centre for microdata methods and practice

Supplement for "A quantile correlated random coefficients panel data model"

Bryan S. Graham
Jinyong Hahn
Alexandre Poirier
James L. Powell

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP34/16

An ESRC Research Centre

Supplemental appendix to A Quantile Correlated Random Coefficients Panel Data Model

Bryan S. Graham[◇], Jinyong Hahn[♮], Alexandre Poirier[†] and James L. Powell[◇*]

August 19, 2016

*Earlier versions of this paper, with an initial draft date of March 2008, were presented under a variety of titles. We would like to thank seminar participants at Berkeley, CEMFI, Duke, UIUC, University of Michigan, Université de Montréal, NYU, Northwestern and at the 2009 North American Winter Meetings of the Econometric Society, the 2009 All-California Econometrics Conference at UC - Riverside, the 2014 Midwest Econometrics Group at the University of Iowa, and the 2nd annual IAAE Conference. We also thank the co-editors and two anonymous referees for their comments and insights. Financial support from the National Science Foundation (SES #0921928) is gratefully acknowledged. All the usual disclaimers apply.

[◇]Department of Economics, University of California - Berkeley, 530 Evans Hall #3880, Berkeley, CA 94720. E-MAIL: bgraham@econ.berkeley.edu, powell@econ.berkeley.edu.

[♮]Department of Economics, University of California - Los Angeles, Box 951477, Los Angeles, CA 90095-1477. E-MAIL: hahn@econ.ucla.edu.

[†]Department of Economics, University of Iowa, S252 John Pappajohn Business Building, Iowa City, IA 52242. E-MAIL: alexandre-poirier@uiowa.edu.

In section 1 of the supplemental appendix, we expand on the connection between our work and that of two recent papers. In section 2, we discuss further some extensions mentioned in section 5 of the original paper. Finally, in sections 3 and 4 we develop the regular "just-identified" ($T = P$) case and prove its results. Theorem and assumption numbering in the supplemental appendix follow that of the main text.

1 Connections to prior work

Our paper is not the first to explore the intersection of quantile regression and panel data. Here we briefly touch on the relationship between our work and two very recent contributions, those of Chernozhukov, Fernández-Val, Hahn and Newey (2013), CFHN for short, and of Arellano and Bonhomme (2016).

CFHN includes some results on quantile effects in the context of a wide-ranging analysis of identification in non-separable panel data models. The simplest case allowing for interesting comparisons between their results and our own is when $P = 2$ with the non-constant covariate binary-valued (i.e., $X_t = (1, X_{2t})'$). For simplicity we ignore time effects in what follows so that the conditional distribution of Y_t given $\mathbf{X} = \mathbf{x}$ is stationary over time and, if (3) is also being maintained, so is the distribution of random coefficients.

Following the notation of CFHN let $T_i(x) = \sum_{t=1}^{T} \mathbf{1}(X_{it} = x)$ and define

$$\bar{G}_i(y, x) = \begin{cases} T_i(x)^{-1} \sum_{t=1}^{T} \mathbf{1}(X_{it} = x) \mathbf{1}(Y_{it} \leq y), & T_i(x) > 0 \\ 0, & T_i(x) = 0 \end{cases}$$

and also

$$\hat{G}_M(y, x) = \frac{\sum_{i=1}^{N} \bar{G}_i(y, x) \mathbf{1}\left(\mathbf{X}_i \in \mathbb{X}^M\right)}{\sum_{i=1}^{N} \mathbf{1}(\mathbf{X}_i \in \mathbb{X}^M)}.$$

Their quantile treatment effect (QTE) estimate is

$$\hat{\lambda}(\tau) = \hat{G}_M^{-1}(\tau, 1) - \hat{G}_M^{-1}(\tau, 0).$$

To understand the relationship between $\hat{\lambda}(\tau)$ and our unconditional quantile effect (UQE) estimand further, we simplify to the case where $T = 2$. Let $\pi_{01} = \Pr\left(X_{21} = 0, X_{22} = 1 \middle| \mathbf{X} \in \mathbb{X}^M\right)$ and $\pi_{10} = \Pr\left(X_{21} = 1, X_{22} = 0 \middle| \mathbf{X} \in \mathbb{X}^M\right)$. Let $\hat{G}_M(y, x) \xrightarrow{p} G_M(y, x)$, under the random coefficients data generating process (3) we have that

$$\begin{aligned} G_M(y,1) &= \pi_{01} F_{B_1+B_2}\left(y \middle| X_{22}=1, \mathbf{X} \in \mathbb{X}^M\right) + \pi_{10} F_{B_1+B_2}\left(y \middle| X_{21}=1, \mathbf{X} \in \mathbb{X}^M\right) \\ G_M(y,0) &= \pi_{01} F_{B_1}\left(y \middle| X_{21}=0, \mathbf{X} \in \mathbb{X}^M\right) + \pi_{10} F_{B_1}\left(y \middle| X_{22}=0, \mathbf{X} \in \mathbb{X}^M\right), \end{aligned}$$

so that $\lambda(\tau) = G_M^{-1}(\tau,1) - G_M^{-1}(\tau,0)$ does not correspond to any quantile of B_2, even if our conditional monotonicity assumption holds. Indeed if we set $F_{B_1+B_2}\left(y \middle| X_{22}=1, \mathbf{X} \in \mathbb{X}^M\right) = F_{B_1}\left(y \middle| X_{22}=0, \mathbf{X} \in \mathbb{X}^M\right)$ and $F_{B_1+B_2}\left(y \middle| X_{21}=1, \mathbf{X} \in \mathbb{X}^M\right) = F_{B_1}\left(y \middle| X_{21}=0, \mathbf{X} \in \mathbb{X}^M\right)$ for all $y \in \mathbb{Y}_t$ we will have $\lambda(\tau)$ identically equal to zero for all $\tau \in (0,1)$ and our movers' UQE, $\beta_2^M(\tau)$, possibly different from zero for all $\tau \in (0,1)$. The difference between CFHN's estimand and our movers' UQE arises from how marginalization over $\mathbf{X}$ occurs. CFHN marginalize over $\mathbf{X}$ before inverting to recover quantile effects, we first recover quantile effects for each possible $\mathbf{x} \in \mathbb{X}^M$ and then 'marginalize'. Under our correlated random coefficients structure the CFHN estimand will, in general, not recover quantiles of B_2.

Arellano and Bonhomme (2016) study identification and estimation of the model

$$Q_{Y_t|\mathbf{X},A}(\tau|\,\mathbf{X},A) = X_t'\beta(\tau) + A\gamma(\tau) \tag{1}$$

for all $\tau \in (0,1)$. Here A corresponds to an unobserved time-invariant regressor. Arellano and Bonhomme (2016) also allow for a particular form of dependence between $\mathbf{X}$ and A, hence their model is a correlated random effects one. If A were observed, then $\theta(\tau) = \left(\beta(\tau)', \gamma(\tau)\right)'$ could be estimated by the τ^{th} linear pooled quantile regression of Y_t onto X_t and A (here the pooling is across all periods of data $t = 1, \ldots, T$). This set-up represents an alternative generalization of the linear panel model of Chamberlain (1984) to the quantile regression setting. Their approach and ours are non-nested.[1] The Arellano and Bonhomme (2016) model effectively includes two-dimensional unobserved heterogeneity. The first component of this heterogeneity vector is A, this component is allowed to covary with $\mathbf{X}$ is a reasonably flexible way. The second component corresponds to the common factor in the random coefficients on X_t and A. This component is independent of both $\mathbf{X}$ and A. Relatedly their set-up also requires a comonotonicity assumption on the random coefficients.

