Quantile Regression Random EffectsAuthor(s): Antonio F. Galvao and Alexandre Poirier

Source: *Annals of Economics and Statistics*, No. 134 (June 2019), pp. 109-148

Published by: GENES on behalf of ADRES

Stable URL: https://www.jstor.org/stable/10.15609/annaeconstat2009.134.0109

Annals of Economics and Statistics, Number 134, June 2019

QUANTILE REGRESSION RANDOM EFFECTS

ANTONIO F. GALVAO[a] AND ALEXANDRE POIRIER[b]

This paper develops a random effects model for quantile regression (QR). We establish identification of the QR coefficients, and develop practical estimation and inference procedures. We employ a simple pooled QR estimator to estimate the coefficients of interest, and derive its statistical properties. The random effects induce cluster dependence hence we use a cluster-robust variance-covariance matrix estimator for inference, and establish its uniform consistency over the set of quantiles. We also develop a new test procedure for uniform testing of linear hypotheses in QR models. This procedure is a modified Wald test applied on a growing number of quantiles such that, asymptotically, the test is uniform over the quantiles. We show this procedure can be applied to test the random effects hypothesis in QR panel data models. Two significant differences between our model and fixed-effects QR models are that effects of time-invariant regressors can be estimated, and that the time-series dimension can be small and finite. We provide Monte Carlo simulations to evaluate the finite sample performance of the estimation and inference procedures. Finally, we apply the proposed methods to study the roles of education and ability in wage determination. We document strong heterogeneity in returns to education along the conditional distribution of earnings.

JEL Codes: C12, C13, C23.

Keywords: Quantile Regression, Panel Data, Random Effects, Hypothesis Testing.

1. INTRODUCTION

Since the seminal work of Koenker and Bassett (1978), quantile regression (QR) has attracted considerable interest in econometrics and statistics. It offers an easy-to-implement method to estimate conditional quantiles. Starting with Koenker (2004), there has been a recent and growing literature on estimation and testing using QR for panel (longitudinal) data models. Panel QR has provided a valuable method of statistical analysis of the heterogeneous effects of policy variables.

Koenker (2004) introduced a general approach to estimation of QR models for longitudinal data. In the model, individual-specific (fixed) effects are treated as pure location shift parameters common to all conditional quantiles. However, the fixed effects (FE) panel QR suffers from the incidental parameters problem (Neyman and Scott (1948)). To overcome this drawback, it has become standard in the literature to employ "large-T" asymptotics where the number of individuals, n, and the number of time periods, T, jointly diverge

The authors would like to express their appreciation to Yonghong An, Badi Baltagi, Stéphane Bonhomme, Matias Cattaneo, Laurent Davezies, Yanqin Fan, Andreas Hagemann, Joel Horowitz, Kengo Kato, Roger Koenker, Suyong Song, Ke-Li Xu, and seminar participants at Texas A&M University, University of Iowa, University of Michigan, University of Notre Dame, University of Washington, EESP-FGV, 2016 Midwest Econometrics Group, New York Camp Econometrics XI, the 2016 North American summer meeting of the Econometric Society and China meeting of the Econometric Society for helpful comments and discussions.

[a]Department of Economics, University of Arizona, McClelland Hall, Room 401 1130 E. Helen Street, Tucson, AZ 85721. agalvao@email.arizona.edu

[b]Corresponding Author. Department of Economics, Georgetown University, Intercultural Center 580, 37th and O Streets, N.W., Washington, DC 20057. alexandre.poirier@georgetown.edu

to infinity. Kato, Galvao, and Montes-Rojas (2012) formally established sufficient conditions for consistency and asymptotic normality of the FE-QR estimator. Unfortunately, existing sufficient conditions under which the asymptotic bias of the FE-QR vanishes require $T >> n$. This requirement on the time series's length is very restrictive and is not satisfied in many empirical applications. Additionally, this method does not accommodate the inclusion of time-invariant regressors due to the presence of the individual-specific effects.

Abrevaya and Dahl (2008) introduced an alternative approach to the FE-QR which estimates panel data QR employing the correlated random effects (CRE) model of Chamberlain (1982). The unobservable individual-specific effect is modeled as a linear function of the observables and a disturbance. Arellano and Bonhomme (2016) introduced a class of QR estimators for short panels, where the conditional quantile response function of the unobserved heterogeneity is also specified as a function of observables. For other recent development in panel data QR see, among others, Wei and He (2006), Wang and Fygenson (2009), Lamarche (2010), Canay (2011), Rosen (2012), Chernozhukov, Fernandez-Val, Hahn, and Newey (2013), Galvao, Lamarche, and Lima (2013), Chernozhukov, Fernandez-Val, Hoderlein, Holzmann, and Newey (2015), Cai, Chen, and Fang (2015), Graham, Hahn, Poirier, and Powell (2018), and Galvao and Kato (2016). Galvao and Kato (2017) provide a review of QR methods for panel (longitudinal) data.

This paper develops yet another alternative to the FE-QR and CRE-QR. We propose a random effects (RE) model for QR panel data with time-invariant regressors.[1] We consider a RE assumption in which the individual-specific components are independent of the regressors. Despite this restriction, we construct a model where the unobserved individual-specific effects affect the dependent variable, which induces heterogeneity across the conditional quantile function of the dependent variable.[2] Thus, although the RE-QR model does not attempt to estimate the individual-specific effects explicitly, their presence generates heterogeneity in the conditional distribution of the dependent variable. We contribute to the literature by establishing identification of the parameters of interest, and developing practical estimation as well as uniform inference procedures. Given the identification result, we apply a simple pooled QR estimator to estimate the coefficients of interest, and establish the uniform consistency and weak convergence of the RE-QR estimator for short panels.

The existence of RE in the model generates within unit cluster-dependence, and thus we employ a cluster-robust variance-covariance matrix to conduct uniform inference. We show that the corresponding variance-covariance estimator is uniformly consistent. We propose inference procedures based on the Wald statistic, which are simple to carry out empirically.

We then develop a new general test procedure for QR models for testing linear hypotheses uniformly over the set of quantiles. The method relies on a modified Wald test applied on a grid of quantiles, which becomes arbitrarily fine as n increases. We show that the test-statistic is asymptotically normal with known variance, and is therefore asymptotically pivotal. This procedure is a general testing procedure for QR models, which include

[1] The literature on RE for mean regression is extensive. We refer the reader to Baltagi, Matyas, and Sevestre (1995) and Baltagi (2013) and references therein.

[2] See equation (1) in Graham and Powell (2012) for a related model where the unobserved heterogeneity affects the distribution of the dependent variable.

our RE-QR model, and we apply the method to test linear hypotheses uniformly over the set of quantiles, a non-trivial task when clustering is present (see Hagemann (2017) for an alternative approach).

In addition, an important contribution of this paper is to develop a novel test for the RE hypothesis in a panel QR model. The test is based on an auxiliary QR which includes time-averages of time-varying regressors. For a fixed quantile-τ the test is a Wald test for the coefficients. We also extend the methods to a new uniform test over the set of quantiles. We derive the limiting distribution of the test statistics and show they are easy to implement in practice. These tests are of practical interest since they can help deciding whether a RE model is appropriate for a given application.

Compared to existing QR panel data procedures, the RE-QR approach has several distinct advantages. First, the estimator is a simple one-step QR procedure and allows for time-invariant regressors. Under the FE-QR model, effects from time-invariant regressors cannot be distinguished from the individual-specific intercepts. Second, inference is based on standard methods and is simple to carry in practice. Finally, the methods allow the time-series dimension, T, to be small and fixed. No assumptions are placed on the time-series dependence of observables and of unobservables. This is in contrast with the FE-QR methods, which require T to diverge to infinity much faster than the cross-sectional dimension n. The fixed time dimension also directly allows for the inclusion of time dummies as regressors, while incorporating time dummies in models with growing T is challenging.

There is a small literature on RE-QR. Kim and Yang (2011) considered a RE-QR analysis of clustered data and propose a semiparametric approach using empirical likelihood. In this case, the slope coefficients of interest depend on the individual index, i, and are different for each individual. The random coefficients are assumed to be independent with a common mean, following parametrically specified distributions. Wang and Fygenson (2009) used a RE approach to develop inference procedures for longitudinal data where some of the measurements are censored by fixed constants. They demonstrate the importance of accounting for the intra-subject dependency induced by the RE. Geraci and Bottai (2007) elaborated on the penalized QR estimator by proposing a linear model that includes RE in order to account for the dependence between serial observations on the same subject. They present a likelihood-based approach that uses the asymmetric Laplace density, and estimate the parameters of interest using a Monte Carlo EM algorithm.

We conduct a Monte Carlo study to evaluate the finite-sample properties of the RE-QR estimator and tests. The results provide evidence that the estimator is approximately unbiased and the clustered variance matrix approximates the true variance. In addition, the results show that empirical levels for the pointwise test for the RE hypothesis approximate well the theoretical level. Moreover, the test possesses large power against selected alternatives. Simulations also show that the proposed uniform test has approximately correct size under most settings.

To illustrate the applicability of the proposed methods, we study the roles of education and ability in wage determination (Card (1999), Harmon and Oosterbeek (2000), Castex and Dechter (2014)). We apply the proposed RE-QR methods to two panel waves (1979 and 1997) of the National Longitudinal Survey of Youth (NLSY) data set. Both waves contain detailed information on the individuals, including their education and cognitive skills. The inclusion of a cognitive skill measure as a regressor makes the RE assumption plausible, as FE are often used to account for skill heterogeneity. The proposed RE-QR methods allow us to capture the heterogeneity in the returns to education and ability along the conditional earnings distribution. Two distinct advantages of the RE-QR approach over FE-QR are that we are able to estimate the effect of time-invariant regressors, such as cognitive skills, and it also allows for short panels such as the two NLSY waves we observe. We estimate the RE-QR model separately for both panels (NLSY79 and NLSY97), as well as for men and women. First, we perform the tests for the RE hypothesis. The pointwise tests do not reject the null of RE for several quantiles for all subsamples. The uniform tests do not reject the null for NLSY79 men and women, and for the NLSY97 women's samples. Second, the resulting estimates document strong heterogeneity in returns to education and ability along the conditional distribution of earnings. In particular, the returns to ability have increased from 1979 to 1997 for women at the top of the distribution of earnings, while they decreased for those at the bottom of the distribution of earnings. Moreover, the returns to education have increased between 1979 to 1997 for both men and women, and especially the returns for higher degrees such as Bachelor's and Master's degrees. The returns are also relatively larger at the top of the distribution.

The paper is organized as follows: Section 2 introduces the model and develops identification and estimation. Section 3 presents the asymptotic theory. Section 4 develops inference procedures and proposes the tests. A Monte Carlo experiment is provided in Section 5. Section 6 presents the empirical example. Section 7 concludes. All proofs are in the Appendix.

2. RANDOM EFFECTS MODEL

This section presents the random effects quantile regression (RE-QR) model. We also discuss the relationship between the RE-QR and fixed effects quantile regression (FE-QR) models. Finally, we present the identification and estimation of the parameters of interest.

2.1. *Model*

Many panel data models can be written as follows

$$y_{it} = m(x_{it}, z_i, \alpha_i, \epsilon_{it}), \quad i = 1, ..., n,\ t = 1, ..., T, \tag{1}$$

where y_{it} is a dependent variable, x_{it} is a k_1-vector of time-varying regressors, z_i is a k_2-vector of time-invariant regressors, α_i are unobserved individual-specific components, and ϵ_{it} is a scalar disturbance. We assume that strict exogeneity holds, in the sense that

$$\epsilon_{it} \perp\!\!\!\perp (x_i, z_i, \alpha_i), \tag{2}$$

where $x_i = [x'_{i1}, \ldots, x'_{iT}]'$. The potentially non-separable function $m(\cdot)$ allows for the effect of x_{it} to depend on the unobserved components $(\alpha_i, \epsilon_{it})$. The number of individuals is denoted by n and the number of time periods is denoted by T.

A linear version of model (1) with scalar individual-specific effect can be written as

$$\begin{aligned} y_{it} &= x'_{it}\beta + z'_i\gamma + \alpha_i + \epsilon_{it} \\ &= x'_{it}\beta + z'_i\gamma + U_{it}, \end{aligned}$$

where $U_{it} \equiv \alpha_i + \epsilon_{it}$. This location model can be generalized to the following location-scale model

$$
\begin{aligned}
y_{it} &= U_{it} + x_{it}'\beta + z_i'\gamma + x_{it}'h_1(U_{it}) + z_i'h_2(U_{it}) \\
&= U_{it} + x_{it}'(\beta + h_1(U_{it})) + z_i'(\gamma + h_2(U_{it})), \qquad (3)
\end{aligned}
$$

where $x_{it}'h_1(\cdot)$ and $z_i'h_2(\cdot)$ are assumed to be non-decreasing.[3] This allows the heterogeneity in responses U_{it} to depend on both the individual effect α_i and idiosyncratic error ϵ_{it}. The individual effect α_i appears both in the intercept, and inside the coefficients on x_{it} and z_i, therefore inducing differences in responses across individuals.

In this paper, we propose a QR version of model (1), which is also a generalization of (3), as

$$
y_{it} = c(U_{it}) + x_{it}'\beta(U_{it}) + z_i'\gamma(U_{it}), \qquad (4)
$$

where U_{it} represents the heterogeneity in responses and can depend on both ϵ_{it} and α_i, as

$$
U_{it} \equiv U(\alpha_i, \epsilon_{it}), \qquad (5)
$$

with $U(\cdot,\cdot)$ being a scalar and unspecified nonparametric function. Note that equation (5) allows the unobserved heterogeneity to depend on both the independent unobserved component, ϵ_{it}, and the individual-specific components, α_i, in an unrestricted form. The functions $c(\cdot)$, $\beta(\cdot)$ and $\gamma(\cdot)$ in (4) quantify the distributional effects for the intercept, and the time-varying and time-invariant regressors, x_{it} and z_i respectively.[4] In this model the unobserved individual-specific effects, α_i, affect the dependent variable through the unobserved heterogeneity $U(\alpha_i, \epsilon_{it})$, which then induces heterogeneity across the conditional quantile function of y_{it}.

We consider a model where the right-hand side of (4) is strictly increasing in U_{it}, and normalize U_{it} to have a uniform distribution on $[0,1]$, $U_{it} \sim$ Unif$[0,1]$, without loss of generality. Under these assumptions, we can write the conditional quantile of y_{it} given $X_i \equiv [X_{i1}', \ldots, X_{iT}']'$, with $X_{it} = [1, x_{it}', z_i']'$ a $K \times 1$ vector of regressors, with $K = k_1 + k_2 + 1$, as

$$
Q_{y_{it}}(\tau|X_i) = c(Q_{U_{it}}(\tau|X_i)) + x_{it}'\beta(Q_{U_{it}}(\tau|X_i)) + z_i'\gamma(Q_{U_{it}}(\tau|X_i)),
$$

where $\tau \in (0,1)$ is the quantile of interest. This follows equation (4) and quantile equivariance.

The RE assumption in standard linear mean panel data models restricts the unobserved component, α_i, to be uncorrelated with all regressors, i.e., $\mathrm{Cov}(\alpha_i, x_{it}) = 0$. We generalize this assumption to the model in (4)–(5) by assuming the following independence condition

$$
\alpha_i \perp\!\!\!\perp X_i. \qquad (6)
$$

Condition (6) is the RE assumption that we consider for the RE QR model.

The independence assumption in (6) is used due to the non-linearity in α_i of equa-

[3]See Arellano and Bonhomme (2011) for a related model.

[4]We can also include the intercept into the time-invariant regressors z_i.

tions (4)–(5). Under this assumption and strict exogeneity, the unobserved heterogeneity $U(\alpha_i, \epsilon_{it})$ is independent from X_i, which gives rise to the following quantile representation

$$Q_{y_{it}}(\tau|X_i) = c(\tau) + x_{it}'\beta(\tau) + z_i'\gamma(\tau) = X_{it}'\theta(\tau), \tag{7}$$

where $\theta(\tau) = [c(\tau), \beta(\tau)', \gamma(\tau)']'$. The presence of τ on the right-hand side follows from our previous normalization of U_{it} and from $Q_{U_{it}}(\tau|X_i) = Q_{U_{it}}(\tau) = \tau$. Equation (7) establishes the linear RE-QR model, given equations (4)–(5), and condition (6). Thus, although the RE-QR model does not attempt to estimate the unobserved individual-specific effects explicitly, these unobserved effects directly generate heterogeneity in the conditional distribution of the dependent variable through the dependence of the coefficients on $U(\alpha_i, \epsilon_{it})$.

The independence assumption in (6) is strong, but is useful to consider for two reasons. First, it allows the model to satisfy a quantile representation that is linear in X_{it} and $\theta(\tau)$. The representation of the model in (7) suggests a simple estimator for the parameters of interest. In particular, a simple pooled linear QR of y_{it} on X_{it} can be applied to consistently estimate the parameters of interest $\theta(\tau)$ under regularity conditions. Second, it allows for consistent estimation in small panels, i.e. $T < \infty$. The QR panel data models with individual-specific FE intercepts suffer from the incidental parameters problem, which can be solved by assuming that the time dimension asymptotically diverges. The RE assumption allows us to considerably expand the number of potential empirical applications with short panels.

2.2. *Relationship between random and fixed effects*

The relationship between random and fixed effects is more delicate for quantile models than for linear, mean-regression models. This section provides intuition on why we model the RE-QR using (4)–(5) together with condition (6). In a linear panel model, traditional conditional mean RE and FE estimation are based from the same linear model which often takes the form

$$y_{it} = x_{it}'\beta + \alpha_i + \epsilon_{it}.$$

Under fixed effects the "within" estimator can recover β without requiring assumptions on the dependence between α_i and x_{it} . Meanwhile, if the RE assumption of $\mathrm{Cov}(\alpha_i, x_{it}) = 0$ is assumed, the pooled regression estimator will also be consistent for β. However, for quantile regression, the FE and RE models differ substantially. We find that the FE-QR estimator may not recover the same parameter that a pooled QR would recover even if assumption (6) holds.

