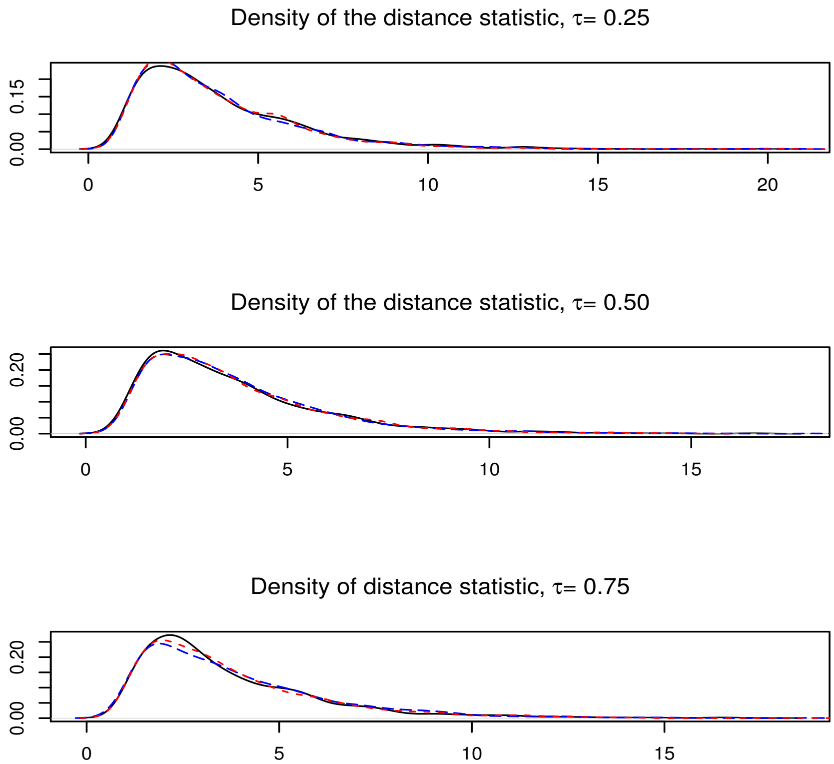

Figure 4: Kernel densities of the distance statistic of Proposition 10, dash dot line corresponds to $b = b/2$ and dash line corresponds to $b = 3/2$.

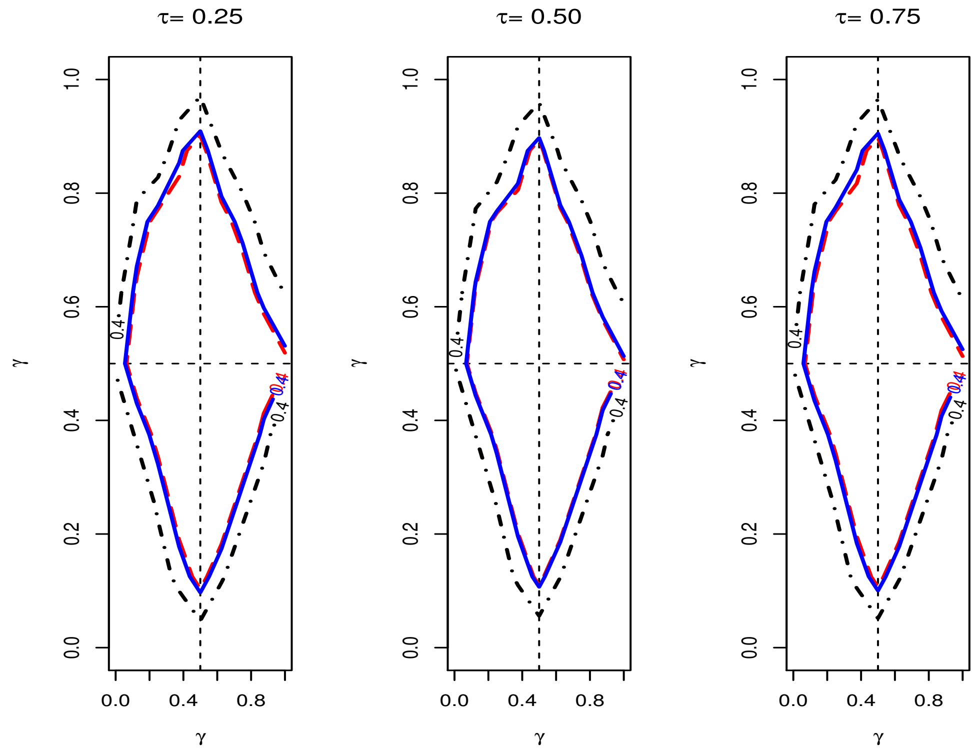

Figure 5: Contour plots of the finite sample power of W^p: The dash dot line corresponds to the complete case estimates, the dash line corresponds to the IPW nonparametric estimates and the continuous line corresponds to the IPW parametric estimates.

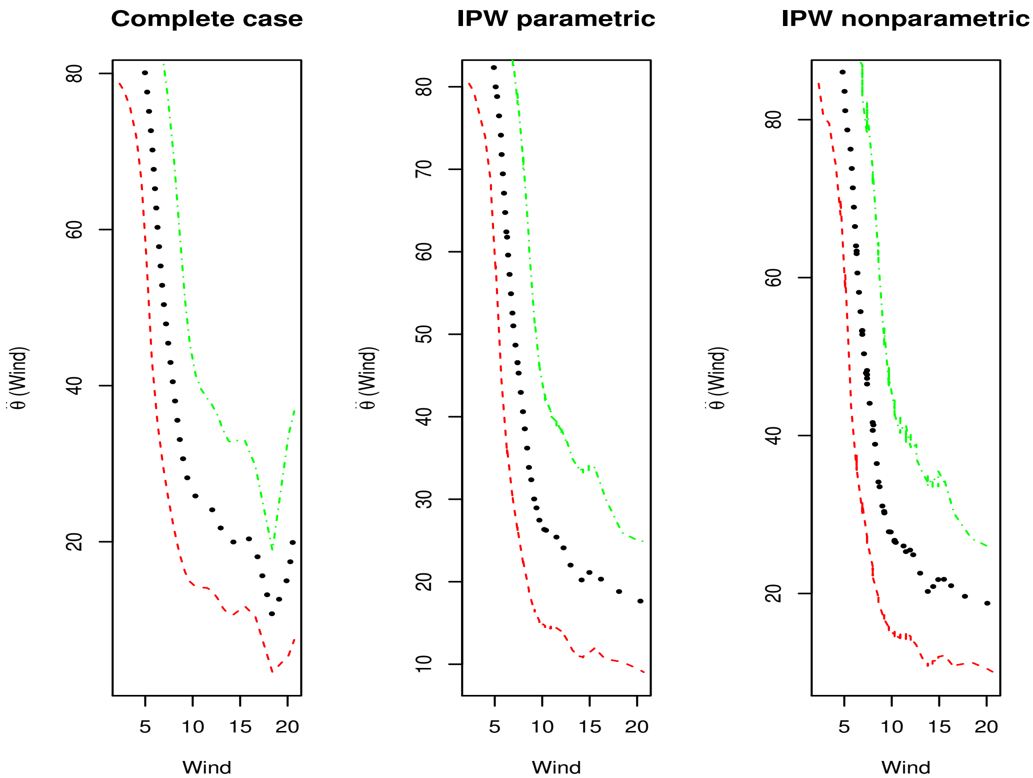

Figure 6 Nonparametric quantile $(\tau = 0.25, 0.50, 0.75)$ estimates of the wind effect on ozone layer

10 Supplemental Appendix

10.1 Proofs

Throughout this appendix we use the following abbreviations: "CLT", "CMT" and "LNN" denote, respectively, central limit theorem, continuous mapping theorem and (possibly uniform) law of large numbers. We also use "CL" and "QAL" to denote, respectively, the convexity lemma (Pollard 1991) and the "quadratic approximation lemma" (Fan & Gijbels 1996). Finally, we use the following identity (Knight 1999)

$$\rho_\tau (x-y) - \rho_\tau (x) = -y(\tau - I(x<0)) + \int_0^y (I(x \le t) - I(x \le 0))\,dt. \tag{10.1}$$

Proof of Theorem 1. Let $\pi(Z_{oi}) := \pi_i$,

$$W_i = \left[X_{1i}^T, X_{2i}^T, X_{2i}^T (X_{3i} - x_3)/h\right]^T,$$

$$\varepsilon_i^* = Y_i - X_{1i}^T \beta_{0\tau} - X_{2i}^T (a_\tau + b_\tau (X_{3i} - x_3)),$$

$$\gamma_\tau = (nh)^{1/2} \left[(\beta_\tau - \beta_{0\tau})^T, (a_\tau - \theta_{0\tau}(x_3))^T, h\left(b_\tau - \theta_{0\tau}'(x_3)\right)^T\right]^T$$

and

$$R_n\left(\gamma_\tau,\widehat{\pi},x_3\right)=\sum_{i=1}^{n}\frac{\delta_i}{\widehat{\pi}_i}\left[\rho_\tau\left(\varepsilon_i^*-\frac{W_i^T\gamma_\tau}{(nh)^{1/2}}\right)-\rho_\tau\left(\varepsilon_i^*\right)\right]K_h\left(X_{3i}-x_3\right)$$

denote the normalized local objective function $Q_n\left(\beta_\tau,a_\tau+b_\tau\left(X_{3i}-x_3\right),\widehat{\pi}\right)K_h\left(X_{3i}-x_3\right)$, which can be written as

$$R_n\left(\gamma_\tau,\widehat{\pi},x_3\right)=R_{1n}\left(\gamma_\tau,\pi_0,x_3\right)-R_{2n}\left(\gamma_\tau,\widehat{\pi},x_3\right),$$

where

$$R_{1n}\left(\gamma_\tau,\pi_0,x_3\right)=\sum_{i=1}^{n}\frac{\delta_i}{\pi_{0i}}\left[\rho_\tau\left(\varepsilon_i^*-\frac{W_i^T\gamma_\tau}{(nh)^{1/2}}\right)-\rho_\tau\left(\varepsilon_i^*\right)\right]K_h\left(X_{3i}-x_3\right),$$

$$R_{2n}\left(\gamma_\tau,\widehat{\pi},x_3\right)=\sum_{i=1}^{n}\frac{\delta_i\left(\widehat{\pi}_i-\pi_{0i}\right)}{\widehat{\pi}_i\pi_{0i}}\left[\rho_\tau\left(\varepsilon_i^*-\frac{W_i^T\gamma_\tau}{(nh)^{1/2}}\right)-\rho_\tau\left(\varepsilon_i^*\right)\right]K_h\left(X_{3i}-x_3\right).$$

By (10.1), we have

$$R_{1n}\left(\gamma_\tau,\pi_0,x_3\right)=\frac{\gamma_\tau^T}{(nh)^{1/2}}\sum_{i=1}^{n}\frac{\delta_i}{\pi_{0i}}W_i\rho_\tau'\left(\varepsilon_i^*\right)K_h\left(X_{3i}-x_3\right)+S_{1n}\left(\gamma,\pi_0,x_3\right),$$

where $\rho_\tau'\left(\cdot\right)=\tau-I\left(\cdot<0\right)$, and

$$S_{1n}\left(\gamma_\tau,\pi_0,x_3\right)=\sum_{i=1}^{n}\frac{\delta_i}{\pi_{0i}}\int_0^{\frac{W_i^T\gamma_\tau}{(nh)^{1/2}}}\left(I\left(\varepsilon_i^*\le t\right)-I\left(\varepsilon_i^*\le 0\right)\right)K_h\left(X_{3i}-x_3\right)dt.$$

By the consistency results for kernel estimators of Masry (1996)

$$S_{1n}\left(\gamma_\tau,\pi_0,x_3\right)=E\left[S_{1n}\left(\gamma_\tau,\pi_0,x_3\right)\right]+O_p\left(\left(\frac{\log n}{nh}\right)^{1/2}\right)\tag{10.2}$$

uniformly for $x_3\in\mathcal{X}_3$. Let $\varsigma_\tau\left(x_3\right)=\theta_{0\tau}\left(X_3\right)-X_2^T\left(a_\tau+b_\tau\left(X_3-x_3\right)\right)$; by iterated expectations $E\left[S_{1n}\left(\gamma,\pi_0,x_3\right)\right]=EE\left[S_{1n}\left(\gamma,\pi_0,x_3\right)|X_i\right]$, so using a Taylor expansion we have

$$E\left[S_{1n}\left(\gamma_\tau,\pi_0,x_3\right)|X_i\right]=\sum_{i=1}^{n}\int_0^{\frac{W_i^T\gamma_\tau}{(nh)^{1/2}}}\left(F_{\varepsilon_i|X_i}\left(\varsigma_{i\tau}\left(x_3\right)+t\right)-F_{\varepsilon_i|X_i}\left(\varsigma_{i\tau}\left(x_3\right)\right)\right)K_h\left(X_{3i}-x_3\right)dt=$$
$$\sum_{i=1}^{n}\int_0^{\frac{W_i^T\gamma_\tau}{(nh)^{1/2}}}f_{\varepsilon_i|X_i}\left(\overline{\varsigma}_{i\tau}\left(x_3\right)\right)tK_h\left(X_{3i}-x_3\right)dt,$$

where $\overline{\varsigma}_{i\tau}\left(x_3\right)$ is the mean value between 0 and $\varsigma_{i\tau}\left(x_3\right)+t$. Adding and subtracting $\sum_{i=1}^{n}\int_0^{\frac{W_i^T\gamma_\tau}{(nh)^{1/2}}}f_{\varepsilon_i|X_i}\left(0\right)\times tK_h\left(X_{3i}-x_3\right)dt$

$$\left|\sum_{i=1}^{n}\int_0^{\frac{W_i^T\gamma_\tau}{(nh)^{1/2}}}f_{\varepsilon_i|X_i}\left(\overline{\varsigma}_{i\tau}\left(x_3\right)\right)tK_h\left(X_{3i}-x_3\right)dt-\sum_{i=1}^{n}\int_0^{\frac{W_i^T\gamma_\tau}{(nh)^{1/2}}}f_{\varepsilon_i|X_i}\left(0\right)tK_h\left(X_{3i}-x_3\right)dt\right|\le$$
$$\sup_{x_3\in\mathcal{X}_3}\frac{C}{2}\frac{\gamma_\tau^T}{nh}\sum_{i=1}^{n}\left|\overline{\varsigma}_{i\tau}\left(x_3\right)\right|W_i^{\otimes 2}\gamma_\tau K_h\left(X_{3i}-x_3\right)=O_p\left(h^2\right),$$

for some $C > 0$, hence

$$E\left[S_n\left(\gamma_\tau, \pi_0, x_3\right)|X_i\right] = \frac{1}{2}\frac{\gamma_\tau^T}{nh}\sum_{i=1}^n f_{\varepsilon_i|X_i}(0)\, W_i^{\otimes 2}\gamma_\tau K_h\left(X_{3i} - x_3\right) + o_p(1), \tag{10.3}$$

and by a standard kernel calculation

$$E\left[E\left[S_n\left(\gamma_\tau, \pi_0, x_3\right)|X\right]\right] = \frac{1}{2}f_{X_3}\left(x_3\right)\gamma_\tau^T\Sigma\left(x_3\right)\gamma_\tau + o(1),$$

where

$$\Sigma\left(x_3\right) = E\left\{f_{\varepsilon|X}(0)\begin{bmatrix} X_1^{\otimes 2} & X_1X_2^T & O_{kp} \\ \left(X_1X_2^T\right)^T & X_2^{\otimes 2} & O_{pp} \\ O_{kp}^T & O_{pp} & \kappa_2 X_2^{\otimes 2}\end{bmatrix}|X_3 = x_3\right\}. \tag{10.4}$$

Combining (10.2) and (10.3), we have that

$$\begin{aligned} R_{1n}\left(\gamma_\tau, \pi_0, x_2\right) = &\frac{\gamma_\tau^T}{(nh)^{1/2}}\sum_{i=1}^n \frac{\delta_i}{\pi_{0i}}W_i\rho_\tau'\left(\varepsilon_i^*\right)K_h\left(X_{3i} - x_3\right) + \frac{1}{2}f_{X_3}\left(x_3\right)\gamma_\tau^T\Sigma\left(x_3\right)\gamma_\tau + \\ &O_p\left(\left(\frac{\log n}{nh}\right)^{1/2} + h^2\right)\end{aligned}$$

uniformly in $x_3 \in \mathcal{X}_3$. Note that for $\widehat{\pi}_i = \pi_i\left(\widehat{\alpha}\right)$ - that is for π_{0i} estimated parametrically-

$$\begin{aligned}\left|R_{2n}\left(\gamma_\tau, \widehat{\pi}_i, x_3\right)\right| &\leq \left\|\widehat{\alpha} - \alpha_0\right\|\left\|\sum_{i=1}^n \frac{\delta_i}{\pi_i^2}\frac{\partial \pi_i\left(\overline{\alpha}\right)}{\partial \alpha}\left[\rho_\tau\left(\varepsilon_i^* - \frac{W_i^T\gamma_\tau}{(nh)^{1/2}}\right) - \rho_\tau\left(\varepsilon_i^*\right)\right]K_h\left(X_{3i} - x_3\right)\right\| + o_p(1) \\ &= O_p\left(n^{-1/2}\right)O_p\left((nh)^{1/2}\right) = o_p(1)\end{aligned}$$

by A4, where $\overline{\alpha}$ is the mean value, whereas for π_{0i} estimated nonparametrically

$$\begin{aligned}\left|R_{2n}\left(\gamma_\tau, \widehat{\pi}_i, x_3\right)\right| &\leq \sup_{Z_{oi}\in\mathcal{Z}}\left\|\widehat{\pi}_i - \pi_{0i}\right\|\left\|\sum_{i=1}^n \frac{\delta_i}{\pi_i^2}\left[\rho_\tau\left(\varepsilon_i^* - \frac{W_i^T\gamma_\tau}{(nh)^{1/2}}\right) - \rho_\tau\left(\varepsilon_i^*\right)\right]K_h\left(X_{3i} - x_3\right)\right\| \\ &= O_p\left(\left(\frac{\log n}{nb^{\dim(Z_o)}}\right)^{1/2} + b^2\right)O_p\left((nh)^{1/2}\right) = o_p(1)\end{aligned}$$

by A2 and A5. Thus $R_n\left(\gamma_\tau, \widehat{\pi}, x_3\right) = R_{1n}\left(\gamma_\tau, \pi_0, x_3\right) + o_p(1)$ and since $R_{1n}\left(\gamma_\tau, \pi_0, x_3\right)$ is convex in γ_τ, by CL and QAL the minimizer $\widehat{\gamma}_\tau$ of $R_n\left(\gamma_\tau, \widehat{\pi}, x_3\right)$ is

$$\begin{aligned}\widehat{\gamma}_\tau = &-\left(f_{X_3}\left(x_3\right)\Sigma\left(x_3\right)\right)^{-1}\frac{1}{(nh)^{1/2}}\sum_{i=1}^n \frac{\delta_i}{\pi_{0i}}W_i\rho_\tau'\left(\varepsilon_i^*\right)K_h\left(X_{3i} - x_3\right) + \\ &O_p\left(\left(\frac{\log n}{nh}\right)^{1/2} + h^2\right) + o_p(1),\end{aligned} \tag{10.5}$$

which corresponds to the Bahadur expansion for the local linear estimator of $\theta_0(x_3)$ that is uniform in $x_3 \in \mathcal{X}_3$. Note that

$$E\left[\frac{\delta}{\pi_0}W\rho_\tau'(\varepsilon)K_h(X_3-x_3)\right]=0,$$

$$Var\left[\frac{\delta}{\pi_0}W\rho_\tau'(\varepsilon)K_h(X_3-x_3)\right]=f_{X_3}(x_3)E\left\{\frac{\tau(1-\tau)}{\pi_0}\begin{bmatrix} v_0X_1^{\otimes 2} & v_0X_1X_2^T & 0 \\ v_0X_2X_1^T & v_0X_2^{\otimes 2} & 0 \\ 0 & 0 & v_2X_2^{\otimes 2}\end{bmatrix}|X_3=x_3\right\}, \\ +o(1),$$

and that by iterated expectations, a Taylor expansion and the fact that $\varsigma_\tau(x_3)=X_2^T\theta_{0\tau}''(x_3)(X_3-x_3)^2/2+o_p(h^2)$

$$E\left[\frac{\delta}{\pi_0}W\left(\rho_\tau'(\varepsilon^*)-\rho_\tau'(\varepsilon)\right)K_h(X_3-x_3)\right]=EE\left[W\left(F_{\varepsilon|X}(\varsigma_\tau(x_3))-F_{\varepsilon|X}(0)\right)K_h(X_3-x_3)|X_3\right] \\ =-\frac{h^2}{2}f_{X_3}(x_3)E\left\{f_{\varepsilon|X}(0|X)\begin{bmatrix}X_1X_2^T\kappa_2 \\ X_2^{\otimes 2}\kappa_2 \\ O_{pp}\end{bmatrix}|X_3=x_3\right\}\theta_{0\tau}''(x_3)+o(1).$$

