

## Article

# Quantile Regression with Generated Regressors

Liqiong Chen <sup>1</sup>, Antonio F. Galvao <sup>2</sup> and Suyong Song <sup>3,\*</sup> <sup>1</sup> Department of Economics, University of Iowa, Iowa City, IA 52242, USA; liqiong-chen@uiowa.edu<sup>2</sup> Department of Economics, University of Arizona, Tucson, AZ 85721, USA; agalvao@email.arizona.edu<sup>3</sup> Department of Economics & Finance, University of Iowa, Iowa City, IA 52242, USA

\* Correspondence: suyong-song@uiowa.edu

**Abstract:** This paper studies estimation and inference for linear quantile regression models with generated regressors. We suggest a practical two-step estimation procedure, where the generated regressors are computed in the first step. The asymptotic properties of the two-step estimator, namely, consistency and asymptotic normality are established. We show that the asymptotic variance-covariance matrix needs to be adjusted to account for the first-step estimation error. We propose a general estimator for the asymptotic variance-covariance, establish its consistency, and develop testing procedures for linear hypotheses in these models. Monte Carlo simulations to evaluate the finite-sample performance of the estimation and inference procedures are provided. Finally, we apply the proposed methods to study Engel curves for various commodities using data from the UK Family Expenditure Survey. We document strong heterogeneity in the estimated Engel curves along the conditional distribution of the budget share of each commodity. The empirical application also emphasizes that correctly estimating confidence intervals for the estimated Engel curves by the proposed estimator is of importance for inference.

**Keywords:** quantile regression; generated regressor; heterogeneity; engel curves

**JEL Classification:** C12; C13; C23



**Citation:** Chen, Liqiong, Antonio F. Galvao and Suyong Song. 2021.

Quantile Regression with Generated Regressors. *Econometrics* 9: 16.

<https://doi.org/10.3390/econometrics9020016>

Academic Editor: In Choi

Received: 16 November 2020

Accepted: 6 April 2021

Published: 12 April 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Since the seminal work of [Koenker and Bassett \(1978\)](#), quantile regression (QR) models have provided a valuable tool in economics, finance, and statistics as a way of capturing heterogeneous effects of covariates on the outcome of interest, exposing a wide variety of forms of conditional heterogeneity under weak distributional assumptions. Importantly, QR also provides a framework for robust inference.

Applied researchers are commonly confronted with the absence of observable regressors in practice. In some cases, proxies for the unobservable variables can be found in data, while in other cases, these regressors need to be estimated. Thus, a very common strategy to deal with unobservable variables is to replace them with estimated values, that is, generated regressors (GRs).<sup>1</sup> These GRs have important implications for the reliability of general standard estimation and inference procedures. [Pagan \(1984\)](#) and [Murphy and Topel \(2002\)](#) point out that even though consistent estimates of parameters of interest are produced when the unobserved regressors are replaced with their estimated values, the conventional ways to estimate standard errors are incorrect. Recently, [Mammen et al. \(2012\)](#) provide a general theory for the impact of GRs on the final estimator's asymptotic properties in nonparametric mean regression. [Hahn and Rider \(2013\)](#) derive the asymptotic distribution of three-step estimators of a finite-dimensional parameter in a semiparametric mean regression models where GR is estimated parametrically or nonparametrically in the first-step.

A number of important statistical applications requires estimation of a conditional quantile function when some of the covariates are not directly observed, but are estimated

in a preliminary step. Examples include, among others, stochastic volatility time-series models, triangular simultaneous equation models, censoring, sample selection models, and treatment effects models.<sup>2</sup>

This paper contributes to the literature by systematically studying and formalizing estimation and inference for linear QR models with a general semiparametric GR. The main contributions are as following. First, we suggest a practical two-step estimation procedure to estimate the parameters of interest in the QR-GR model. The first-step applies a general (semi)-parametric estimator to compute the GRs. The second step uses the GRs as regressor variables in a QR estimation. We establish the asymptotic properties of the two-step QR-GR estimator, namely, consistency and asymptotic normality. We show that the asymptotic variance-covariance matrix needs to be adjusted to account for the first-step estimation error. The basic idea is as follows. Since the first-step estimation of the unobservable regressors produces consistent estimates of the corresponding true parameters, the GRs are consistently estimated. However, the GR are included in the QR model of interest with sampling error, which introduces additional noise into the asymptotic variance-covariance matrix of the coefficients of interest. In other words, the sampling error from the first stage contaminates the second stage estimation. Therefore, the usual way of calculating the QR variance-covariance matrix fails to account for the additional source of error. Under some general conditions, the estimated limiting distribution of the first-step is used to consistently estimate the variance-covariance matrix of the parameters of interest.

The second contribution of the paper is to develop inference procedures for the QR model with GR. Though estimation and inference for conditional average models with GR have been widely studied and used in practice, the literature on QR with GR is more sparse. We develop testing procedures for general linear hypotheses in these models based on Wald-type tests, and derive their associated limiting distributions. To implement the tests, we propose an estimator for the asymptotic variance-covariance of the QR-GR coefficients, and formally establish its consistency. An important advantage of the proposed tests is that they are simple to compute and implement in practice.

Compared to the existing procedures, the QR-GR methods proposed in this paper present several distinctive advantages for applied researchers. First, instead of imposing a linear structure in the first step, as in [Xiao and Koenker \(2009\)](#), or imposing triangular structure between two steps, as in [Ma and Koenker \(2006\)](#) and [Lee \(2007\)](#), we work with the general case of QR-GR where no structural restrictions between the two steps or any specific functional forms and estimation strategies in the first step are imposed. That is useful in applied work since practitioners have a large range of alternatives to construct the unobservable variables by using different estimation strategies or even different data sets. Second, we establish the asymptotic properties of the QR-GR estimator for non-iid data under weak conditions. This is an important generalization for practitioners since it allows for inference in a more general class of models. Third, we develop practical inference procedures. Finally, the weak conditions we imposed for QR-GR allow for simple computational implementation.<sup>3</sup> Linear QR models have been the workhorse of the applied research and the methods lead to a simple algorithm that can be conveniently implemented in empirical applications. Researchers can simply use existing software packages for the first-step estimation and to construct the regressors needed, for example, MLE, OLS, QR or GMM, and then apply the QR procedure with our described variance-covariance matrix adjustment.

Monte Carlo simulations assess the finite-sample properties of the proposed methods. We evaluate the QR-GR estimator in terms of empirical bias and root mean squared error, and compare its performance with methods that are not designed for dealing with GR issues. In addition, we compute the corresponding standard errors of the QR-GR estimator and evaluate its bias. The experiments suggest that the proposed approach performs very well in finite samples and effectively removes the bias of the standard errors induced by the GR. Thus, the proposed variance estimator is approximately unbiased and approximates well the true variance.

Finally, to motivate and illustrate the applicability of the methods, we consider an empirical application to study demand models, using data from the UK Family Expenditure Survey. Demand models (also known as Engel curves) represent the relationship between total expenditure and the share of various commodities to total expenditure. The QR approach is a useful tool in this example because it allows us to capture the heterogeneity in the expenditures of the different commodities along the conditional distribution of each commodity share. The first step in this exercise is to estimate the unobserved motives for budget shares of different commodities using the factor analysis proposed by Barigozzi and Moneta (2016). The unobserved factors consist of the motives of consumption as necessities, luxuries, and unitary elasticity goods. In the second step, we apply the proposed QR-GR estimator to obtain Engel curves for the commodities such as food, housing and leisure, by regressing each commodity share on the estimated motives from the first step. We found that the motives of consumption play different roles for various commodities and contributions of motives to each budget share vary over the total expenditure. Furthermore, the empirical results document important heterogeneity in the Engel curves. The estimated curves present strong heterogeneous effect of the consumption motives on the budget share along the conditional distribution of the budget share in most commodities. Importantly, the empirical study underscores the importance of obtaining correct confidence intervals for the estimated Engel curves by taking into account the GR issue.

The remainder of the article is organized as follows. Section 2 introduces the QR model with GR. Section 3 establishes consistency and asymptotic normality of the two-step estimator. Section 4 proposes a consistent estimator of the variance-covariance matrix. Section 5 presents Monte Carlo simulations. Section 6 presents the empirical application to Engel curves. Section 7 concludes the article. Technical proofs are included in the Appendix A.

## 2. Quantile Regression with Generated Regressors

This section describes the quantile regression (QR) model with generated regressors (GRs) we consider in this paper and the two-step estimation procedure.

### 2.1. Model

For each fixed  $\tau \in (0, 1)$ , we consider the following model

$$y_i = \mathbf{x}_i^\top \boldsymbol{\beta}_0(\tau) + u_i, \quad i = 1, \dots, n, \quad (1)$$

where  $y_i$  is a response variable,  $\mathbf{x}_i = (x_{i1}, \dots, x_{ik})$  is a  $k$ -dimensional vector of explanatory variables,  $\boldsymbol{\beta}_0(\tau)$  is a  $k \times 1$  vector of parameters, and the innovation term  $u_i$  has conditional  $\tau$ -quantile zero, that is  $F_\tau^{-1}(u_i | \mathbf{x}_i) = 0$ .

When all the regressors  $\mathbf{x}_i$  in model (1) are observable, the model can be written as the following standard QR model

$$Q_\tau(y_i | \mathbf{x}_i) = \mathbf{x}_i^\top \boldsymbol{\beta}_0(\tau), \quad (2)$$

where  $Q_\tau(y_i | \mathbf{x}_i)$  is the conditional  $\tau$ -quantile of  $y_i$  given  $\mathbf{x}_i$ . In general,  $\boldsymbol{\beta}_0$  can depend on  $\tau$ . The model is semiparametric in the sense that the functional form of the conditional distribution of  $y_i$  given  $\mathbf{x}_i$  is left unspecified.

The parameter of interest for the researcher is  $\boldsymbol{\beta}_0(\tau)$  in model (2). However, in many applications one or more elements of the vector of regressors  $\mathbf{x}_i$  may not be directly observable, but instead estimated from a model with other given variables, that is, the GRs. In this paper, we assume that some of the regressors in the vector  $\mathbf{x}_i$  are not observable to the researcher, i.e.,  $(x_{i1}, \dots, x_{iq})$  are not directly observable, but  $(x_{iq+1}, \dots, x_{ik})$  are observed, where  $q \leq k$ . In particular, the GR are assumed to have the following form

$$x_{ij} = g_j(\mathbf{w}_i, \boldsymbol{\theta}_j) + v_{ij},$$

where the function  $g_j(\cdot, \cdot)$  is differentiable and known up to the unknown  $p_j \times 1$  parameter vector  $\theta_j$  for  $j = 1, \dots, q$ , and the variable  $\mathbf{w}_i = (w_{i1}, \dots, w_{iq})$  is a  $q$ -dimensional vector of observables, and where  $v_{ij}$  is an error term. Thus, one can still estimate the QR model in (2) by replacing  $x_{ij}$  with the GR  $\hat{x}_{ij}$  for  $j = 1, \dots, q$ . The GR  $\hat{x}_{ij}$  is obtained from the following first-step estimation

$$\hat{x}_{ij} = g_j(\mathbf{w}_i, \hat{\theta}_j), \quad (3)$$

where  $\hat{\theta}_j$  satisfies very general weak conditions as:  $\sqrt{n}(\hat{\theta}_j - \theta_j) = n^{-1/2} \sum_{i=1}^n \mathbf{r}_i(\theta_j) + o_p(1)$ , and  $\mathbf{r}(\cdot)$  is a generic function satisfying  $E[\mathbf{r}_i(\theta_j)] = 0$ . Equation (3) defines the GR and could be satisfied by different models, we provide a few examples below. Notice that most estimators used in empirical applications satisfy this first-order representation. We will discuss these conditions more formally below. To complete the model, together with (2) and (3), we assume that  $F_{\tau}^{-1}(u_i | \mathbf{w}_{i1}, \dots, \mathbf{w}_{iq}, x_{iq+1}, \dots, x_{ik}) = 0$ .

**Remark 1.** One or more elements of the regressors  $\mathbf{x}_i$  may be estimated in the first step. Each of the GRs could be related to different observable variables in different functional forms. For simplicity, each of the GR equations can be estimated separately, and any dependence between the GRs would be captured by the variance covariance matrix.

**Remark 2.** We note that the estimation of quantile regression models with mismeasured regressors leads to inconsistent estimates and a few methodologies have been proposed to overcome this drawback (e.g., Wei and Carroll (2009); Wang et al. (2012); Firpo et al. (2017)). Thus, the problem is different from the quantile regression with generated regressors where the standard quantile estimation is still consistent.

To motivate the existence of GRs for QR models, and illustrate how a QR-GR framework could appear in practice, we include the following motivating examples:

**Example 1** (Two-stage regression with proxy variables). A very important example of GR occurs when the variables  $\mathbf{x}_i$  are not directly observable. For instance, assume that the variables  $(x_{i1}, \dots, x_{iq})$  are related to additional observable variables,  $\mathbf{w}_{ij}$ , as follows

$$\begin{aligned} \tilde{x}_{i1} &= x_{i1} + v_{i1} = g_1(\mathbf{w}_{i1}, \theta_1) + v_{i1}, \\ &\vdots \\ \tilde{x}_{iq} &= x_{iq} + v_{iq} = g_q(\mathbf{w}_{iq}, \theta_q) + v_{iq} \end{aligned}$$

where  $(\tilde{x}_{i1}, \dots, \tilde{x}_{iq})$  are proxy variables, or endogenous observables,  $\mathbf{w}_{ij}$  is a vector of exogenous observable variables,  $\theta_j$  is a  $p_j \times 1$  vector, the functions  $g_j(\cdot)$  are unknown up to the vector  $\theta_j$ , and  $v_{ji}$  are mutually independent innovation terms, for  $j = 1, \dots, q$ .

In this case, to complete the definition of the GR one needs to impose more structure on the innovation term  $v_{ji}$  to estimate the parameters  $\theta_j$ , for  $j = 1, \dots, q$ . As special cases of a more general procedure which generates the regressors, consider a simple but commonly used linear model to generate one regressor,  $x_{i1}$ , as a function of several variables  $\mathbf{w}_i$ , that is,  $\tilde{x}_1 = g(\mathbf{w}, \theta) + v = \mathbf{w}^\top \theta + v$ . The following are two standard examples.

**Example 1.1** (Conditional average). A simple example for the GR model is a linear conditional expectation. In this case, the GR model is defined as

$$E[\tilde{x}_{i1} | \mathbf{w}_i] = \mathbf{w}_i^\top \theta.$$

**Example 1.2** (Conditional quantile). Another simple model for the GR is a linear conditional quantile. In this case, for a given quantile  $\tau'$  specified by the researcher, the GR model is defined as

$$Q_{\tau'}(\tilde{x}_{i1} | \mathbf{w}_i) = \mathbf{w}_i^\top \theta(\tau'),$$

In practice one needs to estimate the parameters  $\theta$  in both examples and compute the GR.

**Example 2** (Quantile regression with endogeneity: control function approach). Consider the following model

$$y_i = x_{i1}\beta_1 + z_i^\top \beta_2 + u_i \quad (4)$$

where  $x_{i1}$  is the endogenous explanatory variable,  $\beta_1$  is a scalar parameter for simplicity, and  $z_i$  is the  $(k-1) \times 1$  vector of exogenous variables,  $\beta_2$  is a  $(k-1) \times 1$  vector of parameters. Assume that  $w_{i1}$  are valid instruments. In the control function approach, one could model the endogenous explanatory variable as following:

$$x_{i1} = g_1(w_{i1}, \theta_1) + v_{1i} \quad (5)$$

where  $w_{i1}$  contains exogenous regressors  $z_i$ . The source of the endogeneity in the model (4) comes from the correlation between  $u_i$  and  $v_{1i}$ . Thus, the endogeneity can be solved by controlling for  $v_{1i}$  in the model. Since we do not observe  $v_{1i}$ , we can replace  $v_{1i}$  with  $\hat{v}_{1i} \equiv x_{i1} - g_1(w_{i1}, \hat{\theta}_1)$ , where  $\hat{v}_{1i}$  could be obtained by either mean or quantile regression. This gives

$$y_i = x_{i1}\beta_1 + z_i^\top \beta_2 + \gamma \hat{v}_{1i} + \eta_i$$

where  $\eta_i = \epsilon_i + \gamma[g_1(w_{i1}, \theta_1) - g_1(w_{i1}, \hat{\theta}_1)]$  which depends on the sampling error in  $\hat{\theta}_1$ . Notice that  $\hat{v}_{1i}$  is a generated regressor. Moreover, it is worthwhile noting that although the control function approach for the endogeneity issue is one of examples for the GR as in e.g., Lee (2007), we are not particularly attempting to solve the endogeneity issue in this paper. Instead, we provide a general framework and establish the asymptotic properties for the quantile models with GR.

## 2.2. Estimation

The estimation of the parameters of interest in model (2) involves a two-step estimation. In the first step one estimates and computes the GR from model (3). In the second step one uses the GR and other regressors, and computes the QR of interest from (2).