To illustrate this difference, consider the additive-in-α_i linear FE-QR model of Kato, Galvao, and Montes-Rojas (2012), which can be represented as

$$y_{it} = \alpha_i + x_{it}'\beta(\epsilon_{it}). \tag{8}$$

We note that under fixed or random effects, the *conditional* model yields $Q_{y_{it}}(\tau|x_i, \alpha_i) = \alpha_i + x_{it}'\beta(\tau)$, a linear function of x_{it} and α_i. Under some regularity conditions, and no restriction on the relationship between α_i and x_{it}, Kato, Galvao, and Montes-Rojas (2012) show that $\beta(\tau)$ can be estimated consistently by a QR with individual-specific dummy variables when T is large. Note that the inclusion of individual-specific dummy variables precludes one from having time-invariant regressors.[5]

Nevertheless, even if $\alpha_i \perp\!\!\!\perp x_i$ holds in (8), a pooled linear QR estimator may not consistently estimate $\beta(\tau)$. This is due to a difference between the *conditional* model, $Q_{y_{it}}(\tau|x_i, \alpha_i)$, and the *marginal* model, $Q_{y_{it}}(\tau|x_i)$. The conditional quantile of y_{it} given x_i might not be linear in x_{it} since $Q_{y_{it}}(\tau|x_i) \neq Q_{\alpha_i}(\tau) + x_{it}'\beta(\tau)$, because the quantile of a sum is generally different from the sum of the quantiles by the non-linearity of the quantile operator.[6] As a result, the pooled QR estimator and the FE-QR estimator estimate different quantities.

In equations (4)–(5), we consider an alternative model which is non-additive in the individual-specific effect, α_i, as

$$y_{it} = x_{it}'\beta(U(\alpha_i, \epsilon_{it})).$$

Under assumption (6), the conditional quantile of y_{it} given x_i in the above equation, i.e. the marginal model, is linear in x_{it} and $\beta(\tau)$. Again, the conditional model will yield a different effect of x_{it} on y_{it}, so a FE estimator with individual-specific dummy variables would not recover the same coefficients on x_{it} due to the non-additivity of α_i. Thus, the FE-QR and RE-QR estimators will converge to different quantities, since they must rely on different modeling assumptions, which stands in contrast to the linear mean-regression panel case where both FE and RE estimators are both consistent. This is due to the fact that under random effects the marginal and conditional models for linear, mean-regression panels yield the same effect of x_{it} on y_{it}[7] while these effects differ for all models considered here.

Another feature of the non-additive RE-QR model is that a failure of the RE assumption (6) will imply that the conditional quantile of y_{it} is no longer linear in X_{it}, because the composite unobserved heterogeneity term U_{it} will generally be correlated with X_i. We can then write the conditional quantile of y_{it} as follows

$$\begin{aligned} Q_{y_{it}}(\tau|X_i) &= X_{it}'\theta(Q_{U_{it}}(\tau|X_i)) \\ &\equiv X_{it}'\tilde{\theta}(\tau; X_i), \end{aligned}$$

where $\tilde{\theta}(\tau; X_i)$ is a nonparametric function of τ and X_i. Graham, Hahn, Poirier, and Powell (2018) discuss the nonparametric identification and estimation of this model, and more specifically of the unconditional quantile effect $\theta(\tau)$. Their approach can be interpreted as the FE version of our model, since they do not restrict the dependence between α_i and x_i. Estimation of their model requires preliminary nonparametric estimates of the conditional quantile function, which are difficult to obtain when X_i are continuously distributed or have many components, whereas the RE assumption will allow us to use standard tech-

[5]A related work which allows for the use of time-invariant regressors is Chetverikov, Larsen, and Palmer (2016), but it requires the existence of valid instruments external to the model and also that T diverges.

[6]The conditional quantile of the sum will be equal to the sum of conditional quantiles if conditional comonotonicity between α_i and $x_{it}'\beta(\epsilon_{it})$ holds conditional on x_i. This is ruled out by the conditional independence of α_i and ϵ_{it} given x_i since comonotonic variables cannot be independent. See Lemma A.1 in the appendix for a proof of this result.

[7]This can be seen from $\mathrm{E}[y_{it}|x_{it}, \alpha_i] = \alpha_i + x_{it}'\beta$ and $\mathrm{E}[y_{it}|x_{it}] = \mathrm{E}[\alpha_i] + x_{it}'\beta$.

niques for linear QR models. They additionally require the number of regressors to be smaller or equal to the number of time periods, while we do not impose any restrictions on the number of regressors relative to T.

These differences between the RE and FE models in the QR case make testing for the presence of RE very important in the QR context. This paper fills this gap by proposing a new test for the RE assumption. In particular, we later propose to test whether conditional quantiles of U_{it} are related to X_i for all τ.

2.3. *Identification and estimation*

Due to the quantile representation in (7), the identification of the coefficients of interest, $\theta(\tau)$, can be analyzed using the traditional identification assumptions for QR. In particular, we require that y_{it} is continuously distributed with unique τ-th quantile, and that the regressors X_{it} are not perfectly collinear. Note that this does not rule out the presence of time-invariant regressors whereas, in a FE model, individual-specific intercepts would "absorb" the effect of any time-invariant regressor. To formally show identification of the QR coefficients, we impose the following assumptions.

A1. The conditional density of y_{it} given X_i $f_{y_t}(y|X_i)$ exists almost surely for every $t = 1, \ldots, T$. Additionally, it is bounded above and continuous in y, uniformly over the support of X_i.

A2. The matrix $\Gamma(\tau) \equiv \mathrm{E}\left[\frac{1}{T}\sum_{t=1}^{T} X_{it}X'_{it}f_{y_t}(X'_{it}\theta(\tau)|X_i)\right]$ is invertible and has minimum eigenvalue bounded away from zero, uniformly in τ over $\mathcal{T} = [\epsilon, 1-\epsilon]$ where $\epsilon \in (0, 1/2)$.

A3'. $\mathrm{E}[\|X_{it}\|] < \infty$ for all t.

A4. Let $y_i = (y_{i1}, \ldots, y_{iT})$ and $X_i = (X'_{i1}, \ldots, X'_{iT})'$. $\{(y_i, X_i)\}_{i=1}^{n}$ are independent and identically distributed (i.i.d.).

Assumption A1 implies that y_{it} is continuously distributed with bounded density. A2 is a rank condition that ensures the parameters $\theta(\tau)$ are uniquely identified. The matrix $\Gamma(\tau)$ will play a crucial role in determining the asymptotic distribution of the QR estimator. A3' assumes the existence of the regressors' first moments. Condition A4 is an assumption on the sampling. Notice that it allows for arbitrary intra-unit dependence and time-series dependence for both X_{it} and y_{it}, but assumes independence across units. The identical distributions of units is also imposed, but could be relaxed using limit theorems for i.n.i.d. data, which is beyond the scope of this paper. These assumptions are standard in the QR literature. The next result formalizes identification.

PROPOSITION 2.1 *Under assumptions A1, A2, A3' and A4, $\theta(\tau)$ is identified in equation* (7).

Given the result in Proposition 2.1, we can propose an estimator for the parameters of interest. The estimator is a pooled linear QR estimator for $[c(\tau), \beta(\tau)', \gamma(\tau)']' = \theta(\tau) \in \mathbb{R}^K$. The estimator is defined as follows:

$$\hat{\theta}(\tau) \equiv \arg\min_{\theta} \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\rho_\tau(y_{it} - X'_{it}\theta), \tag{9}$$

where $\rho_\tau(u) \equiv \{\tau - 1(u \leq 0)\}u$ is the check function (Koenker and Bassett, 1978). We call $\hat{\theta}(\tau)$ the random effects quantile regression (RE-QR) estimator of $\theta(\tau)$. We later show that this estimator is uniformly consistent over a compact range of quantiles $\tau \in \mathcal{T} = [\epsilon, 1-\epsilon]$ for $\epsilon \in (0, 1/2)$, and converges weakly to a tight Gaussian process.

Even though the RE-QR estimator in (9) is consistent under our later assumptions, an additional difficulty in deriving the asymptotic distribution of $\hat{\theta}(\cdot)$ is due to the dependence across time in individual units induced by the unobserved individual effects in (5). To see this, note that (7) implies that

$$\begin{aligned} y_{it} &= c(\tau) + x_{it}'\beta(\tau) + z_i'\gamma(\tau) + v_{it}(\tau) \\ v_{it}(\tau) &= X_{it}'(\theta(U_{it}) - \theta(\tau)), \end{aligned}$$

where $v_{it}(\tau)$ is the QR residual for quantile τ. From the structure of the residual $v_{it}(\tau)$, we can see that the presence of $U_{it} = U(\alpha_i, \epsilon_{it})$ generates dependence of these residuals across time. Variants of this problem have been addressed in the estimation of linear mean panel data models, where the standard-errors are often computed using individual-level clustering.

The particular structure of the asymptotic variance of the QR estimator is important to take into account to obtain correct inference. We use clustered standard error estimates which take into account this time-dependence. They are related to the proposals of Parente and Santos-Silva (2016) and Hagemann (2017) for clustered standard errors in cross-sectional QR models.

3. ASYMPTOTIC THEORY

This section derives the asymptotic properties of the RE-QR estimator over a range of quantiles. To do so, we replace assumption A3' with the following stronger condition.
A3. $\mathrm{E}[\|X_{it}\|^4] < \infty$ for all t.

Existence of fourth moments is used when proving consistency of the standard error, but may be stronger than necessary to establish consistency and asymptotic normality, since only the existence of order $(2+\epsilon)$ moments for some $\epsilon > 0$ is used in our proof.

Using the previous assumptions, we state the main estimation result.

THEOREM 3.1 *Under assumptions A1-A4 and equation* (7), *as $n \to \infty$, $\hat{\theta}(\cdot)$ is uniformly consistent, and*

$$\Gamma(\cdot)\sqrt{n}(\hat{\theta}(\cdot) - \theta(\cdot)) \Rightarrow \mathbf{z}(\cdot),$$

where $\Gamma(\cdot)$ is defined in assumption A2, and $\mathbf{z}(\cdot)$ is a Gaussian process on $\mathcal{T} = [\epsilon, 1-\epsilon]$, $\epsilon \in (0, 1/2)$, with covariance kernel equal to

$$\begin{aligned} \mathrm{E}[\mathbf{z}(\tau)\mathbf{z}(\tau')'] &= V(\tau, \tau') \\ &= \frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathrm{E}\left[(1(v_{it}(\tau) \le 0) - \tau)(1(v_{is}(\tau') \le 0) - \tau')X_{it}X_{is}'\right]. \end{aligned} \tag{10}$$

Theorem 3.1 establishes the limiting distribution of the RE-QR estimator uniformly over the set of quantiles $\tau \in \mathcal{T}$.[8]

[8]We note that these results do not impose restrictions on the relative magnitudes of n and T.

We now turn our attention to the variance-covariance matrix in (10). For given quantiles of interest, the component $V(\tau,\tau)$ can also be written as:

$$\begin{aligned} V(\tau,\tau) &= \mathrm{E}\left[\frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}(1(v_{it}(\tau)\leq 0)-\tau)(1(v_{is}(\tau)\leq 0)-\tau)X_{is}X_{it}'\right] \\ &= \frac{\tau(1-\tau)}{T^2}\sum_{t=1}^{T}\mathrm{E}[X_{it}X_{it}'] \qquad (11) \\ &+ \frac{1}{T^2}\sum_{s\neq t}\mathrm{E}\left[\mathrm{Cov}(1(v_{is}(\tau)\leq 0),1(v_{it}(\tau)\leq 0)|X_{is},X_{it})X_{is}X_{it}'\right], \end{aligned}$$

where the second component disappears if there is no intra-unit dependence of the QR residuals $v_{it}(\tau)$. In other words, dependence of errors is allowed *within* each unit while there is none *between* units. Using simple standard errors for the pooled QR estimator without correcting for the cluster-dependence will produce incorrect inference unless the second term is zero.

We use standard error estimates which are robust to the dependence introduced by the presence of RE and are uniformly consistent over $\mathcal{T}$. To conduct practical inference, we need an estimate of $\Gamma(\tau)$. Let

$$\hat{\Gamma}(\tau) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}X_{it}X_{it}'\frac{1}{h_n}K\left(\frac{\hat{v}_{it}(\tau)}{h_n}\right), \qquad (12)$$

where $\hat{v}_{it}(\tau)=y_{it}-X_{it}'\hat{\theta}(\tau)$ are the estimated residuals, and $K(\cdot)$ be a second order Lipschitz kernel function, and h_n is a bandwidth. This is a variant of the Powell (1991) kernel estimator for QR in cross-sectional models. See also Parente and Santos-Silva (2016) for QR estimators with clustered standard errors. Below, we show that $\hat{\Gamma}(\tau)$ is uniformly consistent without requiring the kernel to be uniform, which we believe is a new result.

The component $V(\tau,\tau')$ can be estimated by

$$\begin{aligned} \hat{V}(\tau,\tau') &= \frac{\min(\tau,\tau')-\tau\tau'}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}X_{it}X_{it}' \\ &+ \frac{1}{nT^2}\sum_{i=1}^{n}\sum_{s\neq t}X_{is}X_{it}'(1(\hat{v}_{is}(\tau)\leq 0)-\tau)(1(\hat{v}_{it}(\tau')\leq 0)-\tau'). \qquad (13) \end{aligned}$$

In the spirit of Angrist, Chernozhukov, and Fernandez-Val (2006), we can also construct standard errors that are robust to misspecification of the conditional quantile function by replacing our estimate of $V(\tau,\tau')$ by

$$\tilde{V}(\tau,\tau') = \frac{1}{nT^2}\sum_{i=1}^{n}\sum_{s=1}^{T}\sum_{t=1}^{T}X_{is}X_{it}'(1(\hat{v}_{is}(\tau)\leq 0)-\tau)(1(\hat{v}_{it}(\tau')\leq 0)-\tau'). \qquad (14)$$

The following result shows that both of these variance matrix estimates are uniformly consistent in $\tau\in\mathcal{T}$.

THEOREM 3.2 *Let K be a second order Lipschitz kernel. Under assumptions A1-A4 and* *equation* (7), $h_n \to 0$ *and* $nh_n^4 \to \infty$ *as* $n \to \infty$,

$$\sup_{\tau \in \mathcal{T}} \left\| \hat{\Gamma}(\tau) - \Gamma(\tau) \right\| = o_p(1),$$
$$\sup_{(\tau,\tau') \in \mathcal{T}^2} \left\| \hat{V}(\tau,\tau') - V(\tau,\tau') \right\| = o_p(1),$$
$$\sup_{(\tau,\tau') \in \mathcal{T}^2} \left\| \tilde{V}(\tau,\tau') - V(\tau,\tau') \right\| = o_p(1).$$

The uniformity of these variance estimates will be important in allowing us to derive tests for hypothesis over many quantiles. These results also allow for inference to be conducted on a single quantile estimate, for example on the median regression estimated coefficients.

Efficient estimation under random effects

In standard linear RE models, the structure of the model implies that there exists estimators that are more efficient than the pooled OLS estimator. It is also possible to extend the idea of efficient estimation for the proposed RE-QR model. Although not pursued in this paper, an efficient estimator for $\theta(\tau)$ can be constructed using existing results. Efficient estimators for QR estimators were developed by Newey and Powell (1990) and Zhao (2001). To obtain a QR estimator that attains the efficiency bound, the sample objective function's terms are weighted by conditional densities of y_{it} given X_i. These weights can be nonparametrically estimated in a first step. A generalization of these results to seemingly unrelated quantile regressions is developed in Jun and Pinkse (2009). Their setup includes that of equation (7) and therefore an efficient estimator can be developed under their assumptions. A *k*-nearest neighbor approach is used to estimate the weight nonparametrically. We leave a more thorough discussion of efficient estimation in our model for future work.

4. INFERENCE

In this section, we turn our attention to inference in the random effects quantile regression (RE-QR) model, and suggest a Wald type test for linear hypotheses. In addition, we propose a novel general uniform test for linear hypotheses in QR models. We also propose a new test of the RE assumption which relies on a set of auxiliary quantile regressions.

4.1. *Testing the coefficients*

General hypotheses on the vector $\theta(\tau) = (\beta(\tau)', \gamma(\tau))'$ can be accommodated by Wald-type tests. It is possible to formulate a wide variety of tests using variants of Wald tests, from simple tests on a single QR coefficient to joint tests involving many covariates and distinct quantiles at the same time. The Wald process and the associated limiting theory provide a natural foundation for testing the hypothesis $H_0 : R\theta(\tau) = r$. Wald tests designed for this purpose were suggested by Koenker and Bassett (1982a), Koenker and Bassett (1982b), and Koenker and Machado (1999). We first consider tests for selected quantiles of interest. Later we introduce a test for linear hypothesis over a range of quantiles $\tau \in \mathcal{T}$.

The Wald statistic for given τ can be constructed as

$$\mathcal{W}_n(\tau) = n(R\hat{\theta}(\tau) - r)'[R\hat{\Omega}(\tau)R']^{-1}(R\hat{\theta}(\tau) - r), \tag{15}$$

where $\hat{\Omega}(\tau) \equiv \hat{\Gamma}(\tau)^{-1}\hat{V}(\tau,\tau)\hat{\Gamma}(\tau)^{-1}$ is a consistent estimator of $\Omega(\tau) = \Gamma(\tau)^{-1}V(\tau,\tau)\Gamma(\tau)^{-1}$, the asymptotic variance of $\hat{\theta}(\tau) - \theta(\tau)$.

If one is interested in testing $H_0 : R\theta(\tau) = r$ at a particular quantile τ, a Chi-square test can be conducted based on the statistic $\mathcal{W}_n(\tau)$. Under H_0, the statistic $\mathcal{W}_n$ is asymptotically χ^2_q with q-degrees of freedom, where q is the rank of the matrix R.

LEMMA 4.1 *(Wald Test Inference). Under $H_0 : R\theta(\tau) = r$, assumptions A1-A4, equation* (7)*, and $h_n \to 0$, $nh_n^4 \to \infty$*

$$\mathcal{W}_n(\tau) \overset{a}{\sim} \chi^2_q,$$

where q is the rank of the matrix R.