Furthermore, it can be showed that

$$Var\left[\frac{\delta}{\pi_0}W\rho_\tau'(\varepsilon^*)K_h(X_3-x_3)-\frac{\delta}{\pi_0}W\rho_\tau'(\varepsilon)K_h(X_3-x_3)\right]=O\left(h^2\right),$$

hence the conclusion follows by CLT and CMT. ■

Proof of Theorem 2. By (10.1) it follows that

$$\left\|\sum_{i=1}^n\frac{\delta_i}{\widehat{\pi}_i}\left[\rho_\tau\left(Y_i-X_{1i}^T\widehat{\beta}-X_{2i}^T(a_\tau-b_\tau(X_{3i}-x_3))\right)- \right.\right. \\ \left.\left.\rho_\tau\left(Y_i-X_{1i}^T\beta_0-X_{2i}^T(a_\tau-b_\tau(X_{3i}-x_3))\right)\right]K_h(X_{3i}-x_3)\right\|=O_p\left(h^{1/2}\right)=o_p(1),$$

hence using the same arguments as those used in the proof of Theorem 1, it is possible to show that, for $W_{2i}=\left[X_{2i}^T,X_{2i}^T(X_{3i}-x_3)/h\right]^T$ and $\gamma_{2\tau}=(nh)^{1/2}\left[(a_\tau-\theta_{0\tau}(x_3))^T,h\left(b_\tau-\theta_{0\tau}'(x_3)\right)^T\right]^T$,

$$R_n(\gamma_{2\tau},\widehat{\pi},x_3)=\sum_{i=1}^n\frac{\delta_i}{\widehat{\pi}_i}\left[\rho_\tau\left(\varepsilon_i^*-\frac{W_{2i}^T\gamma_{2\tau}}{(nh)^{1/2}}\right)-\rho_\tau(\varepsilon_i^*)\right]$$

can be approximated uniformly in $x_3\in\mathcal{X}_3$ as

$$R_n(\gamma_{2\tau},\pi_0,x_2)=\frac{\gamma_{2\tau}^T}{(nh)^{1/2}}\sum_{i=1}^n\frac{\delta_i}{\pi_{0i}}W_{2i}\rho_\tau'(\varepsilon_i^*)K_h(X_{3i}-x_3)+\frac{1}{2}f_{X_3}(x_3)\gamma_{2\tau}^Tdiag(1,\kappa_2)\otimes\Sigma(x_3)\gamma_{2\tau}+ \\ o_p(1),$$

and the conclusion follows as in the proof of Theorem 1. ■

Proof of Theorem 3. Let

$$\begin{aligned}\widehat{\varepsilon}_i^* &= Y_i - X_{1i}^T\beta_{0\tau} - X_{2i}^T\widehat{\theta}_\tau\left(X_{3i}\right),\\ \gamma_{\beta_\tau} &= n^{1/2}\left(\beta_\tau - \beta_{0\tau}\right),\end{aligned}$$

and let

$$R_n\left(\gamma_{\beta\tau},\widehat{\pi}_i\right) = \sum_{i=1}^n \frac{\delta_i}{\widehat{\pi}_i}\left[\rho_\tau\left(\widehat{\varepsilon}_i^* - \frac{X_{1i}^T\gamma_{\beta_\tau}}{n^{1/2}}\right) - \rho_\tau\left(\widehat{\varepsilon}_i^*\right)\right],$$

denote the normalized objective function $Q_n\left(\beta_\tau,\widehat{\theta}_\tau,\widehat{\pi}_i\right)$. Similar to the proof of Theorem 1

$$\begin{aligned}R_n\left(\gamma_{\beta_\tau},\widehat{\pi}_i\right) = \sum_{i=1}^n \frac{\delta_i}{\pi_{0i}}&\left[\rho_\tau\left(\varepsilon_i - X_{2i}^T\left(\widehat{\theta}_\tau\left(X_{3i}\right)-\theta_{\tau 0}\left(X_{3i}\right)\right) - \frac{X_{1i}^T\gamma_{\beta_\tau}}{n^{1/2}}\right) - \right.\\ &\left.\rho_\tau\left(\varepsilon_i - X_{2i}^T\left(\widehat{\theta}_\tau\left(X_{3i}\right)-\theta_{\tau 0}\left(X_{3i}\right)\right)\right)\right] -\\ &\sum_{i=1}^n \frac{\delta_i\left(\widehat{\pi}_i-\pi_{0i}\right)}{\widehat{\pi}_i\pi_{0i}}\left[\rho_\tau\left(\varepsilon_i - X_{2i}^T\left(\widehat{\theta}_\tau\left(X_{3i}\right)-\theta_{\tau 0}\left(X_{3i}\right)\right) - \frac{X_{1i}^T\gamma_{\beta_\tau}}{n^{1/2}}\right) - \right.\\ &\left.\rho_\tau\left(\varepsilon_i - X_{2i}^T\left(\widehat{\theta}_\tau\left(X_{3i}\right)-\theta_{\tau 0}\left(X_{3i}\right)\right)\right)\right]\\ &:= R_{1n}\left(\gamma_{\beta_\tau},\pi_{0i},\widehat{\theta}_\tau\right) + R_{2n}\left(\gamma_{\beta_\tau},\widehat{\pi}_i,\widehat{\theta}_\tau\right).\end{aligned}$$

Again by (10.1)

$$R_{1n}\left(\gamma_{\beta_\tau},\pi_{0i},\widehat{\theta}_\tau\right) = \frac{\gamma_{\beta_\tau}^T}{n^{1/2}}\sum_{i=1}^n \frac{\delta_i}{\pi_{0i}}X_{1i}\rho_\tau'\left(\varepsilon_i\right) + S_{1n}\left(\gamma_{\beta_\tau},\pi_{0i},\widehat{\theta}_\tau\right), \tag{10.6}$$

where

$$S_{1n}\left(\gamma_{\beta_\tau},\pi_{0i},\widehat{\theta}_\tau\right) = \sum_{i=1}^n \int_{X_{2i}^T\left(\widehat{\theta}_\tau(X_{3i})-\theta_{\tau 0}(X_{3i})\right)}^{X_{2i}^T\left(\widehat{\theta}_\tau(X_{3i})-\theta_{\tau 0}(X_{3i})\right)+\frac{X_{1i}^T\gamma_{\beta_\tau}}{n^{1/2}}} \frac{\delta_i}{\pi_{0i}}\left(I\left(\varepsilon_i\le t\right)-I\left(\varepsilon_i\le 0\right)\right)dt.$$

Similarly to (10.3), we can show that

$$E\left[S_{1n}\left(\gamma_{\beta_\tau},\pi_{0i},\widehat{\theta}_\tau\right)\right] = \frac{1}{2}\frac{\gamma_{\beta_\tau}^T}{n}\sum_{i=1}^n f_{\varepsilon_i|X_i}(0)\,X_{1i}^{\otimes 2}\gamma_{\beta_\tau} - \frac{\gamma_{\beta_\tau}^T}{n}\sum_{i=1}^n f_{\varepsilon_i|X_i}(0)\,X_{1i}X_{2i}^T\left(\widehat{\theta}_\tau\left(X_{3i}\right)-\theta_{\tau 0}\left(X_{3i}\right)\right) + o(1), \tag{10.7}$$

so that (10.6) can be written as

$$\begin{aligned}R_{1n}\left(\gamma_{\beta_\tau},\pi_{0i},\widehat{\theta}_\tau\right) = \frac{\gamma_{\beta_\tau}^T}{n^{1/2}}\sum_{i=1}^n \frac{\delta_i}{\pi_{0i}}X_{1i}\rho_\tau'\left(\varepsilon_i\right) + \frac{1}{2}\frac{\gamma_{\beta_\tau}^T}{n}\sum_{i=1}^n f_{\varepsilon_i|X_i}(0)\,X_{1i}^{\otimes 2}\gamma_{\beta_\tau}^T -\\ \frac{\gamma_{\beta\tau}^T}{n}\sum_{i=1}^n f_{\varepsilon_i|X_i}(0)\,X_{2i}^T\left(\widehat{\theta}_\tau\left(X_{3i}\right)-\theta_{\tau 0}\left(X_{3i}\right)\right) + Q_{1n}\left(\gamma_{\beta_\tau},\pi_{0i},\widehat{\theta}_\tau\right) + o_p(1),\end{aligned}$$

where

$$\left|Q_{1n}\left(\gamma_{\beta_\tau},\pi_{0i},\widehat{\theta}_\tau\right)\right| = \left|S_{1n}\left(\gamma_{\beta_\tau},\pi_{0i},\widehat{\theta}_\tau\right) - E\left[S_{1n}\left(\gamma_{\beta_\tau},\pi_{0i},\widehat{\theta}_\tau\right)\right]\right| = o_p(1),$$

since

$$E\left[Q_{1n}\left(\gamma_{\beta_\tau},\pi_{0i},\widehat{\theta}_\tau\right)^2\right]\leq nES_{1i}\left(\gamma_{\beta_\tau},\pi_{0i},\widehat{\theta}_\tau\right)^2= \tag{10.8}$$

$$nE\left[\int_{X_{2i}^T\left(\widehat{\theta}_\tau(X_{3i})-\theta_{\tau 0}(X_{3i})\right)}^{X_{2i}^T\left(\widehat{\theta}_\tau(X_{3i})-\theta_{\tau 0}(X_{3i})\right)+\frac{X_{1i}^T\gamma_{\beta_\tau}}{n^{1/2}}}\Pr\left(0\leq|\varepsilon_i|\leq\max\left(|t|,|u|\right)|X\right)dtdu\right]$$

$$\leq nE\left[\Pr\left(0\leq|\varepsilon_i|\leq\|X_{2i}\|\left\|\widehat{\theta}_\tau\left(X_{3i}\right)-\theta_{\tau 0}\left(X_{3i}\right)\right\|+\left|\frac{X_{1i}^T\gamma_{\beta_\tau}}{n^{1/2}}\right||X_i\right)\left|\frac{\gamma_{\beta_\tau}^TX_{1i}^{\otimes 2}\gamma_{\beta_\tau}}{n}\right|\right]$$

$$=o\left(1\right)$$

as both $\left|X_{1i}^T\gamma_{\beta_\tau}/n^{1/2}\right|$ and $\left\|\widehat{\theta}_\tau\left(X_{3i}\right)-\theta_{\tau 0}\left(X_{3i}\right)\right\|$ are $o_p\left(1\right)$. Let $S=[O_{pk},I_p,O_{pp}]$; then by (10.5) we have

$$R_{1n}\left(\gamma_{\beta_\tau},\pi_{0i},\widehat{\theta}_\tau\right)=\frac{\gamma_{\beta_\tau}^T}{n^{1/2}}\sum_{i=1}^n\frac{\delta_i}{\pi_{0i}}X_{1i}\rho_\tau'\left(\varepsilon_i\right)+\frac{1}{2}\frac{\gamma_{\beta_\tau}^T}{n}\sum_{i=1}^nf_{\varepsilon_i|X_i}\left(0\right)X_{1i}^{\otimes 2}\gamma_{\beta_\tau}-$$
$$\frac{\gamma_{\beta_\tau}^T}{n^{3/2}}\sum_{i=1}^n\sum_{j=1}^nf_{\varepsilon_i|X_i}\left(0\right)X_{1i}X_{2i}^TS\left(f_{X_3}\left(X_{3i}\right)\Sigma\left(X_{3i}\right)\right)^{-1}\times$$
$$\frac{\delta_j}{\pi_{0j}}\left[X_{1j}^T,X_{2j}^T,0_p^T\right]^T\rho_\tau'\left(\varepsilon_j\right)K_h\left(X_{3j}-X_{3i}\right)+O_p\left(n^{1/2}h^{5/2}+\left(\frac{\log n}{nh^2}\right)^2\right),$$

which by LLN and a standard U-statistic projection argument simplifies to

$$R_{1n}\left(\gamma_{\beta_\tau},\pi_{0i},\widehat{\theta}_\tau\right)=\frac{\gamma_{\beta_\tau}^T}{n^{1/2}}\sum_{i=1}^n\frac{\delta_i}{\pi_{0i}}\left(X_{1i}-\varphi\left(X_i\right)\right)\rho_\tau'\left(\varepsilon_i\right)+\gamma_{\beta_\tau}^T\Sigma_2\gamma_{\beta_\tau}+o_p\left(1\right), \tag{10.9}$$

where

$$\varphi\left(X_i\right)=E\left[f_{\varepsilon|X}\left(0\right)X_1X_2^T|X_3=X_{3i}\right]S\Sigma\left(X_{3i}\right)^{-1}\left[X_{1i}^T,X_{2i}^T,0_p^T\right]^T.$$

For $R_{2n}\left(\gamma_{\beta_\tau},\widehat{\pi}_i,\widehat{\theta}_\tau\right)$ note that

$$R_{2n}\left(\gamma_{\beta_\tau},\widehat{\pi}_i,\widehat{\theta}_\tau\right)=\sum_{i=1}^n\frac{\delta_i\left(\widehat{\pi}_i-\pi_{0i}\right)}{\pi_{0i}^2}\left[\frac{\gamma_{\beta_\tau}^T}{n^{1/2}}X_{1i}\rho_\tau'\left(\varepsilon_i\right)+S_{3n}\left(\gamma_{\beta_\tau},\widehat{\theta}_\tau\right)\right]+o_p\left(1\right)$$
$$=\sum_{i=1}^n\frac{\delta_i\left(\widehat{\pi}_i-\pi_{0i}\right)}{\pi_{0i}^2}\left\{\frac{\gamma_{\beta_\tau}^T}{n^{1/2}}X_{1i}\rho_\tau'\left(\varepsilon_i\right)+E\left[S_{3n}\left(\gamma_{\beta_\tau},\widehat{\theta}_\tau\right)\right]\right\}+Q_{2n}\left(\gamma_{\beta_\tau},\widehat{\pi}_i,\widehat{\theta}_\tau\right)+o_p\left(1\right),$$

where

$$S_{3n}\left(\gamma_{\beta_\tau},\widehat{\theta}_\tau\right)=\sum_{i=1}^n\int_{X_{2i}^T\left(\widehat{\theta}_\tau(X_{3i})-\theta_{\tau 0}(X_{3i})\right)}^{X_{2i}^T\left(\widehat{\theta}_\tau(X_{3i})-\theta_{\tau 0}(X_{3i})\right)+\frac{X_{1i}^T\gamma_{\beta_\tau}}{n^{1/2}}}\left(I\left(\varepsilon_i\leq t\right)-I\left(\varepsilon_i\leq 0\right)\right)dt$$

and

$$Q_{2n}\left(\gamma_{\beta_\tau},\widehat{\pi}_i,\widehat{\theta}_\tau\right)=\sum_{i=1}^n\frac{\delta_i\left(\widehat{\pi}_i-\pi_{0i}\right)}{\pi_{0i}^2}\left\{\int_{X_{2i}^T\left(\widehat{\theta}_\tau(X_{3i})-\theta_{\tau 0}(X_{3i})\right)}^{X_{2i}^T\left(\widehat{\theta}_\tau(X_{3i})-\theta_{\tau 0}(X_{3i})\right)+\frac{X_{1i}^T\gamma_{\beta_\tau}}{n^{1/2}}}\left(I\left(\varepsilon_i\leq t\right)-I\left(\varepsilon_i\leq 0\right)\right)dt-\right. \tag{10.10}$$
$$\left.E\left[S_{3n}\left(\gamma_{\beta_\tau},\widehat{\theta}_\tau\right)\right]\right\}.$$

For the case where $\widehat{\pi}_i = \widehat{\pi}_i(\widehat{\alpha})$, the Cauchy-Schwarz inequality, LLN and a similar argument to (10.8) imply that

$$\left|Q_{2n}\left(\gamma_{\beta_\tau}, \widehat{\pi}_i, \widehat{\theta}_\tau\right)\right| \leq 2\left(\|\widehat{\alpha}-\alpha_0\|^2 \sum_{i=1}^{n} \sup_{\alpha\in A}\left\|\frac{1}{\pi_i(\alpha)^2}\frac{\partial \pi_i(\alpha)}{\partial \alpha}\right\|^2\right)^{1/2} \times \tag{10.11}$$

$$\left(\sum_{i=1}^{n}\left|\frac{\delta_i}{\pi_{0i}}\int_{X_{2i}^T\left(\widehat{\theta}_\tau(X_{3i})-\theta_{\tau 0}(X_{3i})\right)}^{X_{2i}^T\left(\widehat{\theta}_\tau(X_{3i})-\theta_{\tau 0}(X_{3i})\right)+\frac{X_{1i}^T\gamma_{\beta_\tau}}{n^{1/2}}}\left(I(\varepsilon_i \leq t) - I(\varepsilon_i \leq 0)\right)dt - E\left[\frac{\delta_i}{\pi_{0i}}S_{3n}\left(\gamma_{\beta_\tau}, \widehat{\theta}_\tau\right)\right]\right|^2\right)^{1/2}$$

$$= O_p(1)\, o_p(1).$$

For the case where $\widehat{\pi}_i$ is estimated nonparametrically, a standard kernel calculation, (10.8) and the Cauchy-Schwarz inequality imply that

$$\left|Q_{2n}\left(\gamma_{\beta_\tau}, \widehat{\pi}_i, \widehat{\theta}_\tau\right)\right| \leq 2\left(\sum_{i=1}^{n}\frac{|\widehat{\pi}_i-\pi_{0i}|^2}{\pi_{0i}^2}\right)^{1/2} \times \tag{10.12}$$

$$\left(\sum_{i=1}^{n}\left|\frac{\delta_i}{\pi_{0i}}\int_{X_{2i}^T\left(\widehat{\theta}_\tau(X_{3i})-\theta_{\tau 0}(X_{3i})\right)}^{X_{2i}^T\left(\widehat{\theta}_\tau(X_{3i})-\theta_{\tau 0}(X_{3i})\right)+\frac{X_{1i}^T\gamma_{\beta_\tau}}{n^{1/2}}}\left(I(\varepsilon_i \leq t) - I(\varepsilon_i \leq 0)\right)dt - E\left[\frac{\delta_i}{\pi_{0i}}S_{3n}\left(\gamma_{\beta_\tau}, \widehat{\theta}_\tau\right)\right]\right|^2\right)^{1/2}$$

$$= O_p\left(n^{1/2}b^2\right) o_p(1) = o_p(1).$$

Combining (10.7), (10.11) and (10.12) yields

$$R_{2n}\left(\gamma_{\beta_\tau}, \widehat{\pi}_i, \widehat{\theta}_\tau\right) = \sum_{i=1}^{n}\frac{\delta_i(\widehat{\pi}_i-\pi_{0i})}{\pi_{0i}^2}\left[\frac{\gamma_{\beta_\tau}^T}{n^{1/2}}\left(X_{1i}\rho_\tau'(\varepsilon_i) - \varphi(X_i)\rho_\tau'(\varepsilon_i)\right) + \gamma_{\beta_\tau}^T\Sigma_2\gamma_{\beta_\tau}\right] + o_p(1).$$