Our model, effectively, includes only a single dimension of unobserved heteroegeneity. How-

[1]Although their estimation procedure is mainly applied to quantile regression specifications like those of equation (1), their identification results are more general and include cases where the time-invariant regressor enters in a nonseparable way.

ever we impose weaker comonotonicity assumptions and leave the dependence structure between the heterogeneity and $\mathbf{X}$ nonparametric. Additionally, we impose restrictions on the number of covariates that can be included, while they do not require such restrictions. Also, additive separability of A and $\mathbf{X}$ is ruled out by our model. Our view is that both approaches are attractive, with the merits of each being context specific.

2 Extensions

Using stayers to estimate time effects when $T > P$

We describe a potential estimation procedure to use stayer realizations to estimate the overidentified (T>P) model.

Let $\mathbf{x}_l$ denote a stayer realization of $\mathbf{X}$ and consider the full rank decomposition $\mathbf{x}_l = \mathbf{u}_l \mathbf{v}_l$. Let $\mathbf{m}_{\mathbf{u}_l} = I_T - \mathbf{u}_l \left(\mathbf{u}_l' \mathbf{u}_l\right)^{-1} \mathbf{u}_l$ and observe that $\mathbf{m}_{\mathbf{u}_l} \mathbf{x} = \mathbf{u}_l \mathbf{v}_l - \mathbf{u}_l \left(\mathbf{u}_l' \mathbf{u}_l\right)^{-1} \mathbf{u}_l \mathbf{v}_l = 0$. We therefore have

$$\mathbf{m}_{\mathbf{u}_l} Q_{\mathbf{Y}|\mathbf{X}} \left(\tau|\mathbf{x}\right) \quad = \quad \mathbf{m}_{\mathbf{u}_l} \mathbf{w} \delta \left(\tau\right)$$

for $l = L+1, \ldots, M$. If we define $\bar{\mathbf{\Pi}} \left(\tau\right) = \left(\mathbf{\Pi} \left(\tau\right)', \mathbf{m}_{\mathbf{u}_{L+1}} Q_{\mathbf{Y}|\mathbf{X}} \left(\tau|\mathbf{x}\right)', \ldots, \mathbf{m}_{\mathbf{u}_M} Q_{\mathbf{Y}|\mathbf{X}} \left(\tau|\mathbf{x}\right)'\right)'$ and $\bar{G} = (G', (G^*)')'$, with G as defined in (25) and G^* equal to

$$\underset{T(M-L)\times(R+PL)}{G^*} = \begin{pmatrix} \mathbf{m}_{\mathbf{u}_{L+1}} \mathbf{w} & \underline{0}_T \underline{0}_P' & \cdots & \underline{0}_T \underline{0}_P' \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{m}_{\mathbf{u}_M} \mathbf{w} & \underline{0}_T \underline{0}_P' & \cdots & \underline{0}_T \underline{0}_P' \end{pmatrix},$$

we have the relationship $\bar{\mathbf{\Pi}} \left(\tau\right) = \bar{G} \gamma \left(\tau\right)$, upon which the obvious analog estimator may be based. In fact we incorporate stayers in this way in the empirical analysis reported in Section 4.

Non-shrinking mass of stayers in the irregular case

In some applications, it is common to observe a positive mass of stayers at $D = 0$ along with a small number of near-stayers. We can model this in our discrete bandwidth framework by letting $\Pr(D = 0) = \pi_0 + 2\phi_0 h_N$ where $\pi_0 > 0$ and keeping $\Pr(D = h_N) = \Pr(D = -h_N) = \phi_0 h_N$.

In this case, it will be possible to estimate $\delta(\tau)$ at a $\sqrt{N}$ rate since the mass of stayers is bounded away from 0. On the other hand, the conditional beta for near-stayer realizations will still be estimated at rate $\sqrt{Nh_N^3}$ since the slowest rate of convergence prevails. Their asymptotic variance will be reduced since the term associated with the estimation of $\delta(\tau)$ vanishes. Estimation of conditional betas for strict-mover realizations can now be performed at the $\sqrt{N}$ rate rather than $\sqrt{Nh_N}$.

Despite these improvements, the movers' ACQE now becomes inconsistent since the fraction of stayers does not vanish asymptotically. We can modify the estimator and include the near-stayers' coefficients as an approximation to the stayers' coefficients in the following way:

$$\hat{\hat{\beta}}_N(\tau) = \sum_{l=1}^{L} \beta(\tau; \mathbf{x}_{lN}) \hat{q}_{lN}^M \left\{ \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}(D_i \neq 0) \right\} + \sum_{l=L_1+1}^{L} \beta(\tau; \mathbf{x}_{lN}) \hat{q}_{l|\pm h} \left\{ \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}(D_i = 0) \right\},$$

where $\hat{q}_{l|\pm h} = \frac{\frac{1}{N}\sum_{i=1}^{N} \mathbf{1}(\mathbf{X}_i = \mathbf{x}_{lN})}{\frac{1}{N}\sum_{i=1}^{N} \mathbf{1}(|D_i| = h_N)}$. The term $\sum_{l=L_1+1}^{L} \beta(\tau; \mathbf{x}_{lN}) \hat{q}_{l|\pm h}$ approximates $\mathbb{E}[\beta(\tau; \mathbf{X}) | \mathbf{X} \in \mathbb{X}_N^S]$ using the near-stayers. The conditional beta for near-stayers is estimated at rate $\sqrt{Nh_N^3}$ but their contribution is now of order $O_p(1)$ rather than $O_p(h_N)$ in the previous case. Therefore, the convergence rate of the ACQE estimator will be $\sqrt{Nh_N^3}$ as well.

The same concern applies in the UQE estimation where the share of stayers does not vanish asymptotically. It is also possible to show that the rate of convergence will deteriorate from $\sqrt{Nh_N}$ to $\sqrt{Nh_N^3}$. The result in this case is analogous to that in Graham and Powell (2012) where the mass at $D = 0$ simplifies the estimation of the common coefficients while complicating the estimation of average partial effects.

In the case when there is a mass of stayers but no near-stayers, i.e. $\Pr(D = 0) = \pi_0$ and $\Pr(|D| = h_N) = 0$, it is not possible to identify the ACQE and UQE. This is a result similar to that found in the overidentified case, which does not contain near-stayers as well. Rates of convergence for the movers' ACQE and the movers' UQE will be of order $N^{-1/2}$, since there is no shrinking mass of near-stayers involved in the rate calculations. It is also possible to show the set identification of the ACQE and the UQE in a way similar to the overidentified case. We present the results for the just-identified case with no near-stayers in section 3 below.