More general hypotheses are also accommodated by the Wald approach. Let $\theta^M \equiv (\theta(\tau_1)', ..., \theta(\tau_M)')$ and define the null hypothesis as $H_0 : R\theta^M = r$. The test statistic is the same Wald test as in (16). By Theorem 3.1, the joint distribution of $\hat{\theta} - \theta$ is equal to:

$$\sqrt{n}(\hat{\theta}^M - \theta^M) \Rightarrow N(0, \Omega^M),$$

where

$$\Omega^M = \begin{pmatrix} \Gamma(\tau_1)^{-1}V(\tau_1,\tau_1)\Gamma(\tau_1)^{-1} & \cdots & \Gamma(\tau_1)^{-1}V(\tau_1,\tau_M)\Gamma(\tau_M)^{-1} \\ \vdots & \ddots & \vdots \\ \Gamma(\tau_M)^{-1}V(\tau_M,\tau_1)\Gamma(\tau_1)^{-1} & \cdots & \Gamma(\tau_M)^{-1}V(\tau_M,\tau_M)\Gamma(\tau_M)^{-1} \end{pmatrix}.$$

We can obtain a consistent estimate of Ω^M using the results of Theorem 3.2. The Wald statistic is asymptotically Chi-squared distributed under H_0 with degrees of freedom equal to the rank of the matrix R. This formulation accommodates, for instance, tests for the equality of several slope coefficients across several quantiles.

4.2. *Uniform tests*

This section introduces a novel uniform test for QR models. The extension of the previous tests to the uniform case is not straightforward as in the standard cross-section QR case. As discussed in Hagemann (2017), Wald inference based on the clustered variance-covariance matrix is not simple in the uniform across quantiles case because the limiting process is no longer a standard Brownian Bridge and depends on nuisance parameters.

As in Portnoy (1984) and Gutenbrunner and Jureckova (1992), in Theorem 3.1 we show that the QR process is tight, and thus, the limiting variate viewed as a function of τ is a Gaussian process over $\tau \in \mathcal{T}$. Nevertheless, unlike in the cross-sectional setting, the presence of clustering generates the following additional term in the variance-covariance matrix (11)

$$\frac{1}{T^2}\sum_{s\neq t} \mathrm{E}\left[\mathrm{Cov}(1(v_{is}(\tau) \leq 0), 1(v_{it}(\tau') \leq 0)|X_{is}, X_{it})X_{is}X_{it}'\right],$$

which makes the limiting Gaussian process of the standard normalized Wald statistic differ from the usual Brownian bridge.

Thus, in this section, we develop an alternative method for uniform testing in QR. We consider a Wald-test on a grid of points where the number of grid points increases to infinity as the sample size increases. The proposed test statistic is normalized in a way that accounts for the increasing number of restrictions being tested.

Suppose we are interested in testing a uniform hypothesis of the form $H_0 : R\theta(\tau) - r = 0$ for all $\tau \in \mathcal{T}$. Let $\mathcal{T}_M = \{\tau_1, \ldots, \tau_M\}$ be a sequence of grids of M quantiles over $\mathcal{T}$ where $\mathcal{T}_M \subseteq \mathcal{T}_{M+1}$ and such that this grid becomes dense in $\mathcal{T}$ as $M \to \infty$. Thus, we can see that $R\theta(\tau) - r$ for all $\tau \in \mathcal{T}$ if and only if it is true for all $\tau \in \mathcal{T}_M$ as $M \to \infty$.

Let $\theta^M = (\theta(\tau_1)', \ldots, \theta(\tau_M)')'$, a $KM \times 1$ vector, $R^M = I_M \otimes R$, an $qM \times KM$ matrix which has rank qM since the rank of R is equal to q, and $r^M = \iota_M \otimes r$, where ι_M denotes a $M \times 1$ vector of ones. The restrictions can be represented in vector form as

$$R^M \theta^M = r^M.$$

Let Ω^M denote the asymptotic variance of $\hat{\theta}^M$ as denoted above.

Recall that, for a given M, the Wald statistic is

$$\mathcal{W}_n^M = n(R^M\hat{\theta}^M - r^M)'(R^M\hat{\Omega}^M R^{M\prime})^{-1}(R^M\hat{\theta}^M - r^M),$$

which asymptotically follows a χ^2_{qM} distribution. This test statistic does not have power against deviations of the uniform null hypothesis at points outside the fixed grid $\mathcal{T}^M$, so it is not a proper uniform test. To allow the test to have power against these deviations, we must allow M to grow with the sample size. Allowing this would mean that $\mathcal{W}_n^M$ diverges to ∞, hence we propose a new normalization which allows the limiting distribution to be normal with known variance.

Since a Chi-squared random variable is a sum of independent squared normals, one can see that $\frac{\chi^2_{qM}}{M} \xrightarrow{p} q$. In an analogy to the central limit theorem, a normalized Chi-squared distribution will asymptotically follow a normal distribution. Based on this observation, we propose the following test-statistic:

$$W_{n,M} = \sqrt{M}\left(\frac{\mathcal{W}_n^M}{M} - q\right).$$

We investigate how this statistic behaves as M and n diverge jointly. To do this, we consider the following two additional assumptions

A5. $K(\cdot)$ is a second order Lipschitz kernel function,

A6. All moments of X_{it} exist for all $t = 1, \ldots, T$.

Assumption A5 is standard and is satisfied by many kernels such as the Gaussian or Epanechnikov kernel. Assumption A6 can be weakened to $\mathrm{E}[\|X_{it}\|^{4+\delta}] < \infty$ for $\delta > 0$ at the cost of more complex rate conditions in the statement of the next theorem. We impose here the existence of all moments to simplify the statement of the theorem. Also note that this is a stronger version of A3. We now state the uniform testing result.

THEOREM 4.1 *Let q be the rank of R in the null hypothesis $H_0 : R\theta(\tau) - r = 0$ for $\tau \in \mathcal{T}$. Let $\mathcal{T}_M$ be a sequence of grid points such that $\mathcal{T}_M \subseteq \mathcal{T}_{M+1} \subseteq \cdots \subseteq \mathcal{T}$ such that $\mathcal{T}_M$ becomes dense in $\mathcal{T}$ as $M \to \infty$. Under assumptions A1-A6, equation* (7)*, and*

$n, M \to \infty$, $h_n \to 0$, $M^3 n^{-1} h_n^{-4} \to 0$, $M^3 h_n^4 \to 0$, *we have*

$$W_{n,M} \Rightarrow N(0, 2q).$$

The test statistic $W_{n,M}$ is asymptotically pivotal since its asymptotic distribution contains no nuisance parameters, which allows for uniform inference without using the bootstrap. The rate conditions provide some guidance on how to select the bandwidth term h_n used in equation (13), and the number of grid points M, which implicitly depends on the sample size n. These restrictions imply that $nh_n^2 \to \infty$ as in Theorem 3.2.

Letting $M = n^a$ and $h_n = n^{-b}$ for $a, b > 0$, we see that our restrictions imply that $0 < b < \frac{1-3a}{4}$ and $b > \frac{3a}{4}$. The fastest rate of growth possible for the number of restrictions is $M = o(n^{1/6})$, a relatively slow rate. The rate conditions we provide are sufficient conditions, but may be significantly different from the necessary rate conditions. In fact, in our Monte-Carlo study the uniform test displays appropriate size for various grids of quantiles. Additional improvements in rate conditions may be obtained if one uses a higher-order kernel to estimate $\Gamma(\tau)$.

In the Monte-Carlo section, we show evidence that our testing procedures possess good finite sample properties, including correct size when the sample size is sufficiently large.

4.3. *Testing the random effects assumption*

Since the FE-QR estimator of Kato, Galvao, and Montes-Rojas (2012) is not consistent for the same coefficients as the RE-QR estimator, it is not appropriate to conduct a Hausman test (Hausman (1978)) for the presence of RE by comparing these two estimators. We instead propose a test based on a Wald test using an auxiliary QR model as

$$y_{it} = c(\tau) + x_{it}'\beta(\tau) + z_i'\gamma(\tau) + x_{i\cdot}'\lambda(\tau) + v_{it}(\tau), \tag{16}$$

where $x_{i\cdot} = \frac{1}{T}\sum_{t=1}^{T} x_{it}$ are time-averages for x_{it}, the set of time-varying regressors. Auxiliary regressions as an alternative to direct Hausman tests were suggested in Mundlak (1978).

Under the null hypothesis of RE, the coefficients $\lambda(\tau)$ in (17) are equal to zero due to the correct specification of the regression model for all τ. Under the alternative, α_i is not independent of X_i and the unobserved $U_{it} = U(\alpha_i, \epsilon_{it})$ will generally be related to X_i which implies that $Q_{U_{it}}(\tau|X_i)$ depends on X_i for at least some $\tau \in (0,1)$. With this insight, we see that the conditional quantile of y_{it} given X_i is equal to

$$Q_{y_{it}}(\tau|X_i) = X_{it}'\theta(Q_{U_{it}}(\tau|X_i)),$$

which then implies that the conditional quantile depends not only on contemporaneous regressors X_{it}, but all time values through X_i. For example, if X_{it} consists of a single, continuously distributed variable, we expect a failure of the RE assumption would lead to $\frac{\partial}{\partial X_{is}} Q_{y_{it}}(\tau|X_i) \neq 0$, for some $s \neq t$, while it would be zero under the null hypothesis of RE. The term $x_{i\cdot}'\lambda(\tau)$ is meant to capture the non-contemporaneous dependence of the quantile function on X_i.

We test this implication of a departure from RE by testing that the coefficients on the linear term $x_{i\cdot}$ in equation (17) are equal to zero using a uniform Wald test. The stated null hypothesis can be written as

$$H_0 : \lambda(\tau) = 0, \text{ for all } \tau \in \mathcal{T}.$$

We can implement this test using our previous results on uniform inference. Let k_1 be the number of time-varying regressors, and k_2 be the number of time-invariant regressors, such that $k_1 + k_2 + 1 = K$. Let R be a $k_1 \times (K + k_1)$ with zeros in the first K columns and an identity matrix of size k_1 in the last k_1 columns. The null hypothesis can then be formulated as $H_0 : R\theta(\tau) = 0$. Letting $\mathcal{T}^M$ denote a grid of M quantiles where M increases with the sample size and $\hat{\lambda}^M = (\hat{\lambda}(\tau_1)', \ldots, \hat{\lambda}(\tau_M)')'$, we can write our test-statistic as

$$W_{n,M} = \sqrt{M}\left(\frac{n}{M}\hat{\lambda}^{M\prime}(R^M \Omega^M R^{M\prime})^{-1}\hat{\lambda}^M - k_1\right) \tag{17}$$

which has a $N(0, 2k_1)$ limiting distribution under H_0 using Theorem (4.1).

This test does not possess power against all alternatives because one can construct data-generating processes where U_{it} depends on X_i nonlinearly such that the linear QR coefficients on $x_{i\cdot}$ are zero despite the failure of the random effects assumption. The power of this test against these specific alternatives can be improved by regressing on x_i and testing all coefficients on x_{is}, $s \neq t$ being equal to zero, uniformly in τ. The cost of this increase in power is a reduced size of the test due to the large number of restrictions being tested. We view the test on coefficients of $x_{i\cdot}$ as a satisfactory compromise.

Also, if the test reject the random effect assumption, one may consider a non-separable model with correlated random coefficients of Graham, Hahn, Poirier, and Powell (2018), as discussed in section 2.2, or a linear in α_i FE QR model. Both impose additional restrictions on the model.

We note that one can also implement a point-wise RE test by testing $H_0 : \lambda(\tau) = 0$ for a fixed τ, which we later use in the Monte-Carlo study and empirical example.

5. MONTE CARLO SIMULATIONS

This section conducts Monte Carlo simulations to investigate the finite sample performance of the proposed methods. We study the random effects quantile regression (RE-QR) estimator, and also the cluster robust variance-covariance matrix estimator for inference. We also consider the performance of the RE assumption test.

5.1. *Design*

We consider the following design for the experiments. We generate data from a version of model (7) with time-invariant regressors as

$$\begin{aligned} y_{it} &= \beta_1 + \beta_2 x_{it} + \beta_3 z_i + (1 + \gamma x_{it} + \gamma z_i)u_{it}, \\ u_{it} &= \alpha_i + \varepsilon_{it}, \end{aligned} \tag{18}$$

where $z_i \sim$ i.i.d. χ^2_3, and $x_{it} = \phi_1 z_i + \phi_2 \alpha_i + w_{it}$ with $w_{it} \sim$ i.i.d. χ^2_3. The parameters ϕ_1 and ϕ_2 are important. We set $\phi_1 = 0.5$ throughout the simulations, which guarantees that x_{it} is correlated with z_i, ensuring the existence of correlation between the time-varying and time-invariant regressors. The second parameter, ϕ_2, controls the correlation between x_{it} and the unobserved effects, α_i. This parameter is important when studying the proposed tests in Section 4. When $\phi_2 = 0$ the RE hypothesis is satisfied, but when $\phi_2 \neq 0$ the RE-QR estimator is not expected to be consistent.

We use two distributions for the unobservables, such that $\alpha_i \sim$ i.i.d. F, $\varepsilon_{it} \sim$ i.i.d. F with $F = N(0, 1)$, and χ^2_3. We set the parameters $\beta_{01} = \beta_{02} = \beta_{03} = 1$, and set $\gamma = 0.5$ to control the heterogeneity. Since γ is different from zero, there is heterogeneity in all coefficients.[9]

In this location-scale shift model, $\beta_{01}(\tau) = \beta_{01} + F_u^{-1}(\tau)$, $\beta_{02}(\tau) = \beta_{02} + \gamma F_u^{-1}(\tau)$, and $\beta_{03}(\tau) = \beta_{03} + \gamma F_u^{-1}(\tau)$, with F_u^{-1} being the quantile function of $u_{it} \sim$ i.i.d. $N(0, 2)$ and $u_{it} \sim$ i.i.d. χ^2_6 when drawing from the standard Normal and Chi-square distributions, respectively.

We consider several sample sizes and quantiles, where $n \in \{100, 200, 500\}$, $T \in \{2, 5, 10\}$ and $\tau \in \{0.25, 0.5, 0.75\}$. The number of repetitions for each experiment is 5,000.

As discussed above, the computation of the cluster robust variance-covariance matrix is a very important step for practical inference. We use the estimators $\hat{V}$ from equation (14) and $\hat{\Gamma}$ from equation (13), with Gaussian kernel, and the bandwidth is selected using the Hall-Sheather rule.[10] First, we evaluate the RE-QR estimator in terms of bias. To do so, we set $\phi_2 = 0$ throughout these experiments. Second, we investigate the small sample properties of the inference procedures. We compute the following estimators: (i) the standard deviations (SD) of RE-QR, which are computed as the sample standard deviation of the vector of estimates of the corresponding parameters; (ii) the clustered robust standard errors (CSE), which are computed from equations (13) and (15) above; (iii) the standard errors without clustering correction (SE), which are computed from the simple pooled QR estimation. We report SE to illustrate the importance of correcting for the correlation induced by the RE.

Third, we evaluate the finite sample performance of the proposed tests for the RE hypothesis. To investigate the empirical size of the test, we set $\phi_2 = 0$. To study the empirical power, we vary the parameter ϕ_2 on the grid $\{0.1, 0.2, ..., 3\}$.

5.2. *Results for bias and standard errors*

First, we discuss the bias results. Tables I and II present the results for the location-scale shift model with Normal and Chi-square distributions, respectively. For this model, the results show a small bias for β_1 for small panels and every τ. However, as expected, the bias disappears as sample size grows. There is a bias reduction when the sample grows in either cross-section or time-series dimension. These results suggest that the RE-QR estimator performs well in small samples even when the time-series dimension is small.

[9]We also considered a location model where $\gamma = 0$. Results were similar and are available upon request.

[10]The Hall-Sheather bandwidth is: $h_n = n^{-1/3} z_\alpha^{2/3} [1.5 s(\tau)/s''(\tau)]^{1/3}$ where z_α satisfies $\Phi(z_\alpha) = 1 - \alpha/2$ and $\alpha = 0.05$, and $s/s'' = f^2/[2(f'/f)^2 + (f'/f - f''/f)]$ with $f = \phi$ and $(f'/f)(F^{-1}(\tau)) = \Phi^{-1}(\tau)$.

Table I: Bias. Location-Scale shift model ($\gamma = 0.5$). $N(0,1)$.

		$\beta_1(\tau)$			$\beta_2(\tau)$			$\beta_3(\tau)$		
		$T=2$	$T=5$	$T=10$	$T=2$	$T=5$	$T=10$	$T=2$	$T=5$	$T=10$
$\tau=0.25$	$n=100$	-0.056	-0.028	-0.030	0.007	0.007	0.003	0.013	0.007	0.012
	$n=200$	-0.014	-0.002	-0.023	0.004	0.002	0.004	0.007	0.003	0.008
	$n=500$	-0.006	-0.008	-0.005	0.004	0.001	-0.001	-0.001	0.003	0.003
$\tau=0.50$	$n=100$	-0.011	-0.005	-0.003	-0.001	0.003	-0.002	0.004	0.001	0.004
	$n=200$	0.017	0.001	-0.005	-0.001	0.004	0.000	-0.001	-0.004	0.003
	$n=500$	0.004	-0.003	-0.002	-0.001	-0.001	-0.001	-0.001	0.002	0.002
$\tau=0.75$	$n=100$	0.042	0.035	0.024	-0.010	-0.006	-0.004	-0.010	-0.007	-0.008
	$n=200$	0.044	0.015	0.006	-0.006	-0.001	-0.002	-0.011	-0.005	-0.001
	$n=500$	0.014	0.004	0.002	-0.003	-0.002	0.000	-0.003	0.000	-0.002

Table II: Bias. Location-Scale shift model ($\gamma = 0.5$). χ^3_2.

		$\beta_1(\tau)$			$\beta_2(\tau)$			$\beta_3(\tau)$		
		$T=2$	$T=5$	$T=10$	$T=2$	$T=5$	$T=10$	$T=2$	$T=5$	$T=10$
$\tau=0.25$	$n=100$	-0.079	-0.065	-0.034	0.024	0.012	0.004	0.042	0.034	0.040
	$n=200$	-0.055	-0.029	-0.001	0.020	0.006	0.005	0.017	0.020	0.008
	$n=500$	0.004	-0.017	-0.009	0.004	0.003	0.005	0.001	0.010	0.006
$\tau=0.50$	$n=100$	0.006	0.008	0.000	0.017	0.002	-0.002	0.013	0.017	0.029
	$n=200$	0.003	0.013	0.013	0.008	0.000	0.004	0.005	0.008	0.005
	$n=500$	0.010	0.000	0.010	0.003	0.000	0.000	-0.004	0.007	0.004
$\tau=0.75$	$n=100$	0.253	0.089	0.073	-0.021	-0.008	-0.012	-0.025	-0.005	0.006
	$n=200$	0.104	0.091	0.054	-0.017	-0.013	-0.004	0.002	-0.011	0.001
	$n=500$	0.066	0.017	0.029	-0.010	-0.006	-0.004	-0.008	0.001	0.001

Second, we discuss the results on the estimation of the standard deviation. Tables III and IV present the results for SD, SE, and CSE for the location-scale shift model with Normal and Chi-square distributions, respectively. In general, Tables III and IV show evidence that the SE underestimates the SD for all quantiles, the downward bias being more pronounced for longer time-series. Thus, by correcting for the induced correlation due to the unobserved heterogeneity, the CSE provides a consistent estimate of the standard deviation.