Estimation of the unobservables usually comes from the same sample of data – may come from a different dataset or even from the parameter estimates by another researcher. Models used to estimate the unknown parameter  $\theta_j$  may generally include linear or nonlinear models. Additionally, the parameters can be estimated by various strategies. The QR-GR two-step estimation procedure is as following:

**Step 1** Estimate  $\theta_j$  from (3) and compute the fitted values  $\hat{x}_{ij} = g_j(w_{ji}, \hat{\theta}_j)$  for  $j = 1, \dots, q$ , and then obtain the generated regressors  $\hat{x}_i = (\hat{x}_{i1}, \dots, \hat{x}_{iq}, x_{iq+1}, \dots, x_{ik})^\top$  for  $i = 1, \dots, n$ .

**Step 2** Compute  $\beta$  from the following QR

$$\hat{\beta}(\tau) = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^n \rho_\tau(y_i - \hat{x}_i^\top \beta), \quad (6)$$

where  $\rho_\tau(u) := \{\tau - I(u \leq 0)\}u$  is the check function as in Koenker and Bassett (1978).

Thus, one uses the estimates of  $\theta_j$ , denoted by  $\hat{\theta}_j$ , in the first step, to obtain the GR  $\hat{x}_i$ . In the examples for the average and quantile models discussed previously we have the following for the first step.

**Example 1.1** (Average continued). In this example, one employs the standard OLS estimator and obtains

$$\hat{\theta} = (w^\top w)^{-1} w^\top \tilde{x}_1$$

and computes  $\hat{x}_{i1} = w^\top \hat{\theta}$ , and also  $\hat{x}_i = (\hat{x}_{i1}, x_{i2}, \dots, x_{ik})^\top$ . Then, the  $\tau$ -th QR estimator  $\hat{\beta}(\tau)$  can be obtained by (6).

**Example 1.2** (Quantile continued). In this case, for a given quantile  $\tau'$  one applies the usual QR procedure to estimate  $\theta(\tau')$  and obtain  $\hat{x}_{i1} = \mathbf{w}_i^\top \hat{\theta}(\tau')$ . Thus, the first-step estimation is given by the following QR

$$\hat{\theta}(\tau') = \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^n \rho_\tau(\tilde{x}_{i1} - \mathbf{w}_i^\top \theta).$$

Then, in the second step, for a given  $\tau$ -quantile, the QR estimator  $\hat{\beta}(\tau)$  can be obtained by (6).

These two practically-common cases illustrate the simple implementation of QR model with GR. We have made R routines for the QR-GR estimator and inference in the QR-GR framework available for the practitioners.

### 3. Asymptotic Properties

We now establish consistency and asymptotic normality of the QR-GR two-step estimator,  $\hat{\beta}(\tau)$ , defined in the previous section. Proofs are collected in the Appendix A. We consider the following regularity conditions:

**A1.**  $\{(y_i, \mathbf{x}_i, \mathbf{w}_i)\}_{i=1}^n$  is independent across  $i$ . The conditional distribution functions of the error term  $\{F_i(u|\mathbf{x}_i)\}$  have continuous densities  $f_i(u|\mathbf{x}_i)$  with a unique conditional  $\tau$ -th quantile equal to 0, and  $f_i(0|\mathbf{x}_i)$  are uniformly bounded away from 0 and  $\infty$ .

**A2.** There exist positive definite matrices  $D_0$  and  $D_1$  such that

- (i)  $D_0 = \operatorname{plim}_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\top,$
- (ii)  $D_1 = \operatorname{plim}_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n f_i(0|\mathbf{x}_i) \mathbf{x}_i \mathbf{x}_i^\top,$
- (iii)  $\max_{i=1, \dots, n} \|\mathbf{x}_i\|/\sqrt{n} \xrightarrow{p} 0.$

**A3.**  $\operatorname{plim}_{n \rightarrow \infty} \hat{\theta} = \theta.$

**A4.**  $\sqrt{n}(\hat{\theta} - \theta) = n^{-1/2} \sum_{i=1}^n \mathbf{r}_i(\theta) + o_p(1)$  where  $\mathbf{r}_i(\cdot)$  is a continuous function which satisfies that  $E[\mathbf{r}_i(\theta)] = 0$  and  $\operatorname{Var}[\mathbf{r}_i(\theta)] = V$ .

**A5.** There exists a positive definite matrix  $D_{12}VD_{12}^\top - 2M$  where

- (i)  $D_{12} = \operatorname{plim}_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n f_i(0|\mathbf{x}_i) \beta(\tau) \bar{\mathbf{x}}_i \nabla_\theta \mathbf{g}(\mathbf{w}_i, \theta)^\top,$  where  $\nabla_\theta \mathbf{g}(\mathbf{w}_i, \theta)$  is  $\sum_{j=1}^q p_j \times q$  Jacobian of  $\mathbf{g}(\mathbf{w}_i, \theta),$
- (ii)  $M = \operatorname{plim}_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \{\psi_\tau(y_i - \mathbf{x}_i^\top \beta_0(\tau)) f_i(0|\mathbf{x}_i) \mathbf{x}_i \mathbf{r}(\theta)^\top \nabla_\theta \mathbf{g}(\mathbf{w}_i, \theta)^\top \bar{\mathbf{x}}_i^\top \beta(\tau)^\top\},$  where  $\psi_\tau(u) := \tau - I(u \leq 0).$

Conditions **A1** and **A2** are the usual conditions in the QR literature. Assumptions **A1** allows for non-iid sampling, and **A2** requires limiting matrices to be well defined. Assumptions **A3**–**A5** refer to the GR estimation in the first step and need to be verified for each empirical application. These conditions are very mild. Assumptions **A3** and **A4** impose consistency and asymptotic normality, respectively, for the first-step estimator of the GR. These conditions hold for most estimators employed in empirical work in the family of M- and Z-estimators. We note that no restrictions are imposed on the functional form of  $\mathbf{g}_j(\cdot)$  except condition **A5** which is a weak smoothness condition needed only for nonlinear models.

The following result states the consistency of the QR-GR estimator.

**Theorem 1** (Consistency). Consider the model in (2) and (3). Under the conditions **A1**–**A3** and **A5**, as  $n \rightarrow \infty$

$$\hat{\beta}(\tau) \xrightarrow{p} \beta_0(\tau).$$

**Remark 3.** A consistent estimate of the unknown parameter  $\theta_j$  for  $j = 1, \dots, q$  in the first step suffices for the consistency of QR with GR. In other words, replacing  $\mathbf{x}$  by  $\hat{\mathbf{x}}$  in a quantile regression still gives us a consistent QR estimator.



The next result establishes the asymptotic normality of the QR-GR two-step estimator.

**Theorem 2** (Asymptotic normality). *Consider the model in (2) and (3). Under conditions A1–A5, as  $n \rightarrow \infty$ , we have that,*

$$\sqrt{n}(\hat{\beta}(\tau) - \beta_0(\tau)) \xrightarrow{d} N(0, \Omega(\tau)),$$

where  $\Omega(\tau) := \tau(1 - \tau)D_1^{-1}D_0D_1^{-1} + D_1^{-1}D_{12}VD_{12}^\top D_1^{-1} - 2D_1^{-1}MD_1^{-1}$ .

The result in Theorem 2 shows that the limiting distribution of  $\sqrt{n}(\hat{\beta}(\tau) - \beta_0(\tau))$  generally depends on the statistical properties of  $\sqrt{n}(\hat{\theta} - \theta)$  since the sampling error of generating regressor in model (3) contaminates the variance-covariance estimation in the second stage. Hence, the asymptotic variance-covariance of the two-step estimator needs to be adjusted when the regressors are generated. The adjusted variance-covariance matrix is larger, since the sampling error  $V$  from the first-step estimation enters variance-covariance matrix. Nevertheless, the sampling variation of  $\hat{\theta}_j$  can be ignored, at least asymptotically, if the coefficients  $\beta_j$  of the GR are zero, i.e.,  $D_{12}^j = 0$  and the corresponding part in  $M$  becomes zero. In that case, replacing the true regressors with GR will not impact the asymptotic distribution of the estimator.

**Remark 4.** *In general, the consistency of QR estimator with GR holds when the true regressor is replaced by GR. However, the asymptotic variance-covariance matrix of the estimator  $\hat{\beta}(\tau)$  needs to be adjusted because of the sampling variation introduced by estimation of  $\hat{\theta}_j$  in the first step.*

**Remark 5.** *Comparing the QR-GR estimator  $\hat{\beta}(\tau)$  with regular QR estimator with the true unobserved regressors, we see that the additional second term appears in the variance-covariance matrix coming from the first-step estimation. The first-step estimation contaminates the variance-covariance matrix of  $\hat{\beta}(\tau)$  in two ways: one is the sampling error  $V_j$  of coefficient estimates  $\theta_j$  in the first stage, the other is the gradient of model specification with respect to the parameter in the first stage. For the simple linear model in the first stage, the gradient is simply the regressors in the first stage. However, for nonlinear model in the first stage, the coefficient estimates show up in  $D_{12}^j$  which makes the variance-covariance matrix of  $\hat{\beta}(\tau)$  larger than when the true regressors are used for estimation.*

The two-step estimation procedure in this paper is easy to implement in practice. First, since weak conditions are needed for QR-GR procedure, different estimation strategies may be used in the first step to construct the GR. Most common estimators in practice satisfy the weak conditions: for example, the simple OLS, QR, MLE methods, etc. Second, both linear and nonlinear model specifications in the first step are allowed. Finally, it is important to notice that weak conditions in the QR step include non-iid models which allow practitioners to proceed inference in a general class of models.

To further illustrate the estimation of QR models with GR, we further discuss the two common models, OLS and QR with GR. We discuss the case where only one regressor is generated in the first step. For both models, after estimating  $\hat{\theta}$  which satisfies the condition  $\sqrt{n}(\hat{\theta} - \theta) = n^{-1/2} \sum_{i=1}^n r_i(\theta) + o_p(1)$  where  $E[r_i(\theta)] = 0$  and  $\text{Var}[r_i(\theta)] = V$  in the first step, one obtains the GR:  $\hat{x} = (\hat{x}_1, x_2, \dots, x_k)^\top$  where  $\hat{x}_1 = g(w, \hat{\theta}) = w^\top \hat{\theta}$ . The following examples derive the asymptotic variance-covariance matrix for both OLS and QR with GR:

**Example 3** (OLS with GR). *In this example,  $\hat{\beta}$  is estimated from the model*

$$E(y_i | \hat{x}_i) = \hat{x}_i^\top \beta.$$

Since the OLS estimator has a closed-form solution, one obtains  $\hat{\beta} = (\sum_{i=1}^n \hat{x}_i \hat{x}_i^\top)^{-1} (\sum_{i=1}^n \hat{x}_i y_i)$ . Finally, the variance-covariance matrix is given by the following result which is proved in the Appendix A, for completeness. Note that according to Proposition A2 in the Appendix A, under conditions C1–C5,

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, H^{-1} \text{Var}[x_i u_i - G r_i(\theta)] H^{-1}),$$

where  $H = \text{plim}_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n (x_i x_i^\top)$ , and  $G = \text{plim}_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n (\beta \otimes x_i^\top)^\top ([1 \ 0 \ \dots \ 0]^\top \otimes w_i^\top)$ . When there is no GR or the coefficients of the GR are zero, the asymptotic variance displayed above simplifies since  $G = 0$ . In the case of  $\text{Var}(u|X) = \sigma_u^2 I_n$ , where  $\sigma_u^2 > 0$  and  $I_n$  is an  $n \times n$  identity matrix, we obtain that  $\sqrt{n}(\hat{\beta} - \beta_0) = \sigma_u^2 H^{-1}$ , since  $E(x_i u_i^2 x_i^\top) = \sigma_u^2 H$ . In addition, in the case of  $\text{Var}(u|X) = \Gamma_0$  and  $\text{plim}_{n \rightarrow \infty} n^{-1} (X^\top \Gamma_0 X) = \Gamma_X$ , where  $\Gamma_X$  is positive definite matrix, since  $E(x_i u_i^2 x_i^\top) = \Gamma_X$ , we have that  $\sqrt{n}(\hat{\beta} - \beta_0) = H^{-1} \Gamma_X H^{-1}$ .

**Example 4 (QR with GR).** In this example,  $\hat{\beta}(\tau)$  is estimated from the model

$$Q_\tau(y_i | \hat{x}_i) = \hat{x}_i^\top \beta(\tau).$$

Unlike the simple closed-form solution in the above OLS estimate, one applies the usual QR procedure to estimate

$$\underset{\beta}{\text{argmin}} \sum_{i=1}^n \rho_\tau(y_i - \hat{x}_i^\top \beta).$$

Theorem 2 shows that

$$\sqrt{n}(\hat{\beta}(\tau) - \beta_0(\tau)) \xrightarrow{d} N(0, D_1^{-1} \{ \text{Var}[x_i \psi_\tau(y_i - \xi_i(\tau))] + \text{Var}[D_{12} r_i(\theta)] - 2M \} D_1^{-1}),$$

where  $D_0$ ,  $D_1$ , and  $D_{12}$  are defined in conditions A2 and A5, and  $M$  is simplified to  $M = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \{ \psi_\tau(y_i - x_i^\top \beta_0(\tau)) f_i(0 | x_i) \beta_1(\tau) w_i^\top r_i(\theta_j) x_i x_i^\top \}$ . In addition, one can notice that the above asymptotic variance-covariance matrix can be simplified to

$$D_1^{-1} \left( \tau(1 - \tau) D_0 + D_{12} V D_{12}^\top - 2M \right) D_1^{-1}.$$

To conclude this section we have two remarks. First, from the second example, one can notice that the variance for the QR with GR has the additional terms  $D_1^{-1} D_{12} V D_{12}^\top D_1^{-1}$  and  $-2D_1^{-1} M D_1^{-1}$  relative to the standard QR model without GR, which has the variance  $\tau(1 - \tau) D_1^{-1} D_0 D_1^{-1}$ . Second, intuitively, as shown in Theorem 1 and Proposition A1, in both OLS and QR frameworks, estimation with GR still produces consistent estimates. However, as shown in Theorem 2 and Proposition A2, the corresponding variance-covariance matrices need to account for two sources of error—the usual estimation error in the OLS or QR method, and the sampling error in generating the regressors.

#### 4. Inference

In this section, we turn our attention to inference in the QR-GR model. First, we suggest an estimator for the asymptotic variance-covariance matrix of the QR-GR estimator. Second, we propose a Wald-type test for general linear hypotheses.

##### 4.1. Variance-Covariance Matrix Estimation

In applications, the variance-covariance matrices are unknown and need to be estimated. Now we suggest an estimator for the corresponding variance of the QR-GR estimator. The estimator is closely related to those suggested by Hendricks and Koenker (1991) and Powell (1991) and given as the following form

$$\hat{\Omega}(\tau) = \tau(1 - \tau) \hat{D}_1^{-1} \hat{D}_0 \hat{D}_1^{-1} + \hat{D}_1^{-1} \hat{D}_{12} \hat{V} \hat{D}_{12}^\top \hat{D}_1^{-1} - 2 \hat{D}_1^{-1} \hat{M} \hat{D}_1^{-1} \quad (7)$$



where

$$\hat{D}_0 = \frac{1}{n} \sum_{i=1}^n \hat{\mathbf{x}}_i \hat{\mathbf{x}}_i^\top \quad (8)$$

$$\hat{D}_1 = \frac{1}{2nc_n} \sum_{i=1}^n I(|y_i - \hat{\mathbf{x}}_i^\top \hat{\boldsymbol{\beta}}(\tau)| < c_n) \hat{\mathbf{x}}_i \hat{\mathbf{x}}_i^\top \quad (9)$$

$$\hat{D}_{12} = (2nc_n)^{-1} \sum_{i=1}^n I(|\hat{\epsilon}_i| < c_n) \hat{\boldsymbol{\beta}}(\tau) \hat{\mathbf{x}}_i \nabla_{\theta} \mathbf{g}(\mathbf{w}_i, \hat{\boldsymbol{\theta}})^\top \quad (10)$$

$$\hat{M} = \frac{1}{2nc_n} \sum_{i=1}^n \{\psi_{\tau}(y_i - \hat{\mathbf{x}}_i^\top \hat{\boldsymbol{\beta}}(\tau)) I(|y_i - \hat{\mathbf{x}}_i^\top \hat{\boldsymbol{\beta}}(\tau)| < c_n) \cdot \hat{\mathbf{x}}_i \mathbf{r}(\hat{\boldsymbol{\theta}})^\top \nabla_{\theta} \mathbf{g}(\mathbf{w}_i, \hat{\boldsymbol{\theta}})^\top \hat{\mathbf{x}}_i^\top \hat{\boldsymbol{\beta}}(\tau)^\top\}. \quad (11)$$

To establish the consistency of  $\hat{\Omega}(\tau)$  we impose the following assumptions:

**B1.** There exists a positive sequence of bandwidths  $\{c_n\}$  such that  $c_n \rightarrow 0$  and  $\sqrt{nc_n} \rightarrow \infty$ .