Continuous regressors

The identification results in the previous section are all based on the conditional quantiles of Y_t given $\mathbf{X} = \mathbf{x}$ for $t = 1, \ldots, T$. There are different methods of estimating these conditional quantiles depending on the structure of the support of $\mathbf{X}$. When the support of $\mathbf{X}$ is discrete, as is assumed in this paper, we can write by standard arguments the following linear representation of our first step conditional quantile estimator:

$$\widehat{Q}_{Y_t|\mathbf{X}}\left(\tau|\,\mathbf{x}\right) - Q_{Y_t|\mathbf{X}}\left(\tau|\,\mathbf{x}\right) = \frac{1}{N}\sum_{i=1}^{N} \frac{\left(\mathbf{1}\left(Y_{it} \leq Q_{Y_t|\mathbf{X}}(\tau|\mathbf{x})\right) - \tau\right)\mathbf{1}(\mathbf{X}_i = \mathbf{x})}{f_{Y_t|\mathbf{X}}(Q_{Y_t|\mathbf{X}}(\tau|\mathbf{x})|\mathbf{x})p_N} + R_N(\tau;\mathbf{x})$$

where p_N is the probability associated with support point $\mathbf{x}$ and the residual $R_N(\tau;\mathbf{x})$ goes to 0 in probability as $N \to \infty$.

A kernel based estimator can be used to compute the analog estimator when $\mathbf{X}$ has continuous support. Examples include local linear regression (Qu and Yoon, 2015), or inversion of conditional CDF estimates (Lee, 2013). The Bahadur representation of these estimators is the following:

$$\begin{aligned}
&\widehat{Q}_{Y_t|\mathbf{X}}\left(\tau|\,\mathbf{x}\right) - Q_{Y_t|\mathbf{X}}\left(\tau|\,\mathbf{x}\right)\\
&= \frac{1}{Nb^{TP}}\sum_{i=1}^{N} \frac{\left(\mathbf{1}\left(Y_{it} \leq Q_{Y_t|\mathbf{X}}(\tau|\mathbf{x})\right) - \tau\right)K\left(B^{-1}\left(\mathrm{vec}(\mathbf{X}_i) - \mathrm{vec}(\mathbf{x})\right)\right)}{f_{Y_t|\mathbf{X}}(Q_{Y_t|\mathbf{X}}(\tau|\mathbf{x})|\mathbf{x})f_{\mathbf{X}}(\mathbf{x})} + R_N(\tau;\mathbf{x})
\end{aligned}$$

where $K(\cdot)$ is a kernel function, and $B = bI_{TP}$ where b is a bandwidth converging to 0.

These representations contain just two differences. The first difference is the presence of the kernel in the continuous case. We note that this kernel function converges to an indicator for $\{\mathbf{X}_i = \mathbf{x}\}$ whenever $K(0) = 1$. This restriction is satisfied by different kernels, including a uniform kernel with support $[-1/2, 1/2]$.

The second difference is the density of $\mathbf{X}$ versus its probability that it is equal to $\mathbf{x}$. The bandwidth term b^{TP} also is present with the density. These two terms can be reconciled if we consider the probability that $\mathbf{X}$ is included in a TP-dimensional neighborhood of $\mathbf{x}$ with diameter equal to b. For exposition, consider $T = P = 1$. Then, the probability that $\mathbf{X} \in [\mathbf{x} - b/2, \mathbf{x} + b/2]$ is approximately equal to $f_X(x)b$. This probability can be approximated with the mass p_N if this mass depends on the sample size through b such that $p_N = O(b)$. That way, discrete probabilities will behave like probabilities for continuous densities, and Np_N will be the approximate sample size for the support point, similar to the Nb in traditional nonparametric density estimation.

We see that there is no substantial difference between the asymptotic representations, except that it is much simpler to impose conditions on the asymptotic convergence of $\widehat{\mathbf{\Pi}}_t(\tau;\mathbf{x}) - \mathbf{\Pi}_t(\tau;\mathbf{x})$ in the discrete case than in the continuous case. For this reason, we conjecture that our rates of convergence and limit distribution results will generalize to the case of continuous regressors (under additional regularity conditions). Such an extension, however, is likely to be difficult and non-trivial.

Choice of Bandwidth

In the spirit of the argument given immediately above, we can use our discrete bandwidth asymptotics to gain some insight into bandwidth selection. As discussed earlier, we need that $h_N \to 0$, $Nh_N \to \infty$, and $Nh_N^3 \to 0$ as $N \to \infty$. These conditions allow us to put shrinking mass on some support points ($h_N \to 0$), while allowing for consistent estimation of objects conditional on these shrinking mass support points ($Nh_N \to \infty$). The third condition is required to eliminate the bias in our estimators for the ACQE and UQE. Let a be a $P \times 1$ vector of constants. The fastest rate of convergence in mean square for either $a'\hat{\bar{\beta}}(\tau)$ to $a'\bar{\beta}(\tau)$ or $\hat{\beta}_p(\tau)$ to $\beta_p(\tau)$ is $N^{-2/3}$ with bandwidth sequences of the form $h_N^* = C_0 N^{-1/3}$ for some constant C_0. Consider first the MSE for the estimator $a'\hat{\bar{\beta}}(\tau)$. Using the results of Theorem 6, the asymptotic bias will be equal to

$$\text{bias} = 2a'(\mathbb{E}[\beta(\tau;\mathbf{X})] - \mathbb{E}[\beta(\tau;\mathbf{X})|D=0])\phi_0 h_N$$

and therefore the asymptotic MSE minimizing bandwidth constant is

$$C_0 = \frac{1}{2}\left(\frac{1}{\phi_0^2}\right)^{1/3}\left(\frac{a'\left(\Upsilon_1(\tau,\tau) + \Xi_0\Sigma_\delta(\tau,\tau)\Xi_0'\right)a}{a'(\mathbb{E}[\beta(\tau;\mathbf{X})] - \mathbb{E}[\beta(\tau;\mathbf{X})|D=0])(\mathbb{E}[\beta(\tau;\mathbf{X})] - \mathbb{E}[\beta(\tau;\mathbf{X})|D=0])'a}\right)^{1/3}$$

which is similar to the one found in Graham and Powell (2012) for average partial effects. Since choosing bandwidth with order exactly equal to $N^{-1/3}$ leads to an asymptotic bias, we can choose a slightly faster bandwidth sequence, say, of order $o(N^{-1/3})$ such that the asymptotic bias disappears. An alternative is to use a plug-in bandwidth using an estimate of C_0 and then bias correct. This approach preserves the rate of convergence of $N^{-1/3}$.