Table III: Standard Deviation, Standard Error, and Clustered Standard Error. Location-scale shift model ($\gamma = 0.5$). $N(0,1)$.

			$\beta_1(\tau)$			$\beta_2(\tau)$			$\beta_3(\tau)$		
			$T=2$	$T=5$	$T=10$	$T=2$	$T=5$	$T=10$	$T=2$	$T=5$	$T=10$
	$n=100$	SD	1.092	0.826	0.709	0.326	0.203	0.150	0.445	0.365	0.319
		SE	1.347	0.765	0.510	0.360	0.216	0.148	0.440	0.267	0.184
		CSE	1.419	0.960	0.783	0.354	0.217	0.154	0.476	0.363	0.321
$\tau=0.25$	$n=200$	SD	0.766	0.577	0.495	0.224	0.142	0.107	0.319	0.252	0.225
		SE	0.874	0.509	0.344	0.244	0.149	0.103	0.300	0.184	0.128
		CSE	0.929	0.645	0.533	0.243	0.151	0.108	0.329	0.256	0.227
	$n=500$	SD	0.486	0.360	0.311	0.141	0.092	0.067	0.198	0.155	0.142
		SE	0.508	0.305	0.210	0.148	0.092	0.064	0.184	0.115	0.081
		CSE	0.545	0.389	0.327	0.149	0.094	0.068	0.205	0.161	0.144
	$n=100$	SD	1.004	0.766	0.674	0.295	0.188	0.141	0.410	0.340	0.303
		SE	1.410	0.780	0.509	0.360	0.209	0.142	0.435	0.257	0.175
		CSE	1.524	1.015	0.814	0.360	0.212	0.149	0.485	0.366	0.319
$\tau=0.50$	$n=200$	SD	0.710	0.533	0.464	0.205	0.134	0.097	0.294	0.235	0.213
		SE	0.896	0.508	0.337	0.239	0.142	0.098	0.291	0.175	0.121
		CSE	0.970	0.662	0.541	0.239	0.145	0.103	0.326	0.251	0.223
	$n=500$	SD	0.447	0.337	0.295	0.132	0.086	0.062	0.184	0.149	0.135
		SE	0.509	0.298	0.202	0.142	0.087	0.060	0.176	0.108	0.075
		CSE	0.552	0.388	0.323	0.143	0.089	0.064	0.198	0.155	0.139
	$n=100$	SD	1.082	0.814	0.720	0.319	0.205	0.151	0.447	0.360	0.323
		SE	1.348	0.763	0.510	0.359	0.215	0.148	0.442	0.265	0.184
		CSE	1.445	0.969	0.785	0.359	0.219	0.156	0.488	0.367	0.321
$\tau=0.75$	$n=200$	SD	0.767	0.572	0.496	0.230	0.145	0.107	0.316	0.251	0.227
		SE	0.871	0.508	0.344	0.242	0.148	0.103	0.299	0.184	0.128
		CSE	0.936	0.647	0.535	0.243	0.151	0.109	0.333	0.257	0.228
	$n=500$	SD	0.482	0.361	0.316	0.143	0.093	0.068	0.197	0.159	0.145
		SE	0.509	0.305	0.210	0.148	0.092	0.064	0.185	0.115	0.080
		CSE	0.548	0.390	0.326	0.149	0.094	0.068	0.206	0.161	0.143

Table IV: Standard Deviation, Standard Error, and Clustered Standard Error. Location-scale shift model ($\gamma = 0.5$). χ^3_2.

			$\beta_1(\tau)$			$\beta_2(\tau)$			$\beta_3(\tau)$		
			$T=2$	$T=5$	$T=10$	$T=2$	$T=5$	$T=10$	$T=2$	$T=5$	$T=10$
	$n=100$	SD	1.885	1.382	1.199	0.536	0.352	0.257	0.741	0.598	0.553
		SE	2.842	1.549	1.004	0.720	0.409	0.273	0.867	0.499	0.337
		CSE	2.975	1.945	1.547	0.705	0.408	0.281	0.924	0.678	0.586
$\tau = 0.25$	$n=200$	SD	1.296	0.980	0.827	0.381	0.246	0.181	0.540	0.419	0.383
		SE	1.787	1.001	0.659	0.467	0.274	0.186	0.569	0.336	0.229
		CSE	1.898	1.266	1.018	0.464	0.276	0.193	0.620	0.463	0.403
	$n=500$	SD	0.823	0.608	0.526	0.238	0.155	0.114	0.336	0.269	0.244
		SE	1.002	0.577	0.387	0.274	0.164	0.113	0.335	0.203	0.141
		CSE	1.071	0.733	0.598	0.274	0.167	0.119	0.370	0.282	0.249
	$n=100$	SD	2.324	1.768	1.541	0.673	0.437	0.313	0.945	0.772	0.715
		SE	3.219	1.767	1.147	0.812	0.471	0.320	0.974	0.577	0.394
		CSE	3.494	2.319	1.863	0.812	0.478	0.335	1.089	0.830	0.732
$\tau = 0.50$	$n=200$	SD	1.625	1.247	1.080	0.476	0.312	0.227	0.681	0.541	0.496
		SE	2.039	1.147	0.763	0.538	0.320	0.221	0.654	0.394	0.274
		CSE	2.219	1.509	1.240	0.540	0.327	0.234	0.737	0.571	0.513
	$n=500$	SD	1.027	0.880	0.683	0.298	0.193	0.143	0.429	0.343	0.320
		SE	1.150	0.768	0.453	0.321	0.196	0.136	0.395	0.243	0.170
		CSE	1.252	0.669	0.736	0.323	0.201	0.145	0.447	0.354	0.320
	$n=100$	SD	3.226	2.473	2.158	0.939	0.612	0.446	1.313	1.071	0.998
		SE	3.352	1.967	1.335	0.930	0.583	0.409	1.148	0.727	0.511
		CSE	3.629	2.542	2.122	0.936	0.598	0.438	1.287	1.034	0.933
$\tau = 0.75$	$n=200$	SD	2.317	1.730	1.514	0.658	0.432	0.316	0.948	0.756	0.705
		SE	2.227	1.336	0.929	0.651	0.410	0.290	0.813	0.513	0.366
		CSE	2.419	1.740	1.492	0.656	0.423	0.312	0.917	0.738	0.678
	$n=500$	SD	1.446	1.092	0.968	0.421	0.272	0.200	0.598	0.483	0.444
		SE	1.336	0.827	0.582	0.411	0.260	0.184	0.516	0.327	0.232
		CSE	1.452	1.083	0.941	0.414	0.269	0.198	0.583	0.475	0.435

5.3. *Results for the random effects test*

5.3.1. *Fixed quantiles*

We present the empirical size and power for the proposed test for the RE hypothesis. To compute the size we set $\phi_2 = 0$ and use a nominal size of 5%. To calculate the empirical power we set $\phi_2 \in \{0.1, 0.2, ..., 3\}$. Table V collects the results for empirical size and Figure 1 displays the empirical power functions when varying ϕ_2. The plot shows the results for the location-scale model and different sample sizes.

In Table V we report the empirical sizes for different samples and location ($\gamma = 0$) and location-scale models ($\gamma = 0.5$). First, we observe that the empirical sizes ($\phi_2 = 0$) are close to the nominal 5%. The size improves with the sample size, in particular one can see that for a small sample $n = 100$ and $T = 2$, we have undersized results. However, the empirical sizes approach the nominal size as samples increase in both direction, and for a large panel $T = 500$ and $T = 10$, sizes are very close to the nominal 5%. Overall, Table V shows that the test for the RE assumption has very good size properties in small samples.

The empirical power functions are displayed in Figure 1. We vary ϕ_2 to compute the empirical power function. The results show that the power of the test improves substan-

TABLE V

SIZE OF THE POINTWISE TEST FOR THE RANDOM EFFECTS HYPOTHESIS ($\phi_2 = 0$) FOR THE LOCATION-SCALE ($\gamma = 0.5$) SHIFT MODEL. NOMINAL SIZE 5%.

		$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$
$n = 100$	$T = 2$	0.029	0.015	0.021
	$T = 5$	0.033	0.022	0.032
	$T = 10$	0.042	0.033	0.043
$n = 200$	$T = 2$	0.027	0.018	0.027
	$T = 5$	0.037	0.028	0.035
	$T = 10$	0.043	0.033	0.041
$n = 500$	$T = 2$	0.040	0.024	0.035
	$T = 5$	0.042	0.033	0.037
	$T = 10$	0.050	0.041	0.043

Figure 1: Empirical power function when varying ϕ_2 for different sample sizes.

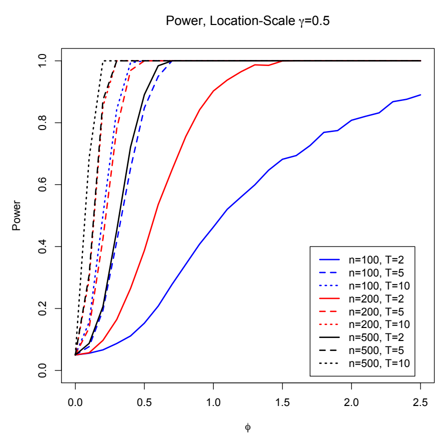


TABLE VI

SIZE OF THE UNIFORM TEST FOR THE RANDOM EFFECTS HYPOTHESIS ($\phi_2 = 0$) FOR THE LOCATION-SCALE ($\gamma = 0.5$) SHIFT MODEL. NOMINAL SIZE 5%.

		τ-grid (1)	τ-grid (2)	τ-grid (3)
$n = 100$	$T = 2$	0.027	0.014	0.089
	$T = 5$	0.067	0.064	0.121
	$T = 10$	0.094	0.101	0.137
$n = 200$	$T = 2$	0.037	0.015	0.030
	$T = 5$	0.053	0.043	0.076
	$T = 10$	0.062	0.064	0.094
$n = 500$	$T = 2$	0.041	0.022	0.018
	$T = 5$	0.052	0.040	0.039
	$T = 10$	0.054	0.051	0.046

Notes: quantile τ-grid (1): $\{0.1, 0.3, 0.5, 0.7, 0.9\}$; τ-grid (2): $\{0.1, 0.2, 0.3, ..., 0.8, 0.9\}$; τ-grid (3): $\{0.10, 0.15, 0.20, ..., 0.85, 0.90\}$.

tially as the sample size increases. The main point is that as the parameter ϕ_2 increases, which measures correlation among the observable covariates and the unobservable heterogeneity, the probability of the test rejecting the RE hypothesis also increases.

5.3.2. *Uniform test*

We now use the uniform test in (18). As in the previous case, to compute the size we set $\phi_2 = 0$ and use nominal size of 5%. The results for empirical size are displayed in Table VI. We present results using following three different grids for the quantiles: (i) $\{0.1, 0.3, 0.5, 0.7, 0.9\}$; (ii) $\{0.1, 0.2, 0.3, ..., 0.8, 0.9\}$; (iii) $\{0.10, 0.15, 0.20, ..., 0.85, 0.90\}$.

The results in Table VI show that the empirical size approximates the nominal size of 5%, especially for large cross-sections. We can observe some size distortions for small cross-sections and small time series, for example when $n = 100$ and $T = 10$. In addition, as the result in Theorem 4.1 describes, the choice of the number of quantiles in the grid is important and depends on the sample size. In this example, we can see that for smaller samples, the coarse grid produces better results, but as the sample increases a finer grid produces results where the empirical sizes are close to the nominal 5%.

Overall the simulations show the usefulness of our methods, estimation, standard inference procedures, and testing for the RE hypothesis. The results suggest the proposed methods have good finite sample performance, leading to reliable, powerful, and computationally attractive inference.

6. APPLICATION

This section illustrates the usefulness of the new proposed methods with an empirical example. We study the roles of education and ability in wage determination. Families and policy makers implement various strategies to enhance an individual's capacity to succeed in the labor market. Investment in human capital is one of the most important channels to achieve this goal. A large literature has documented that workers with higher educational attainment achieve higher earnings, and that this wage differential has been increasing over time. However, very little is known about distributional impacts of education and ability. Our application focuses on documenting the heterogeneity in these changes across quantiles of the conditional distribution of wages. We will later show that the changes in returns to skill for women vary substantially across quantiles, being positive for large quantiles.

The problem of measuring returns to education is an important research area. For examples of comprehensive studies, see Card (1995), Card (1999), and Harmon and Oosterbeek (2000). The large volume of research in this area has been explained by both the interest in the causal effect of education on earnings and the inherent difficulty in measuring this effect. The difficulty arises for several reasons. The classical one is the fact that unobserved factors, such as ability, are probably related to both educational level and earnings. There have been a number of approaches to deal with this problem. There is a literature trying to resolve the problem by using panel data and controlling for individual fixed effects (FE). However, once controlling for individual-specific effects, one cannot include other important time-invariant regressors such as gender and race, for example. Measures of ability are incorporated to proxy for unobserved effects.

Further related work includes Bishop (1991), Blackburn and Neumark (1993), and Grogger and Eide (1995). All of these studies decompose the increasing return to schooling using panel data or repeated cross-sections data and therefore cannot simultaneously identify age, cohort, and time effects. These studies require further parametric assumptions to conclude whether the estimated increase in return to ability is due to changes in the value of cognitive skills or because ability becomes more valuable with work experience. Chay and Lee (2000) and Taber (2001) examine patterns of wage dispersion. More recently, Castex and Dechter (2014) use cross-decade comparisons of the returns to schooling and cognitive ability to estimate changes in returns to formal education and cognitive skills.

This paper follows a specification similar to Castex and Dechter (2014) and estimates wage functions using two panel data waves (the NLSY79 and NLSY97) while controlling for cognitive skills. As in Castex and Dechter (2014), we impose a random effects (RE) assumption on the model. Nevertheless, our study extends the previous works by applying the proposed RE-QR methods, where we are able to uncover heterogeneity on the impacts of skills and education along the conditional quantile function of earnings.

6.1. *Data*

The data are taken from Castex and Dechter (2014), which are collected from the 1979 and 1997 panel waves of the National Longitudinal Survey of Youth (NLSY). NLSY79 is a nationally representative panel data sample of 12,686 young men and women who were 14–22 years old in 1979, and NLSY97 contains 8,984 individuals who were 12–16 years old in 1997. Observations are pooled into panels for 1980–91 for NLSY79 and for 1999–2008 for NLSY97.

The data contain information on individuals' measures of cognitive ability, education, labor market activity, and other family and personal characteristics. Individuals enrolled in school and in military service are excluded. We only consider individuals who have achieved their highest degree of education, work 20 hours or more per week, and earn real hourly wages within the range of \$3–\$100 (in 2007 prices, deflated by CPI).[11] We also

[11]Note that excluding non-workers may lead to sample selection issues, especially for the 1979 female subsample. While important, we do not address this concern in the empirical application. See Arellano and Bonhomme (2017) for a survey of applicable methods.

TABLE VII

DESCRIPTIVE STATISTICS.

	Men				Women			
	NLSY79		NLSY97		NLSY79		NLSY97	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Ln(real hourly wage rate)	2.50	0.45	2.51	0.44	2.34	0.43	2.38	0.42
AFQT score	-0.36	0.99	-0.15	0.98	-0.19	0.92	-0.16	0.97
High School	0.62	0.48	0.72	0.45	0.69	0.46	0.66	0.47
Associate's	0.04	0.19	0.04	0.18	0.06	0.24	0.04	0.21
Bachelor's	0.10	0.30	0.12	0.33	0.12	0.32	0.20	0.40
Master's	0.01	0.08	0.00	0.07	0.01	0.09	0.01	0.12
Experience	6.47	2.70	4.02	2.69	5.96	2.58	3.61	2.57
Black	0.22	0.42	0.19	0.40	0.22	0.42	0.26	0.44
Unemployment	0.07	0.01	0.05	0.00	0.07	0.01	0.05	0.00
Metro status	0.77	0.42	0.55	0.50	0.77	0.42	0.60	0.49
ln(family income)	10.54	0.75	10.79	1.11	10.63	0.73	10.56	1.21
Family intact	0.73	0.44	0.66	0.47	0.72	0.45	0.61	0.49
Mother's education	10.89	2.75	12.58	2.72	10.95	2.58	12.48	2.81
Father's education	10.74	3.35	12.33	2.98	10.93	3.38	12.35	2.98
N	9,108		7,608		7,594		6,757	

exclude individuals with missing information on key variables, and confine the analysis to individuals with ages ranging between 18 and 28. The final samples of men contain 16,716 observations, with 9,108 being in the NLSY79 and 7,608 in the NLSY97. Women samples contain 14,351 observations, 7,594 being in the NLSY79 and 6,757 observations in the NLSY97. Summary statistics are presented in Table VII.

6.2. *Model*

We estimate wage functions for men and women using both NLSY79 and NLSY97, and concentrate in evaluating the changes in effects of cognitive skills and schooling on earnings. We use the RE-QR methods to document possible heterogeneity on these effects. We also report OLS estimates for comparison. We estimate the following conditional quantile function:

$$Q_{w_{it}}(\tau|X_{it}) = educ_i'\beta_1(\tau) + ability_i\beta_2(\tau) + exp_{it}\beta_3(\tau) + exp_{it}^2\beta_4(\tau) + x_{it}'\beta_5(\tau), \tag{19}$$

where w_{it} if the log of $wage_{it}$ (the real hourly wage rate paid to an individual i at time t), $educ_i$ is a vector of four education degree dummy variables (High school, Associate, Bachelor, and Master), $ability_i$ measures cognitive skills using the AFQT score, exp_{it} corresponds to labor market experience, x_{it} is a vector of personal characteristics and family background variables such as race, unemployment, metro status, and family background controls (such as family income, parental education). The five parameters of interest are $(\beta_1(\tau), \beta_2(\tau))$.