**B2.**  $E(\|\mathbf{x}_i\|^4) \leq H_1 < \infty$  for all  $i$  and for some constant  $H_1$ .

**B3.**  $\|\nabla_{\theta_j} g_j(\mathbf{w}_i, \boldsymbol{\theta}_j)\| \leq D_j(\mathbf{w}_i)$  for all  $\boldsymbol{\theta}_j \in \Theta_j$  and for all  $(i, j)$  where  $\max_j E[D_j(\mathbf{w}_i)^2] \leq D_0 < \infty$  and  $\|\nabla_{\theta} \mathbf{g}(\mathbf{w}_i, \boldsymbol{\theta}) - \nabla_{\gamma} \mathbf{g}(\mathbf{w}_i, \boldsymbol{\gamma})\| \leq J \|\boldsymbol{\theta} - \boldsymbol{\gamma}\|$  for some constant  $J < \infty$ .

**B4.** There exists a bounded function  $A_f(\mathbf{x}_i)$  such that  $f_i(\lambda|\mathbf{x}_i) \leq A_f(\mathbf{x}_i)$  for all  $i$  and for some  $\lambda$  near zero a.s.<sup>4</sup>

**B5.**  $f_i(\lambda|\mathbf{x}_i)$  satisfies the Lipschitz condition that  $|f_i(\lambda_1|\mathbf{x}_i) - f_i(\lambda_2|\mathbf{x}_i)| \leq L_i |\lambda_1 - \lambda_2|$  for some constant  $L_i < \infty$  and for all  $i$ .

Assumption **B1** is a restriction on the bandwidth  $c_n$ , which is commonly used to estimate unknown conditional densities. **B2** imposes some moment condition, which is typical in the QR literature. Assumption **B3** imposes smoothness and dominance conditions that are typical for nonlinear models (see, e.g., [Powell \(1991\)](#)). **B4** imposes some local restriction on the conditional density function near zero and satisfies some moment bounds, which can be weakened at the cost of moment bounds for various cross products of the bounding functions. (see, e.g., Assumption ER of [Buchinsky and Hahn \(1998\)](#)). **B5** imposes smoothness condition on the conditional densities and is standard in the QR literature.

The next result states the consistency of the variance-covariance QR-GR estimator.

**Theorem 3** (Variance-covariance matrix estimation). *Under the assumptions **B1–B5** and conditions of Theorem 2, as  $n \rightarrow \infty$*

$$\hat{\Omega}(\tau) \xrightarrow{p} \Omega(\tau).$$

*In other words,  $\hat{D}_0 \xrightarrow{p} D_0$ ,  $\hat{D}_1 \xrightarrow{p} D_1$ ,  $\hat{D}_{12} \xrightarrow{p} D_{12}$  and  $\hat{M} \xrightarrow{p} M$ , where  $D_0$ ,  $D_1$  and  $D_{12}$  are defined in conditions **A2** and **A5**, and  $M$  is defined in Theorem 3.*

As a special case of more general procedure which generates the regressors, consider a linear model  $\tilde{x}_1 = g(\mathbf{w}, \boldsymbol{\theta}) + v = \mathbf{w}^\top \boldsymbol{\theta} + v$ . The gradient of  $g(\mathbf{w}, \boldsymbol{\theta})$ , denoted by  $\nabla_{\theta} g(\mathbf{w}, \boldsymbol{\theta})$ , is reduced to  $\mathbf{w}$ , in the following examples for the average and quantile models:

**Example 1.1** (Average continued). *In this example, one employs the standard OLS estimator in the first step to obtain*

$$\hat{\boldsymbol{\theta}} = (\mathbf{w}^\top \mathbf{w})^{-1} \mathbf{w}^\top \tilde{\mathbf{x}}_1,$$

*and the standard error  $\hat{V}$ . Then, one is able to compute  $\hat{\mathbf{x}}_{1i} = \mathbf{w}_i^\top \hat{\boldsymbol{\theta}}$ , as well as  $\hat{\mathbf{x}}_i = (\hat{x}_{1i}, x_{i2}, \dots, x_{ik})^\top$ . Finally, with the second step estimation  $\hat{\boldsymbol{\beta}}$ , the estimator of the corresponding variance-covariance matrix is simply  $\hat{\Omega}(\tau) := \tau(1 - \tau) \hat{D}_1^{-1} \hat{D}_0 \hat{D}_1^{-1} + \hat{D}_1^{-1} \hat{D}_{12} \hat{V} \hat{D}_{12}^\top \hat{D}_1^{-1} - 2 \hat{D}_1^{-1} \hat{M} \hat{D}_1^{-1}$ .*

$$\begin{aligned}\hat{D}_{12} &= (2nc_n)^{-1} \sum_{i=1}^n I(|y_i - \hat{\mathbf{x}}_i^\top \hat{\boldsymbol{\beta}}(\tau)| < c_n) \hat{\boldsymbol{\beta}}_1(\tau) \hat{\mathbf{x}}_i \mathbf{w}_i^\top \\ \hat{D}_0 &= n^{-1} \sum_{i=1}^n \hat{\mathbf{x}}_i \hat{\mathbf{x}}_i^\top \\ \hat{D}_1 &= (2nc_n)^{-1} \sum_{i=1}^n I(|y_i - \hat{\mathbf{x}}_i^\top \hat{\boldsymbol{\beta}}(\tau)| < c_n) \hat{\mathbf{x}}_i \hat{\mathbf{x}}_i^\top \\ \hat{M} &= (2nc_n)^{-1} \sum_{i=1}^n \left\{ [\tau - I(y_i - \hat{\mathbf{x}}_i^\top \hat{\boldsymbol{\beta}}(\tau) < 0)] I(|y_i - \hat{\mathbf{x}}_i^\top \hat{\boldsymbol{\beta}}(\tau)| < c_n) \hat{\boldsymbol{\beta}}_1(\tau) \mathbf{w}_i^\top (\mathbf{w}^\top \mathbf{w})^{-1} \mathbf{w}^\top \hat{\mathbf{v}} \hat{\mathbf{x}}_i \hat{\mathbf{x}}_i^\top \right\},\end{aligned}$$

with  $\hat{\mathbf{v}}$  being the residual from the first-step estimation.

**Example 1.2** (Quantile continued). In this example, for a given quantile  $\tau'$  one applies the usual QR procedure to estimate  $\boldsymbol{\theta}(\tau')$  and obtains  $\hat{\mathbf{x}}_1$ . Thus, the first-step estimation is given by the following QR

$$\hat{\boldsymbol{\theta}}(\tau') = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \sum_{i=1}^n \rho_\tau(\tilde{x}_{i1} - \mathbf{w}_i^\top \boldsymbol{\theta}).$$

So one obtains the standard error  $\hat{V}$  and computes  $\hat{\mathbf{x}}_{1i} = \mathbf{w}_i^\top \hat{\boldsymbol{\theta}}(\tau')$  to obtain  $\hat{\mathbf{x}}_i = (\hat{x}_{i1}, x_{i2}, \dots, x_{ik})^\top$ . The final formula for the estimator of the variance covariance matrix is analogous to the previous example.

#### 4.2. Testing

In the independent and identically distributed setting, the conditional quantile functions of the response variable, given the covariates, are all parallel, implying that covariate effects shift the location of the response distribution but do not change the scale or shape. However, slope estimates often vary across quantiles, implying that it is important to test for equality of slopes across quantiles. Wald tests designed for this purpose were suggested by [Koenker and Bassett \(1982a\)](#); [Koenker and Bassett \(1982b\)](#); [Koenker and Machado \(1999\)](#). It is possible to formulate a wide variety of tests using variants of the proposed Wald test, from simple tests on a single quantile regression coefficient to joint tests involving many covariates and distinct quantiles at the same time.

General hypotheses on the vector  $\boldsymbol{\beta}(\tau)$  can be accommodated by Wald-type tests. The Wald process and associated limiting theory provide a natural foundation for the hypothesis

$$H_0 : R\boldsymbol{\beta}(\tau) = r,$$

where  $R$  is a full-rank matrix imposing  $s$  number of restrictions on the parameters and  $r$  is a column vector of  $s$  elements. We consider a Wald-type test where we test the coefficients for selected quantiles of interest. For simplicity, we use the model stated in Equation (2) with a single variable in the  $\mathbf{x}$  matrix. The following example is a hypotheses that may be considered in the former framework.

**Example 5** (Test for slope). A hypothesis testing for  $\boldsymbol{\beta}(\tau)$  for given quantile  $\tau$  can be accommodated in the model. For instance,  $H_0 : \beta_1(\tau) = \beta_1$ , so  $R = [1 \ 0 \ \dots \ 0]$  and  $r = \beta_1$ .

In general, for given  $\tau$ , the regression Wald process can be constructed as

$$\mathcal{W}_n = n(R\hat{\boldsymbol{\beta}}(\tau) - r)^\top [R\hat{\Omega}(\tau)R^\top]^{-1} (R\hat{\boldsymbol{\beta}}(\tau) - r), \quad (12)$$

where  $\hat{\Omega}(\tau) = \tau(1 - \tau)\hat{D}_1^{-1}\hat{D}_0\hat{D}_1^{-1} + \hat{D}_1^{-1}\hat{D}_{12}\hat{V}\hat{D}_{12}^\top\hat{D}_1^{-1} - 2\hat{D}_1^{-1}\hat{M}\hat{D}_1^{-1}$ .

In order to implement the test it is necessary to estimate  $\Omega(\tau)$  consistently. It is possible to obtain such an estimator as suggested in Theorem 3 in the previous section, and the main components of  $\hat{\Omega}(\tau)$  can be obtained as in Equations (8)–(10).

Given the results on consistency of  $\hat{\Omega}(\tau)$ , if we are interested in testing  $R\beta(\tau) = r$  at a particular quantile  $\tau = \tau_0$ , a Chi-square test can be conducted based on the statistic  $\mathcal{W}_n(\tau_0)$ . Under  $H_0$ , the statistic  $\mathcal{W}_n$  is asymptotically  $\chi_s^2$  with  $s$ -degrees of freedom, where  $s$  is the rank of the matrix  $R$ . The limiting distribution of the test is summarized in the following theorem.

**Theorem 4** (Wald Test Inference). *Under  $H_0 : R\beta(\tau) = r$ , and conditions A1–A5 and B1–B5, for fixed  $\tau$ ,*

$$\mathcal{W}_n(\tau) \stackrel{a}{\sim} \chi_s^2.$$

## 5. Monte Carlo Simulations

In this section, we evaluate the performance of two-step QR-GR estimator and compare its performance with the usual QR estimator which does not account for the first-step estimation. Additionally, the performance of the proposed variance-covariance estimator is evaluated. The computational results are obtained in the R language.

### 5.1. Monte Carlo Design

We consider the following model as a data generating process:

$$y_i = \beta_0 + \beta_1 x_i^* + (1 + \gamma x_i^*) \epsilon_i,$$

where  $\epsilon \sim N(0, 1)$ , and  $\beta_0$  and  $\beta_1$  are the parameters of interest. We set  $(\beta_0, \beta_1) = (4, 3)$ . The parameter  $\gamma$  captures the heterogeneity, hence we let  $\gamma = \{0, 1\}$ . When  $\gamma = 0$  we have a location-shift model and for  $\gamma = 1$  the location-scale-shift. Thus, for the later model, we have that  $\beta_1(\tau) = \beta_1 + \gamma F_\epsilon^{-1}(\tau)$ .

The regressor  $x^*$  is unobserved but its observable counterpart  $x$  is related to observed variables  $w$  and  $z$  as follows:

$$x_i = x_i^* + v_i = \theta_0 + \theta_1 w_i + \theta_2 z_i + v_i,$$

where  $w, z, v$  are generated as the following:  $z \sim t_5$ ,  $w \sim N(10, 5)$ ,  $v \sim N(0, 25)$ . The parameter vectors are specified as following:  $(\theta_0, \theta_1, \theta_2) = (1, 3, 2)$ . In the simulations, we consider cases where sample size  $n = \{100, 1000\}$ , quantiles  $\tau = \{0.1, 0.3, 0.5, 0.7, 0.9\}$  and we set the number of replications to be 1000.

For comparison, we consider two estimators of  $\beta_1$ : (i) the standard (infeasible) QR using the unobserved regressors  $x^*$ , which we label QR; (ii) the QR estimates with the GR as described above, which is defined as QR-GR. For the two-step QR-GR estimator, the estimation process is as following. In step 1, using the OLS estimation, we obtain the generated regressor (GR) from the model:  $\hat{x} = \hat{\theta}_0 + \hat{\theta}_1 w + \hat{\theta}_2 z$ , where  $(x, w, z)$  are observables. In step 2, for each  $\tau = (0.1, 0.3, 0.5, 0.7, 0.9)$ , we estimate  $\beta_1$  using the QR-GR estimator of  $y$  on  $\hat{x}$ . We also present results for the corresponding standard errors (SE) of the estimators. For the QR-GR estimator we use the estimator in equation  $c_n = k \cdot (\Phi^{-1}(\tau + h_n) - \Phi^{-1}(\tau - h_n))$  where  $k = \min\{SE(\hat{u}), (\text{quantile}(\hat{u}, 0.75) - \text{quantile}(\hat{u}, 0.25))/1.34\}$ ,  $\hat{u}$  is the residual of the QR estimation, and  $h_n$  is the default bandwidth in the R package.  $k$  is a robust estimate of the scale and the bandwidth  $c_n$  is also commonly chosen in the R package.

### 5.2. Simulation Results

#### 5.2.1. Location Shift Model

The results for the location-shift model are provided in Tables 1 and 2. The bias, SE, and root mean squared error (RMSE) of both QR-GR and QR estimators are presented in Table 1. Different sample sizes  $n = 100$  and  $n = 1000$  in the experiments are reported in the

table. According to the theorems in Section 3, both estimators QR-GR and QR should be consistent. As shown in Table 1, both estimators had empirical bias very close to zero, even for small samples.

Table 1 also shows that the standard error decreased as the sample size increases. However, the QR-GR had substantially larger standard error and RMSE than the regular QR with the true regressor. This result was expected from the theoretical results. These observations reflected the fact that having a GR estimation in the first step did not affect the bias performance but induced a substantially larger variance. This confirms that the sampling error of obtaining by the GR contaminates the standard error in the second-stage QR estimation.

In Table 2 we evaluate the performance of proposed variance-covariance estimator discussed in Section 4 for  $n = 1000$ . We report three statistics in Table 2. First, in column 1, we report the sample standard deviation of the QR-GR estimates based on the Monte Carlo repetitions, which approximated the true standard error of the parameter  $\beta_1$  of interest. Second, the average of the proposed standard error of the QR-GR estimator is reported in the column 2. Finally, for comparison, the standard error of the usual (infeasible) QR is given in the column 3.

By comparing columns 1 and 2 of Table 2, we can see that the estimated standard errors (SE) of the proposed QR-GR was very closely to the true value given in column 1. However, the average of the estimated standard error calculated in the usual way was severely biased downwards. This result confirmed the theoretical predictions and reflected that the sampling error from the first-step estimation induced a larger variance-covariance matrix in the second step. Thus, the estimated standard errors from the conventional formula without considering the GR problem underestimated the population counterpart, and in turn severely affected the inference procedure.

**Table 1.** Simulation results: Bias, SE, and root mean squared error (RMSE) for  $n = 100, 1000$ , population value of  $\beta_1 = 3$ .

		$n = 100$			$n = 1000$		
		Bias	SE	RMSE	Bias	SE	RMSE
$\tau = 0.1$	QR	0.000	0.012	0.012	0.000	0.004	0.004
	QR-GR	−0.002	0.107	0.107	−0.002	0.032	0.032
$\tau = 0.3$	QR	0.000	0.009	0.009	0.000	0.003	0.003
	QR-GR	0.003	0.101	0.101	0.000	0.032	0.032
$\tau = 0.5$	QR	0.000	0.008	0.008	0.000	0.003	0.003
	QR-GR	−0.002	0.102	0.102	0.001	0.031	0.031
$\tau = 0.7$	QR	0.000	0.009	0.009	0.000	0.003	0.003
	QR-GR	0.003	0.099	0.099	−0.002	0.032	0.032
$\tau = 0.9$	QR	0.000	0.011	0.011	0.000	0.004	0.004
	QR-GR	0.005	0.102	0.102	0.000	0.032	0.032

**Table 2.** Simulation results for variance and covariance matrix.  $n = 1000$ .

	Empirical SE	Proposed QR-GR SE	Naive QR SE
$\tau = 0.1$	0.032	0.032	0.004
$\tau = 0.3$	0.032	0.032	0.003
$\tau = 0.5$	0.031	0.032	0.003
$\tau = 0.7$	0.032	0.032	0.003
$\tau = 0.9$	0.032	0.032	0.004

### 5.2.2. Location-Scale Shift Model

The results for the bias and RMSE for sample sizes  $n = 100$  and  $n = 1000$  are reported in Table 3. For the location-scale shift model, both QR and QR-GR estimators had small bias

so that they were close to the population value of  $3 + F_{\epsilon}^{-1}(\tau)$ . However, as expected, the QR-GR had larger standard error and root mean square error (RMSE) than the regular QR with the true regressor for both sample sizes. Thus, the results corroborated the theoretical findings that the GR variable had no asymptotic effects in the bias performance but induced a larger variance.