We now consider the bias associated with the estimation of $\beta_p(\tau)$, the UQE associated with the p^{th} regressor. The leading term of the asymptotic bias of $\hat{\beta}_p(\tau)$ is equal to

$$\text{bias} = 2\phi_0 h_N \frac{F_{B_p}(\beta_p(\tau)) - F_{B_p|D=0}(\beta_p(\tau))}{f_{B_p}(\beta_p(\tau))}.$$

Once again, the MSE minimizing rate for the bandwidth is $N^{-1/3}$ and the mean squared error minimizing choice of constant is

$$C_0 = \frac{1}{2}\left(\frac{1}{\phi_0^2}\right)^{1/3}\left(\frac{\Upsilon_3(\tau,\tau)+\Upsilon_4(\tau,\tau)}{\left[F_{B_p}(\beta_p(\tau)) - F_{B_p|D=0}(\beta_p(\tau))\right]^2}\right)^{1/3}.$$

Trimming Extremal Quantiles

Suppose the researcher proceeds with the estimation of conditional quantiles ranging from $\tau = \epsilon$ to $\tau = 1-\epsilon$ so that only these conditional quantiles are identified. We can decompose the moment condition used for the estimation of $\beta_p(\tau)$ for $p = 1, \ldots, P$ in three parts:

$$\begin{aligned}
\tau &= m(b) \\
&= \mathbb{E}\left[\int_0^1 \mathbf{1}(\beta_p(u;\mathbf{X}) \le b)du\right] \\
&= \mathbb{E}\left[\int_0^\epsilon \mathbf{1}(\beta_p(u;\mathbf{X}) \le b)du\right] + \mathbb{E}\left[\int_\epsilon^{1-\epsilon} \mathbf{1}(\beta_p(u;\mathbf{X}) \le b)du\right] + \mathbb{E}\left[\int_{1-\epsilon}^1 \mathbf{1}(\beta_p(u;\mathbf{X}) \le b)du\right].
\end{aligned}$$

The first and third term are uniformly bounded below and above by 0 and ϵ respectively. Therefore,

$$\tau - 2\epsilon \le \mathbb{E}\left[\int_\epsilon^{1-\epsilon} \mathbf{1}(\beta_p(u;\mathbf{X}) \le b)du\right] \le \tau.$$

Let V be uniformly distributed on $[\epsilon, 1-\epsilon]$ independently from $\mathbf{X}$. We can think of this V as the trimmed version of the unobserved heterogeneity U. Then, we see that

$$\frac{\tau - 2\epsilon}{1-2\epsilon} \le F_{\beta_p(V;\mathbf{X})}(b) \le \frac{\tau}{1-2\epsilon},$$

and therefore, the UQE for the p^{th} regressor is partially identified with bounds equal to

$$\beta_p(\tau) \in \left[F^{-1}_{\beta_p(V;\mathbf{X})}\left(\frac{\tau-2\epsilon}{1-2\epsilon}\right), F^{-1}_{\beta_p(V;\mathbf{X})}\left(\frac{\tau}{1-2\epsilon}\right)\right].$$

These lower and upper bounds are identified since we can identify $\beta(\tau;\mathbf{X})$ when $\tau \in [\epsilon, 1-\epsilon]$. We also see that this identification region collapses to a point as ϵ approaches zero. These bounds may be used instead of the point estimates if it is believed that a significant portion of the conditional quantiles need to be trimmed. The researcher may also compare the lower and upper bounds' estimates to get a rough sense of the contribution of these extremal quantiles to identification.

3 Analysis of $T = P$ case with no "near-stayers"

Consider an alternative setup, where there is a non-shrinking point mass of stayers, and no near-stayers. For simplicity, we will assume that all probabilities and support points are fixed. This allows us to satisfy assumption 2 trivially. Let $\pi_0 = \mathbb{P}(D = 0)$ be strictly positive. The estimation of $\delta(\tau)$ will proceed as in the previous case, with a different rate of convergence since there is a point mass of stayers that is bounded away from zero. Proofs can be found in section 4 of this supplemental appendix.

We add the following assumption to guarantee identification.

Assumption 8. (INVERTIBILITY) $\mathbb{E}[\mathbf{W}^{*\prime}\mathbf{W}^{*}|D = 0]$ *is invertible.*

Theorem 8. *Suppose that Assumptions 1 through 6 and 8 are satisfied, then in the $T = P$ case (i) $\sqrt{N}\left(\widehat{\delta}\left(\cdot\right) - \delta\left(\cdot\right)\right)$ converges in distribution to a mean-zero Gaussian process $\mathbf{Z}_{\delta}\left(\cdot\right)$, where $\mathbf{Z}_{\delta}\left(\cdot\right)$ is defined by its covariance function*

$$
\begin{aligned}
\Sigma_{\delta}(\tau, \tau') &= \mathbb{E}[\mathbf{Z}_{\delta}\left(\tau\right)\mathbf{Z}_{\delta}\left(\tau'\right)'] \\
&= \frac{(\min(\tau, \tau') - \tau\tau')}{\pi_0}\mathbb{E}\left[\mathbf{W}^{*\prime}\mathbf{W}^{*}|D = 0\right]^{-1}\mathbb{E}\left[\mathbf{W}^{*\prime}\mathbf{X}^{*}\Lambda(\tau, \tau'; \mathbf{X})\mathbf{X}^{*\prime}\mathbf{W}^{*}|D = 0\right] \times \\
&\quad \mathbb{E}\left[\mathbf{W}^{*\prime}\mathbf{W}^{*}|D = 0\right]^{-1},
\end{aligned}
$$

and (ii) $\sqrt{N}\left(\widehat{\beta}\left(\cdot;\cdot\right) - \beta\left(\cdot;\cdot\right)\right)$ also converges in distribution to a mean zero Gaussian process $\mathbf{Z}\left(\cdot,\cdot\right)$, where $\mathbf{Z}\left(\cdot,\cdot\right)$ is defined by its covariance function

$$
\begin{aligned}
\Sigma(\tau, \mathbf{x}_l, \tau', \mathbf{x}_m) &= \mathbb{E}\left[\mathbf{Z}\left(\tau, \mathbf{x}_l\right)\mathbf{Z}\left(\tau', \mathbf{x}_m\right)'\right] \\
&= \left(\min\left(\tau, \tau'\right) - \tau\tau'\right)\frac{\mathbf{x}_l^{-1}\Lambda\left(\tau, \tau'; \mathbf{x}_l\right)\mathbf{x}_l^{-1\prime}}{p_l} \cdot \mathbf{1}\left(l = m\right) + \mathbf{x}_l^{-1}\mathbf{w}_l\Sigma_{\delta}(\tau, \tau')\mathbf{w}_m'\mathbf{x}_m^{-1\prime},
\end{aligned}
$$

for $l, m = 1, \ldots, L$.

The main difference between this result and the one in the just identified case is that only stayer realizations must be used for estimation of $\delta(\tau)$ in the $T = P$ case, while we chose to use the movers' realizations only for its estimation in the overidentified, $T > P$ case. This is reflected in the $\Sigma_{\delta}(\cdot)$ term.

It is also possible to estimate the movers' ACQE in the just-identified setup, as in the overidentified case.