It is important to highlight that the RE-QR model is an important tool to study this question. The variables of interest, $education_i$ and $ability_i$, are time-invariant, hence a FE model would not be able to estimate the parameters of interest $(\beta_1(\tau), \beta_2(\tau))$. Also, the number of time periods available for estimation is small which also rules out FE-QR estimation. Finally, the RE-QR is able to capture important heterogeneity by distinguishing the effect of education and cognitive ability on different quantiles of the conditional earning distribution.

TABLE VIII

POINTWISE AND UNIFORM TESTS FOR THE NULL HYPOTHESIS OF RANDOM EFFECTS. TEST STATISTICS FOR THE JOINT HYPOTHESIS, AND CORRESPONDING P-VALUES INSIDE BRACKETS.

	Men		Women	
τ	NLSY79	NLSY97	NLSY79	NLSY97
0.1	3.81	4.91	2.65	4.99
	[0.2824]	[0.1787]	[0.4482]	[0.1725]
0.2	7.265	6.37	3.02	5.75
	[0.0639]	[0.0950]	[0.3883]	[0.1246]
0.3	9.97	7.37	4.38	7.52
	[0.0188]	[0.0610]	[0.2237]	[0.0570]
0.4	18.96	11.00	4.84	8.70
	[0.0003]	[0.0117]	[0.1838]	[0.0336]
0.5	21.66	12.55	4.53	10.42
	[0.0001]	[0.0057]	[0.2098]	[0.0153]
0.6	21.80	20.12	6.93	6.868
	[0.0001]	[0.0002]	[0.0074]	[0.0763]
0.7	21.59	23.97	11.40	10.68
	[0.0001]	[0.0000]	[0.0097]	[0.0136]
0.8	20.18	34.24	12.37	7.79
	[0.0002]	[0.0000]	[0.0062]	[0.0505]
0.9	11.63	21.81	14.86	5.86
	[0.0088]	[0.0001]	[0.0019]	[0.1184]
Uniform	4.74	5.06	-1.16	-1.62
	[0.0530]	[0.0390]	[0.6346]	[0.5096]

6.3. *Empirical results*

Before proceeding with the estimation, we apply the tests developed in Section 4 for the RE hypothesis. We estimate the model in equation (20) including the mean of all the time-varying variables as additional regressors, and test the joint hypothesis that the additional coefficients are statistically equal to zero. The results are described in Table VIII for the deciles (pointwise), and at the bottom for the uniform test over the quantiles. The table provides the test statistic and the corresponding p-values. Using the pointwise test we are unable to reject the random effects hypothesis at the 5% level for various quantiles of the four subsamples, especially at lower quantiles, and also relatively more often for the two subsamples of women. Using the uniform test, we are unable to reject the random effects hypothesis for the men NLSY79 subsample, and for the two subsamples of women. The null is rejected for the men NLSY97 subsample.

We now describe the results for estimation of equation (20). We provide results for the proposed RE-QR estimation, and for comparison, we provide results for the RE-OLS.

First, we estimate the model for men with NLSY79. The results are displayed in Figure 2 for cognitive skill and returns to education for the four different degrees. Figure 2 shows estimates for the RE-OLS close to those in Castex and Dechter (2014). Nevertheless, Figure 2 documents very important heterogeneity across the distribution of wages. The returns to ability (left plot in first row), $\beta_2(\tau)$, are slightly larger at the top of the conditional wage distribution. The estimates are statistically positive for all degrees of education. There are relatively larger returns to education for higher degrees as Bachelor and Master. In addition, there is a clear inverted-U shape, which provides evidence that the returns at lower part of the distribution are positive and increasing over the quantiles, however, although still positive, they start to decrease for top quantiles.

Figure 2: Cognitive Skill (left plot first row), High School (middle plot first row), Associate (right plot first row), Bachelor (left plot second row), and Master (right plot second row): NLSY 79 – Men

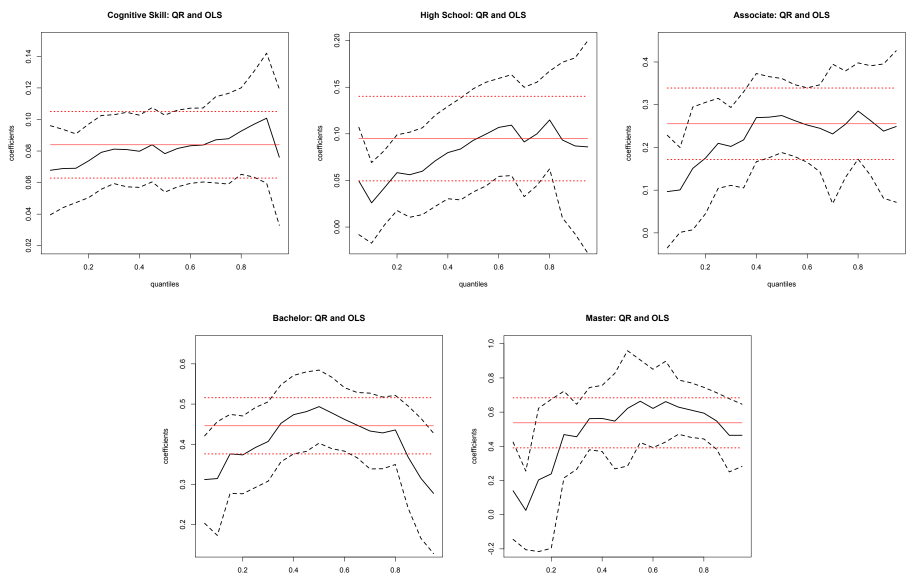


Figure 3: Cognitive Skill (left plot first row), High School (middle plot first row), Associate (right plot first row), Bachelor (left plot second row), and Master (right plot second row): NLSY 97 – Men

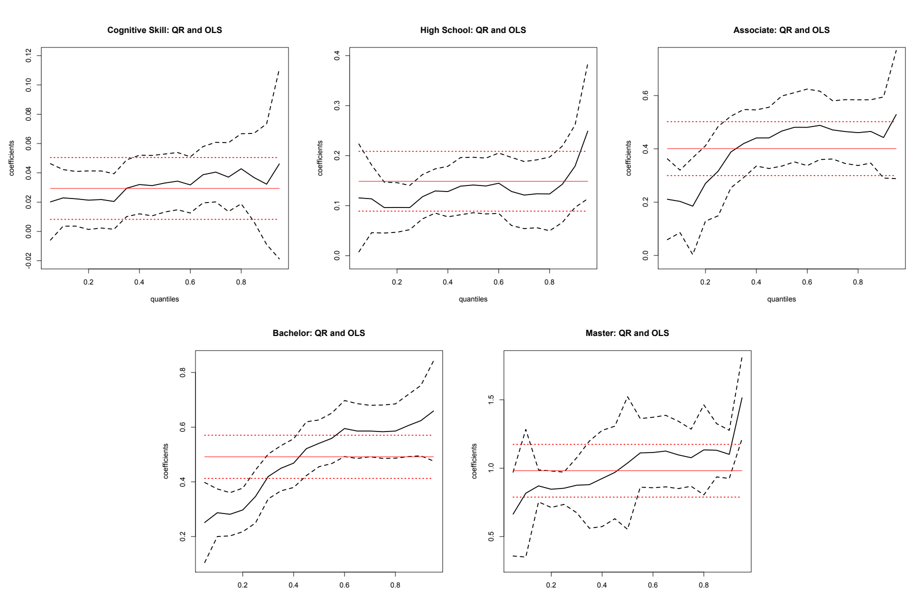
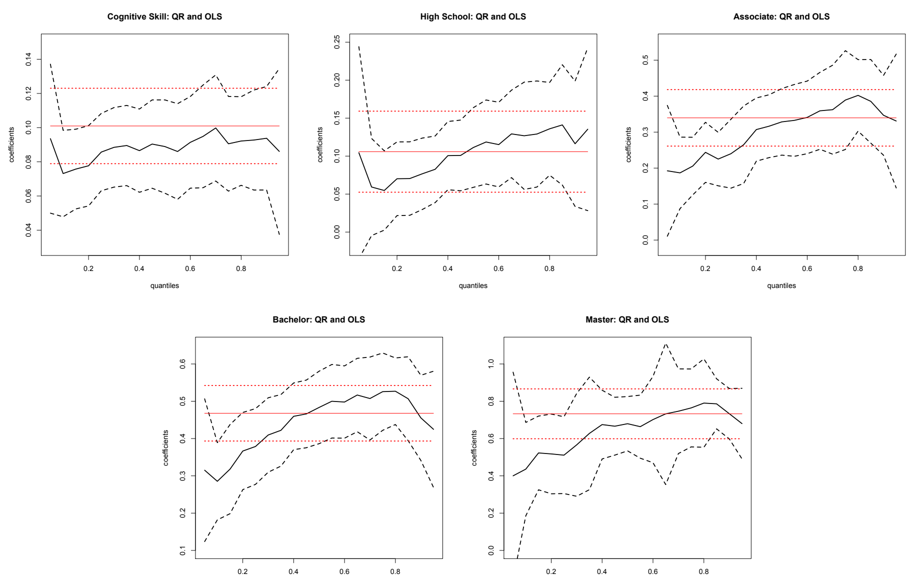


Figure 4: Cognitive Skill (left plot first row), High School (middle plot first row), Associate (right plot first row), Bachelor (left plot second row), and Master (right plot second row): NLSY 79 – Women

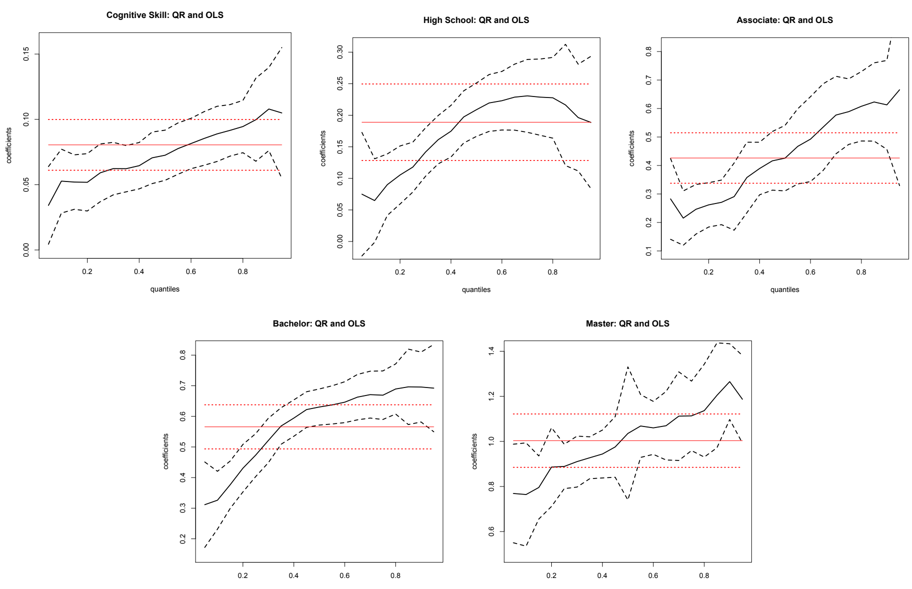


Second, we estimate the model for NLSY97 for men. The results are displayed in Figure 3 for cognitive skill and returns to education. Using NLSY97, the returns to ability are much smaller. In fact, as in Castex and Dechter (2014), we document a significant decline in return to ability over the 20 years. As in the previous case, returns to ability are smaller than the mean return for low parts of the wage distribution. Figure 3 shows that the returns to education for NLSY97 are in general larger relative to NLSY79, for all four degrees. It is important to highlight that these returns have increased substantially for the quantiles at the top of the distribution. For example, for a Master's degree the returns at the top of the distribution have nearly doubled. In addition, the shapes become increasing. Moreover, relative to NLSY79, the results for NLSY97 show relatively larger, and increasing over the quantiles, returns for High School. The returns for both Associate's and Bachelor's degrees in NLSY97 still follow an inverted-U shape, when compared to NLSY79, but they are less noisy and increasing for most part of the earnings distribution. Finally, the results for Master degree is flatter for NLSY97 relative to NLSY79, although larger in NLSY97.

Third, we estimate the model for women with NLSY79. The results are displayed in Figure 4 for cognitive skill and returns to education. It is also interesting to see that the RE-QR estimates from NLSY79 for women are close to those given in Figure 2 using NLSY79 for men. Regarding Figure 4, the results show positive estimates with relatively larger returns for higher degrees as Bachelor's and Master's. In addition, there is some evidence of an inverted-U shape for estimates over the quantiles. However, differently from Figure 2, the estimates are increasing across quantiles, and start decreasing only for top quantiles.

Finally, we estimate the model for NLSY97 for women. The results are displayed in Figure 5 for cognitive skill and returns to education. Compared to the NLSY79 sample of women, we can see a decrease in the RE-OLS returns to skill, but examining this result at a quantile-per-quantile level reveals strong heterogeneity in the change of these

Figure 5: Cognitive Skill (left plot first row), High School (middle plot first row), Associate (right plot first row), Bachelor (left plot second row), and Master (right plot second row): NLSY 97 – Women



responses. Specifically, we see a large decrease in returns to skill at the bottom of the conditional distribution, and an increase at the top of the conditional distribution. Therefore, the reduction in returns to skill for women seems to be mainly driven by women at lower quantiles. Perhaps a change in composition of the female worker pool could explain this decrease at lower quantiles.

Overall the application illustrates that the RE-QR methods are an important tool to examine changes in returns to formal education and cognitive skills. It allows for estimating returns to schooling and ability at different quantiles of the conditional distribution of earnings. Our empirical findings document important and strong heterogeneity along the distributions, as well as changes in the distributions over the 20 years (using the 1979 and 1997 waves of the NLSY). Specifically, compared to Castex and Dechter (2014), we find an increase in returns to skill for women near the top of their conditional distribution.

7. CONCLUSION

This paper develops a random effects (RE) model for quantile regression (QR) panel data with time-invariant regressors. We propose a model where the unobserved individual-specific effects affect the dependent variable through the unobserved heterogeneity, which induces heterogeneity across the conditional quantile function of the dependent variable. Thus, although the RE-QR model does not attempt to estimate the unobserved specific effects, their presence affects the dependent variable's conditional quantiles through the unobserved heterogeneity term U_{it}. We establish identification, and develop practical estimation and inference procedures for the RE model. We apply a simple pooled QR estimator to estimate the coefficients of interest and establish its statistical properties. We employ a cluster robust variance-covariance matrix estimator for inference, and establish its uniform consistency. In addition, we develop novel tests for the random effects hypothesis. Important practical advantages of the procedure are that it allows for estimation of time-invariant effects, as well as for the time-series dimension to be small. We provide

Monte-Carlo simulations to evaluate the finite-sample performance of the estimation and inference procedures. Finally, to illustrate the proposed methods we study the roles of education and ability in wage determination. The results document important heterogeneity in the returns to education and ability across the conditional distribution of wages.

We leave for future work a rigorous treatment of efficient estimators for parameters in RE-QR models. We also leave for future work questions regarding the decomposition of the unobserved heterogeneity (i.e. U_{it}, which has sample analog $\hat{U}_{it}$ defined by $Y_{it} = X_{it}'\hat{\theta}(\hat{U}_{it})$) into permanent and transitory components.

8. APPENDIX: PROOF OF THEOREMS

Throughout the proofs, we let $\mathbb{E}_n X_{it} \equiv \frac{1}{n}\sum_{i=1}^{n} X_{it}$ and $\mathbb{G}_n X_{it} \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_{it} - \mathrm{E}[X_{it}])$. $\Rightarrow$ denotes convergence in distribution.

LEMMA 8.1 *Let $y_{it} = \alpha_i + x_{it}'\beta(\epsilon_{it})$. Assume that assumptions (2.2) and (2.6) hold, that $x_{it}'\beta(\cdot)$ is a strictly increasing function with probability one and that α_i is a non-constant random variable. Then α_i and $x_{it}'\beta(\epsilon_{it})$ are not conditionally comonotonic given X_i.*

PROOF: α_i and $x_{it}'\beta(\epsilon_{it})$ are conditionally comonotonic only if $P(\alpha_i \leq c_1, x_{it}'\beta(\epsilon_{it}) \leq c_2 | X_i) = \min\{P(\alpha_i \leq c_1|X_i), P(x_{it}'\beta(\epsilon_{it}) \leq c_2|X_i)\}$ for any c_1, c_2. By (2.6), $P(\alpha_i \leq c_1|X_i) = P(\alpha_i \leq c_1)$. By assumptions (2.2) and (2.6), $\alpha_i \perp\!\!\!\perp (X_i, \epsilon_{it})$, which also implies $\alpha_i \perp\!\!\!\perp \epsilon_{it}|X_i$. By this conditional independence restriction, $P(\alpha_i \leq c_1, x_{it}'\beta(\epsilon_{it}) \leq c_2|X_i) = P(\alpha_i \leq c_1)P(x_{it}'\beta(\epsilon_{it}) \leq c_2|X_i)$. Therefore, conditional comonotonicity holds if $P(\alpha_i \leq c_1)P(x_{it}'\beta(\epsilon_{it}) \leq c_2|X_i) = \min\{P(\alpha_i \leq c_1), P(x_{it}'\beta(\epsilon_{it}) \leq c_2|X_i)\}$, which implies that either α_i is constant or that $x_{it}'\beta(\epsilon_{it})$ is constant for any fixed x_{it}, both of which are contradictions of the assumptions of lemma 8.1. *Q.E.D.*

PROOF OF PROPOSITION 2.1: Consider the following population objective function, which exists by A3':

$$Q_{\infty}^{T}(\tau,\theta) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\mathrm{E}\left[\rho_\tau(y_{it} - X_{it}'\theta) - \rho_\tau(y_{it} - X_{it}'\theta(\tau))\right]$$
$$= \frac{1}{T}\sum_{t=1}^{T}\mathrm{E}\left[\rho_\tau(y_{it} - X_{it}'\theta) - \rho_\tau(y_{it} - X_{it}'\theta(\tau))\right].$$

By convexity of the check function ρ_τ and by linearity of summations and expectations, we obtain that $Q_{\infty}^{T}(\tau,\theta)$ is convex in θ for any given τ.