As in the previous case, in Table 4 we assess the performance of the proposed variance-covariance estimator discussed in Section 4 for the location-scale model. The results were analogous. We see that the proposed QR-GR standard errors in column 2 closely approximated the true standard error in column 1. However, the estimated standard error calculated in the usual way in column 3 was severely smaller and did not approximate the true standard error well.

**Table 3.** Simulation results: Bias, SE, RMSE for  $n = 100, 1000$ , population value of  $\beta_1 = 3 + F_{\epsilon}^{-1}(\tau)$ .

		$n = 100$			$n = 1000$		
		Bias	SE	RMSE	Bias	SE	RMSE
$\tau = 0.1$	QR	0.053	0.286	0.291	0.138	0.092	0.166
	QR-GR	0.087	0.447	0.455	0.140	0.140	0.197
$\tau = 0.3$	QR	0.014	0.207	0.207	0.028	0.061	0.067
	QR-GR	0.048	0.510	0.512	0.037	0.148	0.152
$\tau = 0.5$	QR	0.001	0.183	0.183	−0.002	0.055	0.055
	QR-GR	0.005	0.573	0.573	0.006	0.165	0.165
$\tau = 0.7$	QR	−0.015	0.205	0.205	−0.028	0.064	0.070
	QR-GR	−0.009	0.635	0.635	−0.047	0.200	0.206
$\tau = 0.9$	QR	−0.048	0.270	0.274	−0.144	0.094	0.172
	QR-GR	−0.109	0.803	0.811	−0.158	0.237	0.285

**Table 4.** Simulation results for variance and covariance matrix.

	Empirical SE	Proposed QR-GR SE	Naive QR SE
$\tau = 0.1$	0.140	0.136	0.092
$\tau = 0.3$	0.148	0.162	0.088
$\tau = 0.5$	0.165	0.185	0.089
$\tau = 0.7$	0.200	0.207	0.088
$\tau = 0.9$	0.237	0.241	0.093

## 6. Application

### 6.1. A Brief Literature Review on Engel Curves

Engel curves describe how household expenditures on particular goods and services depend on household income. The analysis of Engel curves has a long history of estimating the expenditure-income relationship (see Engel (1857); Working (1943)). They are regression functions where the dependent variable is the level or the budget share of total expenses used to purchase a commodity of goods or services, and the explanatory variable, total expenditure, is usually used as a proxy for income. A very robust empirical result referred to as ‘Engel’s law’ states that the poorer a family is, the larger the budget share it spends on food. Other categories of expenditure present a less robust pattern.

Many researchers explored different functional forms for Engel curves which better fit the data. For example, Lewbel (1997) proposed a functional form for Engel curves that contains a linear function of logarithm of total expenditure and some nonlinear function of total expenditure. Nonparametric estimations also have been incorporated in the estimation of Engel curves like Blundell et al. (2007). Methods for comparing different regression functions are discussed in Lewbel (2008). These various shapes of Engel curves suggest a deeper understanding of underlying motives which drive household expenditure decisions. However, it is problematic to assume that only one motive drives the consumption of one particular category of goods or services (Chai and Moneta (2010)). For example, luxury,

as a relative concept, is possible in all sorts of consumption. Barigozzi and Moneta (2016) studied a system of budget share and extracted multiple factors that span the same space of basic Engel curves. To understand how the patterns of consumption may be driven by a mixture of different motives, we use the quantile regression with generated regressor (QR-GR) framework laid out in Section 2, in order to explore the heterogeneous effects of motives over different commodity.

To be specific, we estimate unobserved factors which represent underlying motives of expenditure in the first step using the factor model in Barigozzi and Moneta (2016). We then apply quantile regression (QR) methods to the Engel curve model in the second step where budget share is regressed on the estimated factors. The proposed QR-GR model is used to study how each type of household expenditure is driven by different underlying motives. The QR-GR model has two advantages. First, an important difficulty with the model is that the estimator of the variance-covariance matrix needs to take into account the fact that unobserved factors are estimated. The proposed method is able to provide statistically reliable inference via correct estimation of the variance-covariance matrix. Second, we accommodate possible heterogeneity in the effects of total expenditure over the conditional distribution of budget shares. We find that the estimated Engel curves of budget shares are driven by a mixture of three underlying forces which are motives for a household to consume necessities, luxuries, and goods or services on which is spent the same percentage of the total budget. Additionally, we find heterogeneity in each motive along the conditional distribution of budget shares.

## 6.2. Data Description

The data we used in this paper is the household data from the UK Family Expenditure Survey 1997–2001 and the Expenditure and Food Survey 2002–2006.<sup>5</sup> The data contain the information about household expenditures on different goods and services. About 7000 households were randomly selected and each household's expenditures were recorded for 2 weeks, which enables researchers to explore various household consumption patterns. We used information about the number of family members, total expenditures, and expenditures on 13 aggregated categories: (1) housing (net); (2) fuel, light, and power; (3) food; (4) alcoholic drinks; (5) tobacco; (6) clothing and footwear; (7) household goods; (8) household services; (9) personal goods and services; (10) motoring; (11) fares and other travel; (12) leisure goods; (13) leisure services.<sup>6</sup>

In this paper, we studied a sample of about 4000 households which had two to four family members, and the budget shares of these categories were pooled over 10 years. Table 5 reports some descriptive statistics for total expenditure and 13 categories of expenditures. As shown in the table, on average, about 20% of the budget was spent on food and housing, followed by leisure (about 16%) which included leisure goods and leisure services in our analysis. In order to analyze the deflated data, we picked 2005 as the base year, and use the aggregate price index and the price indices for different categories of expenditure from price indices data (RPI).<sup>7</sup>



**Table 5.** Descriptive statistics.

Sample Size $n = 43,702$	Min	Max	Mean	Std. Dev.
Total expenditure	9.97	1587.95	442.92	253.96
Housing net	−0.18	0.97	0.18	0.12
Fuel light power	−0.15	0.79	0.04	0.04
Food	0.00	0.88	0.19	0.09
Alcoholic drink	0.00	0.53	0.04	0.05
Tobacco	0.00	0.81	0.02	0.05
Clothing and footwear	0.00	0.63	0.05	0.06
Household goods	0.00	0.84	0.07	0.08
Household services	0.00	0.89	0.06	0.05
Personal goods and services	0.00	0.82	0.04	0.04
Motoring	−1.90	0.88	0.13	0.12
Fares and other travel	0.00	0.76	0.02	0.05
Leisure goods	0.00	0.85	0.04	0.05
Leisure services	0.00	1.15	0.12	0.12

### 6.3. Empirical Analysis

Barigozzi and Moneta (2016) study a system of budget shares that are driven by latent factors. Using the factor analysis in Bai and Ng (2002), they determine the number of basic Engel curves (i.e., the rank of the system) and found that budget shares of each expenditure can be approximated by a three-factor model: (i) a decreasing function (necessities), (ii) an increasing function (luxuries), (iii) a constant function over the total expenditure (unitary elasticity goods).

We estimate the following conditional quantile function:

$$Q_{\tau}(y_i|x_{1i}, x_{2i}) = \beta_0(\tau) + \beta_1(\tau)x_{1i} + \beta_2(\tau)x_{2i}, \quad (13)$$

where the quantity  $y_i$  denotes the household budget share of a commodity, and  $x_{1i}$  and  $x_{2i}$  are the first and second motives of total expenditure, respectively. Since the third motive is constant over the total expenditure, a constant is also included.

In order to estimate the quantile model in (13), we first obtain the motives  $x_1$  and  $x_2$  using a regression of factors on total expenditure. We compute these generated regressors by following the functional specifications of each motives established in Barigozzi and Moneta (2016). Denote factors by  $f_1$ ,  $f_2$ , and  $f_3 \equiv 1$ , and denote the total expenditure by  $z$ . Let  $w_1 = (1, \log(z))^{\top}$  and  $w_2 = (1, z \log(z))^{\top}$ . We use the following estimation steps:

**Step 1** Obtain the fitted values (generated regressors) of the first two motives by regressing the corresponding factors on functions of the total expenditure as following:

$$\begin{aligned} \hat{x}_{1i} &= g_1(w_{1i}, \hat{\theta}_1) \equiv w_{1i}^{\top} \hat{\theta}_1 \quad \text{where} \quad \hat{\theta}_1 \equiv \left( \sum_{i=1}^n w_{1i} w_{1i}^{\top} \right)^{-1} \sum_{i=1}^n w_{1i} f_{1i} < 0, \\ \hat{x}_{2i} &= g_2(w_{2i}, \hat{\theta}_2) \equiv w_{2i}^{\top} \hat{\theta}_2 \quad \text{where} \quad \hat{\theta}_2 \equiv \left( \sum_{i=1}^n w_{2i} w_{2i}^{\top} \right)^{-1} \sum_{i=1}^n w_{2i} f_{2i} > 0. \end{aligned}$$

**Step 2** Run a quantile regression of budget shares  $y$  on the three motives where the third motive is associated with a constant:

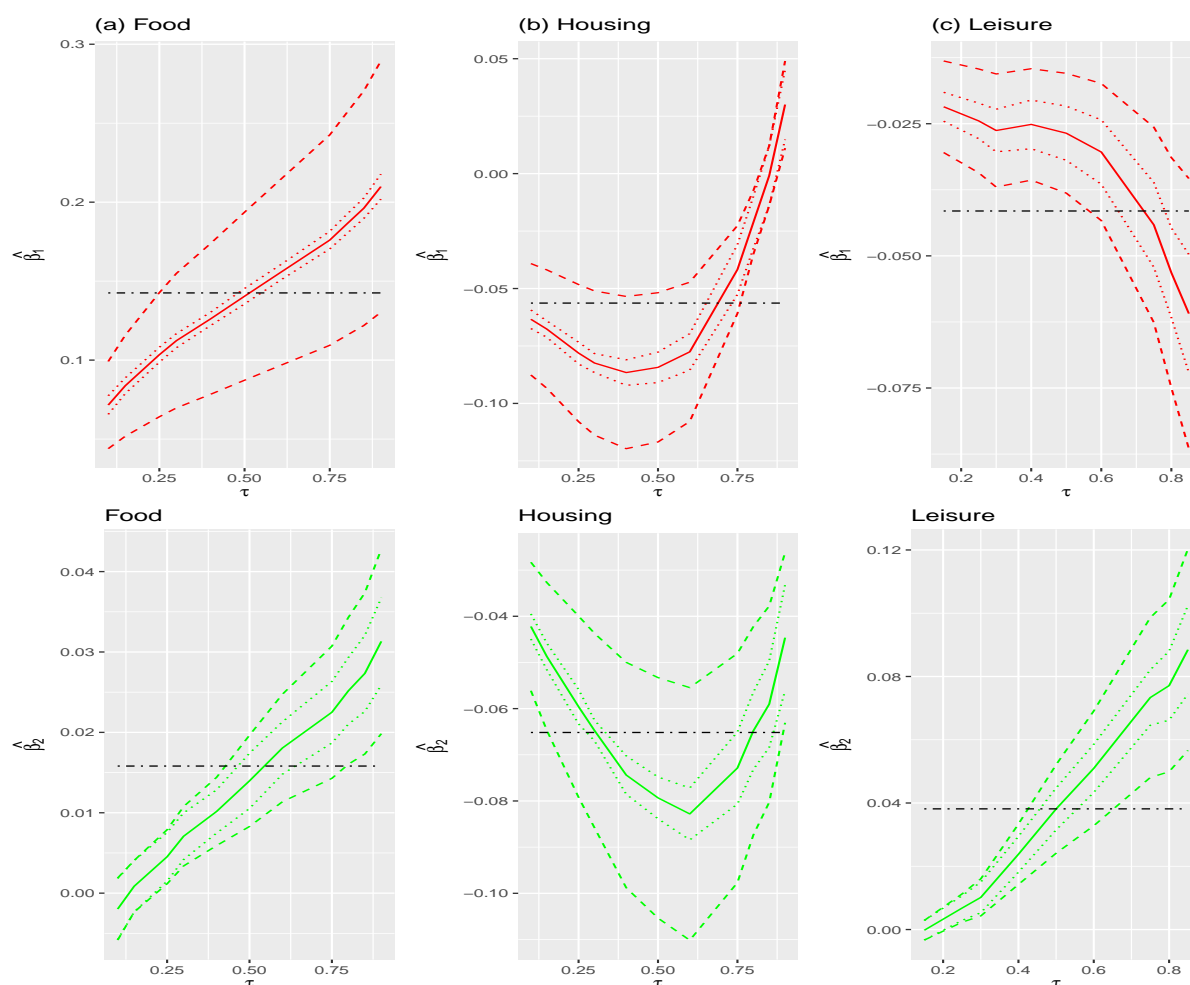
$$\hat{\beta}(\tau) = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^n \rho_{\tau}(y_i - \beta_0 - \beta_1 \hat{x}_1 - \beta_2 \hat{x}_2),$$

with  $\beta = (\beta_0, \beta_1, \beta_2)$ .

**Remark 6.** We note that, in step 1, the motives  $\hat{x}_1$  and  $\hat{x}_2$  are estimated by regressing the factors on the total expenditures where the factors are obtained by the principal component analysis. For the sake of notational simplicity, we assume the factors and total expenditures are observed without errors.

**Remark 7.** In step 2, each motive represents specific reactions to consumption changes. The first one  $x_1$  is interpreted as a motive to consume necessities since it is decreasing in the total expenditure, while the second one  $x_2$  represents the motive to consume luxuries, which is increasing in the total expenditure. The third one  $x_3 \equiv 1$  is related to unitary elasticity goods which represents goods or services allocated the same percentage of total budget by both rich and poor households.

We estimate the coefficients in model (13) across different quantiles. For conciseness, we only report the results for food, housing and leisure. The main results in Figure 1 show how the budget shares of each commodity relate to each factors, respectively. We only show the results for  $x_1$  and  $x_2$  since the third motive  $x_3$  does not vary with the family expenditure. The coefficient estimates over different quantiles reflect marginal impact of each motive on the distribution of budget shares at different quantiles. In all cases, their magnitudes varies over the level of the quantile  $\tau$ . Thus our quantile model well identifies apparent heterogeneous effects of consumption motives on the the budget shares.



**Figure 1.** Coefficient estimates of motives 1 and 2 and their 95% confidence intervals. (a) food budget share; (b) housing budget share; (c) leisure budget share. Upper row: coefficient estimates of factor 1; lower row: coefficient estimates of factor 2. Black line: OLS coefficient estimate; solid line: quantile regression (QR) coefficient estimate; dotted line: 95% confidence intervals obtained by conventional estimation; dashed line: 95% confidence intervals obtained by our proposed estimation.

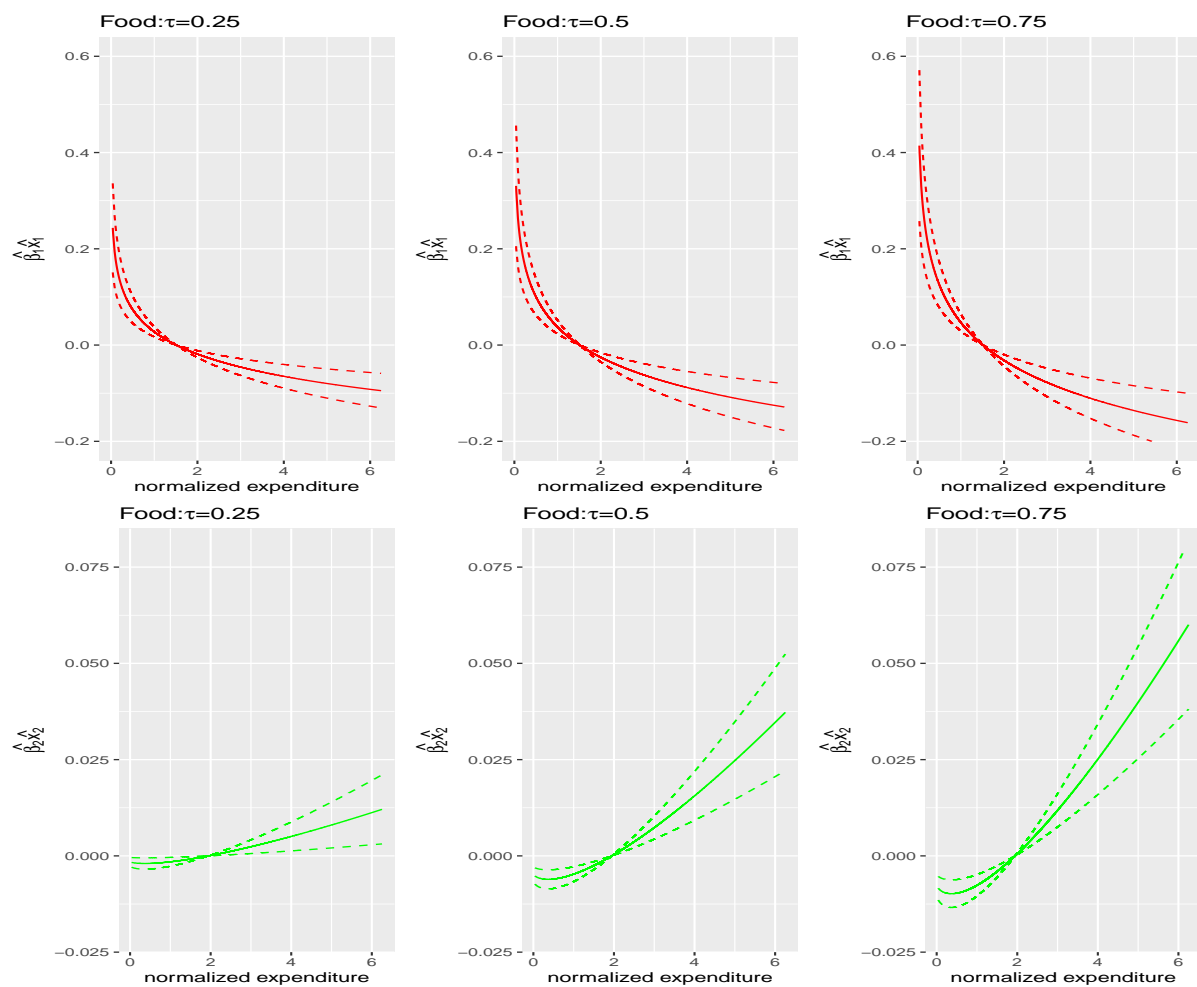
The 95% confidence intervals in dashed lines are based on our proposed estimator of the asymptotic variance-covariance matrix, while the 95% confidence intervals in dotted lines are calculated by conventional QR variance-covariance estimation. Clearly, the adjusted confidence bands for QR-GR are much larger than the conventional bands which do not adjust for the GR issue. This can be observed more clearly in Table 6 which summarizes the estimated standard errors for a subset of quantiles  $\tau \in \{0.25, 0.5, 0.75\}$ . The table reports that the estimated standard errors from the proposed estimator (QR-GR SE) is always larger than those from the conventional estimator (QR SE). These results underscore the importance of estimating standard errors correctly by taking into account the GR issue in Engel curves.