Theorem 9. *Let*

$$\widehat{\bar{\beta}}^M(\tau) = \frac{\frac{1}{N}\sum_{i=1}^N \mathbf{X}_i^{-1}\left(\widehat{Q}_{\mathbf{Y}|\mathbf{X}}\left(\tau|\,\mathbf{X}_i\right) - \mathbf{W}_i\widehat{\delta}(\tau)\right)\mathbf{1}(\mathbf{X}_i \in \mathbb{X}^M)}{\frac{1}{N}\sum_{i=1}^N \mathbf{1}(\mathbf{X}_i \in \mathbb{X}^M)}$$
$$= \sum_{l=1}^L \hat{\beta}(\tau;\mathbf{x}_l)\hat{q}_l^M$$

be the movers' ACQE, where $\hat{q}_l^M = \frac{\frac{1}{N}\sum_{i=1}^N \mathbf{1}(\mathbf{X}_i = \mathbf{x_l})}{\frac{1}{N}\sum_{i=1}^N \mathbf{1}(\mathbf{X}_i \in \mathbb{X}^M)}$ *. Under Assumptions 1 through 6 and 8, we have that:*

$$\sqrt{N}\left(\widehat{\bar{\beta}}^M(\tau) - \bar{\beta}^M(\tau)\right) \xrightarrow{D} \mathbf{Z}_{\bar{\beta}}(\tau),$$

a mean-zero Gaussian process, on $\tau \in (0,1)$*. The variance of the Gaussian process* $\mathbf{Z}_{\bar{\beta}}(\cdot)$ *is defined as*

$$\mathbb{E}\left[\mathbf{Z}_{\bar{\beta}}(\tau)\mathbf{Z}_{\bar{\beta}}(\tau')'\right] = \frac{\mathbb{C}\left(\beta(\tau,\mathbf{X}),\beta(\tau',\mathbf{X})|\mathbf{X}\in\mathbb{X}^M\right)}{\Pr\left(\mathbf{X}\in\mathbb{X}^M\right)} + \Upsilon_1(\tau,\tau') + \Xi_0\Sigma_\delta(\tau,\tau')\Xi_0'$$

with

$$\Upsilon_1(\tau,\tau') = \frac{\min(\tau,\tau') - \tau\tau'}{\Pr\left(\mathbf{X}\in\mathbb{X}^M\right)}\mathbb{E}\left[\mathbf{X}^{-1}\Lambda(\tau,\tau';\mathbf{X})\mathbf{X}^{-1\prime}|\mathbf{X}\in\mathbb{X}^M\right]$$
$$\Xi_0 = \mathbb{E}\left[\mathbf{X}^{-1}\mathbf{W}|\mathbf{X}\in\mathbb{X}^M\right].$$

Finally, we can also recover estimates of the movers' UQE in a similar way to the irregular bandwidth setup.

Theorem 10. *Fix* $p \in \{1,\ldots,P\}$*. Under Assumptions 1 through 6 and 8 we have that*

$$\sqrt{N}\left(\widehat{\beta}_p^M(\tau) - \beta_p^M(\tau)\right) \xrightarrow{D} \mathbf{Z}_{\beta_p}(\tau)$$

on $\tau \in (0,1)$ *with* $\mathbf{Z}_{\beta_p}(\cdot)$ *being a Gaussian process. The covariance of this Gaussian process is equal to:*

$$\mathbb{E}\left[\mathbf{Z}_{\beta_p}(\tau)\mathbf{Z}_{\beta_p}(\tau')'\right] = \frac{\Upsilon_2(\tau,\tau') + \Upsilon_3(\tau,\tau') + \Upsilon_4(\tau,\tau')}{f_{\mathbf{B}_p|\mathbf{X}\in\mathbb{X}^M}(\beta_p^M(\tau))f_{\mathbf{B}_p|\mathbf{X}\in\mathbb{X}^M}(\beta_p^M(\tau'))}$$

with

$$\Upsilon_2(\tau,\tau') = \frac{\mathbb{C}\left(F_{\mathbf{B}_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{X}), F_{\mathbf{B}_p|\mathbf{X}}(\beta_p^M(\tau')|\mathbf{X})|\mathbf{X}\in\mathbb{X}^M\right)}{\Pr(\mathbf{X}\in\mathbb{X}^M)}$$

$$\Upsilon_3(\tau,\tau') = \mathbb{E}\Big[(\min(F_{\mathbf{B}_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{X}), F_{\mathbf{B}_p|\mathbf{X}}(\beta_p^M(\tau')|\mathbf{X})) - F_{\mathbf{B}_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{X})F_{\mathbf{B}_p|\mathbf{X}}(\beta_p^M(\tau')|\mathbf{X}))\times$$
$$e_p'\mathbf{X}^{-1}\Lambda(F_{\mathbf{B}_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{X}), F_{\mathbf{B}_p|\mathbf{X}}(\beta_p^M(\tau')|\mathbf{X}), \mathbf{X})\mathbf{X}^{-1\prime}e_p\times$$
$$f_{\mathbf{B}_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{X})f_{\mathbf{B}_p|\mathbf{X}}(\beta_p^M(\tau')|\mathbf{X})|\mathbf{X}\in\mathbb{X}^M\Big]$$

$$\Upsilon_4(\tau,\tau') = \frac{1}{\pi_0}\mathbb{E}\Big[f_{\mathbf{B}_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{X})f_{\mathbf{B}_p|\mathbf{X}}(\beta_p^M(\tau')|\tilde{\mathbf{X}})e_p\mathbf{X}^{-1}\mathbf{W}\Sigma_\delta(F_{\mathbf{B}_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{X}), F_{\mathbf{B}_p|\mathbf{X}}(\beta_p^M(\tau')|\tilde{\mathbf{X}}))\times$$
$$\tilde{\mathbf{W}}'\tilde{\mathbf{X}}^{-1\prime}e_p|\mathbf{X}\in\mathbb{X}^M, \tilde{\mathbf{X}}\in\mathbb{X}^M\Big]$$

where $\mathbf{X}$ *and* $\tilde{\mathbf{X}}$ *are independent copies.*

Since in this case the UQE is not point identified, we can recover bounds using the movers' UQE as in Theorem 1.

4 Proofs for section 3

Proof of Theorem 8. $\hat{\delta}(\tau) - \delta(\tau)$ has the following linear representation

$$\sqrt{N}\left(\hat{\delta}(\tau) - \delta(\tau)\right) = \left(\frac{1}{N}\sum_{i=1}^N \mathbf{W}_i^{*\prime}\mathbf{W}_i^*\mathbf{1}(D_i=0)\right)^{-1}$$
$$\times\frac{1}{N}\sum_{i=1}^N \mathbf{W}_i^{*\prime}\mathbf{X}_i^*\sqrt{N}\left(\widehat{Q}_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{X}_i) - Q_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{X}_i)\right)$$
$$= \left(\sum_{l=L+1}^M \mathbf{w}_l^{*\prime}\mathbf{w}_l^*\hat{p}_l\right)^{-1}\sum_{l=L+1}^M \mathbf{w}_l^{*\prime}\mathbf{x}_l^*\hat{p}_l\sqrt{N}\left(\widehat{Q}_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{x}_l) - Q_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{x}_l)\right)$$

with
$$\sum_{l=L+1}^M \mathbf{w}_l^{*\prime}\mathbf{w}_l^*\hat{p}_l \xrightarrow{p} \sum_{l=L+1}^M \mathbf{w}_l^{*\prime}\mathbf{w}_l^*p_l = \mathbb{E}\left[\mathbf{W}^{*\prime}\mathbf{W}^*|D=0\right]\pi_0$$