By Knight's identity we can write

$$\frac{1}{T}\sum_{t=1}^{T}\rho_\tau(y_{it} - X_{it}'\theta) - \frac{1}{T}\sum_{t=1}^{T}\rho_\tau(y_{it} - X_{it}'\theta(\tau))$$
$$= -\frac{1}{T}\sum_{t=1}^{T}(\tau - 1(y_{it} \leq X_{it}'\theta(\tau)))X_{it}'(\theta - \theta(\tau))$$
$$+ \frac{1}{T}\sum_{t=1}^{T}\int_{0}^{X_{it}'(\theta-\theta(\tau))}(1(y_{it} \leq X_{it}'\theta(\tau) + s) - 1(y_{it} \leq X_{it}'\theta(\tau)))ds,$$

and therefore

$$Q_{\infty}^{T}(\tau,\theta) = \frac{1}{T}\sum_{t=1}^{T}\mathrm{E}\left[\int_{0}^{X_{it}'(\theta-\theta(\tau))}(1(y_{it} \leq X_{it}'\theta(\tau) + s) - 1(y_{it} \leq X_{it}'\theta(\tau)))ds\right] \geq 0,$$

with equality at $\theta = \theta(\tau)$. From the convexity of $Q_{\infty}^{T}(\tau,\theta)$, we deduce that its set of minimizers is a convex set. We note that for each τ, $\theta(\tau)$ is a local minimum by verifying that

$$
\begin{aligned}
\frac{\partial}{\partial\theta}Q_\infty^T(\tau,\theta) &= \frac{1}{T}\sum_{t=1}^T \mathrm{E}[X_{it}(1(y_{it}\le X_{it}'\theta)-\tau)] \\
\frac{\partial}{\partial\theta}Q_\infty^T(\tau,\theta)|_{\theta=\theta(\tau)} &= \mathbf{0} \\
\frac{\partial^2}{\partial\theta\partial\theta'}Q_\infty^T(\tau,\theta)|_{\theta=\theta(\tau)} &= \frac{1}{T}\sum_{t=1}^T \mathrm{E}[f_{y_t}(X_{it}'\theta(\tau)|X_{it})X_{it}X_{it}'] \\
&= \Gamma(\tau), \qquad (20)
\end{aligned}
$$

where $\Gamma(\tau)$ is positive definite for each τ by Assumption A2. Therefore, its minimizer $\theta(\tau)$ is unique. By our sampling Assumption A4, $Q_\infty^T(\tau,\theta)$ is identified up to a constant so its minimizer, which we showed was unique, is therefore also identified. *Q.E.D.*

PROOF OF THEOREM 3.1: This proof follows that of Theorems 4 and 5 in Angrist, Chernozhukov, and Fernandez-Val (2006).

We first show the uniform consistency of $\hat{\theta}(\cdot)$ over $\mathcal{T}$. Note that for each $\tau\in\mathcal{T}$, the estimator $\hat{\theta}(\tau)$ minimizes the sample objective function

$$
\begin{aligned}
Q_n^T(\tau,\theta) &= \frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T \rho_\tau(y_{it}-X_{it}'\theta) - \frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T \rho_\tau(y_{it}-X_{it}'\theta(\tau)) \\
&= \frac{1}{T}\sum_{t=1}^T \mathbb{E}_n\left[\rho_\tau(y_{it}-X_{it}'\theta)-\rho_\tau(y_{it}-X_{it}'\theta(\tau))\right],
\end{aligned}
$$

with population counterpart equal to

$$
Q_\infty^T(\tau,\theta) = \frac{1}{T}\sum_{t=1}^T \mathrm{E}\left[\rho_\tau(y_{it}-X_{it}'\theta)-\rho_\tau(y_{it}-X_{it}'\theta(\tau))\right].
$$

The result from Proposition 2.1 implies that $\theta(\tau)$ is the unique global minimizer of $Q_\infty(\theta,\tau)$ for each τ. Then, we find that for all θ,

$$
|Q_\infty^T(\tau,\theta)| \le 2\frac{1}{T}\sum_{t=1}^T \mathrm{E}[|X_{it}'(\theta-\theta(\tau))|] \le 2\frac{1}{T}\sum_{t=1}^T \mathrm{E}[\|X_{it}\|]\|\theta-\theta(\tau)\| < \infty.
$$

By a standard law of large numbers and assumption A4, we have that $Q_n^T(\tau,\theta)\xrightarrow{p}Q_\infty^T(\tau,\theta)$ for any $(\tau,\theta)\in\mathcal{T}\times\Theta$, where Θ is any compact subset of the parameter space. Convergence in probability is uniform in $(\tau,\theta)\in\mathcal{T}\times\Theta$ because

$$
\begin{aligned}
\left|Q_n^T(\tau,\theta)-Q_n^T(\tau',\theta')\right| &\le 2\frac{1}{T}\sum_{t=1}^T \mathrm{E}\left[\|X_{it}\|\right]\sup_{\theta\in\Theta}\|\theta\||\tau-\tau'| + 2\frac{1}{T}\sum_{t=1}^T \mathrm{E}\left[\|X_{it}\|\right]\|\theta-\theta'\| \qquad (21) \\
&\le C_1|\tau-\tau'| + C_2\|\theta-\theta'\|,
\end{aligned}
$$

where $C_1, C_2 < \infty$. Therefore the empirical process $(\tau,\theta)\mapsto Q_n(\tau,\theta)$ is stochastically equicontinuous. The pointwise convergence and the compactness of $\mathcal{T}\times\Theta$ ensure the total boundedness. Uniform convergence follows.

Consider a collection of closed balls $B_M(\theta(\tau))$ of radius M and center $\theta(\tau)$, and let $\theta_M(\tau)=\theta(\tau)+\delta_M(\tau)\nu(\tau)$ where $\nu(\tau)=(\nu_1(\tau),\ldots,\nu_d(\tau))'$ is a direction vector with unit norm $\|\nu(\tau)\|=1$ and $\delta_M(\tau)$

is a positive scalar satisfying $\delta_M(\tau) \geq M$. Uniformly in τ, we find that

$$\begin{aligned}\frac{M}{\delta_M(\tau)}(Q_n^T(\tau, \theta_M(\tau)) - Q_n^T(\tau, \theta(\tau))) &\geq Q_n^T(\tau, \theta_M^*(\tau)) - Q_n^T(\tau, \theta(\tau)) \\ &\geq Q_\infty^T(\tau, \theta_M^*(\tau)) - Q_\infty^T(\tau, \theta(\tau)) + o_p(1) \\ &> \epsilon_M + o_p(1),\end{aligned}$$

where the first inequality follows from the convexity in θ of $Q_n^T(\tau, \theta)$ and $\theta_M^*(\tau) = \frac{M}{\delta_M(\tau)}\theta_M(\tau) + \left(1 - \frac{M}{\delta_M(\tau)}\right)\theta(\tau)$ is on the line connecting $\theta_M(\tau)$ and $\theta(\tau)$. The second inequality holds by the uniform convergence which we showed using (22). The strict inequality follows from the identification which we showed in (21). Therefore, for any $M > 0$, $\|\hat{\theta}(\tau) - \theta(\tau)\| \leq M$ uniformly in $\tau \in \mathcal{T}$ with probability approaching one.

We now establish the asymptotic normality of the estimates. By the properties of $\hat{\theta}(\tau)$ (see Theorem 3.3 in Koenker and Bassett (1978)) we have that

$$\begin{aligned}\left\|\mathbb{E}_n\left[\frac{1}{T}\sum_{t=1}^T \varphi_\tau(y_{it} - X_{it}'\hat{\theta}(\tau))X_{it}\right]\right\| &\leq \frac{1}{T}\sum_{t=1}^T \|\mathbb{E}_n[\varphi_\tau(y_{it} - X_{it}'\hat{\theta}(\tau))X_{it}]\| \\ &\leq \frac{1}{T}\sum_{t=1}^T \text{const.} \times \frac{\sup_{i\leq n}\|X_{it}\|}{n},\end{aligned}$$

where $\varphi_\tau(u) = \tau - 1(u < 0)$. The term $\frac{\sup_{i\leq n}\|X_{it}\|}{n}$ will be of order $o_p(n^{-1/2})$ uniformly in t since

$$P\left(\sup_{i\leq n}\|X_{it}\| > n^{1/2}\right) \leq nP\left(\|X_{it}\| > n^{1/2}\right) \leq n\frac{\mathrm{E}[\|X_{it}\|^{2+\epsilon}]}{n^{\frac{2+\epsilon}{2}}} = o(1),$$

since $\mathrm{E}[\|X_{it}\|^{2+\epsilon}] < \infty$ for $\epsilon \leq 2$ by assumption A3 and by T being finite. Therefore, uniformly in $\tau \in \mathcal{T}$,

$$\frac{1}{T}\sum_{t=1}^T \mathbb{E}_n\left[\varphi_\tau(y_{it} - X_{it}'\hat{\theta}(\tau))X_{it}\right] = o_p(n^{-1/2}). \tag{22}$$

Next, we show that $(\tau, \theta) \mapsto \mathbb{G}_n[\frac{1}{T}\sum_{t=1}^T \varphi_\tau(y_{it} - X_{it}'\theta)X_{it}]$ is stochastically equicontinuous on $\Theta \times \mathcal{T}$ with respect to the to the pseudometric

$$\rho((\tau, \theta), (\tau', \theta')) = \max_{j=1,\ldots,K} \frac{1}{T}\sum_{t=1}^T \sqrt{\mathrm{E}[((\varphi_\tau(y_{it} - X_{it}'\theta) - \varphi_{\tau'}(y_{it} - X_{it}'\theta'))X_{itj})^2]}.$$

To show this, note that the function class $\mathcal{F}_t = \{1(y_{it} \leq X_{it}'\theta) : \theta \in \Theta\}$ is a VC subgraph class for $t = 1, \ldots, T$ and is also Donsker with envelope equal to 2. By Theorem 2.10.6 in van der Vaart and Wellner (1996), $\mathcal{T} - \mathcal{F}_t$ is also Donsker with envelope equal to 2 for all t, and by the same theorem, the product of $\mathcal{T} - \mathcal{F}_t$ and X_t is also Donsker with square integrable envelope $2\max_{j=1,\ldots,K}|X_{itj}|$. Finally, we can show that $\frac{1}{T}\sum_{t=1}^T(\mathcal{T} - \mathcal{F}_t)X_{it}$, the average over t of these Donsker classes, is also Donsker with envelope equal to $2\frac{1}{T}\sum_{t=1}^T \max_{j=1,\ldots,K}|X_{itj}|$, which also has bounded second moment.

We will now show that the uniform consistency $\sup_{\tau\in\mathcal{T}}\|\hat{\theta}(\tau) - \theta(\tau)\| = o_p(1)$ implies that

$$\sup_{\tau\in\mathcal{T}} \rho((\tau, b(\tau)), (\tau, \theta(\tau)))|_{b(\tau)=\hat{\theta}(\tau)} = o_p(1),$$

and then by the stochastic equicontinuity of $(\tau, \theta) \mapsto \mathbb{G}_n[\frac{1}{T}\sum_{t=1}^T \varphi_\tau(y_{it} - X_{it}'\theta)X_{it}]$ we have that

$$\mathbb{G}_n\left[\frac{1}{T}\sum_{t=1}^T \varphi_\tau(y_{it} - X_{it}'\hat{\theta}(\tau))X_{it}\right] = \mathbb{G}_n\left[\frac{1}{T}\sum_{t=1}^T \varphi_\tau(y_{it} - X_{it}'\theta(\tau))X_{it}\right] + o_p(1) \tag{23}$$

uniformly in τ.

To show that

$$\sup_{\tau\in\mathcal{T}} \rho((\tau, b(\tau)), (\tau, \theta(\tau)))|_{b(\tau)=\hat{\theta}(\tau)} = o_p(1),$$

we use the following derivations:

$$\begin{aligned}
&\sup_{\tau\in\mathcal{T}} \rho((\tau, b(\tau)), (\tau, \theta(\tau))) \\
&= \sup_{\tau\in\mathcal{T}} \max_{j=1,\dots,K} \frac{1}{T}\sum_{t=1}^{T} \sqrt{\mathrm{E}[((\varphi_\tau(y_{it} - X_{it}'b(\tau)) - \varphi_{\tau'}(y_{it} - X_{it}'\theta(\tau)))X_{itj})^2]} \\
&\leq \sup_{\tau\in\mathcal{T}} \max_{j=1,\dots,K} \frac{1}{T}\sum_{t=1}^{T} \left(\left(\mathrm{E}[|\varphi_\tau(y_{it} - X_{it}'b(\tau)) - \varphi_\tau(y_{it} - X_{it}'\theta(\tau))|^{\frac{2(2+\epsilon)}{\epsilon}}]\right)^{\epsilon} (\mathrm{E}[|X_{itj}|^{2+\epsilon}])^2\right)^{\frac{1}{2(2+\epsilon)}} \\
&\leq \sup_{\tau\in\mathcal{T}} \max_{j=1,\dots,K} \frac{1}{T}\sum_{t=1}^{T} (\mathrm{E}[|1(y_{it} \leq X_{it}'b(\tau)) - 1(y_{it} \leq X_{it}'\theta(\tau))|])^{\frac{\epsilon}{2(2+\epsilon)}} (\mathrm{E}[|X_{itj}|^{2+\epsilon}])^{\frac{1}{2+\epsilon}} \\
&\leq \sup_{\tau\in\mathcal{T}} \frac{1}{T}\sum_{t=1}^{T} (\mathrm{E}|\bar{f}X_{it}'(b(\tau) - \theta(\tau))|)^{\frac{\epsilon}{2(2+\epsilon)}} (\mathrm{E}[\|X_{it}\|^{2+\epsilon}])^{\frac{1}{2+\epsilon}} \\
&\leq \text{const.} \times \bar{f} \times \frac{1}{T}\sum_{t=1}^{T} ((\mathrm{E}[\|X_{it}\|^2])^{\frac{1}{2}} \|b(\tau) - \theta(\tau)\|)^{\frac{\epsilon}{2(2+\epsilon)}}.
\end{aligned}$$

where $\epsilon \leq 2$ and $\bar{f}$ is a bound of $f_{y_t}(y|x)$ which exists by A1. Evaluating this expression at $b(\tau) = \hat{\theta}(\tau)$ gives

$$\sup_{\tau\in\mathcal{T}} \rho((\tau, b(\tau)), (\tau, \theta(\tau)))|_{b(\tau)=\hat{\theta}(\tau)} \leq \text{const.} \times \sup_{\tau\in\mathcal{T}} \|\hat{\theta}(\tau) - \theta(\tau)\|^{\frac{\epsilon}{2(2+\epsilon)}} = o_p(1),$$

by the uniform convergence.

Next we can use a uniform in τ Taylor expansion and find that

$$\mathrm{E}\left[\frac{1}{T}\sum_{t=1}^{T}\varphi_\tau(y_{it} - X_{it}'\theta)X_{it}\right]|_{\theta=\hat{\theta}(\tau)} = \mathrm{E}\left[\frac{1}{T}\sum_{t=1}^{T} f_{y_t}(X_{it}'b(\tau)|X_{it})X_{it}X_{it}'\right]|_{b(\tau)=\theta^*(\tau)}(\hat{\theta}(\tau) - \theta(\tau)),$$

where $\theta^*(\tau)$ is on a line connecting $\hat{\theta}(\tau)$ and $\theta(\tau)$ for all τ. The uniform consistency of $\hat{\theta}(\tau)$ implies the uniform consistency of $\theta^*(\tau)$ to $\theta(\tau)$ over $\tau \in \mathcal{T}$. By the uniform continuity and boundedness of $f_{y_t}(y_{it}|X)$ uniformly in X, we find that

$$\mathrm{E}\left[\frac{1}{T}\sum_{t=1}^{T} f_{y_t}(X_{it}'b(\cdot)|X_{it})X_{it}X_{it}'\right]|_{b(\cdot)=\theta^*(\cdot)} = \mathrm{E}\left[\frac{1}{T}\sum_{t=1}^{T} f_{y_t}(X_{it}'\theta(\cdot)|X_{it})X_{it}X_{it}'\right] + o_p(1) \tag{24}$$

uniformly in $\tau \in \mathcal{T}$. Therefore, we have that

$$\mathrm{E}\left[\frac{1}{T}\sum_{t=1}^{T}\varphi_\tau(y_{it} - X_{it}'\theta)X_{it}\right]|_{\theta=\hat{\theta}(\tau)} = (\Gamma(\tau) + o_p(1))(\hat{\theta}(\tau) - \theta(\tau)) \tag{25}$$

which leads to

$$o_p(1) = [\Gamma(\cdot) + o_p(1)]\sqrt{n}(\hat{\theta}(\cdot) - \theta(\cdot)) + \mathbb{G}_n\left[\frac{1}{T}\sum_{t=1}^{T}\varphi_{\cdot}(y_{it} - X_{it}'\theta(\cdot))X_{it}\right]$$

by combining equation (26) with equations (23) and (24).

The proof is completed using arguments similar to Angrist, Chernozhukov, and Fernandez-Val (2006), leading to

$$\Gamma(\cdot)\sqrt{n}(\hat{\theta}(\cdot)-\theta(\cdot)) = -\mathbb{G}_n\left[\frac{1}{T}\sum_{t=1}^{T}\varphi.(y_{it}-X_{it}'\theta(\cdot))X_t\right] + o_p(1) \Rightarrow \mathbf{z}(\cdot)$$

in $\ell^\infty(\mathcal{T})$ where $\mathbf{z}(\cdot)$ is a Gaussian process with covariance kernel specified in equation (3.1).

Q.E.D.

PROOF OF THEOREM 3.2: Let

$$\hat{\Gamma}(\tau) = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{h_n}\mathbb{E}_n\left[K\left(\frac{y_{it}-X_{it}'\hat{\theta}(\tau)}{h_n}\right)X_{it}X_{it}'\right].$$

By Lemma 8.2 in Wellner (2005), $\mathcal{F}_t = \left\{K\left(\frac{y_{it}-X_{it}'\theta}{h}\right) : \theta \in \Theta, h \in (0, H]\right\}$ is a VC class for $H > 0$ since $K(\cdot)$ is of bounded variation, and $\frac{y_{it}-X_{it}'\theta}{h}$ can be reparameterized as $V_{it}'b$ where $V_{it} = (y_{it}, X_{it}')'$ and $b = (1/h, -\theta'/h)'$. Since $X_{it}X_{it}'$ has finite second moments by A3, the class $\mathcal{G}_t = \left\{K\left(\frac{y_{it}-X_{it}'\theta}{h}\right)X_{it}X_{it}' : \theta \in \Theta, h \in (0, H]\right\}$ will be a Donsker class. By Theorem 2.10.6 in van der Vaart and Wellner (1996), $\frac{1}{T}\sum_{t=1}^{T}\mathcal{G}_t$ will also be a Donsker class.