**Table 6.** Coefficient estimates of motives 1 and 2 and their standard errors at quantiles  $\tau = 0.25, 0.5, 0.75$  for (a) food, (b) housing and (c) leisure. Columns 1 and 4: coefficient estimates; Columns 2 and 5: standard error calculated in adjusted QR estimation; column 3 and 6: standard error calculated in conventional QR estimation.

(a) Food						
	$\hat{\beta}_1$	QR-GR SE	QR SE	$\hat{\beta}_2$	QR-GR SE	QR SE
$\tau = 0.25$	0.103	0.020	0.002	0.005	0.002	0.001
$\tau = 0.5$	0.140	0.027	0.003	0.014	0.003	0.002
$\tau = 0.75$	0.176	0.034	0.003	0.023	0.004	0.002
(b) Housing						
	$\hat{\beta}_1$	QR-GR SE	QR SE	$\hat{\beta}_2$	QR-GR SE	QR SE
$\tau = 0.25$	−0.078	0.015	0.002	−0.060	0.010	0.002
$\tau = 0.5$	−0.084	0.017	0.003	−0.079	0.013	0.002
$\tau = 0.75$	−0.042	0.010	0.006	−0.073	0.013	0.004
(c) Leisure						
	$\hat{\beta}_1$	QR-GR SE	QR SE	$\hat{\beta}_2$	QR-GR SE	QR SE
$\tau = 0.25$	−0.024	0.005	0.002	0.007	0.002	0.002
$\tau = 0.5$	−0.027	0.006	0.003	0.038	0.007	0.003
$\tau = 0.75$	−0.044	0.009	0.004	0.073	0.013	0.005

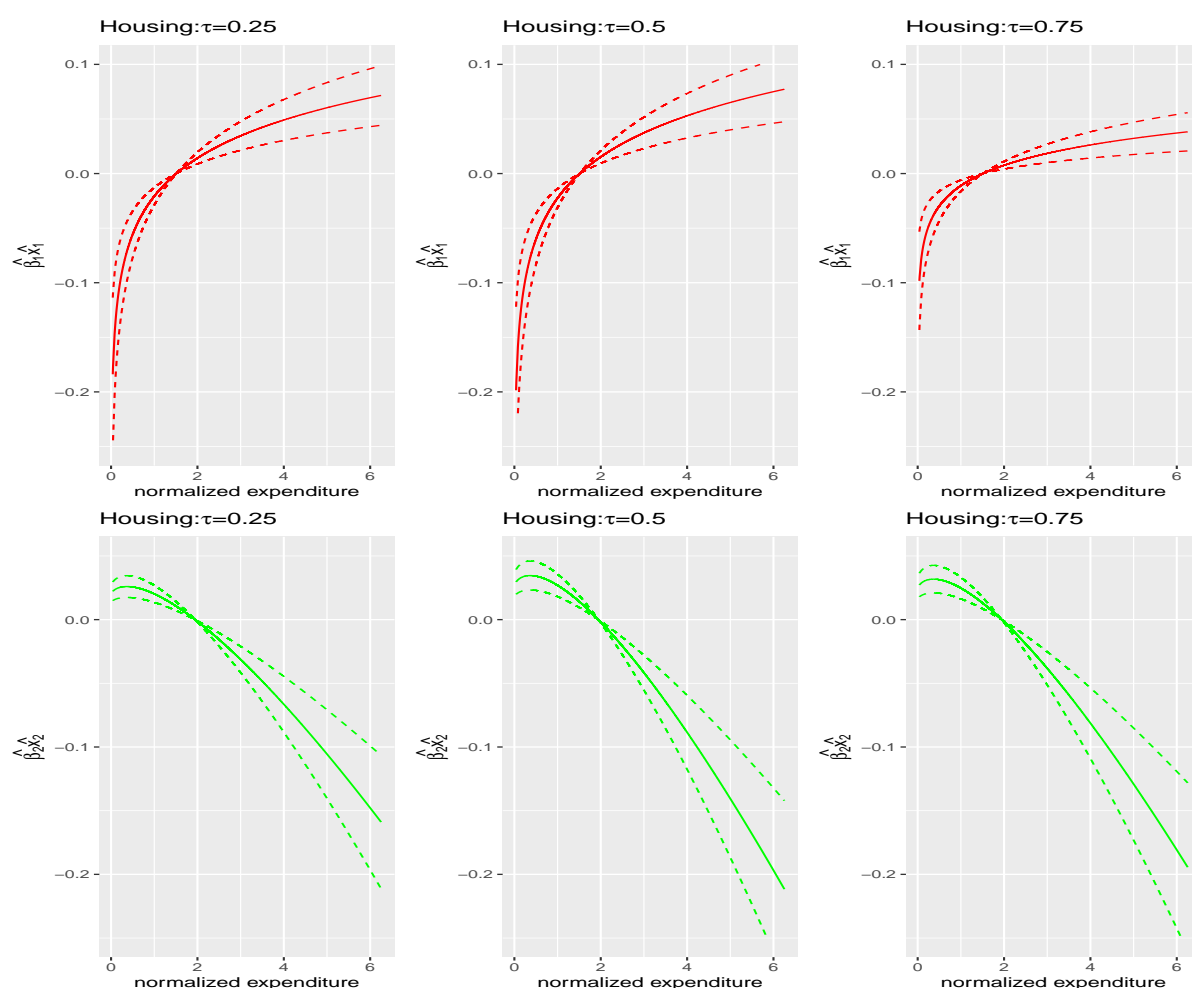
Since both  $x_1$  and  $x_2$  are normalized measures of different motives, comparing the relative magnitudes of their coefficient estimates helps us to tell which motive plays a leading role in a particular commodity. For example, Figure 1a shows that the motive to consume food as necessity clearly plays a leading role relative to luxuries at all quantiles. In Figure 1c, the motive to consume leisure as luxuries is dominating a role as necessity. However, consuming housing in Figure 1b has a mixed result where both necessity and luxuries play a similar role.

To help further interpret how the patterns of budget shares are related to these unobservable motives, we show how the budget share of a particular commodity varies over the total expenditure in Figures 2–4. We estimate the contribution of each motive to the budget share at different quantiles  $\tau \in \{0.25, 0.5, 0.75\}$ . It is interesting to see how the motive to consume a particular commodity as necessity or luxury contributes to its budget share at different levels of total expenditure.



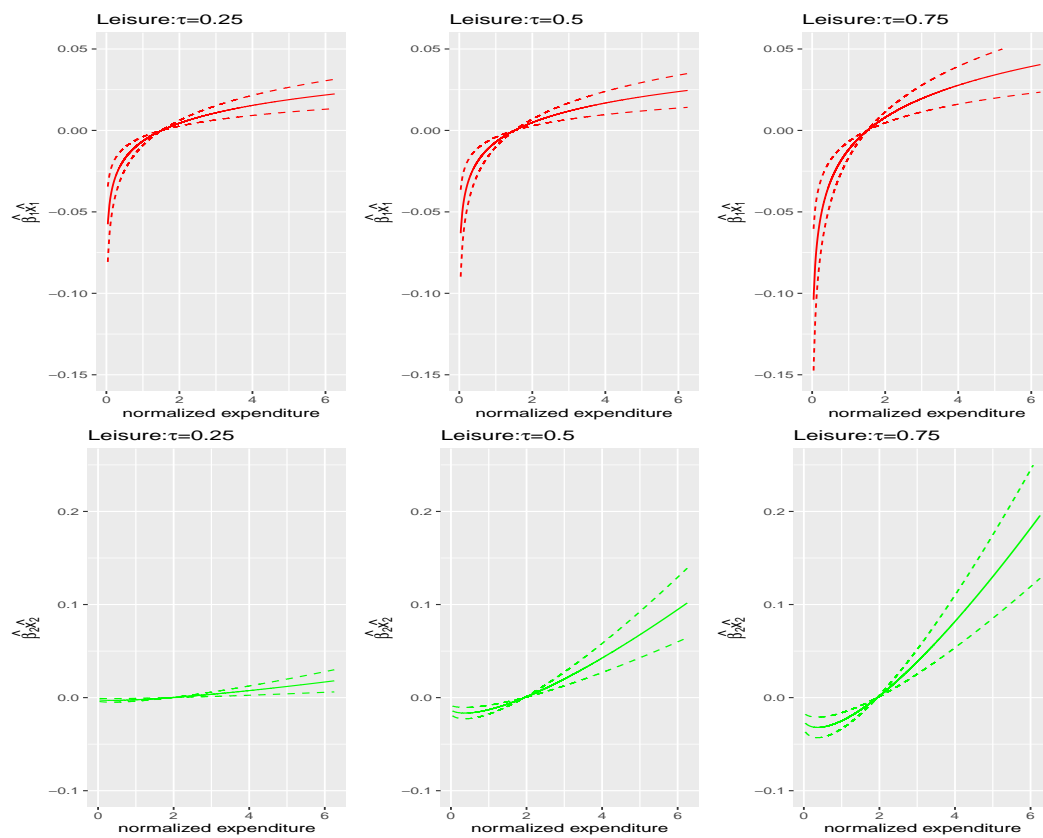
**Figure 2.** Contributions of motive 1 (upper row) and motive 2 (lower row) to food budget share; dotted lines are the corresponding 95% confidence intervals.

For instance, Figure 2 shows that, as total expenditure increases, the contribution of the first motive,  $x_1$  (necessity), to food budget share decreases at all  $\tau$ , while the second motive,  $x_2$  (luxuries), increases, as well documented in the literature. This implies that as total expenditure increases, the motive to consume food as a necessary good decreases, while the motive to consume food as a luxury good increases. The pattern in Figure 3, which is very different from food, reflects that individuals consume housing more as necessity but less as luxury, as their total expenditure increases. The motives to consume food and housing as necessity or luxuries play opposite roles in the budget share as total expenditure changes. However, as shown in Figure 4, both motives to consume leisure increase as total expenditure increases as well. Another interesting point is that the magnitudes of those estimated factors or motives are different across different quantiles of the conditional distribution of budget shares in all three commodities.

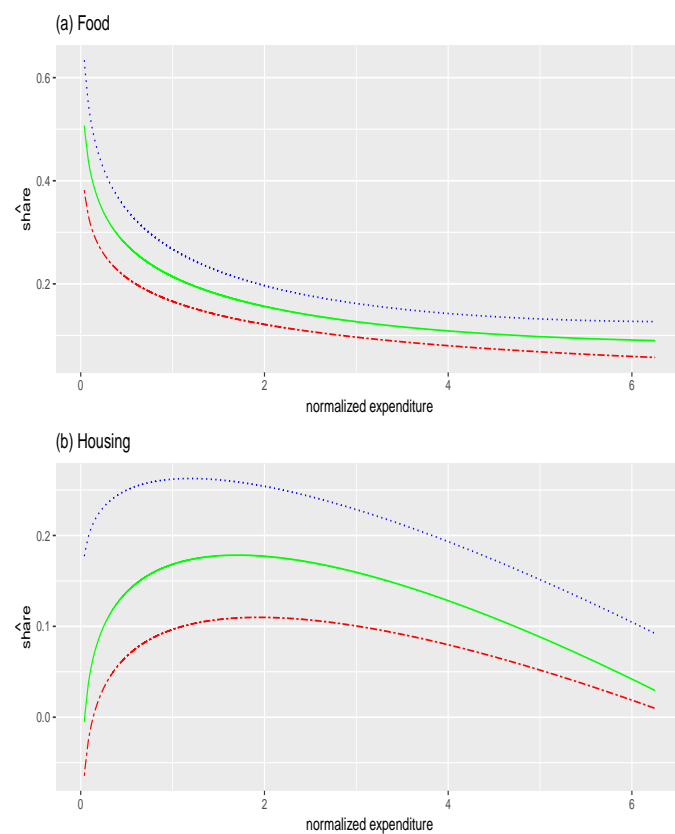


**Figure 3.** Contributions of motive 1 (upper row) and motive 2 (lower row) to housing budget share; dotted lines are the corresponding 95% confidence intervals.

Figures 2–4 suggest that the motive of consumption as necessity or luxury plays different roles over the level of total expenditure. The overall effects of total expenditure on budget shares are presented in Figure 5. The estimated budget shares are calculated by adding all three factors together. Figure 5a shows that the budget share of food decreases as total expenditure increases, which is classically referred to as Engel’s law. In Figure 5b we see a concave shape curve, that is, the budget share of housing increases for low expenditures, but decreases as total expenditure increases. In addition, individuals always raise the budget share of leisure as they consume more, as shown in Figure 5c. Finally, in housing and leisure, the shapes of the curves vary over different quantiles  $\tau$ . There is evidence of substantial asymmetry for the expenditure on housing and leisure, while there is less evidence of asymmetry for food share. For instance, we observe a shift toward leisure at the upper quantile of the leisure share. Thus, the conditional quantile approach would be a useful tool to capture asymmetric patterns of Engel curves across quantiles of budget share.

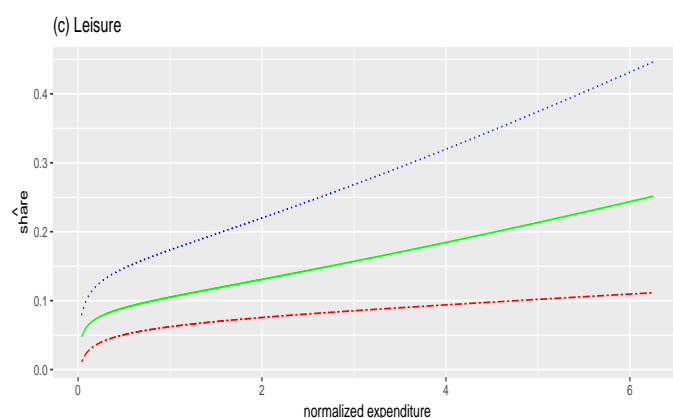


**Figure 4.** Contributions of motive 1 (upper row) and motive 2 (lower row) to leisure budget share; dotted lines are the corresponding 95% confidence intervals.



**Figure 5.** Cont.





**Figure 5.** Fitted values of budget share where  $\widehat{share} = \hat{\beta}_0(\tau) + \hat{\beta}_1(\tau)\hat{x}_1 + \hat{\beta}_2(\tau)\hat{x}_2$ . (a) food; (b) housing; (c) leisure. Red two dashed line: corresponding coefficients estimated at  $\tau = 0.25$ ; green solid line: corresponding coefficients estimated at  $\tau = 0.5$ ; blue dotted line: corresponding coefficients estimated at  $\tau = 0.75$ .

## 7. Conclusions

We study estimation and inference for linear quantile regression (QR) models with generated regressors (GR). We propose a QR-GR two-step estimator for the parameters of interest and an estimator for the corresponding asymptotic variance-covariance matrix. We establish the asymptotic properties of the estimators. Monte Carlo simulation and estimation of the Engel curves using data from the UK Family Expenditure Survey confirm that taking into account the GR problem in the QR framework is essential for correct inference. Furthermore, the empirical application shows strong heterogeneity of the Engel curves over different quantiles of the conditional distribution of budget shares in most commodities.