Thus, by CL and QAL, we have that $\widehat{\gamma}_{\beta_\tau} = \Sigma_2^{-1}\zeta + o_p(1)$, where

$$\zeta = \frac{1}{n^{1/2}}\sum_{i=1}^{n}\left(\frac{\delta_i}{\pi_{0i}} - \frac{\delta_i(\widehat{\pi}_i-\pi_{0i})}{\pi_{0i}^2}\right)\left(X_{1i}-\varphi(X_i)\right)\rho_\tau'(\varepsilon_i). \tag{10.13}$$

For $\widehat{\pi}_i = \widehat{\pi}_i(\widehat{\alpha})$, a mean value expansion, A4 and LLN imply that

$$\frac{1}{n^{1/2}}\sum_{i=1}^{n}\frac{\delta_i(\widehat{\pi}_i-\pi_{0i})}{\pi_{0i}^2}\left(X_{1i}-\varphi(X_i)\right)\rho_\tau'(\varepsilon_i) = \tag{10.14}$$

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\pi_{0i}^2}\left(X_{1i}-\varphi(X_i)\right)\rho_\tau'(\varepsilon_i)\left(\frac{\partial \pi_i(\overline{\alpha})}{\partial \alpha^T}\right)I(\alpha_0)^{-1}n^{1/2}(\widehat{\alpha}-\alpha_0) =$$

$$E\left[\frac{\left(X_1-\varphi(X)\right)\rho_\tau'(\varepsilon_i)}{\pi_0}\frac{\partial \pi_0}{\partial \alpha^T}\right]I(\alpha_0)^{-1}\frac{1}{n^{1/2}}\sum_{i=1}^{n}s(Z_{oi},\alpha_0) + o_p(1)$$

and the conclusion follows by CLT and CMT noting that

$$
\begin{aligned}
&Cov\left(\frac{\delta_i}{\pi_{0i}}\left(X_{1i}-\varphi\left(X_i\right)\right)\rho_\tau'\left(\varepsilon_i\right)-\right. \\
&\left.E\left[\frac{\left(X_1-\varphi\left(X\right)\right)\rho_\tau'\left(\varepsilon_i\right)\left(\partial\pi_0/\partial\alpha^T\right)}{\pi_0}\right]I\left(\alpha_0\right)^{-1}\frac{1}{n^{1/2}}\sum_{i=1}^n s\left(Z_{oi},\alpha_0\right)\right)= \\
&E\left[\frac{\left(\left(X_1-\varphi\left(X\right)\right)\rho_\tau'\left(\varepsilon\right)\right)^{\otimes 2}}{\pi_0}\right]-E\left[\frac{\left(X_1-\varphi\left(X\right)\right)\rho_\tau'\left(\varepsilon\right)}{\pi_0}\frac{\partial\pi_0}{\partial\alpha^T}\right]\times \\
&I\left(\alpha_0\right)^{-1}E\left[\frac{\left(X_1-\varphi\left(X\right)\right)\rho_\tau'\left(\varepsilon\right)}{\pi_0}\frac{\partial\pi_0}{\partial\alpha^T}\right]^T,
\end{aligned}
\tag{10.15}
$$

since by iterated expectations

$$
\begin{aligned}
&E\left\{\frac{\delta_i}{\pi_{0i}}\left(X_{1i}-\varphi\left(X_i\right)\right)\rho_\tau'\left(\varepsilon_i\right)\left[E\left[\frac{\left(X_1-\varphi\left(X\right)\right)\rho_\tau'\left(\varepsilon_i\right)}{\pi_0}\frac{\partial\pi_0}{\partial\alpha^T}\right]\times\right.\right. \\
&\left.\left.I\left(\alpha_0\right)^{-1}\frac{1}{\pi_{0i}}\left(\frac{\partial\pi_{0i}}{\partial\alpha^T}\right)\right]^T\right\}= \\
&E\left[\frac{\left(X_1-\varphi\left(X\right)\right)\rho_\tau'\left(\varepsilon_i\right)}{\pi_0}\frac{\partial\pi_0}{\partial\alpha^T}\right]I\left(\alpha_0\right)^{-1}E\left[\frac{\left(X_1-\varphi\left(X\right)\right)\rho_\tau'\left(\varepsilon_i\right)}{\pi_0}\frac{\partial\pi_0}{\partial\alpha^T}\right]^T.
\end{aligned}
$$

■

Proof of Theorem 4. Given (10.13), for $\widehat{\pi}_i$ estimated nonparametrically, note that

$$
\begin{aligned}
&\frac{1}{n^{1/2}}\sum_{i=1}^n\frac{\delta_i\left(\widehat{\pi}_i-\pi_{0i}\right)}{\pi_{0i}^2}\left(X_{1i}-\varphi\left(X_i\right)\right)\rho_\tau'\left(\varepsilon_i\right)= \\
&\frac{1}{n^{3/2}}\sum_{i=1}^n\left(\delta_i-\pi_{0i}\right)\frac{\sum_{j=1}^n\left(\delta_j-\pi_{0i}\right)L_b\left(Z_{oj}-Z_{oi}\right)}{\pi_{0i}^2 b^{\dim(Z_o)}f\left(Z_{oi}\right)}\left(X_{1i}-\varphi\left(X_i\right)\right)\rho_\tau'\left(\varepsilon_i\right)+ \\
&\frac{1}{n^{3/2}}\sum_{i=1}^n\frac{\left(\delta_j-\pi_{0i}\right)L_b\left(Z_{oj}-Z_{oi}\right)}{\pi_{0i}^2 b^{\dim(Z_o)}f\left(Z_{oi}\right)}\left(X_{1i}-\varphi\left(X_i\right)\right)\rho_\tau'\left(\varepsilon_i\right)= \\
&\frac{1}{n^{1/2}}\sum_{i=1}^n\frac{\left(\delta_i-\pi_{0i}\right)}{\pi_{0i}}E\left[\left(X_{1i}-\varphi\left(X_i\right)\right)\rho_\tau'\left(\varepsilon_i\right)|Z_{oi}\right]+o_p\left(1\right).
\end{aligned}
$$

The conclusion follows by CLT and CMT since by iterated expectations

$$
\begin{aligned}
&E\left[\frac{\delta_i}{\pi_{0i}}\left(X_{1i}-\varphi\left(X_i\right)\right)\rho_\tau'\left(\varepsilon_i\right)\frac{\left(\delta_i-\pi_{0i}\right)}{\pi_{0i}}E\left[\left(X_{1i}-\varphi\left(X_i\right)\right)\rho_\tau'\left(\varepsilon_i\right)|Z_{oi}\right]^T\right]= \\
&E\left[\frac{1-\pi_0}{\pi_0}E\left[\left(X_1-\varphi\left(X\right)\right)\rho_\tau'\left(\varepsilon\right)|Z_o\right]^{\otimes 2}\right].
\end{aligned}
$$

■

Proof of Theorem 5. First note that

$$
\sup_{x_3\in\mathcal{X}_3}||h^k(\frac{\partial^k\widehat{\theta}(x_3)}{\partial\beta_\tau^k}-\frac{\partial^k\theta_0(x_3)}{\partial\beta_\tau^k})||=O_p(h^2+\frac{1}{(nh)^{1/2}})
\tag{10.16}
$$

for $k = 0, 1$, by the same arguments as those used in the proof of Theorem 1, and this rate can be made uniform in $\beta \in B$ by A6'(ii) and of order $o_p(n^{1/4})$ by choosing a suitable b. Next, by the uniform consistency of $\widehat{\pi}_i$ and the boundedness of π_i

$$M_n(\beta_\tau, \widehat{\theta}_{\beta_\tau}, \partial\widehat{\theta}_{\beta_\tau}/\partial\beta_\tau^T, \widehat{\pi}) = M_n(\beta_\tau, \widehat{\theta}_{\beta_\tau}, \partial\widehat{\theta}_{\beta_\tau}/\partial\beta_\tau^T, \pi) + o_p(1) \leq \frac{1}{n}\sum_{i=1}^n M_i(\beta_\tau, \widehat{\theta}_{\beta_\tau}, \partial\widehat{\theta}_{\beta_\tau}/\partial\beta_\tau^T),$$

and that for all $\beta_\tau^\dagger \in B$, $\theta_{\beta\tau}^\dagger \in \Theta_B$

$$\begin{aligned} &||M_i(\beta_\tau^\dagger, \theta^\dagger_{\beta_\tau^\dagger}, \partial\theta^\dagger_{\beta_\tau^\dagger}/\partial\beta_\tau^T) - M_i(\beta_\tau, \theta_{\beta_\tau}, \partial\theta_{\beta_\tau}/\partial\beta_\tau^T)|| \leq \\ &(||X_{1i}||^2 + ||X_{2i}||^2 \left(||\frac{\partial\theta^\dagger_{\beta_\tau^\dagger}(X_{3i})}{\partial\beta_\tau^{\dagger T}} - \frac{\partial\theta^\dagger_{\beta_\tau^\dagger}(X_{3i})}{\partial\beta_\tau^T}||^2\right))|\rho'_\tau(Y_i - X_{1i}^T\beta_\tau^\dagger - X_{2i}^T\theta^\dagger_{\beta_\tau^\dagger}(X_{3i})) - \\ &\rho'_\tau(Y_i - X_{1i}^T\beta_\tau - X_{2i}^T\theta^\dagger_{\beta_\tau^\dagger}(X_{3i}))| + (||X_{1i}||^2 + ||X_{2i}||^2 \left(||\frac{\partial\theta^\dagger_{\beta_\tau}(X_{3i})}{\partial\beta_\tau^{\dagger T}} - \frac{\partial\theta^\dagger_{\beta_\tau}(X_{3i})}{\partial\beta_\tau^T}||^2\right))\times \\ &|\rho'_\tau(Y_i - X_{1i}^T\beta_\tau - X_{2i}^T\theta^\dagger_{\beta_\tau}(X_{3i})) - \rho'_\tau(Y_i - X_{1i}^T\beta_\tau - X_{2i}^T\theta^\dagger_{\beta_\tau}(X_{3i})| + \\ &(||X_{1i}||^2 + ||X_{2i}||^2 \left(||\frac{\partial\theta^\dagger_{\beta_\tau}(X_{3i})}{\partial\beta_\tau^T} - \frac{\partial\theta_{\beta_\tau}(X_{3i})}{\partial\beta_\tau^T}||^2\right))\times \\ &|\rho'_\tau(Y_i - X_{1i}^T\beta_\tau - X_{2i}^T\theta^\dagger_{\beta_\tau^\dagger}(X_{3i}) - \rho'_\tau(Y_i - X_{1i}^T\beta_\tau - X_{2i}^T\theta_{\beta_\tau}(X_{3i}))| := \sum_{j=1}^3 P_j, \end{aligned} \tag{10.17}$$

We only consider P_3 as the two other terms can be dealt in the same way. For all $\epsilon \in (0, 1]$ by iterated expectations and the differentiability of $F_{Y|X}(\cdot)$

$$E \sup_{||\theta^\dagger_{\beta_\tau^\dagger} - \theta_{\beta_\tau}|| \leq \epsilon} \sup_{||\partial\theta^\dagger_{\beta_\tau^\dagger}/\partial\beta_\tau^{\dagger T} - \partial\theta_{\beta_\tau}/\partial\beta_\tau^T|| \leq \epsilon} P_3 \leq E(||X_{1i}||^2 + ||X_{2i}||^2\epsilon^2)(F_{Y|X}(X_{1i}^T\beta_\tau^\dagger + X_{2i}^T\theta^\dagger_{\beta_\tau^\dagger}(X_{3i}) + \epsilon) - F_{Y|X}(X_{1i}^T\beta_\tau + X_{2i}^T\theta_{\beta_\tau}(X_{3i}) - \epsilon)) \leq C\epsilon$$

Notice that by (2.7)

$$\frac{\partial E(M_i(\beta_\tau, \theta_{\beta_\tau}, \partial\theta_{\beta_\tau}/\partial\beta_\tau^T)}{\partial\beta_\tau^T}|_{\beta_\tau = \beta_{0\tau}} = -E(f_{\varepsilon|X}(0)\left(X_1 + \frac{\partial\theta_{0\tau}(X_3)}{\partial\beta_\tau^T}X_2\right)^{\otimes 2}),$$

hence β_τ is uniquely identified. Let $\{\beta_{\tau k} : k = 1, ..., K_1\}$ be an ϵ-cover for $(B, ||\cdot||)$ and $\{\theta_{\tau l}, \partial\theta_{\tau l}/\partial\beta_{\tau k} : k, l = 1, ..., K_2\}$ denote an ϵ-cover for $(\Theta_B, ||\cdot||_{\Theta_B})$. Then by (10.17) for any

$$M_{ij}(\beta_\tau, \theta_{\beta_\tau}, \partial\theta_{\beta_\tau}/\partial\beta^T) = (X_{1ij} + \left(\frac{\partial\theta_{\beta_\tau}(X_{3i})}{\partial\beta_{\tau j}}\right)^T X_{2i})\rho'_\tau(Y_i - X_{1i}^T\beta_\tau - X_{2i}^T\theta_{\beta_\tau}(X_{3i}))$$

$j = 1, ..., k$ there exists $k_1 \in \{1, ..., K_1\}$ and $l_1 \in \{1, ..., K_2\}$ such that

$$|M_{ij}(\beta_\tau, \theta_{\beta_\tau}, \partial\theta_{\beta_\tau}/\partial\beta_\tau^T) - M_{ij}(\beta_{\tau k_1}, \theta_{\beta_\tau l_1}, \partial\theta_{\beta_\tau l_1}/\partial\beta_{\tau k_1})|$$

is bounded by

$$\sup_{\substack{\beta^\dagger_\tau,\theta^\dagger_{\beta^\dagger_\tau},\partial\theta_{\beta^\dagger_\tau}/\partial\beta^{\dagger T}_\tau:||\beta^\dagger_\tau-\beta_{\tau k_1}||<\epsilon,\\ ||\theta^\dagger_{\beta_\tau}-\theta_{\beta_\tau l_1}||_{\Theta_B}<\epsilon, ||\partial\theta^\dagger_{\beta^\dagger_\tau}/\partial\beta^{\dagger T}_\tau-\partial\theta_{\tau l_1}/\partial\beta^T_{\tau k_1}||<\epsilon}} |M_{ij}(\beta^\dagger_\tau,\theta^\dagger_{\beta_\tau},\partial\theta^\dagger_{\beta^\dagger_\tau}/\partial\beta^{\dagger T}_\tau)-M_{ij}(\beta_{\tau k_1},\theta_{\beta_\tau l_1},\partial\theta_{\beta_\tau l_1}/\partial\beta^T_{\tau k_1})| := b_j(\beta_{\tau k_1},\theta_{\beta_\tau l_1},\partial\theta_{\beta_\tau l_1}/\partial\beta_{\tau k_1},\epsilon)$$

hence

$$M_{ij}(\beta_{\tau k_1},\theta_{\beta_\tau l_1},\partial\theta_{\beta_\tau l_1}/\partial\beta_{\tau k_1})-b_j(\beta_{\tau k_1},\theta_{\beta_\tau l_1},\partial\theta_{\beta_\tau l_1}/\partial\beta_{\tau k_1},\epsilon)\le M_{ij}(\beta_\tau,\theta_{\beta_\tau},\partial\theta_{\beta_\tau}/\partial\beta^T_\tau)\le M_{ij}(\beta_{\tau k_1},\theta_{\beta_\tau l_1},\partial\theta_{\beta_\tau l_1}/\partial\beta_{\tau k_1})+b_j(\beta_{\tau k_1},\theta_{\beta_\tau l_1},\partial\theta_{\beta_\tau l_1}/\partial\beta_{\tau k_1},\epsilon)$$

and $(E[b_j(\beta_{\tau k_1},\theta_{\beta_\tau l_1},\partial\theta_{\beta_\tau l_1}/\partial\beta_{\tau k_1},\epsilon)]^2)^{1/2}\le C_j\epsilon^{1/2}$ for all $\beta_{\tau k_1}$, $\theta_{\tau l_1}$, $\partial\theta_{\tau l_1}/\partial\beta_{\tau k_1}$ and all $\epsilon=o(1)$. Therefore

$$\{[M_{ij}(\beta_{\tau k_1},\theta_{\beta_\tau l_1},\partial\theta_{\beta_\tau l_1}/\partial\beta_{\tau k_1})-b_j(\beta_{\tau k_1},\theta_{\beta_\tau l_1},\partial\theta_{\beta_\tau l_1}/\partial\beta_{\beta_\tau k_1},\epsilon)\\ M_{ij}(\beta_{\tau k_1},\theta_{\beta_\tau l_1},\partial\theta_{\beta_\tau l_1}/\partial\beta_{k_1})+b_j(\beta_{\tau k_1},\theta_{\beta_\tau l_1},\partial\theta_{\beta_\tau l_1}/\partial\beta_{\tau k_1},\epsilon)]:k_1\in\{1,...,K_1\},l_1\in 1,...,K_2\}$$

forms a $\delta=2C_j\epsilon^{1/2}$ bracket for the function class $\{\mathcal{Q}_j=M_{ij}(\beta_\tau,\theta_{\beta_\tau},\partial\theta_{\beta_\tau}/\partial\beta^T_\tau):\beta_\tau\in B,\theta_{\beta_\tau},\partial\theta_{\beta_\tau}/\partial\beta^T_\tau\in\Theta_B\}$, hence

$$N_{[]}(\delta,\mathcal{Q}_j,||\cdot||_{L_2(P)})\le N\left([\frac{\delta}{2C_j}]^2,B,||\cdot||\right)N\left([\frac{\delta}{2C_j}]^2,\Theta_B,||\cdot||_{\Theta_B}\right),$$

where $N_{[]}(\cdot)$ and $N(\cdot)$ are, respectively, the bracketing and covering numbers (see Van der Vaart & Wellner (1996) for a definition). Since $N\left([\frac{\delta}{2C_j}]^2,B,||\cdot||\right)=O(\exp(C_{1j}\delta^{1/s_1})$ and $N\left([\frac{\delta}{2C_j}]^2,\Theta_B,||\cdot||_\infty\right)=O(\exp(C_{2j}\delta^{1/s_2}))$ for $\Theta_B=C^\alpha_M(\mathcal{X}_3)$, the bracketing integral $\int_0^\infty(\log(N_{[]}(\delta,\mathcal{Q}_j,||\cdot||_{L_2(P)}))^{1/2}d\delta$ is finite, hence by the L_2 boundedness of the brackets $b_j(\beta_{\tau k_1},\theta_{\beta_\tau l_1},\partial\theta_{\beta_\tau l_1}/\partial\beta_{\tau k_1},\epsilon)$ imply that for all $\epsilon_n=o(1)$

$$\sup_{\substack{||\beta_\tau-\beta_{0\tau}||\le\epsilon_n,||\theta_{\beta_\tau}-\theta_{0\tau}||\le\epsilon_n\\ ||\partial\theta_{\beta_\tau}/\partial\beta^T_\tau-\partial\theta_{0\tau}/\partial\beta^T_\tau||\le\epsilon_n}}||M_n(\beta_\tau,\theta_{\beta_\tau},\partial\theta_{\beta_\tau}/\partial\beta^T_\tau)-M(\beta_\tau,\theta_{\beta_\tau},\partial\theta_{\beta_\tau}/\partial\beta^T_\tau)- M_n(\beta_{0\tau},\theta_{0\tau},\partial\theta_{0\tau}/\partial\beta^T_\tau)||=o_p(n^{-1/2}) \tag{10.18}$$