and
$$\sum_{l=L+1}^M \mathbf{w}_l^{*\prime}\mathbf{x}_l^*\hat{p}_l\sqrt{N}\left(\widehat{Q}_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{x}_l) - Q_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{x}_l)\right) \xrightarrow{d} \sum_{l=L+1}^M \mathbf{w}_l^{*\prime}\mathbf{x}_l^*\sqrt{p_l}\mathbf{Z}_Q(\tau,\mathbf{x}_l),$$

which has asymptotic covariance equal to

$$\begin{aligned}
&\mathbb{E}\left[\sum_{l=L+1}^{M}\mathbf{w}_l^{*\prime}\mathbf{x}_l^{*}\sqrt{p_l}\mathbf{Z}_Q(\tau,\mathbf{x}_l)\left(\sum_{l'=L+1}^{M}\mathbf{w}_{l'}^{*\prime}\mathbf{x}_{l'}^{*}\sqrt{p_{l'}}\mathbf{Z}_Q(\tau',\mathbf{x}_{l'})\right)'\right]\\
&=\sum_{l=L+1}^{M}\sum_{l'=L+1}^{M}\mathbf{w}_l^{*\prime}\mathbf{x}_l^{*}(\min(\tau,\tau')-\tau\tau')\Lambda(\tau,\tau';\mathbf{x}_l)\cdot\mathbf{1}(l=l')\,\mathbf{x}_l^{*\prime}\mathbf{w}_l^{*}p_lp_{l'}\\
&=(\min(\tau,\tau')-\tau\tau')\mathbb{E}\left[\mathbf{W}^{*\prime}\mathbf{X}^{*}\Lambda(\tau,\tau';\mathbf{X})\mathbf{X}^{*\prime}\mathbf{W}^{*}|D=0\right]\pi_0.
\end{aligned}$$

To derive the asymptotic distribution of $\sqrt{N}\left(\widehat{\beta}(\cdot;\cdot)-\beta(\cdot;\cdot)\right)$ we note that

$$\begin{aligned}
\sqrt{N}\left(\widehat{\beta}(\tau;\mathbf{x}_l)-\beta(\tau;\mathbf{x}_l)\right)&=\mathbf{x}_l^{-1}\sqrt{N}\left(\widehat{Q}_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{x}_l)-Q_{\mathbf{Y}|\mathbf{X}}(\tau|\mathbf{x}_l)\right)+\mathbf{x}_l^{-1}\mathbf{w}_l\sqrt{N}\left(\hat{\delta}(\tau)-\delta(\tau)\right)\\
&\xrightarrow{d}\mathbf{x}_l^{-1}\mathbf{Z}_Q(\tau,\mathbf{x}_l)+\mathbf{x}_l^{-1}\mathbf{w}_l\mathbf{Z}_\delta(\tau).
\end{aligned}$$

$\mathbf{Z}_Q(\tau,\mathbf{x}_l)$ and $\mathbf{Z}_\delta(\tau)$ are independent processes since they are computed using disjoint sub-populations: $\mathbf{x}_l$ for $l=1,\ldots,L$ are not used in the computation of $\hat{\delta}(\tau)$. Therefore, the asymptotic variance of $\sqrt{N}\left(\widehat{\beta}(\cdot;\cdot)-\beta(\cdot;\cdot)\right)$ is the sum of the variance of its terms. □

Proof of Theorem 9. We see that

$$\sqrt{N}\left(\widehat{\bar{\beta}}^{M}(\tau)-\bar{\beta}^{M}(\tau)\right)=\sum_{l=1}^{L}\beta(\tau;\mathbf{x}_l)\sqrt{N}\left(\hat{q}_l^M-q_l^M\right) \tag{2}$$

$$+\sum_{l=1}^{L}\sqrt{N}\left(\hat{\beta}(\tau;\mathbf{x}_l)-\beta(\tau;\mathbf{x}_l)\right)\hat{q}_l^M. \tag{3}$$

By a result similar to that in (76) in the main text, term (1) converges to a mean zero Gaussian process with covariance equal to $\frac{\mathbb{C}\left(\beta(\tau,\mathbf{X}),\beta(\tau',\mathbf{X})|X\in\mathbb{X}^M\right)}{\Pr(\mathbf{X}\in\mathbb{X}^M)}$. Term (3) converges to

$$\sum_{l=1}^{L}\sqrt{N}\left(\hat{\beta}(\tau;\mathbf{x}_l)-\beta(\tau;\mathbf{x}_l)\right)\hat{q}_l^M\xrightarrow{d}\sum_{l=1}^{L}\mathbf{Z}(\tau,\mathbf{x}_l)q_l^M$$

which has a covariance kernel equal to

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{l=1}^{L}\sum_{l'=1}^{L}\mathbf{Z}(\tau,\mathbf{x}_l)\mathbf{Z}(\tau,\mathbf{x}_{l'})q_l^M q_{l'}^M\right] \\
&= \mathbb{E}\left[\mathbf{Z}(\tau,\mathbf{x}_l)\mathbf{Z}(\tau,\mathbf{x}_{l'})\right] q_l^M q_{l'}^M \\
&= \left(\min\left(\tau,\tau'\right)-\tau\tau'\right)\sum_{l=1}^{L}\sum_{l'=1}^{L}\frac{\mathbf{x}_l^{-1}\Lambda\left(\tau,\tau';\mathbf{x}_l\right)\mathbf{x}_l^{-1\prime}}{p_l}\cdot\mathbf{1}\left(l=l'\right)q_l^M q_{l'}^M \\
&+\sum_{l=1}^{L}\sum_{l'=1}^{L}\mathbf{x}_l^{-1}\mathbf{w}_l\Sigma_\delta(\tau,\tau')\mathbf{w}_{l'}'\mathbf{x}_{l'}^{-1\prime}q_l^M q_{l'}^M \\
&= \frac{\min\left(\tau,\tau'\right)-\tau\tau'}{\Pr\left(\mathbf{X}\in\mathbb{X}^M\right)}\mathbb{E}\left[\mathbf{X}^{-1}\Lambda(\tau,\tau',\mathbf{X})\mathbf{X}^{-1\prime}|\mathbf{X}\in\mathbb{X}^M\right]+\sum_{l=1}^{L}\mathbf{x}_l^{-1}\mathbf{w}_l q_l^M\Sigma_\delta(\tau,\tau')\sum_{l'=1}^{L}\mathbf{w}_{l'}'\mathbf{x}_{l'}^{-1\prime}q_{l'}^M \\
&= \Upsilon_1(\tau,\tau')+\Xi_0\Sigma_\delta(\tau,\tau')\Xi_0'.
\end{aligned}
$$

Since terms (1) and (2) are uncorrelated, the asymptotic covariance of $\sqrt{N}\left(\widehat{\bar{\beta}}^M(\tau)-\bar{\beta}^M(\tau)\right)$ is equal to the sum of the covariance of its two terms. □