**Author Contributions:** The authors equally contributed to the conceptualization, methodology, formal analysis, and writing. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

### Appendix A.1. Proof of the Main Results

**Proof of Theorem 1 (Consistency).** Let  $\hat{\beta}(\tau)$  be the two-step quantile regression (QR) with generated regressors (GR) estimator, and  $\check{\beta}(\tau)$  be the usual (infeasible) QR estimator with true unobservable variables. Consider

$$\hat{\beta}(\tau) - \beta_0(\tau) = (\hat{\beta}(\tau) - \check{\beta}(\tau)) + (\check{\beta}(\tau) - \beta_0(\tau)). \quad (A1)$$

Under the standard QR conditions A1 and A2, we have that the second term in Equation (A1) satisfies that

$$\check{\beta}(\tau) - \beta_0(\tau) \xrightarrow{p} 0, \quad (A2)$$

and the following linear representation

$$\check{\beta}(\tau) - \beta_0(\tau) = D_1^{-1} \frac{1}{n} \sum_{i=1}^n x_i \psi_\tau(y_i - x_i^\top \beta_0(\tau)) + o_p(1) \quad (A3)$$

where  $\psi_\tau(u) := \tau - I(u \leq 0)$ .

Let  $\bar{x}_q = (x_1, \dots, x_q)^\top$  and  $g = (g_1, \dots, g_q)^\top$ . By noticing that  $\hat{\beta} = \hat{\beta}(\hat{x}_1, \dots, \hat{x}_q, x_{q+1}, \dots, x_k)$  and  $\check{\beta} = \check{\beta}(x_1, \dots, x_q, x_{q+1}, \dots, x_k)$ , and by expanding the first term in Equation (A1)  $\hat{\beta}(\tau) - \check{\beta}(\tau)$ , we have

$$\begin{aligned}\hat{\beta}(\tau) - \check{\beta}(\tau) &= \left( \frac{\partial \check{\beta}(\tau)}{\partial x_1}^\top, \dots, \frac{\partial \check{\beta}(\tau)}{\partial x_q}^\top \right) \begin{bmatrix} \hat{x}_1 - x_1 \\ \vdots \\ \hat{x}_q - x_q \end{bmatrix} + o_p(1) \\ &= \left( \frac{\partial \check{\beta}(\tau)}{\partial x_1}^\top, \dots, \frac{\partial \check{\beta}(\tau)}{\partial x_q}^\top \right) \begin{bmatrix} g_1(w_1, \hat{\theta}_1) - g_1(w_1, \theta_1) \\ \vdots \\ g_q(w_q, \hat{\theta}_q) - g_q(w_q, \theta_q) \end{bmatrix} + o_p(1) \\ &= \left( \frac{\partial \check{\beta}(\tau)}{\partial x_1}^\top, \dots, \frac{\partial \check{\beta}(\tau)}{\partial x_q}^\top \right) \begin{bmatrix} \nabla_{\theta_1} g_1(w_1, \theta_1)^\top (\hat{\theta}_1 - \theta_1) \\ \vdots \\ \nabla_{\theta_q} g_q(w_q, \theta_q)^\top (\hat{\theta}_q - \theta_q) \end{bmatrix} + o_p(1) \\ &= \left( \frac{\partial \check{\beta}(\tau)}{\partial \bar{x}_q^\top} \right)^\top \nabla_{\theta} g(w, \theta)^\top (\hat{\theta} - \theta) + o_p(1)\end{aligned}$$

where the third equality follows from a Taylor expansion for  $g_j(\cdot)$  for  $j = 1, \dots, q$  and where  $\frac{\partial \check{\beta}(\tau)}{\partial \bar{x}_q^\top}$  is  $q \times k$  matrix,  $\nabla_{\theta} g(w, \theta)$  is  $\sum_{j=1}^q p_j \times q$  matrix, and  $(\hat{\theta} - \theta)$  is  $\sum_{j=1}^q p_j \times 1$  vector.

As seen in (A3),  $\check{\beta}(\tau)$  can be rewritten as

$$\check{\beta}(\tau) = \left( \frac{1}{n} \sum_{i=1}^n f_i(0|x_i) x_i x_i^\top \right)^{-1} \frac{1}{n} \sum_{i=1}^n x_i \left[ f_i(0|x_i) x_i^\top \beta_0(\tau) + \psi_\tau(y_i - x_i^\top \beta_0(\tau)) \right] + o_p(1).$$

By taking the derivative of  $\check{\beta}(\tau)$  with respect to  $x_j$ , one can notice that the contribution of the  $\psi_\tau(y_i - x_i^\top \beta_0(\tau))$  term is  $o_p(1)$ .

Now we apply Lemma 2 about G-inverse from Ma and Koenker (2006), and obtain that for  $j = 1, \dots, q$

$$\begin{aligned}\left( \frac{\partial \check{\beta}(\tau)}{\partial \bar{x}_q^\top} \right)^\top \nabla_{\theta} g(w, \theta)^\top &= -D_1^{-1} \left( \frac{1}{n} \sum_{i=1}^n f_i(0|x_i) \beta(\tau) \bar{x}_i \nabla_{\theta} g(w, \theta)^\top \right) + o_p(1) \\ &= -D_1^{-1} D_{12} + o_p(1),\end{aligned}$$

where

$$D_{12} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n f_i(0|x_i) \beta(\tau) \bar{x}_i \nabla_{\theta} g(w, \theta)^\top.$$

It follows that

$$\begin{aligned}\hat{\beta}(\tau) - \check{\beta}(\tau) &= \left( \frac{\partial \check{\beta}(\tau)}{\partial \bar{x}_q^\top} \right)^\top \nabla_{\theta} g(w, \theta)^\top (\hat{\theta} - \theta) + o_p(1) \\ &= (-D_1^{-1} D_{12} + o_p(1)) (\hat{\theta} - \theta) + o_p(1) \\ &= [-D_1^{-1} D_{12} (\hat{\theta} - \theta)] + o_p(1).\end{aligned}$$

Note that  $\text{plim}_{n \rightarrow \infty} \hat{\theta} = \theta$  by assumption A3 and  $D_1^{-1} D_{12}$  is bounded by assumptions A2 and A5, we have that  $-D_1^{-1} D_{12} (\hat{\theta} - \theta) = o_p(1)$ . Therefore, as  $n \rightarrow \infty$ , we have

$$\hat{\beta}(\tau) - \check{\beta}(\tau) \xrightarrow{p} 0. \quad (\text{A4})$$

Combining (A2) and (A4), we have

$$\hat{\beta}(\tau) - \beta_0(\tau) = (\hat{\beta}(\tau) - \check{\beta}(\tau)) + (\check{\beta}(\tau) - \hat{\beta}_0(\tau)) \xrightarrow{p} 0.$$

□

**Proof of Theorem 2 (Asymptotic normality).** Recall that  $\hat{\beta}(\tau)$  is the two-step QR-GR estimator and  $\check{\beta}(\tau)$  is the infeasible QR estimator. Consider the following

$$\sqrt{n}(\hat{\beta}(\tau) - \beta_0(\tau)) = \sqrt{n}(\hat{\beta}(\tau) - \check{\beta}(\tau)) + \sqrt{n}(\check{\beta}(\tau) - \beta_0(\tau)). \quad (\text{A5})$$

Under the usual QR conditions A1 and A2, we have that the second term in Equation (A5) has the standard QR expansion as

$$\sqrt{n}(\check{\beta}(\tau) - \beta_0(\tau)) = D_1^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i \psi_\tau(y_i - x_i^\top \beta_0(\tau)) + o_p(1), \quad (\text{A6})$$

and satisfies a Central Limit Theorem such that

$$\sqrt{n}(\check{\beta}(\tau) - \beta_0(\tau)) \xrightarrow{d} N(0, \tau(1-\tau)D_1^{-1}D_0D_1^{-1}), \quad (\text{A7})$$

where  $\psi_\tau(u) := \tau - I(u \leq 0)$ .

By noticing that  $\hat{\beta} = \hat{\beta}(\hat{x}_1, \dots, \hat{x}_q, x_{q+1}, \dots, x_k)$  and  $\check{\beta} = \check{\beta}(x_1, \dots, x_q, x_{q+1}, \dots, x_k)$ , and by expanding the first term in Equation (A5)  $\sqrt{n}(\hat{\beta}(\tau) - \check{\beta}(\tau))$ , we have

$$\begin{aligned} \sqrt{n}(\hat{\beta}(\tau) - \check{\beta}(\tau)) &= \sqrt{n} \left( \frac{\partial \check{\beta}(\tau)}{\partial x_1}^\top, \dots, \frac{\partial \check{\beta}(\tau)}{\partial x_q}^\top \right) \begin{bmatrix} \hat{x}_1 - x_1 \\ \vdots \\ \hat{x}_q - x_q \end{bmatrix} + o_p(1) \\ &= \sqrt{n} \left( \frac{\partial \check{\beta}(\tau)}{\partial x_1}^\top, \dots, \frac{\partial \check{\beta}(\tau)}{\partial x_q}^\top \right) \begin{bmatrix} g_1(w_1, \hat{\theta}_1) - g_1(w_1, \theta_1) \\ \vdots \\ g_q(w_q, \hat{\theta}_q) - g_q(w_q, \theta_q) \end{bmatrix} + o_p(1) \\ &= \sqrt{n} \left( \frac{\partial \check{\beta}(\tau)}{\partial x_1}^\top, \dots, \frac{\partial \check{\beta}(\tau)}{\partial x_q}^\top \right) \begin{bmatrix} \nabla_{\theta_1} g_1(w_1, \theta_1)^\top (\hat{\theta}_1 - \theta_1) \\ \vdots \\ \nabla_{\theta_q} g_q(w_q, \theta_q)^\top (\hat{\theta}_q - \theta_q) \end{bmatrix} + o_p(1), \\ &= \sqrt{n} \left( \frac{\partial \check{\beta}(\tau)}{\partial \bar{x}_q^\top} \right)^\top \nabla_{\theta} g(w, \theta)^\top (\hat{\theta} - \theta) + o_p(1). \end{aligned}$$

As seen in (A6), the Bahadur representation for  $\sqrt{n}\check{\beta}(\tau)$  can be rewritten as

$$\sqrt{n}\check{\beta}(\tau) = \left( \frac{1}{n} \sum_{i=1}^n f_i(0|x_i) x_i x_i^\top \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i \left[ f_i(0|x_i) x_i^\top \beta_0(\tau) + \psi_\tau(y_i - x_i^\top \beta_0(\tau)) \right] + o_p(1).$$

Taking the derivative of  $\sqrt{n}\check{\beta}(\tau)$  with respect to  $x_j$  and noticing that the contribution of the  $\psi_\tau(y_i - x_i^\top \beta_0(\tau))$  term is  $o_p(1)$ , we get

$$\begin{aligned} \left( \frac{\partial \check{\beta}(\tau)}{\partial \bar{x}_q^\top} \right)^\top \nabla_{\theta} g(w, \theta)^\top &= -D_1^{-1} \left( \frac{1}{n} \sum_{i=1}^n f_i(0|x_i) \beta(\tau) \bar{x}_i \nabla_{\theta} g(w, \theta)^\top \right) + o_p(1) \\ &= -D_1^{-1} D_{12} + o_p(1), \end{aligned}$$

where

$$D_{12} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n f_i(0|x_i) \beta(\tau) \bar{x}_i \nabla_{\theta} g(w, \theta)^{\top}. \quad (\text{A8})$$

Thus, we have the following representation

$$\sqrt{n}(\hat{\beta}(\tau) - \check{\beta}(\tau)) = -D_1^{-1} D_{12} \sqrt{n}(\hat{\theta} - \theta) + o_p(1). \quad (\text{A9})$$

Note that since  $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, V)$  by assumption **A4** and  $D_1^{-1} D_{12}$  is bounded by assumptions **A2** and **A5**, we have that  $-D_1^{-1} D_{12} \sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, D_1^{-1} D_{12} V D_{12}^{\top} D_1^{-1})$ . Therefore, we have by assumption **A3** that

$$\sqrt{n}(\hat{\beta}(\tau) - \check{\beta}(\tau)) \xrightarrow{d} N\left(0, D_1^{-1} D_{12} V D_{12}^{\top} D_1^{-1}\right). \quad (\text{A10})$$

Recall from (A6) that  $\sqrt{n}(\check{\beta}(\tau) - \beta_0(\tau)) = D_1^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i \psi_{\tau}(y_i - x_i^{\top} \beta_0(\tau)) + o_p(1)$ . In addition, (A9) together with (A8) implies that

$$\sqrt{n}(\hat{\beta}(\tau) - \check{\beta}(\tau)) = -D_1^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n f_i(0|x_i) \beta(\tau) \bar{x}_i \nabla_{\theta} g(w_i, \theta)^{\top} (\hat{\theta} - \theta) + o_p(1).$$

Substituting **A4** where  $\hat{\theta} - \theta = \frac{1}{n} \sum_{i=1}^n r(\theta)$  in the above equation, we obtain that

$$\begin{aligned} & \text{Cov}(\sqrt{n}(\check{\beta}(\tau) - \beta_0(\tau)), \sqrt{n}(\hat{\beta}(\tau) - \check{\beta}(\tau))) \\ &= -D_1^{-1} E \left\{ \frac{1}{n} \sum_{i=1}^n \psi_{\tau}(y_i - x_i^{\top} \beta_0(\tau)) f_i(0|x_i) x_i r(\theta)^{\top} \nabla_{\theta} g(w_i, \theta)^{\top} \bar{x}_i^{\top} \beta(\tau)^{\top} \right\} D_1^{-1} \\ &= -D_1^{-1} M D_1^{-1} + o_p(1) \end{aligned}$$

where  $M = \text{plim}_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \{\psi_{\tau}(y_i - x_i^{\top} \beta_0(\tau)) f_i(0|x_i) x_i r(\theta)^{\top} \nabla_{\theta} g(w_i, \theta)^{\top} \bar{x}_i^{\top} \beta(\tau)^{\top}\}$ .

Finally, combining the above covariance term with (A7) and (A10) with (A5), we obtain

$$\sqrt{n}(\hat{\beta}(\tau) - \beta_0(\tau)) \xrightarrow{d} N\left(0, \tau(1-\tau) D_1^{-1} D_0 D_1^{-1} + D_1^{-1} D_{12} V D_{12}^{\top} D_1^{-1} - 2 D_1^{-1} M D_1^{-1}\right).$$

□

**Proof of Theorem 3 (Consistency of variance-covariance matrix).** (i) Claim that  $\hat{D}_0 \xrightarrow{p} D_0$ , in other words,

$$n^{-1} \sum_{i=1}^n \hat{x}_i \hat{x}_i^{\top} - n^{-1} \sum_{i=1}^n x_i x_i^{\top} = o_p(1).$$

By definition:

$$\hat{x}_i = (\hat{x}_{i1}, \dots, \hat{x}_{iq}, x_{iq+1}, \dots, x_{ik})^{\top}, x_i = (x_{i1}, \dots, x_{iq}, x_{iq+1}, \dots, x_{ik})^{\top}.$$

We show that  $n^{-1} \sum_{i=1}^n \hat{x}_i \hat{x}_i^{\top}$  and  $n^{-1} \sum_{i=1}^n x_i x_i^{\top}$  are close to each other element by element:

(a) For  $j = 1, \dots, q, j' = 1, \dots, q$ , we have

$$\hat{x}_{ij} = g_j(w_{ji}, \hat{\theta}_j),$$

$$\hat{\theta}_j \xrightarrow{p} \theta_j.$$

Assumption **B3** implies the following uniform convergence by applying the uniform weak law of large number (UWLLN): ( $E(\cdot)$  means expectation in terms of joint distribution)

$$n^{-1} \sum_{i=1}^n x_{ij} x_{ij'} = n^{-1} \sum_{i=1}^n g_j(w_{ji}, \theta_j) g_{j'}(w_{ji'}, \theta_{j'}) \xrightarrow{p} E[g_j(w_j, \theta_j) g_{j'}(w_{j'}, \theta_{j'})],$$

and assumption **A3** that  $\hat{\theta}_j \rightarrow \theta_j$  for  $j = 1, \dots, q$  implies that

$$n^{-1} \sum_{i=1}^n \hat{x}_{ij} \hat{x}_{ij'} = n^{-1} \sum_{i=1}^n g_j(w_{ji}, \hat{\theta}_j) g_{j'}(w_{ji'}, \hat{\theta}_{j'}) \xrightarrow{p} E[g_j(w_j, \theta_j) g_{j'}(w_{j'}, \theta_{j'})].$$

It follows that

$$n^{-1} \sum_{i=1}^n \hat{x}_{ij} \hat{x}_{ij'} - n^{-1} \sum_{i=1}^n x_{ij} x_{ij'} = o_p(1).$$

(b) For  $j = 1, \dots, q$ ,  $s = q + 1, \dots, k$ , similarly by applying UWLLN, we have

$$n^{-1} \sum_{i=1}^n x_{ij} x_{is} = n^{-1} \sum_{i=1}^n g_j(w_{ji}, \theta_j) x_{is} \xrightarrow{p} E[g_j(w_j, \theta_j) x_{is}],$$

and the consistency of  $\theta_j$  implies that

$$n^{-1} \sum_{i=1}^n \hat{x}_{ij} x_{is} = n^{-1} \sum_{i=1}^n g_j(w_{ji}, \hat{\theta}_j) x_{is} \xrightarrow{p} E[g_j(w_j, \theta_j) x_{is}].$$

It follows that

$$n^{-1} \sum_{i=1}^n \hat{x}_{ij} x_{is} - n^{-1} \sum_{i=1}^n x_{ij} x_{is} = o_p(1).$$

After claiming (a) and (b), we have

$$n^{-1} \sum_{i=1}^n \hat{\mathbf{x}}_i \hat{\mathbf{x}}_i^\top - n^{-1} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\top = o_p(1).$$

(ii) Claim that  $\hat{D}_1 \xrightarrow{p} D_1$ .