We now establish the $n^{1/2}$ consistency of $\widehat{\beta}^p_\tau$. Let $\epsilon_n=o(1)$ such that $\Pr(||\widehat{\beta}^p_\tau-\beta_{0\tau}||\ge\epsilon_n,||\widehat{\theta}_{\beta_\tau}-\theta_{0\tau}||\ge\epsilon_n,||\partial\widehat{\theta}_{\beta_\tau}/\partial\beta^T_\tau-\partial\theta_{0\tau}/\partial\beta^T_\tau||\ge\epsilon_n)\to 1$, hence it is enough to consider the restricted parameter spaces $B_\epsilon,\Theta_{B_\epsilon}$. By A6'(iv) and the triangle inequality there exists a $C>0$ such that $||\widehat{\beta}^p_\tau-\beta_{0\tau}||$ is bounded by

$$||M(\widehat{\beta}^p_\tau,\theta_{0\tau},\partial\theta_{0\tau}/\partial\beta^T_\tau,\pi)-M(\widehat{\beta}^p_\tau,\widehat{\theta}_{\beta_\tau},\partial\widehat{\theta}_{\beta_\tau}/\partial\beta^T_\tau,\pi)||+\\ ||M(\widehat{\beta}^p_\tau,\widehat{\theta}_{\beta_\tau},\partial\widehat{\theta}_{\beta_\tau},/\partial\beta^T_\tau,\pi)-M_n(\widehat{\beta}^p_\tau,\widehat{\theta}_{\beta_\tau},\partial\widehat{\theta}_{\beta_\tau}/\partial\beta^T_\tau,\pi_i)+\\ M_n(\beta_{0\tau},\theta_{0\tau},\partial\theta_{0\tau}/\partial\beta^T_\tau,\pi_i)||+||M_n(\widehat{\beta}^p_\tau,\widehat{\theta}_{\beta_\tau},\partial\widehat{\theta}_{\beta_\tau}/\partial\beta^T_\tau,\pi_i)||+O_p(n^{-1/2}), \tag{10.19}$$

where the $O_p(n^{-1/2})$ term comes from the asymptotic normality of $M_n(\beta_{0\tau}, \theta_{0\tau}, \partial\theta_{0\tau}/\partial\beta_\tau^T, \pi_i)$ shown at (10.20) below. Since $\Pr(\widehat{\beta}_\tau^p, \widehat{\theta}_{\beta_\tau}, \partial\widehat{\theta}_{\beta_\tau}/\partial\beta_\tau^T \in B_\epsilon \times \Theta_{B_\epsilon}) \to 1$, the smoothness of $M(\cdot)$ in β_τ, θ_{β_τ} and $\partial\theta_{\beta_\tau}/\partial\beta_\tau^T$ implies that the first term in (10.19) is bounded by

$$C(||\widehat{\theta}_{\beta_\tau} - \theta_{0\tau}||^2_{\Theta_B} + ||\frac{\partial\widehat{\theta}_{\beta_\tau}}{\partial\beta_\tau^T} - \frac{\partial\theta_{0\tau}}{\partial\beta_\tau^T}||^2_{\Theta_B})+$$
$$||E(f_{Y|X}(X_{1i}^T\widehat{\beta}_\tau + X_{2i}^T\theta_{0\tau}(X_{3i}) - f_{\varepsilon|X}(0))(X_{1i} + \frac{\partial\theta_{0\tau}(X_{3i})}{\partial\beta_\tau^T})^{\otimes 2}||||\widehat{\beta}_\tau^p - \beta_{0\tau}|| =$$
$$o_p(n^{-1/2}) + o_p(1)||\widehat{\beta}_\tau^p - \beta_{0\tau}||,$$

By (10.18) the second term in (10.19) is bounded by

$$o_p(1)(\frac{1}{n^{1/2}} + ||M_n(\widehat{\beta}_\tau^p, \widehat{\theta}_{\beta_\tau}, \partial\widehat{\theta}_{\beta_\tau}/\partial\beta_\tau^T, \pi_i)|| + ||M_n(\widehat{\beta}_\tau^p, \theta_{0\tau}, \partial\theta_{0\tau}/\partial\beta_\tau^T, \pi_i)||(1 + o_p(1) + O_p(n^{-1/2})).$$

Since $M(\beta_{0\tau}, \theta_{0\tau}, \partial\theta_{0\tau}/\partial\beta_\tau^T, \pi) = 0$, it follows from the above that

$$||M_n(\beta_\tau, \widehat{\theta}_{\beta_\tau}, \partial\widehat{\theta}_{\beta_\tau}/\partial\beta_\tau^T, \pi_i)||(1 - o_p(1)) \leq o_p(1)||M(\beta_\tau, \theta_{0\tau}, \partial\theta_{0\tau}/\partial\beta_\tau^T) -$$
$$M(\beta_{0\tau}, \theta_{0\tau}, \partial\theta_{0\tau}/\partial\beta_\tau^T)|| + O_p(n^{-1/2}),$$

where all the $o_p(1)$ and $O_p(n^{-1/2})$ terms hold uniformly for $\beta_\tau \in B_\epsilon$, hence

$$||\widehat{\beta}_\tau^p - \beta_{0\tau}||C \leq ||M(\widehat{\beta}_\tau^p, \theta_{0\tau}, \partial\theta_{0\tau}/\partial\beta_\tau^T)|| \leq O_p(n^{-1/2}).$$

Let

$$\Gamma_n(\beta_\tau, \pi) = \frac{1}{n}\sum_{i=1}^n \frac{\delta_i}{\pi_i}\rho_\tau'(Y_i - X_{1i}^T\beta_\tau - X_{2i}^T\theta_{0\tau}(X_{3i}))(X_{1i}'\beta_\tau + \left(\frac{\partial\theta_{0\tau}(X_{3i})}{\partial\beta_\tau^T}\right)^T X_{2i}) + \Sigma_4(\beta_\tau - \beta_{0\tau});$$

by the $n^{1/2}$ consistency of $\widehat{\beta}_\tau^p$ and (10.16)

$$||M_n(\widehat{\beta}_\tau^p, \widehat{\theta}_{\beta tau}, \partial\widehat{\theta}_{\beta_\tau}/\partial\beta_\tau^T, \widehat{\pi}_i) - \Gamma_n(\widehat{\beta}_\tau, \widehat{\pi}_i)|| \leq ||M(\widehat{\beta}_\tau^p, \widehat{\theta}_{\beta tau}, \partial\widehat{\theta}_{\beta_\tau}/\partial\beta_\tau^T) - M(\widehat{\beta}_\tau^p, \theta_{0\tau}, \partial\theta_{0\tau}/\partial\beta_\tau^T)||+$$
$$||M(\widehat{\beta}_\tau^p, \theta_{0\tau}, \partial\theta_{0\tau}/\partial\beta_\tau^T) - \Sigma_4(\widehat{\beta}_\tau^p - \beta_{0\tau})|| + ||M_n(\widehat{\beta}_\tau^p, \widehat{\theta}_{\beta tau}, \partial\widehat{\theta}_{\beta_\tau}/\partial\beta_\tau^T) - M(\widehat{\beta}_\tau^p, \widehat{\theta}_{\beta tau}, \partial\widehat{\theta}_{\beta_\tau}/\partial\beta_\tau^T) -$$
$$M_n(\beta_\tau, \theta_{0\tau}, \partial\theta_{0\tau}/\partial\beta_\tau^T)|| = o_p(n^{1/2}).$$

Similarly,

$$||M_n(\overline{\beta}_\tau^p, \widehat{\theta}_{\beta_\tau}, \partial\widehat{\theta}_{\beta_\tau}/\partial\beta_\tau^T, \widehat{\pi}_i) - \Gamma_n(\overline{\beta}_\tau, \widehat{\pi}_i)|| = o_p(n^{1/2}),$$

where

$$n^{1/2}(\overline{\beta}_\tau^p - \beta_{0\tau}) = -\Sigma_4\frac{1}{n^{1/2}}\sum_{i=1}^n\left(\frac{\delta_i}{\pi_i} - \frac{\delta_i(\widehat{\pi}_i - \pi_i)}{\pi_i^2}\right)(\rho_\tau'(\varepsilon_i)(X_{1i} + \left(\frac{\partial\theta_{0\tau}}{\partial\beta_\tau^T}\right)^T X_{2i}), \tag{10.20}$$

where $\overline{\beta}_\tau^p$ is the minimiser of $\Gamma_n(\beta_\tau, \widehat{\pi}_i)$. The asymptotic normality of (10.20) follows by the same arguments used in the proof of Theorem 3. Next we show that $n^{1/2}(\widehat{\beta}_\tau^p - \overline{\beta}_\tau^p) = o_p(1)$. Since $\widehat{\beta}_\tau^p$ almost minimizes $||\Gamma_n(\beta_\tau, \widehat{\pi}_i)||$ and $\overline{\beta}_\tau^p$ is the minimizer of $||\Gamma_n(\beta_\tau, \widehat{\pi}_i)||$, we have $||\Gamma_n(\widehat{\beta}_\tau^p, \widehat{\pi}_i)|| = ||\Gamma_n(\overline{\beta}_\tau, \widehat{\pi}_i)|| + o_p(n^{-1/2})$, so squaring both sides and using a simple expansion

$$||\Gamma_n(\widehat{\beta}_\tau, \widehat{\pi}_i)||^2 = ||\Gamma_n(\overline{\beta}_\tau, \widehat{\pi}_i)||^2 + ||\Sigma_4(\widehat{\beta}_\tau^p - \overline{\beta}_\tau^p)||^2 + o_p(n^{-1})$$

which implies $||\Sigma_4(\widehat{\beta}_\tau^p - \overline{\beta}_\tau^p)|| = o_p(n^{-1})$ and by A6'(i) $||\widehat{\beta}_\tau^p - \overline{\beta}_\tau^p)|| = o_p(n^{-1/2})$. The conclusion follows, since it can be easily seen that

$$\frac{\partial \theta_{0\tau}(X_{3i})}{\partial \beta_\tau} = -E(f_{\varepsilon|X}(0) X_2 X_2^T | X_3 = X_{3i})^{-1} E(f_{\varepsilon|X}(0) X_2 X_1^T | X_3 = X_{3i}).$$

■

Proof of Theorem 6. By the same arguments used in the proof of Theorem 3 we have that, conditionally on $\left(Y_i, \delta_i, X_i^T\right)_{i=1}^n$

$$\begin{aligned} R_{\xi n}\left(\gamma_{\beta_\tau}, \widehat{\pi}_i\right) = & \sum_{i=1}^n \frac{\delta_i \xi_i}{\pi_{0i}} \left[\rho_\tau\left(\widehat{\varepsilon}_i^* - \frac{X_{1i}^T \gamma_{\beta_\tau}}{n^{1/2}}\right) - \rho_\tau\left(\widehat{\varepsilon}_i^*\right)\right] - \\ & \sum_{i=1}^n \frac{\delta_i \xi_i \left(\widehat{\pi}_i - \pi_{0i}\right)}{\widehat{\pi}_{\xi i} \pi_{0i}} \left[\rho_\tau\left(\widehat{\varepsilon}_i^* - \frac{X_{1i}^T \gamma_{\beta_\tau}}{n^{1/2}}\right) - \rho_\tau\left(\widehat{\varepsilon}_i^*\right)\right] \\ := & R_{\xi 1n}\left(\gamma_{\beta_\tau}, \pi_0\right) + R_{\xi 2n}\left(\gamma_{\beta_\tau}, \widehat{\pi}_\xi\right). \end{aligned}$$

Using the same arguments as those used in the proof of Theorem 3, we have by CL and QAL that $\widehat{\gamma}_{\xi\beta_\tau} = \Sigma_2^{-1}\zeta_\xi$, where

$$\zeta_\xi = \frac{1}{n^{1/2}} \sum_{i=1}^n \xi_i \left(\frac{\delta_i}{\pi_{0i}} - \frac{\delta_i\left(\widehat{\pi}_i - \pi_{0i}\right)}{\pi_{0i}^2}\right)\left(X_{1i} - \varphi\left(X_i\right)\right) \rho_\tau'\left(\varepsilon_i\right).$$

For $\widehat{\pi}_i$ estimated parametrically it follows that

$$\begin{aligned} n^{1/2}\left(\widehat{\beta}_\tau^* - \widehat{\beta}_\tau\right) = & \Sigma_2^{-1} \frac{1}{n^{1/2}} \sum_{i=1}^n \left(\xi_i - 1\right)\left\{\left(X_{1i} - \varphi\left(X_i\right)\right)\rho_\tau'\left(\varepsilon_i\right) - \right. \\ & \left. E\left[\frac{\left(X_1 - \varphi\left(X\right)\right)\rho_\tau'\left(\varepsilon\right)}{\pi_0} \frac{\partial \pi_0}{\partial \alpha^T}\right] I\left(\alpha_0\right)^{-1} \frac{1}{n^{1/2}} \sum_{i=1}^n s\left(Z_{oi}, \alpha_0\right)\right\} + o_p(1), \end{aligned}$$

since $\left\|\Sigma_2^{*-1} - \Sigma_2^{-1}\right\| = o_p(1)$ by LLN and CMT, where $\Sigma_2^* = E^*\left[\xi_i f_{\varepsilon_i|X_i}(0) X_{1i}^{\otimes 2}\right]$ and E^* denote expectation conditional on $\left(\left[Y_i, \delta_i, X_i^T\right]^T\right)_{i=1}^n$, whereas for $\widehat{\pi}_i$ estimated nonparametrically it follows that

$$\begin{aligned} n^{1/2}\left(\widehat{\beta}_\tau^* - \widehat{\beta}_\tau\right) = & \Sigma_2^{-1} \frac{1}{n^{1/2}} \sum_{i=1}^n \left(\xi_i - 1\right)\left\{\left(X_{1i} - \varphi\left(X_i\right)\right)\rho_\tau'\left(\varepsilon_i\right) - \right. \\ & \left. \frac{1}{n^{1/2}} \sum_{i=1}^n \frac{\left(\delta_i - \pi_{0i}\right)}{\pi_{0i}} E\left[\left(X_{1i} - \varphi\left(X_i\right)\right)\rho_\tau'\left(\varepsilon_i\right) | Z_{oi}\right]\right\} + o_p(1), \end{aligned}$$

and the first conclusion follows by CMT and Lemma 2.9.5 of Van der Vaart & Wellner (1996). For the profile estimator, first note that by (i), the uniform consistency of $\widehat{\pi}_i$, the c_r and triangle inequalities and LLN show that

$$||\widehat{\Sigma}_4-\Sigma_4|| \leq \sup_{X_i\in\mathcal{X}}|\widehat{f}_{\widehat{\varepsilon}|X_i}(0)-f_{\varepsilon_i|X_i}(0)|\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\widehat{\pi}_i}(||X_{1i}||^2+\sup_{X_{3i}\in\mathcal{X}_3}||\frac{\partial\widehat{\theta}_\tau(X_{3i})}{\partial\beta_\tau^T}-\frac{\partial\theta_{0\tau}(X_{3i})}{\partial\beta_\tau^T}||^2||X_{2i}||^2)+$$

$$||\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\widehat{\pi}_i}(X_{1i}+\left(\frac{\partial\theta_{0\tau}(X_{3i})}{\partial\beta_\tau^T}\right)^T X_{2i}))^{\otimes 2}-\Sigma_4||=o_p(1),$$

hence by CMT $||\widehat{\Sigma}_4^{-1}-\Sigma_4^{-1}||=o_p(1)$. By the uniform consistency of kernel estimators $||\widehat{\varphi}^p(X_i)-\varphi(X_i)||=o_p(1)$, hence

$$n^{1/2}(\widehat{\beta}_\tau^{p*}-\widehat{\beta}_\tau^{p})=\Sigma_4^{-1}\frac{1}{n^{1/2}}\sum_{i=1}^{n}\xi_i\left(\frac{\delta_i}{\pi_{0i}}-\frac{\delta_i\left(\widehat{\pi}_i-\pi_{0i}\right)}{\pi_{0i}^2}\right)\left(X_{1i}-\varphi^p\left(X_i\right)\right)\rho_\tau'\left(\varepsilon_i\right).$$

and the rest of the proof follows by the same arguments as those used above. ■

Proof of Corollary 7. Let E^* denote expectation conditional on $\left(\left[Y_i,\delta_i,X_i^T\right]^T\right)_{i=1}^n$ and let $q=2+\epsilon$. Given Theorem 6, it is sufficient to show that $E^*\left(n^{q/2}\left\|\widehat{\beta}_\tau^*-\widehat{\beta}_\tau\right\|^q\right)=O_p(1)$ and $E^*\left(n^{q/2}\left\|\widehat{\beta}_\tau^{p*}-\widehat{\beta}_\tau^{p}\right\|^q\right)=O_p(1)$ For $\widehat{\pi}_i$ estimated parametrically, the c_r inequality implies that

$$\begin{aligned}E^*\left(\left\|\widehat{\beta}_\tau^*-\widehat{\beta}_\tau\right\|^q\right)\leq 2^{q-1}&\left(E^*\left\|\Sigma_2^{*-1}\frac{1}{n}\sum_{i=1}^{n}\left(\xi_i-1\right)\left(X_{1i}-\varphi\left(X_i\right)\right)\rho_\tau'\left(\varepsilon_i\right)\right\|^q+\right.\\ &\left.E^*\left\|\Sigma_2^{*-1}E^*\left[\frac{\left(X_1-\varphi\left(X\right)\right)\rho_\tau'\left(\varepsilon\right)}{\pi_0}\frac{\partial\pi_0}{\partial\alpha^T}\right]I^*\left(\alpha_0\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}s\left(Z_{oi},\alpha_0\right)\right\|^q\right)\\ &:=V_1+V_2.\end{aligned}$$

For V_1 note that by Jensen, Holder and the c_r inequalities

$$\begin{aligned}V_1&\leq\left\|\Sigma_2^{*-1}\right\|^q\left|\left(E^*\frac{1}{n}\sum_{i=1}^{n}\left|\left(\xi_i-1\right)\right|^q\left\|\left(X_{1i}-\varphi\left(X_i\right)\right)\rho_\tau'\left(\varepsilon_i\right)\right\|^q\right)^{2/q}\right|^{q/2}\\ &\leq C\left\|\Sigma_2^{*-1}\right\|^q\left|\left(\frac{1}{n}\sum_{i=1}^{n}\left\|X_{1i}\right\|^q+\frac{1}{n}\sum_{i=1}^{n}\left\|\varphi\left(X_i\right)\right\|^q\right)^{2/q}O\left(\frac{1}{n}\right)\right|^{q/2}=O_p\left(\frac{1}{n^{q/2}}\right)\end{aligned}$$

by A7 and LLN. A similar argument can be used to show that $V_2=O_p\left(n^{-q/2}\right)$, hence $E^*\left(n^{q/2}\left\|\widehat{\beta}_\tau^*-\widehat{\beta}_\tau\right\|^q\right)=O_p(1)$. For $\widehat{\pi}_i$ estimated nonparametrically, again by the c_r inequality

$$\begin{aligned}E^*\left(\left\|\widehat{\beta}_\tau^*-\widehat{\beta}_\tau\right\|^q\right)\leq 2^{q-1}&\left(E^*\left\|\Sigma_2^{*-1}\frac{1}{n}\sum_{i=1}^{n}\left(\xi_i-1\right)\left(X_{1i}-\varphi\left(X_i\right)\right)\rho_\tau'\left(\varepsilon_i\right)\right\|^q+\right.\\ &\left.E^*\left\|\Sigma_2^{*-1}\frac{1}{n}\sum_{i=1}^{n}\left(\xi_i-1\right)\frac{\left(\delta_i-\pi_{0i}\right)}{\pi_{0i}}E^*\left[\left(X_{1i}-\varphi\left(X_i\right)\right)\rho_\tau'\left(\varepsilon_i\right)|Z_{oi}\right]\right\|^q\right)\\ &=V_1+V_3.\end{aligned}$$