Proof of Theorem 10. We start by deriving the asymptotic distribution of the sample cumulative distribution function of $\widehat{\beta}_p(U;\mathbf{X})$ with U distributed uniformly on $[0,1]$ independently from $\mathbf{X}$, while conditioning on $\mathbf{X}\in\mathbb{X}^M$. The CDF estimand at $c\in\mathbb{R}$ is denoted as $F_{B_p|\mathbf{X}\in\mathbb{X}^M}(c)$ and the estimator is

$$
\begin{aligned}
\widehat{F}_{\widehat{\beta}_p(U;\mathbf{X})|\mathbf{X}\in\mathbb{X}^M}(c) &= \frac{\frac{1}{N}\sum_{i=1}^N\int_0^1\mathbf{1}(\widehat{\beta}_p(u,\mathbf{X}_i)\le c)du\mathbf{1}(\mathbf{X}_i\in\mathbb{X}^M)}{\frac{1}{N}\sum_{i=1}^N\mathbf{1}(\mathbf{X}_i\in\mathbb{X}^M)} \\
&= \sum_{l=1}^{L}\left(\int_0^1\mathbf{1}(\widehat{\beta}_p(u,\mathbf{x}_l)\le c)du\right)\widehat{q}_l^M.
\end{aligned}
$$

The integration over $u\in(0,1)$ can be done exactly since $\widehat{\beta}_p(u,\mathbf{x}_l)$ is piecewise linear for each $l\in\{1,\dots,L\}$ with finitely many pieces. This asymptotic distribution can be written as the sum of two terms:

$$
\begin{aligned}
\widehat{F}_{\widehat{\beta}_p(U;\mathbf{X})|\mathbf{X}\in\mathbb{X}^M}(c)-F_{B_p|\mathbf{X}\in\mathbb{X}^M}(c) &= \sum_{l=1}^{L}\left(\int_0^1\mathbf{1}(\widehat{\beta}_p(u,\mathbf{x}_l)\le c)du-\int_0^1\mathbf{1}(\beta_p(u,\mathbf{x}_l)\le c)du\right)\widehat{q}_l^M \\
&+\sum_{l=1}^{L}\int_0^1\mathbf{1}(\beta_p(u,\mathbf{x}_l)\le c)du\left(\widehat{q}_l^M-q_l^M\right).
\end{aligned}
$$

We will show that these two terms both converge in uniformly over $c\in\mathbb{R}$. For the first term, we have that $\sqrt{N}\left(\hat{\beta}_p(\tau;\mathbf{x}_l)-\hat{\beta}_p(\tau;\mathbf{x}_l)\right)\xrightarrow{d}\left(\mathbf{Z}(\tau,\mathbf{x}_l)\right)_p=\mathbf{Z}_p(\tau,\mathbf{x}_l)$ over $\tau\in(0,1)$ and all $l=1,\ldots,L$, and $(\cdot)_p$ denotes the p^{th} element of the vector. By the same argument as in (79), we have

$$
\begin{aligned}
&\sqrt{N}\left(\int_0^1 \mathbf{1}(\hat{\beta}_p(u;\mathbf{x}_l)\leq c)du-\int_0^1 \mathbf{1}(\beta_p(u;\mathbf{x}_l)\leq c)du\right)\\
&=\sqrt{N}\left(\hat{\beta}_p(F_{B_p|\mathbf{X}}(c|\mathbf{x}_l);\mathbf{x}_l)-\hat{\beta}_p(F_{B_p|\mathbf{X}}(c|\mathbf{x}_l);\mathbf{x}_l)\right)f_{B_p|\mathbf{X}}(c|\mathbf{x}_l)+o_p(1)\\
&\xrightarrow{d}\mathbf{Z}_p(F_{B_p|\mathbf{X}}(c|\mathbf{x}_l),\mathbf{x}_l)f_{B_p|\mathbf{X}}(c|\mathbf{x}_l).
\end{aligned}
$$

This convergence is uniform in $c\in\mathbb{R}$ since $F_{B_p|\mathbf{X}}(c|\mathbf{x}_l)$ ranges between 0 and 1, and uniform in $\mathbf{x}_l$ since its support is finite. Therefore,

$$
\sum_{l=1}^{L}\sqrt{N}\left(\int_0^1 \mathbf{1}(\hat{\beta}_p(u;\mathbf{x}_l)\leq c)du-\int_0^1 \mathbf{1}(\beta_p(u;\mathbf{x}_l)\leq c)du\right)\hat{q}_l^M\xrightarrow{d}\sum_{l=1}^{L}\mathbf{Z}_p(F_{B_p|\mathbf{X}}(c|\mathbf{x}_l),\mathbf{x}_l)f_{B_p|\mathbf{X}}(c|\mathbf{x}_l)q_l^M
$$

for $c\in\mathbb{R}$. Also, the second term will converge over $c\in\mathbb{R}$ to a Gaussian process $\mathbf{Z}_{2p}(c)$ with asymptotic covariance of

$$
\mathbb{E}\left[\mathbf{Z}_{2p}(c)\mathbf{Z}_{2p}(c')'\right]=\frac{\mathbb{C}\left(F_{B_p|\mathbf{X}}(c|\mathbf{X}),F_{B_p|\mathbf{X}}(c'|\mathbf{X})|\mathbf{X}\in\mathbb{X}^M\right)}{\Pr\left(\mathbf{X}\in\mathbb{X}^M\right)}.
$$

Note that $\mathbf{Z}_{2p}(c)$ and $\sum_{l=1}^{L}\mathbf{Z}_p(F_{B_p|\mathbf{X}}(c|\mathbf{x}_l),\mathbf{x}_l)f_{B_p|\mathbf{X}}(c|\mathbf{x}_l)q_l^M$ are uncorrelated since the variation in the latter is conditional on $\mathbf{X}$ while that in the former depends on $\mathbf{X}$ only. Therefore,

$$
\widehat{F}_{\widehat{\beta}_p(U;\mathbf{X})|\mathbf{X}\in\mathbb{X}^M}(c)-F_{B_p|\mathbf{X}\in\mathbb{X}^M}(c)\xrightarrow{d}\sum_{l=1}^{L}\mathbf{Z}_p(F_{B_p|\mathbf{X}}(c|\mathbf{x}_l),\mathbf{x}_l)f_{B_p|\mathbf{X}}(c|\mathbf{x}_l)q_l^M+\mathbf{Z}_{2p}(c)
$$

for $c\in\mathbb{R}$.

Using the same invertibility argument as in (82), we see that

$$
\begin{aligned}
\sqrt{N}\left(\hat{\beta}_p^M(\tau)-\beta_p^M(\tau)\right)&\xrightarrow{d}\frac{\sum_{l=1}^{L}\mathbf{Z}_p(F_{B_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{x}_l),\mathbf{x}_l)f_{B_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{x}_l)q_l^M+\mathbf{Z}_{2p}(\beta_p^M(\tau))}{f_{B_p|\mathbf{X}\in\mathbb{X}^M}(\beta_p^M(\tau))}\\
&=\mathbf{Z}_{\beta_p}(\tau)
\end{aligned}
$$

uniformly over $\tau\in(0,1)$.