To obtain a consistent estimator of  $D_1$  which contains unknown conditional density function, the estimator  $\hat{D}_1$  replaces the conditional density functions by uniform kernel weights.

Define

$$\tilde{D}_1 \equiv (2nc_n)^{-1} \sum_{i=1}^n I(|\epsilon_i| < c_n) \mathbf{x}_i \mathbf{x}_i^\top.$$

Denote  $\hat{\epsilon}_i \equiv y_i - \hat{\mathbf{x}}_i^\top \hat{\boldsymbol{\beta}}(\tau)$  and  $\epsilon_i \equiv y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_0(\tau)$ .

(a) Consider

$$\begin{aligned} \|\hat{D}_1 - \tilde{D}_1\| = & \left\| (2nc_n)^{-1} \times \sum_{i=1}^n \left\{ [I(|\hat{\epsilon}_i| < c_n) - I(|\epsilon_i| < c_n)] \hat{\mathbf{x}}_i \hat{\mathbf{x}}_i^\top \right. \right. \\ & \left. \left. + I(|\epsilon_i| < c_n) [\hat{\mathbf{x}}_i - \mathbf{x}_i] \hat{\mathbf{x}}_i^\top + I(|\epsilon_i| < c_n) \mathbf{x}_i [\hat{\mathbf{x}}_i - \mathbf{x}_i]^\top \right\} \right\|. \end{aligned}$$

Notice that

$$|I(|\hat{\epsilon}_i| < c_n) - I(|\epsilon_i| < c_n)| \leq I(|\epsilon_i - c_n| < |\hat{\mathbf{x}}_i^\top \hat{\boldsymbol{\beta}}(\tau) - \mathbf{x}_i^\top \boldsymbol{\beta}_0(\tau)|) \\ + I(|\epsilon_i + c_n| < |\hat{\mathbf{x}}_i^\top \hat{\boldsymbol{\beta}}(\tau) - \mathbf{x}_i^\top \boldsymbol{\beta}_0(\tau)|), \quad (\text{A11})$$

and

$$\|\hat{\mathbf{x}}_i - \mathbf{x}_i\| \leq \max_j \{\|\nabla g_j(\mathbf{w}_{ji}, \boldsymbol{\theta}_j^*)\| \cdot \|\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j\|\}. \quad (\text{A12})$$

Combining (A11) and (A12), we have

$$\|\hat{D}_1 - \tilde{D}_1\| \leq (2nc_n)^{-1} \times \sum_{i=1}^n \left\{ \left[ I(|\epsilon_i - c_n| < |\hat{\mathbf{x}}_i^\top \hat{\boldsymbol{\beta}}(\tau) - \mathbf{x}_i^\top \boldsymbol{\beta}_0(\tau)|) \right. \right. \\ \left. \left. + I(|\epsilon_i + c_n| < |\hat{\mathbf{x}}_i^\top \hat{\boldsymbol{\beta}}(\tau) - \mathbf{x}_i^\top \boldsymbol{\beta}_0(\tau)|) \right] \|\hat{\mathbf{x}}_i\|^2 \right. \\ \left. + I(|\epsilon_i| < c_n) \max_j \{\|\nabla g_j(\mathbf{w}_{ji}, \boldsymbol{\theta}_j^*)\| \cdot \|\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j\|\} \|\hat{\mathbf{x}}_i\| \right. \\ \left. + I(|\epsilon_i| < c_n) \|\hat{\mathbf{x}}_i\| \max_j \{\|\nabla g_j(\mathbf{w}_{ji}, \boldsymbol{\theta}_j^*)\| \cdot \|\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j\|\} \right\} \\ \equiv T_1 + 2T_2.$$

For any  $\delta > 0$  and any  $j \in \{1, \dots, q\}$ ,  $n$  is large enough that

$$c_n < 1, \\ c_n^{-1} \cdot \|\boldsymbol{\beta}_0(\tau)\| \cdot \|\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j\| < \delta, \\ c_n^{-1} \|\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\| < \delta. \quad (\text{A13})$$

We can find a bound for  $\|\hat{D}_1 - \tilde{D}_1\|$  such that it can be arbitrarily small in probability if  $\delta$  is chosen to be sufficiently small. Since the bound for  $T_2$  is easy to derive, we only show that  $E(T_1) = O(\delta)$ .

To show this, applying assumption B3, (A12) and (A13), we have

$$|\hat{\mathbf{x}}_i^\top \hat{\boldsymbol{\beta}}(\tau) - \mathbf{x}_i^\top \boldsymbol{\beta}_0(\tau)| = |\hat{\mathbf{x}}_i^\top \hat{\boldsymbol{\beta}}(\tau) - \hat{\mathbf{x}}_i^\top \boldsymbol{\beta}_0(\tau) + \hat{\mathbf{x}}_i^\top \boldsymbol{\beta}_0(\tau) - \mathbf{x}_i^\top \boldsymbol{\beta}_0(\tau)| \\ \leq |\hat{\mathbf{x}}_i^\top (\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau))| + |(\hat{\mathbf{x}}_i - \mathbf{x}_i)^\top \boldsymbol{\beta}_0(\tau)| \\ \leq \|\hat{\mathbf{x}}_i\| \cdot \|\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\| + \max_j \{\|\nabla g_j(\mathbf{w}_{ji}, \boldsymbol{\theta}_j^*)\| \cdot \|\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j\|\} \cdot \|\boldsymbol{\beta}_0(\tau)\| \\ \leq \delta c_n \|\hat{\mathbf{x}}_i\| + \delta c_n \max_j D_j(\mathbf{w}_j) \\ = \delta c_n [\|\hat{\mathbf{x}}_i\| + \max_j D_j(\mathbf{w}_j)].$$

Inequality (A11) can be rewritten as

$$|I(|\hat{\epsilon}_i| < c_n) - I(|\epsilon_i| < c_n)| \leq I(|\epsilon_i - c_n| < \delta c_n [\|\hat{\mathbf{x}}_i\| + \max_j D_j(\mathbf{w}_j)]) \\ + I(|\epsilon_i + c_n| < \delta c_n [\|\hat{\mathbf{x}}_i\| + \max_j D_j(\mathbf{w}_j)]). \quad (\text{A14})$$



Then we have

$$\begin{aligned}
 E(T_1) &\leq E \left\{ (2nc_n)^{-1} \sum_{i=1}^n \left[ I(|\epsilon_i - c_n| < \delta c_n [\|\hat{\mathbf{x}}_i\| + \max_j D_j(\mathbf{w}_j)]) \right. \right. \\
 &\quad \left. \left. + I(|\epsilon_i + c_n| < \delta c_n [\|\hat{\mathbf{x}}_i\| + \max_j D_j(\mathbf{w}_j)]) \right] \|\hat{\mathbf{x}}_i\|^2 \right\} \\
 &\leq E \left\{ \|\hat{\mathbf{x}}_i\|^2 (2nc_n)^{-1} \sum_{i=1}^n 2 \int_{c_n - \delta c_n [\|\hat{\mathbf{x}}_i\| + \max_j D_j(\mathbf{w}_j)]}^{c_n + \delta c_n [\|\hat{\mathbf{x}}_i\| + \max_j D_j(\mathbf{w}_j)]} f_i(\lambda | \mathbf{x}_i) d\lambda \right\} \\
 &\leq E \left\{ \|\hat{\mathbf{x}}_i\|^2 (2nc_n)^{-1} 4\delta c_n \sum_{i=1}^n [\|\hat{\mathbf{x}}_i\| + \max_j D_j(\mathbf{w}_j)] A_f(\mathbf{x}_i) \right\} \\
 &\leq n^{-1} \sum_{i=1}^n 2 \cdot \delta \cdot A = 2A \cdot \delta = O(\delta),
 \end{aligned}$$

where  $A$  is some bound for  $E[(\|\hat{\mathbf{x}}_i\|^3 + \max_j D_j(\mathbf{w}_j) \|\hat{\mathbf{x}}_i\|^2) A_f(\mathbf{x}_i)]$  by assumption **B4**.

(b) We now show  $\tilde{D}_1 - D_1 = o_p(1)$ . Note that

$$\begin{aligned}
 \tilde{D}_1 - D_1 &= (2nc_n)^{-1} \sum_{i=1}^n \left\{ I(|\epsilon_i| < c_n) \mathbf{x}_i \mathbf{x}_i^\top - E[I(|\epsilon_i| < c_n) | \mathbf{x}_i] \mathbf{x}_i \mathbf{x}_i^\top \right\} \\
 &\quad + n^{-1} \sum_{i=1}^n \left\{ (2c_n)^{-1} E[I(|\epsilon_i| < c_n) | \mathbf{x}_i] \mathbf{x}_i \mathbf{x}_i^\top - [f_i(0 | \mathbf{x}_i) \mathbf{x}_i \mathbf{x}_i^\top] \right\}.
 \end{aligned}$$

For the first term, it has zero expectation and if  $\mathbf{a}$  and  $\mathbf{b}$  are  $k$ -dimensional vectors with  $\|\mathbf{a}\| = \|\mathbf{b}\| = 1$  and consider the variance of the first term,

$$\begin{aligned}
 &Var[\mathbf{a}^\top \{ (2nc_n)^{-1} \sum_{i=1}^n I(|\epsilon_i| < c_n) \mathbf{x}_i \mathbf{x}_i^\top - E[I(|\epsilon_i| < c_n) | \mathbf{x}_i] \mathbf{x}_i \mathbf{x}_i^\top \} \mathbf{b}] \\
 &= E \left\{ (2nc_n)^{-1} \sum_{i=1}^n I(|\epsilon_i| < c_n) \mathbf{a}^\top \mathbf{x}_i \mathbf{x}_i^\top \mathbf{b} - E[I(|\epsilon_i| < c_n) | \mathbf{x}_i] \mathbf{a}^\top \mathbf{x}_i \mathbf{x}_i^\top \mathbf{b} \right\}^2 \\
 &= (2nc_n)^{-2} E \left\{ \sum_{i=1}^n \{ I(|\epsilon_i| < c_n) - E[I(|\epsilon_i| < c_n) | \mathbf{x}_i] \}^2 \times [\mathbf{a}^\top \mathbf{x}_i \mathbf{x}_i^\top \mathbf{b}]^2 \right\} \\
 &\leq (2nc_n)^{-2} \sum_{i=1}^n E[\|\mathbf{x}_i\| \|\mathbf{x}_i\|]^2 \\
 &= (4nc_n^2)^{-1} H_1 = o(1),
 \end{aligned}$$

where the second equality holds because all the cross-product terms are zero by the law of iterated expectations, and  $H_1$  is some bound for  $E(\|\mathbf{x}_i\|^4)$  by assumption **B2**. Therefore, the first term converges in quadratic mean to zero, which implies that it converges to zero in probability.

For the second term,

$$\begin{aligned}
 |(2c_n)^{-1} E[I(|\epsilon_i| < c_n) | \mathbf{x}_i] - f_i(0 | \mathbf{x}_i)| &= \left| (2c_n)^{-1} \int_{-c_n}^{c_n} f_i(\lambda | \mathbf{x}_i) d\lambda - f_i(0 | \mathbf{x}_i) \right| \\
 &\leq |(2c_n)^{-1} 2c_n f_i(\lambda^* | \mathbf{x}_i) - f_i(0 | \mathbf{x}_i)| \\
 &\leq L_i |c_n| = o_p(1),
 \end{aligned}$$

where  $f_i(\lambda^* | \mathbf{x}_i) \equiv \max_{\lambda \in [-c_n, c_n]} f_i(\lambda | \mathbf{x}_i)$  and the last inequality uses assumption **B5**.

Combining claim (a) and (b), we have  $\hat{D}_1 - D_1 = o_p(1)$ .

(iii) Claim that  $\hat{D}_{12} - D_{12} = o_p(1)$ :

$$\begin{aligned}\hat{D}_{12} &= (2nc_n)^{-1} \sum_{i=1}^n I(|\hat{\epsilon}_i| < c_n) \hat{\beta}(\tau) \hat{\mathbf{x}}_i \nabla_{\theta} \mathbf{g}(\mathbf{w}_i, \hat{\theta})^{\top}, \\ D_{12} &= n^{-1} \sum_{i=1}^n f_i(0|x_i) \beta(\tau) \bar{\mathbf{x}}_i \nabla_{\theta} \mathbf{g}(\mathbf{w}_i, \theta)^{\top}.\end{aligned}$$

Define

$$\tilde{D}_{12} \equiv (2nc_n)^{-1} \sum_{i=1}^n I(|\epsilon_i| < c_n) \beta(\tau) \bar{\mathbf{x}}_i \nabla_{\theta} \mathbf{g}(\mathbf{w}_i, \theta)^{\top},$$

where  $\hat{\epsilon}_i \equiv y_i - \hat{\mathbf{x}}_i^{\top} \hat{\beta}(\tau)$  and  $\epsilon_i \equiv y_i - \mathbf{x}_i^{\top} \beta_0(\tau)$ . Similar to Part (ii), we need to show that  $\|\hat{D}_{12} - \tilde{D}_{12}\| = o_p(1)$  and  $\|\tilde{D}_{12} - D_{12}\| = o_p(1)$ .  $\|\tilde{D}_{12} - D_{12}\| = o_p(1)$  is easy to derive by following the idea in Part (ii)(b), we only show that  $\|\hat{D}_{12} - \tilde{D}_{12}\| = o_p(1)$ .

$$\begin{aligned}\|\hat{D}_{12} - \tilde{D}_{12}\| &= \left\| (2nc_n)^{-1} \times \sum_{i=1}^n \left\{ [I(|\hat{\epsilon}_i| < c_n) - I(|\epsilon_i| < c_n)] \hat{\beta}(\tau) \hat{\mathbf{x}}_i \nabla_{\theta} \mathbf{g}(\mathbf{w}_i, \hat{\theta})^{\top} \right. \right. \\ &\quad + I(|\epsilon_i| < c_n) (\hat{\beta}(\tau) - \beta(\tau)) \hat{\mathbf{x}}_i \nabla_{\theta} \mathbf{g}(\mathbf{w}_i, \hat{\theta})^{\top} \\ &\quad + I(|\epsilon_i| < c_n) \beta(\tau) [\hat{\mathbf{x}}_i - \bar{\mathbf{x}}_i] \nabla_{\theta} \mathbf{g}(\mathbf{w}_i, \hat{\theta})^{\top} \\ &\quad \left. \left. + I(|\epsilon_i| < c_n) \beta(\tau) \bar{\mathbf{x}}_i [\nabla_{\theta} \mathbf{g}(\mathbf{w}_i, \hat{\theta})^{\top} - \nabla_{\theta} \mathbf{g}(\mathbf{w}_i, \theta)^{\top}] \right\} \right\|.\end{aligned}$$

Applying assumption **B3**, (A12) and (A13), we have for  $j' = 1, \dots, q$

$$\begin{aligned}\|\beta(\tau) [\hat{\mathbf{x}}_i - \bar{\mathbf{x}}_i] \nabla_{\theta} \mathbf{g}(\mathbf{w}_i, \hat{\theta})^{\top}\| &= \|\beta(\tau)\| \max_{j'} \{ \|\hat{\theta}_{j'} - \theta_{j'}\| \cdot \|\nabla_{\theta_{j'}} g_{j'}(\mathbf{w}_{j'}, \theta_{j'}^*)\| \} \cdot \|\nabla_{\theta_j} g_j(\mathbf{w}_{j'}, \hat{\theta}_j)\| \\ &\leq \delta c_n \max_{j'} D_{j'}(\mathbf{w}_{j'}) \cdot D_j(\mathbf{w}_j),\end{aligned}$$

and using assumption **B3** and (A13), we have

$$\begin{aligned}\|I(|\epsilon_i| < c_n) \beta(\tau) \bar{\mathbf{x}}_i [\nabla_{\theta} \mathbf{g}(\mathbf{w}_i, \hat{\theta})^{\top} - \nabla_{\theta} \mathbf{g}(\mathbf{w}_i, \theta)^{\top}]\| &\leq I(|\epsilon_i| < c_n) \|\beta(\tau)\| \cdot \|\bar{\mathbf{x}}_i\| \cdot J \|\hat{\theta} - \theta\| \\ &\leq I(|\epsilon_i| < c_n) \delta c_n J \|\bar{\mathbf{x}}_i\|.\end{aligned}$$