This implies that

$$\sup_{\tau\in\mathcal{T}}\left\|\mathbb{E}_n\left[K\left(\frac{y_{it}-X_{it}'\hat{\theta}(\tau)}{h_n}\right)X_{it}X_{it}'\right] - \mathrm{E}\left[K\left(\frac{y_{it}-X_{it}'\theta}{h_n}\right)X_{it}X_{it}'\right]|_{\theta=\hat{\theta}(\tau)}\right\| = O_p(n^{-1/2}),$$

and also

$$\begin{aligned}\sup_{\tau\in\mathcal{T}}\left\|\hat{\Gamma}(\tau) - \frac{1}{h_n}\mathrm{E}\left[K\left(\frac{y_{it}-X_{it}'\theta}{h_n}\right)X_{it}X_{it}'\right]|_{\theta=\hat{\theta}(\tau)}\right\| &= O_p(h_n^{-1}n^{-1/2})\\ &= o_p(1),\end{aligned}$$

since $nh_n^2 \to \infty$.

By an argument similar to that of equation (25) and using the Lipschitz property of $K(\cdot)$,

$$\frac{1}{h_n}\mathrm{E}\left[K\left(\frac{y_{it}-X_{it}'\theta}{h_n}\right)X_{it}X_{it}'\right]|_{\theta=\hat{\theta}(\tau)} = \frac{1}{h_n}\mathrm{E}\left[K\left(\frac{y_{it}-X_{it}'\theta(\tau)}{h_n}\right)X_{it}X_{it}'\right] + O\left(\frac{1}{h_n^2}(\hat{\theta}(\tau)-\theta(\tau)\right)$$

where $O(h_n^{-2}(\hat{\theta}(\tau)-\theta(\tau)) = o_p(h_n^{-2}n^{-1/2}) = o_p(1)$ uniformly in τ by $nh_n^4 \to \infty$ and the uniform convergence in distribution of $\sqrt{n}(\hat{\theta}(\cdot)-\theta(\cdot))$. Finally, we see that $\frac{1}{T}\sum_{t=1}^{T}\frac{1}{h_n}\mathrm{E}\left[K\left(\frac{y_{it}-X_{it}'\theta(\tau)}{h_n}\right)X_{it}X_{it}'\right] = \Gamma(\tau)+O(h_n^2)$ uniformly in τ by the boundedness of the densities, the fact that $K(\cdot)$ is a second order kernel, and by $h_n \to 0$.

Also, let

$$\begin{aligned}\tilde{V}(\tau,\tau') &= \frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathbb{E}_n[(\tau - 1(y_{is} \le X_{is}'\hat{\theta}(\tau)))(\tau' - 1(y_{it} \le X_{it}'\hat{\theta}(\tau')))]\\ &= \frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\tilde{V}_{s,t}(\tau,\tau').\end{aligned}$$

By Angrist, Chernozhukov, and Fernandez-Val (2006) A.1.4, $\tilde{V}_{s,t}(\tau,\tau')$ converges uniformly over $(\tau,\tau') \in \mathcal{T}^2$ to $V_{s,t}(\tau,\tau') = \mathrm{E}[(1(v_{is}(\tau) \le 0) - \tau)(1(v_{it}(\tau') \le 0) - \tau')X_{is}X_{it}']$ for $s,t = 1,\ldots,T$. Therefore, $\tilde{V}(\tau,\tau')$ converges uniformly to $V(\tau,\tau')$.

To show that $\hat{V}(\tau,\tau')$ also converges uniformly to $V(\tau,\tau')$, note that

$$\begin{aligned}
&\hat{V}(\tau,\tau') - \tilde{V}(\tau,\tau') \\
&= \frac{\min(\tau,\tau') - \tau\tau'}{T^2}\sum_{t=1}^{T}\mathbb{E}_n[X_{it}X_{it}'] - \frac{1}{T^2}\sum_{t=1}^{T}\frac{1}{n}\sum_{i=1}^{n}X_{it}X_{it}'(1(\hat{v}_{it}(\tau)\le 0)-\tau)(1(\hat{v}_{it}(\tau')\le 0)-\tau').
\end{aligned}$$

By the uniform convergence of $\tilde{V}_{t,t}(\tau,\tau')$, we have

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}X_{it}X_{it}'(1(\hat{v}_{it}(\tau)\le 0)-\tau)(1(\hat{v}_{it}(\tau')\le 0)-\tau') \\
&\xrightarrow{p} \mathrm{E}[(1(v_{it}(\tau)\le 0)-\tau)(1(v_{it}(\tau')\le 0)-\tau')X_{it}X_{it}'] \\
&= (\min(\tau,\tau')-\tau\tau')\mathrm{E}[X_{it}X_{it}']
\end{aligned}$$

for each t, and therefore

$$\frac{1}{T^2}\sum_{t=1}^{T}\frac{1}{n}\sum_{i=1}^{n}X_{it}X_{it}'(1(\hat{v}_{it}(\tau)\le 0)-\tau)(1(\hat{v}_{it}(\tau')\le 0)-\tau') \xrightarrow{p} \frac{\min(\tau,\tau')-\tau\tau'}{T^2}\sum_{t=1}^{T}\mathrm{E}[X_{it}X_{it}']$$

uniformly in τ.

Also,

$$\begin{aligned}
&\sup_{(\tau,\tau')\in\mathcal{T}^2}\left\|\frac{\min(\tau,\tau')-\tau\tau'}{T^2}\sum_{t=1}^{T}\mathbb{E}_n[X_{it}X_{it}'] - \frac{\min(\tau,\tau')-\tau\tau'}{T^2}\sum_{t=1}^{T}\mathrm{E}[X_{it}X_{it}']\right\| \\
&\le \sup_{(\tau,\tau')\in\mathcal{T}^2}\frac{\min(\tau,\tau')-\tau\tau'}{T^2}\sum_{t=1}^{T}\|\mathbb{E}_n[X_{it}X_{it}'] - \mathrm{E}[X_{it}X_{it}']\| \\
&= \frac{1}{4T^2}\sum_{t=1}^{T}\|\mathbb{E}_n[X_{it}X_{it}'] - \mathrm{E}[X_{it}X_{it}']\| \\
&= o_p(1)
\end{aligned}$$

by the law of large numbers and assumption A1. Combining these results, we find that

$$\begin{aligned}
&\sup_{(\tau,\tau')\in\mathcal{T}^2}\|\hat{V}(\tau,\tau')-\tilde{V}(\tau,\tau')\| \\
&\le \sup_{(\tau,\tau')\in\mathcal{T}^2}\left\|\frac{\min(\tau,\tau')-\tau\tau'}{T^2}\sum_{t=1}^{T}(\mathbb{E}_n[X_{it}X_{it}'] - \mathrm{E}[X_{it}X_{it}'])\right\| \\
&+ \sup_{(\tau,\tau')\in\mathcal{T}^2}\left\|\frac{1}{T^2}\sum_{t=1}^{T}\tilde{V}_{t,t}(\tau,\tau') - \frac{\min(\tau,\tau')-\tau\tau'}{T^2}\sum_{t=1}^{T}\mathrm{E}[X_{it}X_{it}']\right\| \\
&= o_p(1)+o_p(1) = o_p(1),
\end{aligned}$$

which then implies that

$$\begin{aligned}
&\sup_{(\tau,\tau')\in\mathcal{T}^2}\|\hat{V}(\tau,\tau')-V(\tau,\tau')\| \\
&\le \sup_{(\tau,\tau')\in\mathcal{T}^2}\|\hat{V}(\tau,\tau')-\tilde{V}(\tau,\tau')\| + \sup_{(\tau,\tau')\in\mathcal{T}^2}\|\tilde{V}(\tau,\tau')-V(\tau,\tau')\| \\
&= o_p(1).
\end{aligned}$$

Therefore, both covariance matrix estimators are uniformly convergent. *Q.E.D.*

PROOF OF LEMMA 4.1: By Theorem 3.1,

$$\sqrt{n}(\hat{\theta}(\tau)-\theta(\tau)) \Rightarrow N(0,\Omega(\tau))$$

and under the null hypothesis,

$$\sqrt{n}(R\hat{\theta}(\tau) - r) \Rightarrow N(0, R\Omega(\tau)R').$$

By Theorem 3.2, $\hat{\Omega}(\tau)$ is a consistent estimator of $\Omega(\tau)$, and by Slutsky's theorem,

$$\mathcal{W}_n(\tau) = n(R\hat{\theta}(\tau) - r)'[R\hat{\Omega}(\tau)R']^{-1}(R\hat{\theta}(\tau) - r) \Rightarrow \chi^2_q.$$

Q.E.D.

PROOF OF THEOREM 4.1: Let

$$W_{n,M} = \sqrt{M}\left(\frac{n}{M}(R^M\hat{\theta}^M - r^M)'(R^M\hat{\Omega}^M R^{M\prime})^{-1}(R^M\hat{\theta}^M - r^M) - q\right)$$

denote the test statistic. For each $m = 1, \ldots, M$, by Theorem 3.1

$$\begin{aligned}\sqrt{n}(\hat{\theta}(\tau_m) - \theta(\tau_m)) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Gamma(\tau_m)^{-1}\frac{1}{T}\sum_{t=1}^{T}\varphi_{\tau_m}(y_{it} - X_{it}'\theta(\tau_m))X_{it} + o_p(1) \\ &\equiv \frac{1}{\sqrt{n}}\sum_{i=1}^{n}a_i(\tau_m) + T_n(\tau_m)\end{aligned}$$

where, $a_i(\tau_m)$ is the influence function, and $T_n(\tau_m)$ is the residual. By Theorem 3.1, $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}a_i(\cdot) \Rightarrow \mathbf{Z}(\cdot) \equiv \Gamma(\cdot)^{-1}\mathbf{z}(\cdot)$ and $T_n(\cdot) \Rightarrow \mathbf{0}$ in $\ell^\infty(\mathcal{T})$.

By the assumption that $R\theta(\tau) = r$, we have that

$$\sqrt{n}(R\hat{\theta}(\tau_m) - r) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}Ra_i(\tau_m) + RT_n(\tau_m). \tag{26}$$

Using the results of Jurecková and Sen (1996) (see also Knight (2001), Portnoy (2012)), the residual $T_n(\cdot)$ satisfies $\sup_{\tau\in\mathcal{T}}\|T_n(\tau)\| = O_p(n^{-1/4})$.

To aid with the asymptotic analysis of $W_{n,M}$, we also consider the following infeasible test statistics:

$$\begin{aligned}\tilde{W}_{n,M} &= \sqrt{M}\left(\frac{n}{M}(R^M\hat{\theta}^M - r^M)'(R^M\Omega^M R^{M\prime})^{-1}(R^M\hat{\theta}^M - r^M) - q\right) \\ \bar{W}_{n,M} &= \sqrt{M}\left(\frac{1}{M}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Ra_i(\tau_m)\right)'(R^M\Omega^M R^{M\prime})^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Ra_i(\tau_m)\right) - q\right).\end{aligned}$$

We will show that all three of these statistics are asymptotically equivalent. We have that

$$\begin{aligned}\sup_{\tau\in\mathcal{T}}\left\|\hat{\Gamma}(\tau) - \Gamma(\tau)\right\| &\le \sup_{\tau\in\mathcal{T}}\left\|\hat{\Gamma}(\tau) - \frac{1}{T}\sum_{t=1}^{T}\frac{1}{h_n}\mathrm{E}\left[K\left(\frac{y_{it} - X_{it}'\theta}{h_n}\right)X_{it}X_{it}'\right]|_{\theta=\hat{\theta}(\tau)}\right\| \\ &\quad + \sup_{\tau\in\mathcal{T}}\left\|\frac{1}{T}\sum_{t=1}^{T}\frac{1}{h_n}\mathrm{E}\left[K\left(\frac{y_{it} - X_{it}'\theta}{h_n}\right)X_{it}X_{it}'\right]|_{\theta=\hat{\theta}(\tau)}\right. \\ &\quad \left. - \frac{1}{T}\sum_{t=1}^{T}\frac{1}{h_n}\mathrm{E}\left[K\left(\frac{y_{it} - X_{it}'\theta(\tau)}{h_n}\right)X_{it}X_{it}'\right]\right\| \\ &\quad + \sup_{\tau\in\mathcal{T}}\left\|\frac{1}{T}\sum_{t=1}^{T}\frac{1}{h_n}\mathrm{E}\left[K\left(\frac{y_{it} - X_{it}'\theta(\tau)}{h_n}\right)X_{it}X_{it}'\right] - \Gamma(\tau)\right\|.\end{aligned}$$

We have already shown in the proof of Theorem 3.2 that

$$\sup_{\tau\in\mathcal{T}}\left\|\hat{\Gamma}(\tau) - \frac{1}{T}\sum_{t=1}^{T}\frac{1}{h_n}\mathrm{E}\left[K\left(\frac{y_{it} - X_{it}'\theta}{h_n}\right)X_{it}X_{it}'\right]|_{\theta=\hat{\theta}(\tau)}\right\| = O_p(n^{-1/2}h_n^{-1}).$$

By the same theorem and the second order of the Kernel, the third term

$$\sup_{\tau\in\mathcal{T}}\left\|\frac{1}{T}\sum_{t=1}^{T}\frac{1}{h_n}\mathrm{E}\left[K\left(\frac{y_{it}-X_{it}'\theta(\tau)}{h_n}\right)X_{it}X_{it}'\right]-\Gamma(\tau)\right\| = O(h_n^2).$$

By the Lipschitz assumption A5 on kernel $K(\cdot)$,

$$\begin{aligned}
&\sup_{\tau\in\mathcal{T}}\left\|\mathrm{E}\left[\frac{1}{h_n}K\left(\frac{y_{it}-X_{it}'\theta}{h_n}\right)X_{it}X_{it}'\right]|_{\theta=\hat\theta(\tau)}-\mathrm{E}\left[\frac{1}{h_n}K\left(\frac{y_{it}-X_{it}'\theta(\tau)}{h_n}\right)X_{it}X_{it}'\right]\right\| \\
&\le \sup_{\tau\in\mathcal{T}}\frac{1}{h_n}\mathrm{E}\left[\left|K\left(\frac{y_{it}-X_{it}'\theta}{h_n}\right)-K\left(\frac{y_{it}-X_{it}'\theta(\tau)}{h_n}\right)\right|\|X_{it}X_{it}'\|\right]|_{\theta=\hat\theta(\tau)} \\
&\le \sup_{\tau\in\mathcal{T}}\frac{1}{h_n^2}\mathrm{E}\left[\text{const.}\times|X_{it}'(\theta-\theta(\tau))|\|X_{it}X_{it}'\|\right]|_{\theta=\hat\theta(\tau)} \\
&= O_p(h_n^{-2}n^{-1/2}),
\end{aligned}$$

by $\sup_{\tau\in\mathcal{T}}\|\hat\theta(\tau)-\theta(\tau)\| = O_p(n^{-1/2})$.

By Theorem A.1.4 in Angrist, Chernozhukov, and Fernandez-Val (2006),

$$\|\hat V(\tau,\tau')-V(\tau,\tau')\| = O_p(n^{-1/2})$$

uniformly in $(\tau,\tau')\in\mathcal{T}^2$. Combining these results,

$$\left\|\hat\Omega(\tau,\tau')-\Omega(\tau,\tau')\right\| = O_p(h_n^{-2}n^{-1/2})+O(h_n^2) \tag{27}$$

uniformly in τ,τ'.

Denote by $F(m,m')$ the $q\times q$ block at position (m,m') of $(R^M\Omega^M R^{M\prime})^{-1}$ a $qM\times qM$ matrix. Similarly, let $\hat F(m,m')$ be the $q\times q$ block at position (m,m') $(R^M\hat\Omega^M R^{M\prime})^{-1}$. By the continuity of the inverse and result (28), the blocks satisfy

$$\sup_{m,m'\in\{1,\ldots,M\}}\|\hat F(m,m')-F(m,m')\| = O_p(h_n^{-2}n^{-1/2})+O(h_n^2).$$

Therefore,

$$\begin{aligned}
|W_{n,M}-\tilde W_{n,M}| &= \left|\frac{1}{\sqrt{M}}\sum_{m=1}^{M}\sum_{m'=1}^{M}\sqrt{n}(R\hat\theta(\tau_m)-r)'(\hat F(m,m')-F(m,m'))\sqrt{n}(R\hat\theta(\tau_{m'})-r)\right| \\
&\le M^{3/2}\left(\sup_{m\in\{1,\ldots,M\}}\sqrt{n}\|R\hat\theta(\tau_m)-r\|\right)^2\sup_{m,m'\in\{1,\ldots,M\}}\|\hat F(m,m')-F(m,m')\| \\
&= M^{3/2}O_p(1)\left(O_p(h_n^{-2}n^{-1/2})+O(h_n^2)\right) \\
&= O_p\left(\frac{M^{3/2}}{n^{1/2}h_n^2}\right)+O_p(M^{3/2}h_n^2) \\
&= o_p(1),
\end{aligned}$$

since $M^3n^{-1}h_n^{-4}=o(1)$, $M^3h_n^4=o(1)$ by the theorem assumptions.

The vector $(R^M\Omega^M R^{M\prime})^{-1/2}\sqrt{n}(R^M\hat\theta^M-r^M)$ is a $qM\times 1$ vector, and by equation (27), we write the $q\times 1$ vector in position m as follows

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde a_i(m)+\tilde T_n(m),$$

where $\mathrm{E}[\tilde a_i(m)]=\mathbf{0}_q$, $\mathrm{E}[\tilde a_i(m)\tilde a_i(m')']=\mathbf{I}_q 1(m=m')$ by the properties of the inverse. We also have $\sup_{m\in\{1,\ldots,M\}}\|\tilde T_n(m)\| = O_p(n^{-1/4})$ and $\sup_{m\in\{1,\ldots,M\}}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde a_i(m)\right\| = O_p(1)$ by our earlier results about Donsker classes.