Note that by Jensen and Holder inequalities

$$V_3 \leq C \left\| \Sigma_2^{*-1*} \left[\left(X_{1i} - \varphi\left(X_i\right) \right) \rho_\tau' \left(\varepsilon_i\right) | Z_{oi} \right] \right\|^q \left| \left(\frac{1}{n} \sum_{i=1}^n \left| \left(\delta_i - \pi_{0i}\right) \right|^q \right)^{2/q} O\left(\frac{1}{n}\right) \right|^{q/2} = O_p\left(\frac{1}{n^{q/2}}\right),$$

hence $E^* \left(n^{q/2} \left\| \widehat{\beta}_\tau^* - \widehat{\beta}_\tau \right\|^q \right) = O_p(1)$ by A7 and LLN. Similar arguments can be used to show that $E^*(n^{q/2} || \widehat{\beta}_\tau^{p*} - \widehat{\beta}_\tau^p ||^q) = O_p(1)$. ■

Proof of Proposition 8. The uniform consistency assumptions and the triangle inequality show that

$$\begin{aligned}
\left\| \widehat{\Sigma}_3\left(x_3^*\right) - \Sigma_3\left(x_3^*\right) \right\| \leq & \sup_{x_3^* \in \mathcal{X}_3} \left| \widehat{f}_{X_3}\left(x_3^*\right) - f_{X_3}\left(x_3^*\right) \right| \left\| \frac{1}{nh} \sum_{i=1}^n \frac{\delta_i}{\widehat{\pi}\left(Z_{oi}\right)} \widehat{f}_{\widehat{\varepsilon}_i | X_i}(0) X_{2i}^{\otimes 2} K_h\left(X_{3i} - x_3^*\right) \right\| + \\
& \sup_{x_3^* \in \mathcal{X}_3} \left| f_{X_3}\left(x_3^*\right) \right| \left[\sup_{X_i \in \mathcal{X}} \left| \widehat{f}_{\widehat{\varepsilon}_i | X_i}(0) - f_{\varepsilon_i | X_i}(0) \right| \left\| \frac{1}{nh} \sum_{i=1}^n \frac{\delta_i}{\widehat{\pi}\left(Z_{oi}\right)} X_{2i}^{\otimes 2} K_h\left(X_{3i} - x_3^*\right) \right\| + \right. \\
& \sup_{Z_{oi} \in \mathcal{Z}} \left| \widehat{\pi}\left(Z_{oi}\right) - \pi_0\left(Z_{oi}\right) \right| \left\| \frac{1}{nh} \sum_{i=1}^n \frac{\delta_i}{\widehat{\pi}\left(Z_{oi}\right) \pi_0\left(Z_{oi}\right)} X_{2i}^{\otimes 2} K_h\left(X_{3i} - x_3^*\right) \right\| + \\
& \left\| \frac{1}{nh} \sum_{i=1}^n f_{\varepsilon | X}\left(0 | X_i\right) X_{2i}^{\otimes 2} K_h\left(X_{3i} - x_3^*\right) - E\left[f_{\varepsilon | X}(0) X_2^{\otimes 2} | X_3 = x_3^* \right] \right\| + o_p(1) = \\
& o_p(1) O_p(1) + O_p(1) o_p(1) = o_p(1).
\end{aligned}$$

Similarly, we have that $\left\| \widehat{\Sigma}_{3\widehat{\pi}}\left(x_3^*\right) - \Sigma_{3\pi}\left(x_3^*\right) \right\| = o_p(1)$. Under (4.1), the same arguments as those used in Theorem 2 and CMT show that

$$(nh)^{1/2} R\left(\widehat{\theta}_\tau\left(x_3^*\right) - \theta_{0\tau}\left(x_3^*\right) \right) \xrightarrow{d} N\left(\gamma_\tau\left(x_3^*\right), R \Sigma_3\left(x_3^*\right)^{-1} \Sigma_{3\pi}\left(x_3^*\right) \Sigma_3\left(x_3^*\right)^{-1} R^T \right)$$

hence the first conclusion follows by standard results on quadratic forms in non zero mean normal random vectors. The consistency of $W_l\left(x_3^*\right)$ under the assumption that $(nh)^{1/2} \gamma_{\tau n}\left(x_3^*\right) \to \infty$ is a direct consequence of the previous conclusion. ■

Proof of Theorem 9. The proof relies on similar arguments used by Fan et al. (2001), and consists of showing that $D_\pi\left(\theta_{0\tau}\right)$ can be approximated by a U-statistic, which, after being appropriately standardised, converges to a standard normal variate. Note that the same arguments of Theorem 2 imply that

$$\begin{aligned}
& D_{\widehat{\pi}}\left(\theta_{0\tau}\right) = \sum_{i=1}^n \frac{\delta_i}{\pi_i} \rho_\tau\left(Y_i - X_{1i}^T \beta_{0\tau} - X_{2i}^T \widehat{\theta}_{\tau - i}\left(X_{3i}\right) \right) - \sum_{i=1}^n \frac{\delta_i}{\pi_i} \rho_\tau\left(Y_i - X_{1i}^T \beta_{0\tau} - X_{3i}^T \theta_{0\tau}\left(X_{3i}\right) \right) + \\
& \sum_{i=1}^n \frac{\delta_i\left(\widehat{\pi}_i - \pi_{0i}\right)}{\widehat{\pi}_i \pi_{0i}} \left(\rho_\tau\left(Y_i - X_{1i}^T \beta_{0\tau} - X_{3i}^T \widehat{\theta}_{\tau - i}\left(X_{3i}\right) \right) - \rho_\tau\left(Y_i - X_{1i}^T \beta_{0\tau} - X_{3i}^T \theta_{0\tau}\left(X_{3i}\right) \right) \right) + o_p(1) := \\
& D_{1\pi} + D_{2\pi} + o_p(1),
\end{aligned}$$

where

$$\widehat{\theta}_{\tau - i}\left(X_{3i}\right) - \theta_{0\tau}\left(X_{3i}\right) = \left(f_{X_3}\left(X_{3i}\right) \Sigma_3\left(X_{3i}\right) \right)^{-1} \frac{1}{nh} \sum_{j \neq i}^n \frac{\delta_j}{\pi_{0j}} X_{2j} \rho_\tau'\left(\varepsilon_j^*\right) K_h\left(X_{3j} - X_{3i}\right) + o_p\left((nh)^{-1/2} \right). \tag{10.21}$$

By (10.1)

$$D_{1\pi} = -\sum_{i=1}^{n} \frac{\delta_i}{\pi_i} X_{2i}^T \left(\widehat{\theta}_{\tau-i}(X_{3i}) - \theta_{0\tau}(X_{3i})\right) \rho'_\tau(\varepsilon_i) +$$
$$\sum_{i=1}^{n} \frac{\delta_i}{\pi_{0i}} \int_0^{X_{2i}^T\left(\widehat{\theta}_{\tau-i}(X_{3i})-\theta_{0\tau}(X_{3i})\right)} \left(I(\varepsilon_i \le t) - I(\varepsilon_i \le 0)\right) dt := D_{11\pi} + D_{12\pi}.$$

Using (10.21)

$$D_{11\pi} = -\sum_{i=1}^{n} \frac{\delta_i}{\pi_{0i}} X_{2i}^T \rho'_\tau(\varepsilon_i) \left(f_{X_3}(X_{3i}) \Sigma_3(X_{3i})\right)^{-1} \frac{1}{nh} \sum_{j\neq i}^{n} \frac{\delta_j}{\pi_{0j}} X_{2j} \rho'_\tau\left(\varepsilon_j^*\right) K_h(X_{3j} - X_{3i}) =$$
$$-\sum_{i=1}^{n} \frac{\delta_i}{\pi_{0i}} X_{2i}^T \rho'_\tau(\varepsilon_i) \left(f_{X_3}(X_{3i}) \Sigma_3(X_{3i})\right)^{-1} \frac{1}{nh} \sum_{j\neq i}^{n} \frac{\delta_j}{\pi_{0j}} X_{2j} \rho'_\tau(\varepsilon_j) K_h(X_{3j} - X_{3i}) -$$
$$\sum_{i=1}^{n} \frac{\delta_i}{\pi_{0i}} X_{2i}^T \rho'_\tau(\varepsilon_i) \left(f_{X_3}(X_{3i}) \Sigma_3(X_{3i})\right)^{-1} \frac{1}{nh} \sum_{j\neq i}^{n} \frac{\delta_j}{\pi_{0j}} X_{2j} \left(\rho'_\tau\left(\varepsilon_j^*\right) - \rho'_\tau(\varepsilon_j)\right) K_h(X_{3j} - X_{3i}) :=$$
$$D_{111\pi} + D_{112\pi}$$

and

$$D_{112\pi} = -\sum_{i=1}^{n} \frac{\delta_i}{\pi_{0i}} X_{2i}^T \rho'_\tau(\varepsilon_i) \left(f_{X_3}(X_{3i}) \Sigma_3(X_{3i})\right)^{-1} \frac{1}{nh} \sum_{j\neq i}^{n} E\left[\frac{\delta_j}{\pi_{0j}} X_{2j} \left(\rho'_\tau\left(\varepsilon_j^*\right) - \rho'_\tau(\varepsilon_j)\right) K_h(X_{3j} - X_{3i})\right] -$$
$$\sum_{i=1}^{n} \frac{\delta_i}{\pi_{0i}} X_{2i}^T \rho'_\tau(\varepsilon_i) \left(f_{X_3}(X_{3i}) \Sigma_3(X_{3i})\right)^{-1} \frac{1}{nh} \sum_{j\neq i}^{n} \left\{\frac{\delta_j}{\pi_{0j}} X_{2j} \left(\rho'_\tau\left(\varepsilon_j^*\right) - \rho'_\tau(\varepsilon_j)\right) K_h(X_{3j} - X_{3i}) -\right.$$
$$\left. E\left[\frac{\delta_j}{\pi_{0j}} X_{2j} \left(\rho'_\tau\left(\varepsilon_j^*\right) - \rho'_\tau(\varepsilon_j)\right) K_h(X_{3j} - X_{3i})\right]\right\} := D_{1121\pi} + D_{1122\pi}.$$

By the results of Theorem 2 and a standard kernel calculation we have that

$$D_{1121\pi} = \frac{h^2}{2} \sum_{i=1}^{n} \frac{\delta_i}{\pi_{0i}} X_{2i}^T \rho'_\tau(\varepsilon_i) \theta''_{0\tau}(X_{3i}) \kappa_2 (1 + o_p(1)), \tag{10.22}$$
$$E(D_{1122\pi})^2 = O(h),$$

so that

$$D_{112\pi} = n^{1/2} h^2 \sum_{i=1}^{n} \frac{\delta_i}{\pi_{0i}} X_{2i}^T \rho'_\tau(\varepsilon_i) \theta''_{0\tau}(X_{3i}) \kappa_2 (1 + o_p(1)) / n^{1/2} := n^{1/2} h^2 T_{1\pi} = O_p\left(n^{1/2} h^2\right).$$

Next, by iterated expectations

$$E(D_{12\pi}) = \sum_{i=1}^{n} E \int_0^{X_{2i}^T\left(\widehat{\theta}_{\tau-i}(X_{3i})-\theta_{0\tau}(X_{3i})\right)} \left(F_{\varepsilon_i|X_i}(t) - F_{\varepsilon_i|X_i}(0)\right) dt =$$
$$\frac{1}{2} \sum_{i=1}^{n} E\left[f_{\varepsilon_i|X_i}(0) \left(\widehat{\theta}_{\tau-i}(X_{3i}) - \theta_{0\tau}(X_{3i})\right)^T X_{2i}^{\otimes 2} \left(\widehat{\theta}_{\tau-i}(X_{3i}) - \theta_{0\tau}(X_{3i})\right)\right],$$

and

$$E\left(D_{12\pi}-E\left(D_{12\pi}\right)\right)^{2}\leq nE\left(D_{12i}\right)^{2}\leq$$

$$nE\left(\Pr\left(\varepsilon_{i}\left(0\leq\left|\varepsilon_{i}\right|\leq\left\|X_{2}\right\|\left\|\widehat{\theta}_{\tau-i}\left(X_{3i}\right)-\theta_{0\tau}\left(X_{3i}\right)\right\|\left|X_{i}\right)\right)\left\|X_{2i}\right\|^{2}\left\|\widehat{\theta}_{\tau-i}\left(X_{3i}\right)-\theta_{0\tau}\left(X_{3i}\right)\right\|^{2}\right)=o\left(1/h\right)$$

hence $D_{12\pi}=E\left(D_{12\pi}\right)+o_{p}\left(h^{-1/2}\right)$. By (10.21)

$$\begin{aligned}
E\left(D_{12\pi}\right)=&\frac{1}{2}\sum_{i=1}^{n}E\left[f_{\varepsilon_{i}|X_{i}}\left(0\right)\sum_{j\neq i}^{n}\frac{1}{nh}\frac{\delta_{j}}{\pi_{0j}}X_{2j}^{T}\rho_{\tau}'\left(\varepsilon_{j}\right)\left(f_{X_{3}}\left(X_{3i}\right)\Sigma_{12}\left(X_{3i}\right)\right)^{-1}X_{2i}^{\otimes2}\times\right.\\
&\left.\sum_{k\neq i}^{n}\frac{1}{nh}\frac{\delta_{k}}{\pi_{0k}}\left(f_{X_{3}}\left(X_{3i}\right)\Sigma_{3}\left(X_{3i}\right)\right)^{-1}X_{2k}\rho_{\tau}'\left(\varepsilon_{k}\right)K_{h}\left(X_{3j}-X_{3i}\right)K_{h}\left(X_{3k}-X_{3i}\right)\right]+\\
&\frac{1}{2}\sum_{i=1}^{n}E\left[f_{\varepsilon_{i}|X_{i}}\left(0\right)\sum_{j\neq i}^{n}\frac{1}{nh}\frac{\delta_{j}}{\pi_{0j}}X_{2j}^{T}\left(\rho_{\tau}'\left(\varepsilon_{j}^{*}\right)-\rho_{\tau}'\left(\varepsilon_{j}\right)\right)\left(f_{X_{3}}\left(X_{3i}\right)\Sigma_{3}\left(X_{3i}\right)\right)^{-1}X_{2i}^{\otimes2}\times\right.\\
&\left.\sum_{k\neq i}^{n}\frac{1}{nh}\frac{\delta_{k}}{\pi_{0k}}\left(f_{X_{3}}\left(X_{3i}\right)\Sigma_{3}\left(X_{3i}\right)\right)^{-1}X_{2k}\left(\rho_{\tau}'\left(\varepsilon_{k}^{*}\right)-\rho_{\tau}'\left(\varepsilon_{k}\right)\right)K_{h}\left(X_{3j}-X_{3i}\right)K_{h}\left(X_{3k}-X_{3i}\right)\right]+\\
&\sum_{i=1}^{n}E\left[f_{\varepsilon_{i}|X_{i}}\left(0\right)\sum_{j\neq i}^{n}\frac{1}{nh}\frac{\delta_{j}}{\pi_{0j}}X_{2j}^{T}\rho_{\tau}'\left(\varepsilon_{j}\right)\left(f_{X_{3}}\left(X_{3i}\right)\Sigma_{3}\left(X_{3i}\right)\right)^{-1}X_{2i}^{\otimes2}\times\right.\\
&\left.\sum_{k\neq i}^{n}\frac{1}{nh}\frac{\delta_{k}}{\pi_{0k}}\left(f_{X_{3}}\left(X_{3i}\right)\Sigma_{3}\left(X_{3i}\right)\right)^{-1}X_{2k}\left(\rho_{\tau}'\left(\varepsilon_{k}^{*}\right)-\rho_{\tau}'\left(\varepsilon_{k}\right)\right)K_{h}\left(X_{3j}-X_{3i}\right)K_{h}\left(X_{3k}-X_{3i}\right)\right]:=\\
&D_{121\pi}+D_{122\pi}+D_{123\pi}.
\end{aligned}$$

For $D_{122\pi}$ and $D_{123\pi}$, similar to (10.22), we have that

$$\begin{aligned}
D_{122\pi}=&-\frac{nh^{4}}{8}E\left[f_{\varepsilon|X}\left(0\right)\theta_{0\tau}''\left(X_{3}\right)^{T}X_{2}^{\otimes2}\theta_{0\tau}''\left(X_{3}\right)\right]\int\int t^{2}\left(t+s\right)^{2}K\left(t\right)K\left(t+s\right)dtds+o_{p}\left(1\right)\\
&:=-nh^{4}T_{2\pi}=O_{p}\left(nh^{4}\right),\\
D_{123\pi}=&-\frac{n^{1/2}h^{2}}{2}\sum_{i=1}^{n}\frac{\delta_{i}}{\pi_{0i}}X_{2i}^{T}\rho_{\tau}'\left(\varepsilon_{i}\right)\theta_{0\tau}''\left(X_{3i}\right)\int\int t^{2}\left(t+s\right)^{2}K\left(t\right)K\left(t+s\right)dtds\left(1+o_{p}\left(1\right)\right)/n^{1/2}\\
&:=-n^{1/2}h^{2}T_{3\pi}.
\end{aligned}$$

For $D_{121\pi}$,

$$D_{121\pi} = \frac{1}{2(nh)^2}\sum_{j\neq i}^{n}\left(\frac{\delta_j}{\pi_{0j}}\right)^2 E\left[X_{2j}^T\rho_\tau'(\varepsilon_j)^2\left(f_{X_3}(X_{3i})\Sigma_3(X_{3i})\right)^{-1}\sum_{i=1}^{n} f_{\varepsilon_i|X_i}(0)X_{2i}^{\otimes 2}\times\right.$$
$$\left.\left(f_{X_3}(X_{3i})\Sigma_3(X_{3i})\right)^{-1}X_{2j}K_h^2(X_{3j}-X_{3i})\right]+$$
$$\frac{1}{2(nh)^2}\sum_{\substack{j\neq k\\ j,k\neq i}}^{n} E\left[\frac{\delta_j}{\pi_{0j}}X_{2j}^T\rho_\tau'(\varepsilon_j)\left(f_{X_3}(X_{3i})\Sigma_3(X_{3i})\right)^{-1}\sum_{i=1}^{n} f_{\varepsilon_i|X_i}(0)X_{2i}^{\otimes 2}\times\right.$$
$$\left.\frac{\delta_k}{\pi_{0k}}\left(f_{X_3}(X_{3i})\Sigma_3(X_{3i})\right)^{-1}X_{2k}\rho_\tau'(\varepsilon_k)K_h(X_{3j}-X_{3i})K_h(X_{3k}-X_{3i})\right] := D_{1211\pi}+D_{1212\pi}.$$