To conclude this proof, we evaluate $\mathbb{E}\left[\mathbf{Z}_{\beta_p}(\tau)\,\mathbf{Z}_{\beta_p}(\tau')'\right]$, the asymptotic covariance of the estimated UQE:

$$\mathbb{E}\left[\mathbf{Z}_{\beta_p}(\tau)\,\mathbf{Z}_{\beta_p}(\tau')'\right] = \frac{\sum_{l=1}^{L}\sum_{l'=1}^{L} f_{B_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{x}_l) f_{B_p|\mathbf{X}}(\beta_p^M(\tau')|\mathbf{x}_{l'}) q_l^M q_{l'}^M}{f_{B_p|\mathbf{X}\in\mathbb{X}^M}(\beta_p^M(\tau)) f_{B_p|\mathbf{X}\in\mathbb{X}^M}(\beta_p^M(\tau'))}$$
$$\times\, \mathbb{E}\left[\mathbf{Z}_p(F_{B_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{x}_l),\mathbf{x}_l)\mathbf{Z}_p(F_{B_p|\mathbf{X}}(\beta_p^M(\tau')|\mathbf{x}_{l'}),\mathbf{x}_{l'})\right]$$
$$+\, \frac{\mathbb{E}\left[\mathbf{Z}_{2p}(\beta_p^M(\tau))\mathbf{Z}_{2p}(\beta_p^M(\tau'))\right]}{f_{B_p|\mathbf{X}\in\mathbb{X}^M}(\beta_p^M(\tau)) f_{B_p|\mathbf{X}\in\mathbb{X}^M}(\beta_p^M(\tau'))}$$

where

$$\mathbb{E}\left[\mathbf{Z}_p(F_{B_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{x}_l),\mathbf{x}_l)\mathbf{Z}_p(F_{B_p|\mathbf{X}}(\beta_p^M(\tau')|\mathbf{x}_{l'}),\mathbf{x}_{l'})\right]$$
$$= \left(\min\left(F_{B_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{x}_l), F_{B_p|\mathbf{X}}(\beta_p^M(\tau')|\mathbf{x}_{l'})\right) - F_{B_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{x}_l)F_{B_p|\mathbf{X}}(\beta_p^M(\tau')|\mathbf{x}_{l'})\right)$$
$$\times\, e_p' \frac{\mathbf{x}_l^{-1}\Lambda\left(F_{B_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{x}_l), F_{B_p|\mathbf{X}}(\beta_p^M(\tau')|\mathbf{x}_{l'});\mathbf{x}_l\right)\mathbf{x}_l^{-1\prime}}{p_l} e_p \cdot \mathbf{1}\left(l=l'\right)$$
$$+\, e_p'\mathbf{x}_l^{-1}\mathbf{w}_l\Sigma_\delta(F_{B_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{x}_l), F_{B_p|\mathbf{X}}(\beta_p^M(\tau')|\mathbf{x}_{l'}))\mathbf{w}_{l'}'\mathbf{x}_{l'}^{-1\prime}e_p$$

and

$$\sum_{l=1}^{L}\sum_{l'=1}^{L} f_{B_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{x}_l) q_l^M \mathbb{E}\left[\mathbf{Z}_p(F_{B_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{x}_l),\mathbf{x}_l)\mathbf{Z}_p(F_{B_p|\mathbf{X}}(\beta_p^M(\tau')|\mathbf{x}_{l'}),\mathbf{x}_{l'})\right] f_{B_p|\mathbf{X}}(\beta_p^M(\tau')|\mathbf{x}_{l'}) q_{l'}^M$$
$$= \mathbb{E}\Bigg[\frac{\left(\min\left(F_{B_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{X}), F_{B_p|\mathbf{X}}(\beta_p^M(\tau')|\mathbf{X})\right) - F_{B_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{X})F_{B_p|\mathbf{X}}(\beta_p^M(\tau')|\mathbf{X})\right)}{\Pr\left(\mathbf{X}\in\mathbb{X}^M\right)}$$
$$\times\, e_p'\;\mathbf{X}^{-1}\Lambda\left(F_{B_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{X}), F_{B_p|\mathbf{X}}(\beta_p^M(\tau')|\mathbf{X});\mathbf{X}\right)\mathbf{X}^{-1\prime}e_p f_{B_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{X}) f_{B_p|\mathbf{X}}(\beta_p^M(\tau')|\mathbf{X})\Big|\,\mathbf{X}\in\mathbb{X}^M\Bigg]$$
$$+\, e_p'\mathbb{E}\Big[f_{B_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{X}) f_{B_p|\mathbf{X}}(\beta_p^M(\tau')|\tilde{\mathbf{X}})\mathbf{X}^{-1}\mathbf{W}\Sigma_\delta(F_{B_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{X}), F_{B_p|\mathbf{X}}(\beta_p^M(\tau')|\tilde{\mathbf{X}}))$$
$$\times\tilde{\mathbf{W}}'\tilde{\mathbf{X}}^{-1\prime}|\mathbf{X}\in\mathbb{X}^M, \tilde{\mathbf{X}}\in\mathbb{X}^M\Big]\, e_p$$
$$= \Upsilon_3(\tau,\tau') + \Upsilon_4(\tau,\tau'),$$

where $\tilde{\mathbf{X}}$ is an independent copy of $\mathbf{X}$. Finally,

$$\mathbb{E}\left[\mathbf{Z}_{2p}(\beta_p^M(\tau))\mathbf{Z}_{2p}(\beta_p^M(\tau'))\right] = \frac{\mathbb{C}\left(F_{B_p|\mathbf{X}}(\beta_p^M(\tau)|\mathbf{X}), F_{B_p|\mathbf{X}}(\beta_p^M(\tau')|\mathbf{X})|\mathbf{X}\in\mathbb{X}^M\right)}{\Pr\left(\mathbf{X}\in\mathbb{X}^M\right)}$$
$$= \Upsilon_2(\tau,\tau').$$

□

References

[1] Arellano, Manuel and Stéphane Bonhomme. (2016). "Nonlinear Panel Data Estimation via Quantile Regressions," *Econometrics Journal*, forthcoming.

[2] Chamberlain, Gary. (1984). "Panel data," *Handbook of Economics* 2: 1247 - 1318 (Z. Griliches & M. Intriligator, Eds.). Amsterdam: North-Holland.

[3] Chernozhukov, Victor, Iván Fernández-Val, Jinyong Hahn and Whitney Newey. (2013). "Average and quantile effects in nonseparable panel models," *Econometrica* 81 (2): 535 - 580.

[4] Graham, Bryan S. and James L. Powell. (2012). "Identification and estimation of average partial effects in 'irregular' correlated random coefficient panel data models," *Econometrica* 80 (5): 2105 - 2152.

[5] Lee, Ying-Ying. (2013). "Nonparametric weighted average quantile derivative," *Mimeo*, Nuffield College, Oxford University.

[6] Qu, Zhongjun and Jungmo Yoon. (2015) "Nonparametric estimation and inference on conditional quantile processes," *Journal of Econometrics* 185 (1): 1 - 19.