We can also write

$$
\begin{aligned}
|\tilde{W}_{n,M} - \bar{W}_{n,M}| &= \left| -2\frac{1}{\sqrt{M}} \sum_{m=1}^{M} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{a}_i(m) \right)' \tilde{T}_n(m) - \frac{1}{\sqrt{M}} \sum_{m=1}^{M} \tilde{T}_n(m)'\tilde{T}_n(m) \right| \\
&\leq 2\sqrt{M} \sup_{m\in\{1,\ldots,M\}} \left(\left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{a}_i(m) \right\| \|\tilde{T}_n(m)\| \right) + \sqrt{M} \sup_{m\in\{1,\ldots,M\}} \|\tilde{T}_n(m)\|^2 \\
&\leq 2\sqrt{M} \sup_{m\in\{1,\ldots,M\}} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{a}_i(m) \right\| \sup_{m\in\{1,\ldots,M\}} \|\tilde{T}_n(m)\| + \sqrt{M} O_p(n^{-1/2}) \\
&= O_p(\sqrt{M} n^{-1/4}) + O_p(\sqrt{M} n^{-1/2}) \\
&= o_p(1),
\end{aligned}
$$

if $M^2 n^{-1} = o(1)$. Note that combining $M^3 h_n^4 = o(1)$ and $M^3 n^{-1} h_n^{-4} = o(1)$ yields $M^6 n^{-1} = o(1)$, which implies that $M^2 n^{-1} = o(1)$.

Finally, let

$$
\begin{aligned}
\bar{W}_{n,M} &= \sqrt{M} \left(\frac{1}{M} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} R a_i(\tau_m) \right)' (R^M \Omega^M R^{M\prime})^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} R a_i(\tau_m) \right) - q \right) \\
&= \frac{1}{\sqrt{M}} \sum_{m=1}^{M} \left(\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{a}_i(m) \right)' \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{a}_i(m) \right) - q \right) \\
&\equiv \frac{1}{\sqrt{M}} \sum_{m=1}^{M} V_m^n,
\end{aligned}
$$

where

$$
\begin{aligned}
V_m^n &= \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{a}_i(m) \right)' \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{a}_i(m) \right) - q \\
&\equiv n\overline{\tilde{a}(m)}'\overline{\tilde{a}(m)} - q,
\end{aligned}
$$

where $\overline{\tilde{a}(m)} = \frac{1}{n}\sum_{i=1}^{n} \tilde{a}_i(m)$. By previous results about the process convergence of $a_i(\tau)$, we have that

$$
\begin{aligned}
\mathrm{E}[V_m^n] &= \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}[\tilde{a}_i(m)'\tilde{a}_i(m)] - q \\
&= \mathrm{trace}(\mathrm{E}[\tilde{a}_i(m)\tilde{a}_i(m)']) - q \\
&= \mathrm{trace}(\mathbf{I}_q) - q \\
&= 0,
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{E}[(V_m^n)^2] &= \mathrm{E}[(n\overline{\tilde{a}(m)}'\overline{\tilde{a}(m)} - q)^2] \\
&= n^2\mathrm{E}[(\overline{\tilde{a}(m)}'\overline{\tilde{a}(m)})^2] - 2q^2 + q^2 \\
&= \frac{1}{n^2}\mathrm{E}[(\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{p=1}^{q}\tilde{a}_i^p(m)\tilde{a}_j^p(m))^2] - q^2 \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{i'=1}^{n}\sum_{j'=1}^{n}\sum_{p=1}^{q}\sum_{p'=1}^{q}\mathrm{E}[\tilde{a}_i^p(m)\tilde{a}_j^p(m)\tilde{a}_{i'}^{p'}(m)\tilde{a}_{j'}^{p'}(m)] - q^2 \\
&= \frac{1}{n^2}(n\sum_{p=1}^{q}\sum_{p'=1}^{q}\mathrm{E}[\tilde{a}_i^p(m)^2\tilde{a}_i^{p'}(m)^2] + n(n-1)\sum_{p=1}^{q}\sum_{p'=1}^{q}\mathrm{E}[\tilde{a}_i^p(m)^2]\mathrm{E}[\tilde{a}_i^{p'}(m)^2] \\
&+ 2n(n-1)\sum_{p=1}^{q}\sum_{p'=1}^{q}\mathrm{E}[\tilde{a}_i^p(m)\tilde{a}_i^{p'}(m)]^2) - q^2 \\
&= \frac{\text{const.}}{n} + q^2 + 2q - q^2 \\
&= \frac{\text{const.}}{n} + 2q,
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{E}[V_m^n V_{m'}^n] &= n^2\mathrm{E}[(\overline{\tilde{a}(m)}'\overline{\tilde{a}(m')})^2] - q^2 \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{i'=1}^{n}\sum_{j'=1}^{n}\sum_{p=1}^{q}\sum_{p'=1}^{q}\mathrm{E}[\tilde{a}_i^p(m)\tilde{a}_j^p(m)\tilde{a}_{i'}^{p'}(m')\tilde{a}_{j'}^{p'}(m')] - q^2 \\
&= \frac{1}{n^2}(n\sum_{p=1}^{q}\sum_{p'=1}^{q}\mathrm{E}[\tilde{a}_i^p(m)^2\tilde{a}_i^{p'}(m')^2] + n(n-1)\sum_{p=1}^{q}\sum_{p'=1}^{q}\mathrm{E}[\tilde{a}_i^p(m)^2]\mathrm{E}[\tilde{a}_i^{p'}(m')^2] \\
&+ 2n(n-1)\sum_{p=1}^{q}\sum_{p'=1}^{q}\mathrm{E}[\tilde{a}_i^p(m)\tilde{a}_i^{p'}(m')]^2) - q^2 \\
&= \frac{\text{const.}}{n} + \frac{n-1}{n}q^2 + 0 - q^2 \\
&= \frac{\text{const.}}{n}
\end{aligned}
$$

for $m \neq m'$.

Additionally, for any finite number k, $(V_1^n, V_2^n, \ldots, V_k^n) \Rightarrow (\chi_q^2 - q, \chi_q^2 - q, \ldots, \chi_q^2 - q)$ where the χ_q^2 distributions are mutually independent. Therefore, we have asymptotic independence of $\{V_{m_l}^n\}_{l=1,\ldots,k}$ for any indices $m_1, \ldots, m_k$ and any finite k. By the multivariate version of the Berry-Esseen Theorem applied to the square of sample averages,[12] (see Gotze (1991)),

$$
\begin{aligned}
\sup_{v_1,v_2}|P(V_m^n \le v_1, V_{m'}^n \le v_2) - P(\chi_q^2 - q \le v_1)P(\chi_q^2 - q \le v_2)| &= O(n^{-1/2}) \\
\sup_{v_1}|P(V_m^n \le v_1) - P(\chi_q^2 - q \le v_1)| &= O(n^{-1/2}),
\end{aligned}
$$

[12]The sup-distance between CDFs of continuous transformations of the normalized averages will also be of order $O(n^{-1/2})$ by continuity.

for all m, m'. Therefore,

$$\begin{aligned}
&\sup_{v_1,v_2} |P(V_m^n \le v_1, V_{m'}^n \le v_2) - P(V_m^n \le v_1)P(V_{m'}^n \le v_2)| \\
&= \sup_{v_1,v_2} |P(V_m^n \le v_1, V_{m'}^n \le v_2) - P(\chi_q^2 - q \le v_1)P(\chi_q^2 - q \le v_2) \\
&- (P(V_m^n \le v_1)P(V_{m'}^n \le v_2) - P(\chi_q^2 - q \le v_1)P(\chi_q^2 - q \le v_2))| \\
&\le \sup_{v_1,v_2} |P(V_m^n \le v_1, V_{m'}^n \le v_2) - P(\chi_q^2 - q \le v_1)P(\chi_q^2 - q \le v_2)| \\
&+ 2 \sup_{v_1,v_2} P(\chi_q^2 - q \le v_1)|P(V_{m'}^n \le v_2) - P(\chi_q^2 - q \le v_2)| \\
&+ \sup_{v_1,v_2} |P(V_m^n \le v_1) - P(\chi_q^2 - q \le v_1)||P(V_{m'}^n \le v_2) - P(\chi_q^2 - q \le v_2)| \\
&= O(n^{-1/2})
\end{aligned}$$

for all m, m'.

Because of this, the sequence $\{V_m^n\}_{m=1}^M$ is α-mixing with $\alpha(M) = O(n^{-1/2})$. Define

$$\begin{aligned}
\sigma^2 &= \lim_{M\to\infty} \text{Var}\left(\frac{1}{\sqrt{M}} \sum_{m=1}^M V_m^n\right) \\
&= \lim_{M\to\infty} \left(2q + \text{const.} \times n^{-1} + 2(M-1) \times \text{const.} \times n^{-1}\right) \\
&= 2q,
\end{aligned}$$

since $Mn^{-1} = o(1)$. Under A6 and the assumption that $\sum_{m=1}^M \alpha(M)^{1/2} < \infty$, which is equivalent to $M^4 n^{-1} = O(1)$, we have

$$\bar{W}_{n,M} \Rightarrow N(0, 2q)$$

by Theorem 7.8 in Durrett (2005). Finally, since $W_{n,M} - \bar{W}_{n,M} = o_p(1)$, we conclude that

$$W_{n,M} \Rightarrow N(0, 2q).$$

Q.E.D.

REFERENCES

ABREVAYA, J., AND C. M. DAHL (2008): "The effects of birth inputs on birthweight: evidence from quantile estimation on panel data," *Journal of Business and Economic Statistics*, 26, 379–397.

ANGRIST, J., V. CHERNOZHUKOV, AND I. FERNANDEZ-VAL (2006): "Quantile regression under misspecification, with an application to the U.S. wage structure," *Econometrica*, 74, 539–563.

ARELLANO, M., AND S. BONHOMME (2011): "Nonlinear panel data analysis," *Annual Review of Economics*, 3, 395–424.

——— (2016): "Nonlinear panel data estimation via quantile regressions," *Econometrics Journal*, 19, C61–C94.

ARELLANO, M., AND S. BONHOMME (2017): "Sample Selection in Quantile Regression: A Survey," in *Handbook of Quantile Regression*, ed. by R. Koenker, V. Chernozhukov, X. He, and L. Peng. Chapman and Hall/CRC.

BALTAGI, B. H. (2013): *Econometric Analysis of Panel Data*. Wiley, 5th Edition, West Sussex, United Kingdom.

BALTAGI, B. H., L. MATYAS, AND P. SEVESTRE (1995): "Error components models," in *The Econometrics of Panel Data*, ed. by L. Matyas, and P. Sevestre, pp. 49–88. Springer.

BISHOP, J. H. (1991): "Achievement, test scores, and relative wages," in *Workers and their wages: Changing patterns in the United States*, ed. by M. H. Kosters. Washington, DC: American Enterprise Institute Press.

BLACKBURN, M. L., AND D. NEUMARK (1993): "Are OLS estimates of the return to schooling biased downward? Another look," Discussion paper, NBER.

CAI, Z., L. CHEN, AND Y. FANG (2015): "Semiparametric estimation of partially varying-coefficient dynamic panel data models," *Econometric Reviews*, 34(6-10), 695–719.

CANAY, I. A. (2011): "A simple approach to quantile regression for panel data," *Econometrics Journal*, 14, 368–386.

CARD, D. (1995): "Earnings, schooling, and ability revisited," in *Research in Labor Economics*, ed. by S. Polachek, vol. 14. JAI Press.

——— (1999): "The causal effect of education on earnings," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, vol. 3. Amsterdam: Elsevier.

CASTEX, G., AND E. K. DECHTER (2014): "The changing roles of education and ability in wage determination," *Journal of Labor Economics*, 32, 685–710.

CHAMBERLAIN, G. (1982): "Multivariate regression models for panel data," *Journal of Econometrics*, 18, 5–46.

CHAY, K. Y., AND D. S. LEE (2000): "Changes in relative wages in the 1980s Returns to observed and unobserved skills and black–white wage differentials," *Journal of Econometrics*, 99(1), 1–38.

CHERNOZHUKOV, V., I. FERNANDEZ-VAL, J. HAHN, AND W. NEWEY (2013): "Average and quantile effects in nonseparable panel models," *Econometrica*, 81, 535–580.

CHERNOZHUKOV, V., I. FERNANDEZ-VAL, S. HODERLEIN, H. HOLZMANN, AND W. NEWEY (2015): "Nonparametric identification in panels using quantiles," *Journal of Econometrics*, 188(2), 378–392.

CHETVERIKOV, D., B. LARSEN, AND C. PALMER (2016): "IV quantile regression for group-level treatments, with an application to the effects of trade on the distribution of wages," *Econometrica*, 84, 809–833.

DURRETT, R. (2005): *Probability: Theory and Examples*. Cambridge university press.

GALVAO, A. F., AND K. KATO (2016): "Smoothed quantile regression for panel data," *Journal of Econometrics*, 193, 92–112.

——— (2017): "Quantile Regression Methods for Longitudinal Data," in *Handbook of Quantile Regression*, ed. by R. Koenker, V. Chernozhukov, X. He, and L. Peng. CRC/Chapman-Hall.

GALVAO, A. F., C. LAMARCHE, AND L. LIMA (2013): "Estimation of censored quantile regression for panel data with fixed effects," *Journal of the American Statistical Association*, 108, 1075–1089.

GERACI, M., AND M. BOTTAI (2007): "Quantile regression for longitudinal data using the asymmetric Laplace distribution," *Biostatistics*, 8, 140–154.

GOTZE, F. (1991): "On the Rate of convergence in the multivariate CLT," *The Annals of Probability*, 19, 724–739.

GRAHAM, B. S., J. HAHN, A. POIRIER, AND J. L. POWELL (2018): "A quantile correlated random coefficients panel data model," *Journal of Econometrics*, 206, 303–335.

GRAHAM, B. S., AND J. L. POWELL (2012): "Identification and estimation of average partial effects in

"irregular" correlated random coefficient panel data models," *Econometrica*, 80, 2105–2152.
GROGGER, J., AND E. EIDE (1995): "Changes in college skills and the rise in the college wage premium," *Journal of Human Resources*, 30, 280–310.
GUTENBRUNNER, C., AND J. JURECKOVA (1992): "Regression rank scores and regression quantiles," *Annals of Statistics*, 20, 305–330.
HAGEMANN, A. (2017): "Cluster-robust bootstrap inference in quantile regression models," *Journal of the American Statistical Association*, 112, 446–456.
HARMON, C. P., AND H. OOSTERBEEK (2000): "The returns to education: a review of the evidence, issues and deficiencies in the literature," CEE, LSE.
HAUSMAN, J. A. (1978): "Specification tests in econometrics," *Econometrica*, 46, 1251–1271.
JUN, S. J., AND J. PINKSE (2009): "Efficient semiparametric seemingly unrelated quantile regression estimation," *Econometric Theory*, 25(05), 1392–1414.
JURECKOVÁ, J., AND P. K. SEN (1996): *Robust Statistical Procedures: Asymptotics and Interrelations*, vol. 311. John Wiley & Sons.
KATO, K., A. F. GALVAO, AND G. MONTES-ROJAS (2012): "Asymptotics for panel quantile regression models with individual effects," *Journal of Econometrics*, 170, 76–91.
KIM, M.-O., AND Y. YANG (2011): "Semiparametric approach to a random effects quantile regression model," *Journal of the American Statistical Association*, 106, 1405–1417.
KNIGHT, K. (2001): "Comparing conditional quantile estimators: first and second order considerations," Mimeo.
KOENKER, R. (2004): "Quantile regression for longitudinal data," *Journal of Multivariate Analysis*, 91, 74–89.
KOENKER, R., AND G. W. BASSETT (1978): "Regression quantiles," *Econometrica*, 46, 33–49.
——— (1982a): "Robust tests for heteroscedasticity based on regression quantile," *Econometrica*, 50, 43–61.
——— (1982b): "Tests of linear hypothesis and l_1 estimation," *Econometrica*, 50, 1577–1584.
KOENKER, R., AND J. A. F. MACHADO (1999): "Godness of fit and related inference processes for quantile regression," *Journal of the American Statistical Association*, 94, 1296–1310.
LAMARCHE, C. (2010): "Robust Penalized Quantile Regression Estimation for Panel Data," *Journal of Econometrics*, 157, 396–408.
MUNDLAK, Y. (1978): "On the pooling of time series and cross section data," *Econometrica*, 46, 69–85.
NEWEY, W. K., AND J. L. POWELL (1990): "Efficient estimation of linear and type I censored regression models under conditional quantile restrictions," *Econometric Theory*, 6(03), 295–317.
NEYMAN, J., AND E. L. SCOTT (1948): "Consistent estimates based on partially consistent observations," *Econometrica*, 16, 1–32.
PARENTE, P. M., AND J. M. SANTOS-SILVA (2016): "Quantile regression with clustered data," *Journal of Econometric Methods*, 5, 1–16.
PORTNOY, S. (1984): "Tightness of the sequence of empiric cdf processes defined from regression fractiles," in *Robust and Nonlinear Time Series Analysis*, ed. by J. Frank, W. Hardle, and D. Martin. Spring-Verlag: New York.
——— (2012): "Nearly root-n approximation for regression quantile processes," *The Annals of Statistics*, 40(3), 1714–1736.
POWELL, J. L. (1991): "Estimation of monotonic regression models under quantile regressions," in *Nonparametric and Semiparametric Models in Econometrics*, ed. by W. Barnett, J. Powell, and G. Tauchen. Cambridge University Press, Cambridge.
ROSEN, A. (2012): "Set Identification via Quantile Restrictions in Short Panels," *Journal of Econometrics*, 166, 127–137.
TABER, C. R. (2001): "The rising college premium in the eighties: Return to college or return to unobserved ability?," *The Review of Economic Studies*, 68(3), 665–691.
VAN DER VAART, A., AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes*. Springer-Verlag, New York, New York.
WANG, H. J., AND M. FYGENSON (2009): "Inference for censored quantile regression models in longitudinal studies," *Annals of Statistics*, 37, 756–781.
WEI, Y., AND X. HE (2006): "Conditional Growth Charts," *Annals of Statistics*, 34, 2069–2097.
WELLNER, J. A. (2005): "Empirical processes: theory and applications," Mimeo.
ZHAO, Q. (2001): "Asymptotically efficient median regression in the presence of heteroskedasticity of unknown form," *Econometric Theory*, 17(04), 765–784.