Note that $D_{1211\pi}$ can be re-written as

$$D_{1211\pi} = \frac{1}{2(nh)^2}\sum_{j\neq i}^{n}\sum_{i=1}^{n}\left(\frac{\delta_j}{\pi_{0j}}\right)^2 X_{2j}^T\rho_\tau'(\varepsilon_j)^2\int\left[\left(f_{X_3}(X_{3i})\Sigma_3(X_{3i})\right)^{-1}\Sigma_3(X_{3i})\times\right.$$
$$\left.\left(f_{X_3}(X_{3i})\Sigma_3(X_{3i})\right)^{-1}X_{2j}K_h^2(X_{3j}-X_{3i})\right]f_{X_3}(X_{3i})\,dX_{3i},$$

and that

$$Var(D_{1211\pi}) \leq \frac{n^3}{n^4h^2}\int\int trE\left[\left(\frac{\tau(1-\tau)}{\pi_{0j}}\left(f_{X_3}(X_{3i})\Sigma_3(X_{3i})\right)^{-1}X_{2j}^{\otimes 2}|X_{3j}\right)\times\right.$$
$$\left.K_h^2(X_{3j}-X_{3i})\right]^2 f_{X_3}^2(X_{3j})\,dX_{3i}dX_{3j} = O\left(\frac{1}{nh^2}\right),$$

hence $D_{1211\pi} = E(D_{1211\pi}) + o_p\left(h^{-1/2}\right)$, and by iterated expectations and a standard kernel calculation we have that

$$E(D_{1211\pi}) = \frac{1}{2h^2}\int\int tr\left(E\left[\left(\frac{\delta_j}{\pi_{0j}}\right)^2\rho_\tau'(\varepsilon_j)^2\left(f_{X_3}(X_{3i})\Sigma_3(X_{3i})\right)^{-1}X_{2j}^{\otimes 2}\times\right.\right.$$
$$\left.\left.|X_{3i},X_{3j}\right]K_h^2(X_{3j}-X_{3i})\right)f_{X_3}(X_{3i})f_{X_3}(X_{3j})\,dX_{3i}dX_{3j}\Big) =$$
$$\frac{1}{2h^2}\int\int tr\left[\left(f_{X_3}(X_{3i})\Sigma_3(X_{3i})\right)^{-1}\left(f_{X_3}(X_{3i})\Sigma_3(X_{3i})\right)\left(f_{X_3}(X_{3i})\Sigma_3(X_{3i})\right)^{-1}\times\right.$$
$$\left.E\left[\frac{\tau(1-\tau)}{\pi_{0j}}X_{2j}^{\otimes 2}|X_{3j}\right]K_h^2(X_{3j}-X_{3i})f_{X_3}(X_{3j})\,dX_{3i}dX_{3j}\right. =$$
$$\frac{1}{2h}\int\int tr\left(E\left[\frac{\tau(1-\tau)}{\pi_{0j}}\left(f_{X_3}(X_{3i})\Sigma_3(X_{3i})\right)^{-1}X_{2j}^{\otimes 2}|X_{3j}=X_{3i}+th\right]K^2(t)\,dtdX_{3i}\right) =$$
$$\frac{1}{2h}\int tr\left(E\left[\frac{\tau(1-\tau)}{\pi_0}\Sigma_3(X_3)^{-1}X_2^{\otimes 2}|X_3\right]\kappa_2 dX_3\left(1+O(h)\right)\right) =$$
$$\frac{tr}{2h}E\left[\frac{\tau(1-\tau)}{\pi_0 f_{X_3}(X_3)}\Sigma_3(X_3)^{-1}X_2^{\otimes 2}\right]\kappa_2\left(1+o(1)\right).$$

For $D_{1212\pi}$ since $K(\cdot)$ is symmetric we have by a standard U statistic argument

$$D_{1212\pi} = \frac{2}{nh^2}\sum_{i<j}\frac{\delta_i}{\pi_{0i}}X_{2i}^T\rho_\tau'(\varepsilon_i)\left(f_{X_3}(X_{3i})\Sigma_3(X_{3i})\right)^{-1}h\int f_{X_3}(X_{3i})\Sigma_{12}(X_{3i})\times$$
$$K_h(t)K\left(\frac{X_{3i}-X_{3j}}{h}+t\right)dt\frac{\delta_j}{\pi_{0j}}X_{2j}^T\rho_\tau'(\varepsilon_j)(1+O(h)) =$$
$$\frac{2}{nh}\sum_{i<j}\frac{\delta_i}{\pi_{0i}}X_{2i}^T\rho_\tau'(\varepsilon_i)\left(f_{X_3}(X_{3i})\Sigma_3(X_{3i})\right)^{-1}K_h * K_h(X_{3i}-X_{3j})\frac{\delta_j}{\pi_{0j}}X_{2j}^T\rho_\tau'(\varepsilon_j)+o_p(1).$$

Thus

$$D_{1\pi} = \frac{U_\pi}{2h^{1/2}}+\frac{1}{2h}E\left[\frac{\tau(1-\tau)}{\pi_0 f_{X_3}(X_3)}\Sigma_3(X_3)^{-1}X_2^{\otimes 2}\right]\kappa_2+n^{1/2}h^2(T_{1\pi}-T_{3\pi})-nh^4T_{2\pi}+o_p\left(h^{-1/2}\right),$$

where

$$U_\pi = \sum_{1\le i<j\le n}U_{ijn},$$

$$U_{ijn} = \frac{\sqrt{h}}{n}\rho_\tau'(\varepsilon_i)\rho_\tau'(\varepsilon_j)U_{ij\pi}$$
$$U_{ij\pi} = U_{1ij\pi}+U_{1ji\pi}+U_{2ij\pi}+U_{2ji\pi}$$

and

$$U_{1ij\pi} = 2\frac{\delta_i}{\pi_{0i}}\frac{\delta_j}{\pi_{0j}}X_{2i}^T\left(f_{X_3}(X_{3i})\Sigma_3(X_{3i})\right)^{-1}X_{2j}K_h(X_{3j}-X_{3i}),$$
$$U_{2ij\pi} = \frac{\delta_i}{\pi_{0i}}\frac{\delta_j}{\pi_{0j}}X_{2i}^T\left(f_{X_3}(X_{3i})\Sigma_3(X_{3i})\right)^{-1}X_{2j}K_h * K_h(X_{3j}-X_{3i}).$$

To show the asymptotic normality of U_π, we check conditions C(i)-C(iv) of Proposition 3.2 of de Jong (1987), that is C(i) $E(U_{ijn}) = 0$, C(ii) $Var(U_\pi)$ converges to a finite quantity as $n\to\infty$, C(iii) $G_I = \sum_{1\le i<j\le n}E\left(U_{ijn}^4\right)$ is of smaller order than $\lim_{n\to\infty}Var(U_\pi)$, C(iv)

$$G_{II} = \sum_{1\le i<j<k\le n}\left(E\left(U_{ijn}^2U_{ikn}^2\right)+E\left(U_{jin}^2U_{jkn}^2\right)+E\left(U_{kin}^2U_{kjn}^2\right)\right)$$

is of smaller order than $\lim_{n\to\infty}Var(U_\pi)$ and C(v)

$$G_{III} = \sum_{1\le i<j<k<l\le n}\left(E\left(U_{ijn}U_{ikn}U_{ljn}U_{lkn}\right)+E\left(U_{ijn}U_{iln}U_{kjn}U_{kln}\right)+E\left(U_{ikn}U_{iln}U_{jkn}U_{jln}\right)\right)$$

is of smaller order than $\lim_{n\to\infty} Var(U_\pi)$. C(i) is true by definition; to show C(ii) note that

$$
\begin{aligned}
E\left(U_{1ij\pi}\right)^2 &= \frac{4}{h} tr\left(E\left(\frac{\Sigma_3\left(X_3\right)^{-1}}{\pi_0 f_{X_3}\left(X_3\right)} X_2^{\otimes 2}\right)^2 \kappa_2\left(1+O\left(h\right)\right)\right), \\
E\left(U_{1ij\pi}\right)^2 &= E\left(U_{1ji}\right)^2, \\
E\left(U_{2ij\pi}\right)^2 &= \frac{1}{h} tr\left(E\left(\frac{\Sigma_3\left(X_3\right)^{-1}}{\pi_0 f_{X_3}\left(X_3\right)} X_2^{\otimes 2}\right)^2 \int\left(K_h * K_h\left(t\right)\right)^2 dt\left(1+O\left(h\right)\right)\right), \\
E\left(U_{2ij\pi}\right)^2 &= E\left(U_{2ji}\right)^2, \\
E\left(U_{1ij\pi}U_{2ij\pi}\right) &= \frac{2}{h} tr\left(E\left(\frac{\Sigma_3\left(X_3\right)^{-1}}{\pi_0 f_{X_3}\left(X_3\right)} X_2^{\otimes 2}\right)^2 \int\left(K_h * K_h * K_h\left(t\right)\right) dt\left(1+O\left(h\right)\right)\right),
\end{aligned}
$$

so that

$$
Var\left(U_\pi\right) := \sigma_\pi^2 = \frac{2}{h} tr\left(E\left(\frac{\tau\left(1-\tau\right)}{\pi_0 f_{X_3}\left(X_3\right)}\Sigma_3\left(X_3\right)^{-1} X_2^{\otimes 2}\right)^2 \int\left(2K_h\left(t\right) - K_h * K_h\left(t\right)\right)^2 dt\right) + o\left(1\right).
$$

To show C(iii), note that by a direct calculation

$$
\begin{aligned}
E\left(U_{1ij\pi}\rho_\tau'\left(\varepsilon_i\right)\rho_\tau'\left(\varepsilon_j\right)\right) &= O\left(h^{-3}\right) \\
E\left(U_{2ij\pi}\rho_\tau'\left(\varepsilon_i\right)\rho_\tau'\left(\varepsilon_j\right)\right) &= O\left(h^{-2}\right),
\end{aligned}
$$

which implies that $E\left(U_{ijn}^4\right) = h^2 O\left(h^{-3}\right)/n^4 = O\left(1/n^4 h\right) = o\left(1\right)$. To show condition C(iv), note that $E\left(U_{ijn}^2 U_{ikn}^2\right) = O\left(E\left(U_{ijn}^4\right)\right) = o\left(1\right)$. Finally, to show C(v) note that for $i \neq j \neq k \neq l$,

$$
\begin{aligned}
E\left(U_{1ij\pi}U_{1jk\pi}U_{1kl\pi}U_{1li\pi}\rho_\tau'\left(\varepsilon_i\right)\rho_\tau'\left(\varepsilon_j\right)\rho_\tau'\left(\varepsilon_k\right)\rho_\tau'\left(\varepsilon_l\right)\right) &= O\left(\frac{1}{h}\right), \\
E\left(U_{1ij\pi}U_{1jk\pi}U_{1kl\pi}U_{2li\pi}\rho_\tau'\left(\varepsilon_i\right)\rho_\tau'\left(\varepsilon_j\right)\rho_\tau'\left(\varepsilon_k\right)\rho_\tau'\left(\varepsilon_l\right)\right) &= O\left(\frac{1}{h}\right), \\
E\left(U_{1ij\pi}U_{1jk\pi}U_{2kl\pi}U_{2li\pi}\rho_\tau'\left(\varepsilon_i\right)\rho_\tau'\left(\varepsilon_j\right)\rho_\tau'\left(\varepsilon_k\right)\rho_\tau'\left(\varepsilon_l\right)\right) &= O\left(\frac{1}{h}\right), \\
E\left(U_{1ij\pi}U_{2jk\pi}U_{2kl\pi}U_{2li\pi}\rho_\tau'\left(\varepsilon_i\right)\rho_\tau'\left(\varepsilon_j\right)\rho_\tau'\left(\varepsilon_k\right)\rho_\tau'\left(\varepsilon_l\right)\right) &= O\left(\frac{1}{h}\right), \\
E\left(U_{2ij\pi}U_{2jk\pi}U_{2kl\pi}U_{2li\pi}\rho_\tau'\left(\varepsilon_i\right)\rho_\tau'\left(\varepsilon_j\right)\rho_\tau'\left(\varepsilon_k\right)\rho_\tau'\left(\varepsilon_l\right)\right) &= O\left(\frac{1}{h}\right),
\end{aligned}
$$

so that $E\left(U_{ijn}U_{jkn}U_{kln}U_{lin}\right) = h^2 O\left(1/h\right)/n^4 = O\left(\left(h/n^4\right)\right) = o\left(1\right)$; hence by Proposition 3.2 of de Jong (1987) we have that $U_\pi \xrightarrow{d} N\left(0, Var\left(U_\pi\right)\right)$. To deal with the second term $D_{2\pi}$, note that

$$
\left|D_{2\pi}\right| \leq \sup_i \left|\left(\widehat{\pi}_i - \pi_{0i}\right)\right| \left|\frac{D_{1\pi}}{\pi_{0i}}\right| + o_p\left(1\right) = \sup_i \left|\left(\widehat{\pi}_i - \pi_{0i}\right)\right| O_p\left(1\right) + o_p\left(1\right). \tag{10.23}
$$

For π_{0i} estimated parametrically

$$
\sup_i \left|\left(\widehat{\pi}_i - \pi_{0i}\right)\right| \leq \left\|\widehat{\alpha} - \alpha_0\right\| \sup_{\alpha\in A}\sup_i \left\|\frac{\partial \pi_i}{\partial \alpha}\right\| = O_p\left(n^{-1/2}\right) o_p\left(n^{1/\delta}\right) = o_p\left(1\right),
$$

whereas for π_{0i} estimated nonparametrically, $\sup_{Z_{oi}\in Z}|(\widehat{\pi}_i-\pi_{0i})|=o_p(1)$ by standard results on the uniform convergence of kernel estimators. Thus

$$D_\pi(\theta_{0\tau})=D_{1\pi}+o_p(1)$$

and the conclusion follows. ■

Proof of Proposition 10. First note that without MAR

$$\begin{aligned}
&\widehat{\theta}_\tau(X_{3i})-\theta_{0\tau}(X_{3i})=(f_{X_3}(X_{3i})\Sigma_3(X_{3i}))^{-1}\times\\
&\frac{1}{nh}\left[\sum_{j=i}^n X_{2j}\rho_\tau'(\varepsilon_j^*)K_h(X_{3j}-X_{3i})+\sum_{j\neq i}^n X_{2j}\rho_\tau'(\varepsilon_j^*)K_h(X_{3j}-X_{3i})\right]+o_p\left((nh)^{-1/2}\right),\\
&=(f_{X_3}(X_{3i})\Sigma_3(X_{3i}))^{-1}\\
&\frac{1}{nh}\left[\sum_{j=i}^n X_{2j}\rho_\tau'(\varepsilon_j^*)K_h(0)+\sum_{j\neq i}^n X_{2j}\rho_\tau'(\varepsilon_j^*)K_h(X_{3j}-X_{3i})\right]+o_p\left((nh)^{-1/2}\right);
\end{aligned}$$

then, similar to the proof of Theorem 9

$$\begin{aligned}
D(\theta_{0\tau})=&-\frac{1}{nh}\sum_{i=1}^n X_{2i}^T\rho_\tau'(\varepsilon_i)^2(f_{X_3}(X_{3i})\Sigma_3(X_{3i}))^{-1}X_{2i}K_h(0)-\\
&\frac{1}{nh}\sum_{i=1}^n X_{2i}^T\rho_\tau'(\varepsilon_i)(f_{X_3}(X_{3i})\Sigma_3(X_{3i}))^{-1}\frac{1}{nh}\sum_{j\neq i}^n X_{2j}\rho_\tau'(\varepsilon_j^*)K_h(X_{3j}-X_{3i})+o_p\left((nh)^{-1/2}\right)\\
&=D_1+D_2.
\end{aligned}$$

For D_1 the LLN implies that $E(D_1)=E(\tau(1-\tau)p/f_{X_3}(X_3))K(0)/h$ while by a standard calculation $Var(D_1)=O\left(1/nh^2\right)$, hence $E(D_1)=E(\tau(1-\tau)p/f_{X_3}(X_3))K(0)/h+o_p\left(h^{-1/2}\right)$. For D_2 the same arguments of Theorem 9 show that

$$\begin{aligned}
\mu&=\frac{tr}{2h}E\left[\frac{\tau(1-\tau)}{f_{X_3}(X_3)}\Sigma_3(X_3)^{-1}X_2^{\otimes 2}\right]\kappa_2=\frac{p}{2h}E\left(\frac{\tau(1-\tau)}{f(X_3)}\right)\kappa_2,\\
\sigma^2&=\frac{2}{h}tr\left(E\left(\frac{\tau(1-\tau)}{f_{X_3}(X_3)}\Sigma_3(X_3)^{-1}X_2^{\otimes 2}\right)^2\int(2K_h(t)-K_h*K_h(t))^2dt\right)\\
&=\frac{2p^2}{h}E\left(\frac{\tau(1-\tau)}{f_{X_3}(X_3)}\right)^2\int(2K_h(t)-K_h*K_h(t))^2.
\end{aligned}$$

The conclusion follows as in Fan et al. (2001). ■

Proof of Proposition 11. Note that by the triangle inequality and LLN

$$\begin{aligned}
|\widehat{\mu}_{\widehat{\pi}}-\mu_\pi|\leq&\frac{1}{2h}\sup_i\left|\frac{\pi(Z_{oi})f_{X_3}(X_{3i})}{\widehat{\pi}(Z_{oi})\widehat{f}_{X_3}(X_{3i})}\right|\left(\left\|\widehat{\Sigma}_3(X_{3_i})^{-1}-\Sigma_3(X_{3_i})^{-1}\right\|\frac{1}{n}\sum_{i=1}^n\left[\frac{\tau(1-\tau)}{\pi(Z_{oi})f_{X_3}(X_{3i})}X_2^{\otimes 2}\right]\kappa_2+\right.\\
&\left.\left\|\frac{1}{n}\sum_{i=1}^n\left[\frac{\tau(1-\tau)}{\pi(Z_{oi})f_{X_3}(X_{3i})}X_2^{\otimes 2}\Sigma_3(X_{3_i})^{-1}\right]\right\|\kappa_2-\mu_\pi\right)=O_p(1)(o_p(1)O_p(1)+o_p(1)),
\end{aligned}$$

where $\left\|\widehat{\Sigma}_3\left(X_{3_i}\right)^{-1}-\Sigma_3\left(X_{3_i}\right)^{-1}\right\|=o_p(1)$ by the same arguments used in the proof of Proposition (8). For $\widehat{T}_{1\pi}$ the triangle inequality implies that

$$\left|\widehat{T}_{1\widehat{\pi}}-T_{1\pi}\right|\leq\sup_i\left|\frac{\widehat{\pi}\left(Z_{oi}\right)-\pi\left(Z_{oi}\right)}{\widehat{\pi}\left(Z_{oi}\right)}\right|\left\|\frac{1}{n^{1/2}}\sum_{i=1}^n\frac{\delta_i}{\pi\left(Z_{oi}\right)}X_{2i}\rho_\tau'\left(\widehat{\varepsilon}_i\right)\right\|\sup_{X_{3i}\in\mathcal{X}_3}\left(\left\|\theta_\tau''\left(X_{3i}\right)\right\|+\right.$$
$$\left.+\left\|\widehat{\theta}_\tau''\left(X_{3i}\right)-\theta_\tau''\left(X_{3i}\right)\right\|\right)\|\kappa_2+$$
$$\sup_{X_{3i}\in\mathcal{X}_3}\left\|\widehat{\theta}_\tau''\left(X_{3i}\right)-\theta_\tau''\left(X_{3i}\right)\right\|\left\|\frac{1}{n^{1/2}}\sum_{i=1}^n\frac{\delta_i}{\pi\left(Z_{oi}\right)}X_{2i}^T\rho_\tau'\left(\widehat{\varepsilon}_i\right)\right\|\kappa_2+$$
$$\left\|\frac{1}{n^{1/2}}\sum_{i=1}^n\frac{\delta_i}{\pi\left(Z_{oi}\right)}X_{2i}\left(\rho_\tau'\left(\widehat{\varepsilon}_i\right)-\rho_\tau'\left(\varepsilon_i\right)\right)\right\|\kappa_2:=V_{1\pi}+V_{2\pi}+V_{3\pi},$$

and $V_{1\pi}=o_p(1)\,O_p(1)\,(1+o_p(1))$ and $V_{2\pi}=o_p(1)\,O_p(1)$ by the assumptions, whereas the same arguments of Theorem (1), the triangle inequality and CLT imply that

$$V_{3\pi}\leq\left(\left\|\widehat{\beta}_\tau-\beta_{0\tau}\right\|+\sup_{X_{3i}\in\mathcal{X}_3}\left\|\widehat{\theta}_\tau''\left(X_{3i}\right)-\theta_\tau''\left(X_{3i}\right)\right\|\right)\left|\frac{1}{n^{1/2}}\sum_{i=1}^n\frac{\delta_i}{\pi\left(Z_{oi}\right)}X_{2i}\rho_\tau'\left(\varepsilon_i\right)\right|\kappa_2=o_p(1)\,O_p(1)\,.$$

For $\widehat{T}_{2\widehat{\pi}}$ again by the triangle inequality

$$\left|\widehat{T}_2-T_2\right|\leq\frac{1}{8}\sup_i\left|\widehat{f}_{\varepsilon_i|X_i}(0)-f_{\varepsilon_i|X_i}(0)\right|\left(\sup_{X_{3i}\in\mathcal{X}_3}\left\|\widehat{\theta}_\tau''\left(X_{3i}\right)-\theta_\tau''\left(X_{3i}\right)\right\|^2+1\right)\left\|\frac{1}{n}\sum_{i=1}^nX_{2i}^{\otimes2}\right\|\times$$
$$\left|\int\int t^2(t+s)^2K(t)\,K(t+s)\,dtds\right|+$$
$$\sup_{X_{3i}\in\mathcal{X}_3}\left\|\widehat{\theta}_\tau''\left(X_{3i}\right)-\theta_\tau''\left(X_{3i}\right)\right\|^2\left|\frac{1}{8n}\sum_{i=1}^nf_{\varepsilon_i|X_i}(0)\,X_{2i}^{\otimes2}\right|\left|\int\int t^2(t+s)^2K(t)\,K(t+s)\,dtds\right|+$$
$$\left|\frac{1}{8}\frac{1}{n}\sum_{i=1}^nf_{\varepsilon_i|X_i}(0)\,\theta_\tau''\left(X_{3i}\right)^TX_{2i}^{\otimes2}\theta_\tau''\left(X_{3i}\right)\int\int t^2(t+s)^2K(t)\,K(t+s)\,dtds-T_2\right|=o_p(1)$$

by the assumptions and LLN, and similarly for $\widehat{T}_{3\pi}$ and $\widehat{\sigma}^2_{\widehat{\pi}}$. ■

Proof of Proposition 12. We consider only the case $\theta_\tau^c=\widetilde{\theta}_\tau$; let $\widetilde{\phi}_\tau-\phi_{0\tau}=\left[\left(\widetilde{\beta}_\tau-\beta_{0\tau}\right)^T,\left(\widetilde{\theta}_\tau-\theta_{0\tau}\right)^T\right]^T$ and note that

$$D_{\widehat{\pi}}\left(\theta_\tau^c\right)=\sum_{i=1}^n\frac{\delta_i}{\pi_i}\left[\rho_\tau\left(\varepsilon_i-X_{2i}^T\left(\widehat{\theta}_{\tau-i}\left(X_{3i}\right)-\theta_{0\tau-i}\left(X_{3i}\right)\right)\right)-\rho_\tau\left(\varepsilon_i\right)\right]+$$
$$\sum_{i=1}^n\frac{\delta_i\left(\widehat{\pi}_i-\pi_{0i}\right)}{\widehat{\pi}_i\pi_{0i}}\left[\rho_\tau\left(\varepsilon_i-X_{2i}^T\left(\widehat{\theta}_{\tau-i}\left(X_{3i}\right)-\theta_{0\tau-i}\left(X_{3i}\right)\right)\right)-\rho_\tau\left(\varepsilon_i\right)\right]-$$
$$\sum_{i=1}^n\frac{\delta_i}{\pi_i}\left[\rho_\tau\left(\varepsilon_i-X_i^T\left(\widetilde{\phi}_\tau-\phi_{0\tau}\right)\right)-\rho_\tau\left(\varepsilon_i\right)\right]-$$
$$\sum_{i=1}^n\frac{\delta_i\left(\widehat{\pi}_i-\pi_{0i}\right)}{\widehat{\pi}_i\pi_{0i}}\left[\rho_\tau\left(\varepsilon_i-X_i^T\left(\widetilde{\phi}_\tau-\phi_{0\tau}\right)\right)-\rho_\tau\left(\varepsilon_i\right)\right]$$
$$=D_{3\pi}+D_{4\pi}+D_{5\pi}+D_{6\pi}.$$

By the same arguments of Theorem 9 $D_{3\pi} = D_{1\pi} + o_p(1)$ and $|D_{4\pi}| = o_p(1)$; for $D_{5\pi}$ (10.1), QAL and standard results on parametric quantile regression (Koenker & Machado 1999*b*) imply that

$$D_{5\pi} = n^{-1} \sum_{i=1}^{n} \frac{\delta_i}{\pi_i} X_i^T \rho'(\varepsilon_i) \left(E\left(f_{\varepsilon|X}(0) X^{\otimes 2} \right) \right)^{-1} \sum_{i=1}^{n} \frac{\delta_i}{\pi_i} X_i^T \rho'(\varepsilon_i) \xrightarrow{d} \sum_{j=1}^{k} \lambda_j \chi_j^2(1) = O_p(1),$$

where λ_j are the eigenvalues of the matrix $E\left(\tau(1-\tau) X^{\otimes 2}/\pi_0\right) \left(E\left(f_{\varepsilon|X}(0) X^{\otimes 2}\right)\right)^{-1}$. Finally, again by the same arguments used in Theorem 9 $|D_{6\pi}| = o_p(1)$, hence the conclusion follows as in Theorem 9 of Fan et al. (2001). ∎

Proof of Theorem 13. Note that under (5.2)

$$\begin{aligned} D_{\widehat{\pi}}(\theta_{0\tau}) = & \sum_{i=1}^{n} \frac{\delta_i}{\pi_i} \left(\rho_\tau \left(Y_i - X_{1i}^T \beta_{0\tau} - X_{3i}^T \widehat{\theta}_{\tau - i}(X_{3i}) \right) - \rho_\tau \left(Y_i - X_{1i}^T \beta_{0\tau} - X_{3i}^T \theta_{n\tau}(X_{3i}) \right) \right) - \\ & \sum_{i=1}^{n} \frac{\delta_i}{\pi_i} \left(\rho_\tau \left(Y_i - X_{1i}^T \beta_{0\tau} - X_{3i}^T \theta_{n\tau}(X_{3i}) \right) - \rho_\tau \left(Y_i - X_{1i}^T \beta_{0\tau} - X_{3i}^T \theta_{0\tau}(X_{3i}) \right) \right) + \\ & \sum_{i=1}^{n} \frac{\delta_i(\widehat{\pi}_i - \pi_{0i})}{\widehat{\pi}_i \pi_{0i}} \left(\rho_\tau \left(Y_i - X_{1i}^T \beta_{0\tau} - X_{3i}^T \widehat{\theta}_{\tau - i}(X_{3i}) \right) - \rho_\tau \left(Y_i - X_{1i}^T \beta_{0\tau} - X_{3i}^T \theta_{n\tau}(X_{3i}) \right) \right. \\ & \rho_\tau \left(Y_i - X_{1i}^T \beta_{0\tau} - X_{3i}^T \theta_{n\tau}(X_{3i}) \right) - \rho_\tau \left(Y_i - X_{1i}^T \beta_{0\tau} - X_{3i}^T \theta_{0\tau} \right)) + o_p(1) := \\ & D_{7\pi} + D_{8\pi} + D_{9\pi} + o_p(1), \end{aligned}$$

and that $\widehat{\theta}_{\tau-i}(\cdot)$ (centred at $\theta_{n\tau}(\cdot)$) admits the same asymptotic representation as that given in (10.21). For $D_{7\pi}$, the same arguments as those used in the proof of Theorem 9 show that

$$D_{7\pi} = U_\pi - \frac{nh^4}{8} E\left[f_{\varepsilon|X}(0) \gamma_{n\tau}''(X_3)^T X_2^{\otimes 2} \gamma_{n\tau}''(X_3) \right] \int \int t^2 (t+s)^2 K(t) K(t+s)\, dtds + o_p\left(h^{-1/2} \right).$$

For $D_{8\pi}$, (10.1) shows that

$$\begin{aligned} D_{8\pi} = & -\sum_{i=1}^{n} \frac{\delta_i}{\pi_{0i}} X_{2i}^T \gamma_{n\tau}(X_{3i}) \rho_\tau'(\varepsilon_i) + \sum_{i=1}^{n} \int_0^{X_{2i}^T \gamma_{n\tau}(X_{3i})} \frac{\delta_i}{\pi_{0i}} \left(I(\varepsilon_i \le t) - I(\varepsilon_i \le 0) \right) dt = \\ & -\sum_{i=1}^{n} \frac{\delta_i}{\pi_{0i}} X_{2i}^T \gamma_{n\tau}(X_{3i}) \rho_\tau'(\varepsilon_i) + \frac{n}{2} E\left(f_{\varepsilon|X}(0) \gamma_n(X_3)^T X_2^{\otimes 2} \gamma_n(X_3) \right), \end{aligned}$$

and finally, similarly to (10.23), $D_{9\pi} = o_p(1)$. Thus

$$\begin{aligned} D_\pi(\theta_{0\tau}) = & U_\pi - \sum_{i=1}^{n} \frac{\delta_i}{\pi_{0i}} X_{2i}^T \gamma_{n\tau}(X_{3i}) \rho_\tau'(\varepsilon_i) + \frac{n}{2} E\left(f_{\varepsilon|X}(0|X) \gamma_{n\tau}(X_3)^T X_2^{\otimes 2} \gamma_{n\tau}(X_3) \right) \\ & -\frac{nh^4}{8} E\left[f_{\varepsilon|X}(0) \gamma_{n\tau}''(X_3)^T X_2^{\otimes 2} \gamma_{n\tau}''(X_3) \right] \int \int t^2 (t+s)^2 K(t) K(t+s)\, dtds + o_p\left(h^{-1/2} \right) + o_p(1), \end{aligned}$$

and the first conclusion follows by the same arguments as those used in the proof of Theorem 9, noting that

$$Var\left(\sum_{i=1}^{n} \frac{\delta_i}{\pi_{0i}} X_{2i}^T \gamma_{n\tau}(X_{3i}) \rho_\tau'(\varepsilon_i) \right) = nE\left(\frac{\tau(1-\tau)}{\pi_0} X_2^T \gamma_{n\tau}(X_3)^{2\otimes} X_2 \right).$$

The consistency of $D_{\widehat{\pi}}(\theta_{0\tau})$ follows directly from the assumption that

$$nhE\left(f_{\varepsilon|X}(0)\,\gamma_{n\tau}(X_3)^T X_2^{\otimes 2}\gamma_{n\tau}(X_3)\right)\to\infty.$$

■

Proof of Proposition 14. The same arguments as those used in Theorem 3 and CMT show that

$$n^{1/2}R\left(\widehat{\beta}_\tau-\beta_{0\tau}\right)\xrightarrow{d}N\left(\gamma_\tau, R\Sigma_2^{-1}\Sigma_{2*}\Sigma_2^{-1}R^T\right)$$

hence the first conclusion follows by standard results on quadratic forms in non zero mean Normal random vectors. The consistency of W under the assumption that $n^{1/2}\gamma_{\tau n}\to\infty$ is a direct consequence of the previous conclusion. ■

10.2 Additional simulations results

This section considers the additional case where only the responses are missing. The missing mechanism is specified as

$$\pi_0(Z_{oi}) = \frac{\exp(\alpha_{10} + \alpha_{20}X_{11i} + \alpha_{30}X_{21i} + \alpha_{40}X_{22i} + \alpha_{50}X_{3i})}{1 + \exp(\alpha_{10} + \alpha_{20}X_{11i} + \alpha_{30}X_{21i} + \alpha_{40}X_{22i} + \alpha_{50}X_{3i})} \tag{10.24}$$

and as in the main paper the percentage of missing at the τ quantile are chosen to be at approximately 10% and 40%.

Table 8a $\varepsilon_\tau \sim N(0,1)$, $\tau = 0.25$

n	100				400			
	bias	*se*	*length*	*cov*	*bias*	*se*	*length*	*cov*
	MAR (10.24) 10%				MAR (10.24) 10%			
$\widehat{\beta}_{1\tau c}$	.090	.181	.419	.890	.073	.094	.254	.903
$\widehat{\beta}_{2\tau c}$	.105	.843	.895	.898	.083	.438	.484	.905
$\widehat{\beta}^p_{1\tau c}$	.088	.190	.423	.892	.078	.100	.253	.890
$\widehat{\beta}^p_{2\tau c}$	.110	.869	.902	.891	.085	.440	.490	.902
$\widehat{\beta}_{1\tau p}$	.030	.189	.428	.941	.015	.095	.220	.943
$\widehat{\beta}_{2\tau p}$	.070	.825	.903	.943	.032	.448	.490	.945
$\widehat{\beta}^p_{1\tau p}$	.030	.192	.910	.945	.020	.450	.488	.948
$\widehat{\beta}^p_{2\tau p}$	.075	.190	.438	.947	.040	.440	.483	.947
$\widehat{\beta}_{1\tau np}$	.030	.193	.431	.942	.016	.099	.227	.943
$\widehat{\beta}_{2\tau np}$	.072	.830	.912	.942	.036	.495	.493	.944
$\widehat{\beta}^p_{1\tau np}$	.035	.200	.433	.946	.020	.095	.225	.947
$\widehat{\beta}^p_{2\tau np}$	.080	.840	.920	.950	.040	.473	.478	.948
	MAR (10.24) 40%				MAR (10.24) 40%			
$\widehat{\beta}_{1\tau c}$	.110	.190	.415	.882	.105	.124	.260	.886
$\widehat{\beta}_{2\tau c}$	.120	.890	.917	.880	.110	.481	.510	.901
$\widehat{\beta}^p_{1\tau c}$	.129	.210	.457	.880	.101	.120	.253	.888
$\widehat{\beta}^p_{2\tau c}$	.110	.869	.902	.891	.110	.475	.505	.902
$\widehat{\beta}_{1\tau p}$	.035	.201	.431	.942	.025	.095	.220	.944
$\widehat{\beta}_{2\tau p}$	.079	.838	.905	.945	.030	.445	.481	.946
$\widehat{\beta}^p_{1\tau p}$	.038	.196	.912	.946	.021	.448	.490	.947
$\widehat{\beta}^p_{2\tau p}$	.078	.830	.810	.946	.038	.438	.480	.946
$\widehat{\beta}_{1\tau np}$	.040	.205	.445	.943	.018	.097	.226	.942
$\widehat{\beta}_{2\tau np}$	.080	.846	.916	.945	.035	.493	.490	.942
$\widehat{\beta}^p_{1\tau np}$	.037	.205	.437	.947	.021	.093	.223	.946
$\widehat{\beta}^p_{2\tau np}$	.085	.843	.923	.952	.039	.470	.473	.946

Table 8b $\varepsilon_\tau \sim N(0,1)$, $\tau = 0.50$

n	100				400			
	bias	*se*	*length*	*cov*	*bias*	*se*	*length*	*cov*
	MAR (10.24) 10%				MAR (10.24) 10%			
$\widehat{\beta}_{1\tau c}$	.096	.170	.390	.894	.075	.100	.170	.900
$\widehat{\beta}_{2\tau c}$	.101	.787	.850	.896	.082	.530	.452	.901
$\widehat{\beta}^p_{1\tau c}$	.090	.164	.386	.897	.078	.102	.172	.696
$\widehat{\beta}^p_{2\tau c}$	.094	.778	.835	.893	.081	.515	.445	.905
$\widehat{\beta}_{1\tau p}$	.038	.170	.400	.943	.024	.080	.215	.943
$\widehat{\beta}_{2\tau p}$	.031	.790	.885	.944	.025	.350	.465	.945
$\widehat{\beta}^p_{1\tau p}$	.040	.168	.832	.945	.025	.086	.215	.943
$\widehat{\beta}^p_{2\tau p}$	.040	.164	.930	.945	.030	.363	.210	.943
$\widehat{\beta}_{1\tau np}$	.041	.175	.400	.943	.033	.087	.210	.943
$\widehat{\beta}_{2\tau np}$	.033	.785	.890	.942	.025	.368	.472	.943
$\widehat{\beta}_{1\tau np}$	.043	.173	.401	.944	.035	.096	.476	.942
$\widehat{\beta}^p_{2\tau np}$	.035	.779	.405	.945	.025	.365	.210	.943
	MAR (10.24) 40%				MAR (10.24) 40%			
$\widehat{\beta}_{1\tau c}$	.121	.185	.399	.881	.112	.129	.199	.894
$\widehat{\beta}_{2\tau c}$	.128	.805	.836	.883	.109	.596	.503	.898
$\widehat{\beta}^p_{1\tau c}$	.125	.189	.403	.884	.110	.131	.202	.893
$\widehat{\beta}^p_{2\tau c}$	.130	.832	.841	.889	.112	.584	.509	.896
$\widehat{\beta}_{1\tau p}$	.045	.183	.407	.941	.030	.091	.224	.942
$\widehat{\beta}_{2\tau p}$	.038	.792	.841	.940	.028	.378	.593	.943
$\widehat{\beta}^p_{1\tau p}$	.044	.181	.400	.942	.031	.087	.221	.942
$\widehat{\beta}^p_{2\tau p}$	.036	.799	.843	.942	.027	.375	.590	.943
$\widehat{\beta}_{1\tau np}$	.047	.184	.403	.942	.032	.094	.219	.944
$\widehat{\beta}_{2\tau np}$	.040	.801	.883	.942	.031	.396	.496	.943
$\widehat{\beta}^p_{1\tau np}$	.048	.187	.405	.941	.033	.090	.212	.948
$\widehat{\beta}^p_{2\tau np}$	.041	.803	.875	.943	.030	.399	.491	.944

Table 8c $\varepsilon_\tau \sim N(0,1)$, $\tau = 0.75$

n	100				400			
	bias	*se*	*length*	*cov*	*bias*	*se*	*length*	*cov*
	MAR (10.24) 10%				MAR (10.24) 10%			
$\widehat{\beta}_{1\tau c}$	.110	.209	.542	.888	.088	.138	.284	.898
$\widehat{\beta}_{2\tau c}$	.105	.803	.833	.892	.085	.501	.449	.896
$\widehat{\beta}^p_{1\tau c}$	.112	.214	.548	.886	.091	.132	.280	.895
$\widehat{\beta}^p_{2\tau c}$	.105	.798	.836	.895	.085	.494	.445	.893
$\widehat{\beta}_{1\tau p}$	.055	.208	.418	.943	.030	.110	.238	.946
$\widehat{\beta}_{2\tau p}$	.068	.808	.836	.942	.036	.406	.439	.944
$\widehat{\beta}^p_{1\tau p}$	.056	.210	.421	.946	.031	.105	.227	.943
$\widehat{\beta}^p_{2\tau p}$	.070	.203	.832	.945	.034	.102	.431	.943
$\widehat{\beta}_{1\tau np}$	.050	.210	.460	.942	.029	.112	.241	.944
$\widehat{\beta}_{2\tau np}$	.069	.811	.836	.944	.035	.403	.433	.945
$\widehat{\beta}^p_{1\tau np}$	.051	.205	.454	.941	.030	.109	.243	.946
$\widehat{\beta}^p_{2\tau np}$	.065	.809	.832	.947	.034	.399	.436	.947
	MAR (10.24) 40%				MAR (10.24) 40%			
$\widehat{\beta}_{1\tau c}$	.118	.228	.501	.880	.102	.181	.303	.889
$\widehat{\beta}_{2\tau c}$	.110	.822	.890	.885	.099	.452	.491	.895
$\widehat{\beta}^p_{1\tau c}$	.121	.225	.497	.878	.110	.179	.312	.887
$\widehat{\beta}^p_{2\tau c}$	.111	.819	.887	.880	.102	.447	.490	.894
$\widehat{\beta}_{1\tau p}$	.058	.210	.421	.943	.029	.118	.263	.941
$\widehat{\beta}_{2\tau p}$	.072	.818	.842	.945	.034	.429	.442	.946
$\widehat{\beta}^p_{1\tau p}$	.060	.208	.419	.942	.031	.110	.256	.943
$\widehat{\beta}^p_{2\tau p}$	.070	.809	.840	.943	.033	.420	.440	.945
$\widehat{\beta}_{1\tau np}$	.059	.213	.456	.945	.031	.113	.244	.946
$\widehat{\beta}_{2\tau np}$	.071	.817	.819	.943	.035	.407	.436	.945
$\widehat{\beta}^p_{1\tau np}$	.061	.210	.449	.947	.030	.110	.232	.944
$\widehat{\beta}^p_{2\tau np}$	.073	.815	.817	.942	.036	.401	.435	.943

.

Table 9a $\varepsilon_\tau \sim t\,(5)$, $\tau = 0.25$

n	100				400			
	bias	*se*	*length*	*cov*	*bias*	*se*	*length*	*cov*
	MAR (10.24) 10%				MAR (10.24) 10%			
$\widehat{\beta}_{1\tau c}$	.094	.195	.420	.898	.084	.110	.224	.896
$\widehat{\beta}_{2\tau c}$	.092	.832	.870	.898	.077	.510	.496	.891
$\widehat{\beta}^p_{1\tau c}$	.092	.198	.418	.893	.080	.105	.220	.891
$\widehat{\beta}^p_{2\tau c}$	.095	.834	.875	.897	.075	.505	.493	.893
$\widehat{\beta}_{1\tau p}$	.043	.200	.425	.941	.024	.101	.210	.945
$\widehat{\beta}_{2\tau p}$	.040	.812	.882	.942	.026	.424	.430	.944
$\widehat{\beta}^p_{1\tau p}$	.045	.197	.428	.941	.022	.100	.210	.945
$\widehat{\beta}^p_{2\tau p}$	.043	.815	.888	.943	.026	.417	.426	.947
$\widehat{\beta}_{1\tau np}$	.042	.204	.424	.941	.020	.106	.210	.945
$\widehat{\beta}_{2\tau np}$	.044	.820	.884	.941	.021	.504	.432	.944
$\widehat{\beta}^p_{1\tau np}$	.045	.204	.434	.942	.022	.100	.206	.945
$\widehat{\beta}^p_{2\tau np}$	.044	.821	.885	.943	.023	.496	.431	.944
	MAR (10.24) 40%				MAR (10.24) 40%			
$\widehat{\beta}_{1\tau c}$	.125	.219	.455	.878	.080	.115	.228	.817
$\widehat{\beta}_{2\tau c}$	.124	.850	.882	.872	.079	.502	.414	.882
$\widehat{\beta}^p_{1\tau c}$	.129	.215	.450	.878	.084	.113	.230	.888
$\widehat{\beta}^p_{2\tau c}$	.127	.845	.880	.872	.083	.508	.410	.885
$\widehat{\beta}_{1\tau p}$	.043	.216	.428	.941	.026	.116	.231	.944
$\widehat{\beta}_{2\tau p}$	.044	.818	.880	.944	.027	.440	.448	.942
$\widehat{\beta}^p_{1\tau p}$	.043	.208	.429	.945	.028	.112	.230	.946
$\widehat{\beta}^p_{2\tau p}$	.045	.810	.878	.944	.030	,434	.433	.944
$\widehat{\beta}_{1\tau np}$	.043	.209	.437	.942	.024	.110	.230	.943
$\widehat{\beta}_{2\tau np}$	.045	.825	.886	.943	.024	.508	.450	.945
$\widehat{\beta}^p_{1\tau np}$	.043	.204	.430	.944	,022	.110	.227	.945
$\widehat{\beta}^p_{2\tau np}$	.044	.823	.885	.943	.028	.505	.444	.944

Table 9b $\varepsilon_\tau \sim t\,(5)\,,\ \tau = 0.50$

n	100				400			
	bias	*se*	*length*	*cov*	*bias*	*se*	*length*	*cov*
	MAR (10.24) 10%				MAR (10.24) 10%			
$\widehat{\beta}_{1\tau c}$	.100	.182	.410	.901	.074	.100	.212	.902
$\widehat{\beta}_{2\tau c}$	.093	.814	.880	.895	.068	.495	.414	.901
$\widehat{\beta}^p_{1\tau c}$	.108	.176	.410	.902	.075	.098	.210	.903
$\widehat{\beta}^p_{2\tau c}$	.095	.810	.878	.890	.065	.492	.410	.902
$\widehat{\beta}_{1\tau p}$	.050	.194	.415	.943	.030	.094	.216	.945
$\widehat{\beta}_{2\tau p}$	.045	.778	.863	.942	.031	.391	.428	.946
$\widehat{\beta}^p_{1\tau p}$	.053	.190	.412	.946	.030	.090	.210	.947
$\widehat{\beta}^p_{2\tau p}$	.048	.779	.862	.944	.032	.380	.424	.947
$\widehat{\beta}_{1\tau np}$	.052	.200	.418	.942	.030	.095	.210	.945
$\widehat{\beta}_{2\tau np}$	.043	.793	.872	.943	.024	.384	.414	.945
$\widehat{\beta}^p_{1\tau np}$	.052	.198	.415	.946	.031	.090	.208	.946
$\widehat{\beta}^p_{2\tau np}$	.047	.784	.865	.946	.025	.379	.410	.947
	MAR (10.24) 40%				MAR (10.24) 40%			
$\widehat{\beta}_{1\tau c}$	.115	.220	.420	.885	.109	.125	.220	.896
$\widehat{\beta}_{2\tau c}$	.124	.872	.890	.885	.115	.480	.455	.898
$\widehat{\beta}^p_{1\tau c}$	.115	.215	.418	.880	.110	.126	.220	.897
$\widehat{\beta}^p_{2\tau c}$	.124	.870	.890	.889	.117	.480	.452	.898
$\widehat{\beta}_{1\tau p}$	.053	.195	.418	.945	.036	.103	.215	.941
$\widehat{\beta}_{2\tau p}$	.047	.790	.875	.947	.038	.410	.460	.943
$\widehat{\beta}^p_{1\tau p}$	.052	.192	.416	.948	.034	.097	.211	.943
$\widehat{\beta}^p_{2\tau p}$	.050	.787	.871	.947	.035	.411	.451	.946
$\widehat{\beta}_{1\tau np}$	.052	.202	.420	.945	.038	.105	.222	.941
$\widehat{\beta}_{2\tau np}$	.045	.801	.892	.947	.036	.420	.428	.946
$\widehat{\beta}^p_{1\tau np}$	.054	.200	.413	.946	.036	.095	.218	.944
$\widehat{\beta}^p_{2\tau np}$	.049	.801	.884	.948	.037	.415	.424	.944

Table 9c $\varepsilon_\tau \sim t\,(5)\,,\ \tau = 0.75$

n	100				400			
	bias	*se*	*length*	*cov*	*bias*	*se*	*length*	*cov*
	MAR (10.24) 10%				MAR (10.24) 10%			
$\widehat{\beta}_{1\tau c}$	.110	.208	.401	.893	.087	.109	.210	.898
$\widehat{\beta}_{2\tau c}$	.105	.825	.870	.892	.070	.448	.426	.902
$\widehat{\beta}^p_{1\tau c}$	.110	.205	.400	.892	.088	.105	.206	.901
$\widehat{\beta}^p_{2\tau c}$	.103	.823	.873	.890	.074	.448	.420	.895
$\widehat{\beta}_{1\tau p}$	.054	.212	.410	.944	.034	.114	.214	.946
$\widehat{\beta}_{2\tau p}$	.070	.833	.879	.943	.029	.455	.434	.947
$\widehat{\beta}^p_{1\tau p}$	.058	.212	.412	.942	.039	.109	.211	.946
$\widehat{\beta}^p_{2\tau p}$	.072	.832	.875	.943	.031	.450	.429	.946
$\widehat{\beta}_{1\tau np}$	.055	.216	.412	.942	.039	.117	.217	.943
$\widehat{\beta}_{2\tau np}$	.074	.834	.884	.943	.030	.458	.499	.946
$\widehat{\beta}^p_{1\tau np}$	.062	.212	.414	.941	.037	.114	.214	.944
$\widehat{\beta}^p_{2\tau np}$	.070	.825	.870	.944	.031	.452	.495	.956
	MAR (10.24) 40%				MAR (10.24) 40%			
$\widehat{\beta}_{1\tau c}$	.120	.234	.412	.881	.105	.121	.231	.890
$\widehat{\beta}_{2\tau c}$	.118	.893	.880	.882	.091	.455	.433	.888
$\widehat{\beta}^p_{1\tau c}$	.125	.228	.409	.880	.107	.118	.229	.893
$\widehat{\beta}^p_{2\tau c}$	.119	.893	.883	.882	.092	.449	.429	.886
$\widehat{\beta}_{1\tau p}$	.060	.219	.419	.942	.039	.124	.220	.944
$\widehat{\beta}_{2\tau p}$	.079	.839	.888	.938	.032	.457	.435	.943
$\widehat{\beta}^p_{1\tau p}$	.062	.214	.415	.943	.041	.120	.218	.943
$\widehat{\beta}^p_{2\tau p}$	.077	.837	.889	.942	.033	.455	.437	.944
$\widehat{\beta}_{1\tau np}$	.053	.220	.423	.944	.040	.125	.220	.943
$\widehat{\beta}_{2\tau np}$	.076	.840	.886	.942	.034	.459	.441	.942
$\widehat{\beta}^p_{1\tau np}$	.053	.218	.419	.943	.042	.122	.219	.942
$\widehat{\beta}^p_{2\tau np}$	.075	.830	.882	.942	.036	.452	.439	.943

Table 10a $\varepsilon_\tau \sim \chi^2(4) - 4$, $\tau = 0.25$

n	100				400			
	bias	*se*	*length*	*cov*	*bias*	*se*	*length*	*cov*
	MAR (10.24) 10%				MAR (10.24) 10%			
$\widehat{\beta}_{1\tau c}$	.095	.190	.420	.902	.088	.105	.235	.904
$\widehat{\beta}_{2\tau c}$	.096	.836	.903	.899	.092	.446	.475	.902
$\widehat{\beta}^p_{1\tau c}$	.095	.198	.419	.903	.082	.105	.234	.903
$\widehat{\beta}^p_{2\tau c}$	.097	.836	.903	.902	.092	.444	.474	.905
$\widehat{\beta}_{1\tau p}$	.045	.200	.425	.944	.033	.105	.220	.946
$\widehat{\beta}_{2\tau p}$	.043	.806	.855	.943	.032	.452	.470	.945
$\widehat{\beta}^p_{1\tau p}$	.052	.193	.425	.946	.030	.107	.223	.944
$\widehat{\beta}^p_{2\tau p}$	.043	.798	.852	.947	.035	.449	.466	.945
$\widehat{\beta}_{1\tau np}$	.044	.239	.438	.945	.030	.105	.233	.943
$\widehat{\beta}_{2\tau np}$	.045	.814	.850	.943	.033	.454	.473	.943
$\widehat{\beta}^p_{1\tau np}$	.046	.230	.428	.945	.036	.108	.235	.944
$\widehat{\beta}^p_{2\tau np}$	.055	.804	.854	.945	.038	.448	.475	.945
	MAR (10.24) 40%				MAR (10.24) 40%			
$\widehat{\beta}_{1\tau c}$	.118	.214	.425	.888	.105	.115	.250	.890
$\widehat{\beta}_{2\tau c}$	.124	.850	.886	.892	.112	.465	.494	.894
$\widehat{\beta}^p_{1\tau c}$	.123	.214	.420	.890	.113	.118	.250	.897
$\widehat{\beta}^p_{2\tau c}$	.124	.850	.890	.892	.104	.460	.490	.898
$\widehat{\beta}_{1\tau p}$	.046	.223	.430	.943	.033	.120	.255	.944
$\widehat{\beta}_{2\tau p}$	.042	.854	.862	.944	.035	.460	.485	.945
$\widehat{\beta}^p_{1\tau p}$	.052	.213	.432	.944	.035	.117	.250	.945
$\widehat{\beta}^p_{2\tau p}$	.043	.845	.850	.942	.037	.460	.480	.946
$\widehat{\beta}_{1\tau np}$	.045	.220	.433	.943	.033	.115	.260	.944
$\widehat{\beta}_{2\tau np}$	.052	.860	.855	.942	.035	.470	.495	.945
$\widehat{\beta}^p_{1\tau np}$	.043	.222	.432	.944	.036	.110	.262	.944
$\widehat{\beta}^p_{2\tau np}$	.050	.862	.853	.945	.038	.473	.492	.946

Table 10b $\varepsilon_\tau \sim \chi^2(4) - 4$, $\tau = 0.5$

n	100				400				
	bias	*se*	*length*	*cov*	*bias*	*se*	*length*	*cov*	
	MAR (10.24) 10%				MAR (10.24) 10%				
$\widehat{\beta}_{1\tau c}$	.094	.204	.447	.903	.089	.116	.230	.901	
$\widehat{\beta}_{2\tau c}$	.093	.830	.859	.902	.078	.448	.491	.903	
$\widehat{\beta}^p_{1\tau c}$	.102	.198	.441	.904	.085	.113	.227	.903	
$\widehat{\beta}^p_{2\tau c}$	.095	.835	.864	.902	.073	.432	.482	.900	
$\widehat{\beta}_{1\tau p}$	.052	.207	.419	.942	.034	.111	.230	.944	
$\widehat{\beta}_{2\tau p}$	.043	.830	.868	.943	.041	.448	.494	.945	
$\widehat{\beta}^p_{1\tau p}$	.058	.199	.413	.945	.038	.108	.234	.944	
$\widehat{\beta}^p_{2\tau p}$	.045	.833	.860	.946	.041	.445	.489	.946	
$\widehat{\beta}_{1\tau np}$	.054	.200	.424	.942	.036	.114	.239	.945	
$\widehat{\beta}_{2\tau np}$	.046	.832	.865	.943	.042	.434	.493	.942	
$\widehat{\beta}^p_{1\tau np}$	.057	.198	.414	.941	.034	.108	.236	.945	
$\widehat{\beta}^p_{2\tau np}$	.045	.829	.860	.944	.043	.435	.493	.946	
	MAR (10.24) 40%				MAR (10.24) 40%				
$\widehat{\beta}_{1\tau c}$	.123	.209	.435	.880	.106	.134	.259	.891	
$\widehat{\beta}_{2\tau c}$	.128	.845	.872	.884	.108	.492	.496	.892	.105
$\widehat{\beta}^p_{1\tau c}$	.132	.204	.436	.889	.132	.257	.897		
$\widehat{\beta}^p_{2\tau c}$	.129	.847	.874	.891	.109	.491	.499	.901	
$\widehat{\beta}_{1\tau p}$	.059	.211	.438	.946	.041	.123	.264	.945	
$\widehat{\beta}_{2\tau p}$	.049	.844	.874	.945	.039	.463	.504	.946	
$\widehat{\beta}^p_{1\tau p}$	.054	.209	.434	.948	.042	.118	.261	.946	
$\widehat{\beta}^p_{2\tau p}$	.049	.844	.869	.946	.041	.460	.495	.947	
$\widehat{\beta}_{1\tau np}$	.055	.213	.431	.947	.081	.129	.272	.948	
$\widehat{\beta}_{2\tau np}$	.050	.842	.875	.943	.041	.470	.507	.946	
$\widehat{\beta}^p_{1\tau np}$	.057	.200	.419	.946	.083	.121	.269	.943	
$\widehat{\beta}^p_{2\tau np}$	.052	.832	.871	.945	.045	.473	.503	.942	

Table 10c $\varepsilon_\tau \sim \chi^2(4) - 4$, $\tau = 0.75$

n	100				400			
	bias	*se*	*length*	*cov*	*bias*	*se*	*length*	*cov*
	MAR (10.24) 10%				MAR (10.24) 10%			
$\widehat{\beta}_{1\tau c}$	.103	.209	.421	.896	.085	.118	.216	.902
$\widehat{\beta}_{2\tau c}$	.095	.863	.872	.895	.081	.469	.485	.900
$\widehat{\beta}^p_{1\tau c}$	.106	.208	.416	.899	.086	.108	.209	.905
$\widehat{\beta}^p_{2\tau c}$	.097	.859	.869	.903	.084	.465	.479	.904
$\widehat{\beta}_{1\tau p}$	.059	.204	.414	.943	.039	.117	.218	.945
$\widehat{\beta}_{2\tau p}$	.048	.865	.864	.944	.030	.454	.422	.947
$\widehat{\beta}^p_{1\tau p}$	.061	.208	.413	.942	.036	.115	.215	.945
$\widehat{\beta}^p_{2\tau p}$	.058	.862	.865	.946	.038	.449	.413	.946
$\widehat{\beta}_{1\tau np}$	.055	.215	.424	.944	.036	.128	.208	.944
$\widehat{\beta}_{2\tau np}$	.052	.868	.875	.943	.035	.459	.489	.946
$\widehat{\beta}^p_{1\tau np}$	.061	.209	.419	.948	.039	.124	.208	.943
$\widehat{\beta}^p_{2\tau np}$	.046	.863	.876	.945	.038	.454	.482	.947
	MAR (10.24) 40%				MAR (10.24) 40%			
$\widehat{\beta}_{1\tau c}$	.133	.218	.425	.895	.108	.139	.242	.894
$\widehat{\beta}_{2\tau c}$	.125	.875	.878	.897	.109	.478	.496	.900
$\widehat{\beta}^p_{1\tau c}$	.134	.215	.421	.885	.107	.134	.238	.899
$\widehat{\beta}^p_{2\tau c}$	.132	.891	.873	.899	.103	.474	.499	.903
$\widehat{\beta}_{1\tau p}$	.065	.218	.428	.944	.040	.118	.245	.942
$\widehat{\beta}_{2\tau p}$	.052	.879	.882	.947	.038	.469	.504	.946
$\widehat{\beta}^p_{1\tau p}$	.060	.214	.424	.945	.041	.115	.239	.947
$\widehat{\beta}^p_{2\tau p}$	.054	.874	.879	.946	.035	.465	.496	.945
$\widehat{\beta}_{1\tau np}$	.063	.219	.432	.942	.039	.121	.248	.942
$\widehat{\beta}_{2\tau np}$	.053	.872	.884	.944	.043	.461	.502	.940
$\widehat{\beta}^p_{1\tau np}$	.060	.213	.428	.946	.036	.119	.235	.948
$\widehat{\beta}^p_{2\tau np}$	.053	.861	.878	.943	.040	.465	.499	.947