Combining the above two inequalities, assumption **B3** and (A14), we have

$$\begin{aligned}\|\hat{D}_{12} - \tilde{D}_{12}\| &\leq (2nc_n)^{-1} \sum_{i=1}^n \left\{ \left[ I(|\epsilon_i - c_n| < \delta c_n [\|\hat{\mathbf{x}}_i\| + \max_{j'} D_{j'}(\mathbf{w}_{j'})]) \right. \right. \\ &\quad + I(|\epsilon_i + c_n| < \delta c_n [\|\hat{\mathbf{x}}_i\| + \max_{j'} D_{j'}(\mathbf{w}_{j'})]) \left. \right] \|\beta(\tau)\| \cdot \|\hat{\mathbf{x}}_i\| D_j(\mathbf{w}_j) \\ &\quad + I(|\epsilon_i| < c_n) \delta c_n \|\hat{\mathbf{x}}_i\| D_j(\mathbf{w}_j) \\ &\quad + I(|\epsilon_i| < c_n) \delta c_n \max_{j'} D_{j'}(\mathbf{w}_{j'}) D_j(\mathbf{w}_j) \\ &\quad \left. + I(|\epsilon_i| < c_n) \delta c_n J \|\hat{\mathbf{x}}_i\| \right\} \\ &\equiv N_1 + N_2 + N_3 + N_4.\end{aligned}$$

Similar to Part (ii)(a), we can find a bound for  $\|\hat{D}_{12} - \tilde{D}_{12}\|$ , such that it can be arbitrarily small in probability if  $\delta$  is chosen sufficiently small. Since the bounds for  $N_2$ ,  $N_3$  and  $N_4$  are easy to derive, we only show that  $E(N_1) = O(\delta)$  as follows:

$$\begin{aligned} E(N_1) &= E \left\{ (2nc_n)^{-1} \sum_{i=1}^n \left[ I(|\epsilon_i - c_n| < \delta c_n [\|\hat{\mathbf{x}}_i\| + \max_{j'} D_{j'}(\mathbf{w}_{j'})]) \right. \right. \\ &\quad \left. \left. + I(|\epsilon_i + c_n| < \delta c_n [\|\hat{\mathbf{x}}_i\| + \max_{j'} D_{j'}(\mathbf{w}_{j'})]) \right] \|\boldsymbol{\beta}(\tau)\| \cdot \|\hat{\mathbf{x}}_i\| D_j(\mathbf{w}_j) \right\} \\ &\leq E \left\{ \|\boldsymbol{\beta}(\tau)\| \cdot \|\hat{\mathbf{x}}_i\| D_j(\mathbf{w}_j) (2nc_n)^{-1} \sum_{i=1}^n 2 \int_{c_n - \delta c_n [\|\hat{\mathbf{x}}_i\| + \max_{j'} D_{j'}(\mathbf{w}_{j'})]}^{c_n + \delta c_n [\|\hat{\mathbf{x}}_i\| + \max_{j'} D_{j'}(\mathbf{w}_{j'})]} f_i(\lambda | \mathbf{x}_i) d\lambda \right\} \\ &\leq E \left\{ \|\boldsymbol{\beta}(\tau)\| \cdot \|\hat{\mathbf{x}}_i\| D_j(\mathbf{w}_j) (2nc_n)^{-1} 4\delta c_n \sum_{i=1}^n [\|\hat{\mathbf{x}}_i\| + \max_{j'} D_{j'}(\mathbf{w}_{j'})] A_f(\mathbf{x}_i) \right\} \\ &\leq n^{-1} \sum_{i=1}^n 2\|\boldsymbol{\beta}(\tau)\| \cdot \delta \cdot B_j = O(\delta), \end{aligned}$$

for some  $B_j$  satisfying  $E[D_j(\mathbf{w}_j)\{\|\mathbf{x}_i\|^2 + \max_{j'} D_{j'}(\mathbf{w}_{j'})\|\mathbf{x}_i\|] A_f(\mathbf{x}_i)] < B_j < \infty$ , where the last inequality holds by assumption **B4**.

(iv) Claim that  $\widehat{M} \xrightarrow{p} M$ .

$$\begin{aligned} M &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \left\{ \psi_\tau(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_0(\tau)) f_i(0 | \mathbf{x}_i) \mathbf{x}_i \mathbf{r}(\boldsymbol{\theta})^\top \nabla_{\boldsymbol{\theta}} \mathbf{g}(\mathbf{w}_i, \boldsymbol{\theta})^\top \hat{\mathbf{x}}_i^\top \boldsymbol{\beta}(\tau)^\top \right\} \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \left\{ [\tau - I(\epsilon_i < 0)] f_i(0 | \mathbf{x}_i) \mathbf{x}_i \mathbf{r}(\boldsymbol{\theta})^\top \nabla_{\boldsymbol{\theta}} \mathbf{g}(\mathbf{w}_i, \boldsymbol{\theta})^\top \hat{\mathbf{x}}_i^\top \boldsymbol{\beta}(\tau)^\top \right\} \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \left\{ [\tau f_i(0 | \mathbf{x}_i) - I(\epsilon_i < 0) f_i(0 | \mathbf{x}_i)] \mathbf{x}_i \mathbf{r}(\boldsymbol{\theta})^\top \nabla_{\boldsymbol{\theta}} \mathbf{g}(\mathbf{w}_i, \boldsymbol{\theta})^\top \hat{\mathbf{x}}_i^\top \boldsymbol{\beta}(\tau)^\top \right\}. \end{aligned}$$

Denote  $\hat{\epsilon}_i \equiv y_i - \hat{\mathbf{x}}_i^\top \hat{\boldsymbol{\beta}}(\tau)$  and  $\epsilon_i \equiv y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_0(\tau)$ . To obtain a consistent estimator of  $M$  which contains unknown conditional density function, the estimator  $\widehat{M}$  replaces the conditional density functions by uniform kernel weights:

$$\begin{aligned} \widehat{M} &= \frac{1}{2nc_n} \sum_{i=1}^n \left\{ [\tau - I(\hat{\epsilon}_i < 0)] I(|\hat{\epsilon}_i| < c_n) \hat{\mathbf{x}}_i \mathbf{r}(\hat{\boldsymbol{\theta}})^\top \nabla_{\boldsymbol{\theta}} \mathbf{g}(\mathbf{w}_i, \hat{\boldsymbol{\theta}})^\top \hat{\mathbf{x}}_i^\top \hat{\boldsymbol{\beta}}(\tau)^\top \right\} \\ &= \frac{1}{2nc_n} \sum_{i=1}^n \left\{ [\tau I(|\hat{\epsilon}_i| < c_n) - I(\hat{\epsilon}_i < 0) I(|\hat{\epsilon}_i| < c_n)] \hat{\mathbf{x}}_i \mathbf{r}(\hat{\boldsymbol{\theta}})^\top \nabla_{\boldsymbol{\theta}} \mathbf{g}(\mathbf{w}_i, \hat{\boldsymbol{\theta}})^\top \hat{\mathbf{x}}_i^\top \hat{\boldsymbol{\beta}}(\tau)^\top \right\}. \end{aligned}$$

Define

$$\begin{aligned} \widetilde{M} &= \frac{1}{2nc_n} \sum_{i=1}^n \left\{ [\tau - I(\epsilon_i < 0)] I(|\epsilon_i| < c_n) \mathbf{x}_i \mathbf{r}(\boldsymbol{\theta})^\top \nabla_{\boldsymbol{\theta}} \mathbf{g}(\mathbf{w}_i, \boldsymbol{\theta})^\top \hat{\mathbf{x}}_i^\top \boldsymbol{\beta}(\tau)^\top \right\} \\ &= \frac{1}{2nc_n} \sum_{i=1}^n \left\{ [\tau I(|\epsilon_i| < c_n) - I(\epsilon_i < 0) I(|\epsilon_i| < c_n)] \mathbf{x}_i \mathbf{r}(\boldsymbol{\theta})^\top \nabla_{\boldsymbol{\theta}} \mathbf{g}(\mathbf{w}_i, \boldsymbol{\theta})^\top \hat{\mathbf{x}}_i^\top \boldsymbol{\beta}(\tau)^\top \right\}. \end{aligned}$$

Similar to Part (ii), we need to show that  $\|\widehat{M} - \widetilde{M}\| = o_p(1)$  and  $\|\widetilde{M} - M\| = o_p(1)$ .  $\|\widetilde{M} - M\| = o_p(1)$  is easy to derive by following the idea in Part (ii)(b), we only show that  $\|\widehat{M} - \widetilde{M}\| = o_p(1)$ .

Consider

$$\begin{aligned}
\|\widehat{M} - \widetilde{M}\| &= \|(2nc_n)^{-1} \times \sum_{i=1}^n \{[\tau I(|\widehat{\epsilon}_i| < c_n) - I(\widehat{\epsilon}_i < 0)I(|\widehat{\epsilon}_i| < c_n)]\widehat{\mathbf{x}}_i \mathbf{r}(\widehat{\boldsymbol{\theta}})^\top \nabla_{\boldsymbol{\theta}} \mathbf{g}(\mathbf{w}_i, \widehat{\boldsymbol{\theta}})^\top \widehat{\mathbf{x}}_i^\top \widehat{\boldsymbol{\beta}}(\tau)^\top \\
&\quad - [\tau I(|\epsilon_i| < c_n) - I(\epsilon_i < 0)I(|\epsilon_i| < c_n)]\mathbf{x}_i \mathbf{r}(\boldsymbol{\theta})^\top \nabla_{\boldsymbol{\theta}} \mathbf{g}(\mathbf{w}_i, \boldsymbol{\theta})^\top \mathbf{x}_i^\top \boldsymbol{\beta}(\tau)^\top\} \| \\
&= \|(2nc_n)^{-1} \times \sum_{i=1}^n \{([\tau I(|\widehat{\epsilon}_i| < c_n) - \tau I(|\epsilon_i| < c_n)] \\
&\quad - [I(\widehat{\epsilon}_i < 0)I(|\widehat{\epsilon}_i| < c_n) - I(\epsilon_i < 0)I(|\epsilon_i| < c_n)]) \times \widehat{\mathbf{x}}_i \mathbf{r}(\widehat{\boldsymbol{\theta}})^\top \nabla_{\boldsymbol{\theta}} \mathbf{g}(\mathbf{w}_i, \widehat{\boldsymbol{\theta}})^\top \widehat{\mathbf{x}}_i^\top \widehat{\boldsymbol{\beta}}(\tau)^\top \\
&\quad + [\tau I(|\epsilon_i| < c_n) - I(\epsilon_i < 0)I(|\epsilon_i| < c_n)] \\
&\quad \times [\widehat{\mathbf{x}}_i \mathbf{r}(\widehat{\boldsymbol{\theta}})^\top \nabla_{\boldsymbol{\theta}} \mathbf{g}(\mathbf{w}_i, \widehat{\boldsymbol{\theta}})^\top \widehat{\mathbf{x}}_i^\top \widehat{\boldsymbol{\beta}}(\tau)^\top - \mathbf{x}_i \mathbf{r}(\boldsymbol{\theta})^\top \nabla_{\boldsymbol{\theta}} \mathbf{g}(\mathbf{w}_i, \boldsymbol{\theta})^\top \mathbf{x}_i^\top \boldsymbol{\beta}(\tau)^\top]\} \| \\
&\leq \|(2nc_n)^{-1} \times \sum_{i=1}^n \{[\tau I(|\widehat{\epsilon}_i| < c_n) - \tau I(|\epsilon_i| < c_n)] \\
&\quad - [I(\widehat{\epsilon}_i < 0)I(|\widehat{\epsilon}_i| < c_n) - I(\epsilon_i < 0)I(|\epsilon_i| < c_n)]\} \\
&\quad \times \widehat{\mathbf{x}}_i \mathbf{r}(\widehat{\boldsymbol{\theta}})^\top \nabla_{\boldsymbol{\theta}} \mathbf{g}(\mathbf{w}_i, \widehat{\boldsymbol{\theta}})^\top \widehat{\mathbf{x}}_i^\top \widehat{\boldsymbol{\beta}}(\tau)^\top \| \\
&\quad + \|(2nc_n)^{-1} \times \sum_{i=1}^n [\tau I(|\epsilon_i| < c_n) - I(\epsilon_i < 0)I(|\epsilon_i| < c_n)] \times [\widehat{\mathbf{x}}_i \mathbf{r}(\widehat{\boldsymbol{\theta}})^\top \nabla_{\boldsymbol{\theta}} \mathbf{g}(\mathbf{w}_i, \widehat{\boldsymbol{\theta}})^\top \widehat{\mathbf{x}}_i^\top \widehat{\boldsymbol{\beta}}(\tau)^\top \\
&\quad - \mathbf{x}_i \mathbf{r}(\boldsymbol{\theta})^\top \nabla_{\boldsymbol{\theta}} \mathbf{g}(\mathbf{w}_i, \boldsymbol{\theta})^\top \mathbf{x}_i^\top \boldsymbol{\beta}(\tau)^\top] \| \\
&\leq \|(2nc_n)^{-1} \times \sum_{i=1}^n \{[\tau I(|\widehat{\epsilon}_i| < c_n) - \tau I(|\epsilon_i| < c_n)] \times \widehat{\mathbf{x}}_i \mathbf{r}(\widehat{\boldsymbol{\theta}})^\top \nabla_{\boldsymbol{\theta}} \mathbf{g}(\mathbf{w}_i, \widehat{\boldsymbol{\theta}})^\top \widehat{\mathbf{x}}_i^\top \widehat{\boldsymbol{\beta}}(\tau)^\top\} \| \\
&\quad + \|(2nc_n)^{-1} \times \sum_{i=1}^n [I(\widehat{\epsilon}_i < 0)I(|\widehat{\epsilon}_i| < c_n) - I(\epsilon_i < 0)I(|\epsilon_i| < c_n)] \\
&\quad \times \widehat{\mathbf{x}}_i \mathbf{r}(\widehat{\boldsymbol{\theta}})^\top \nabla_{\boldsymbol{\theta}} \mathbf{g}(\mathbf{w}_i, \widehat{\boldsymbol{\theta}})^\top \widehat{\mathbf{x}}_i^\top \widehat{\boldsymbol{\beta}}(\tau)^\top \| \\
&\quad + \|(2nc_n)^{-1} \times \sum_{i=1}^n [\tau I(|\epsilon_i| < c_n) - I(\epsilon_i < 0)I(|\epsilon_i| < c_n)] \times [\widehat{\mathbf{x}}_i \mathbf{r}(\widehat{\boldsymbol{\theta}})^\top \nabla_{\boldsymbol{\theta}} \mathbf{g}(\mathbf{w}_i, \widehat{\boldsymbol{\theta}})^\top \widehat{\mathbf{x}}_i^\top \widehat{\boldsymbol{\beta}}(\tau)^\top \\
&\quad - \mathbf{x}_i \mathbf{r}(\boldsymbol{\theta})^\top \nabla_{\boldsymbol{\theta}} \mathbf{g}(\mathbf{w}_i, \boldsymbol{\theta})^\top \mathbf{x}_i^\top \boldsymbol{\beta}(\tau)^\top] \| \\
&= Z_1 + Z_2 + Z_3.
\end{aligned}$$

Notice that

$$I(\widehat{\epsilon}_i < 0)I(|\widehat{\epsilon}_i| < c_n) - I(\epsilon_i < 0)I(|\epsilon_i| < c_n) = I(-c_n < \widehat{\epsilon}_i < 0) - I(-c_n < \epsilon_i < 0)$$

and

$$\begin{aligned}
|I(-c_n < \widehat{\epsilon}_i < 0) - I(-c_n < \epsilon_i < 0)| &\leq I(|\epsilon_i + c_n| < |\widehat{\mathbf{x}}_i^\top \widehat{\boldsymbol{\beta}}(\tau) - \mathbf{x}_i^\top \boldsymbol{\beta}_0(\tau)|) \\
&\quad + I(|\epsilon_i| < |\widehat{\mathbf{x}}_i^\top \widehat{\boldsymbol{\beta}}(\tau) - \mathbf{x}_i^\top \boldsymbol{\beta}_0(\tau)|).
\end{aligned}$$

From (A12), we already have

$$|\widehat{\mathbf{x}}_i^\top \widehat{\boldsymbol{\beta}}(\tau) - \mathbf{x}_i^\top \boldsymbol{\beta}_0(\tau)| \leq \delta c_n [\|\widehat{\mathbf{x}}_i\| + \max_j D_j(w_j)].$$

We can find a bound for  $\|\widehat{M} - \widetilde{M}\|$  such that it can be arbitrarily small in probability if  $\delta$  is chosen sufficiently small. Since the bounds for  $Z_1$  and  $Z_3$  are easy to derive using the similar argument we used in part (iii), we only show that  $E(Z_2) = O(\delta)$  as following:

$$\begin{aligned}
 E(Z_2) &\leq E\{(2nc_n)^{-1} \times \sum_{i=1}^n [I(|\epsilon_i + c_n| < \delta c_n[\|\widehat{x}_i\| + \max_j D_j(w_j)])] \\
 &\quad + I(|\epsilon_i| < \delta c_n[\|\widehat{x}_i\| + \max_j D_j(w_j)])] \|\widehat{x}_i r(\widehat{\theta})^\top \nabla_{\theta} g(w_i, \widehat{\theta})^\top \widehat{x}_i^\top \widehat{\beta}(\tau)^\top\| \} \\
 &\leq E\{ \|\widehat{x}_i r(\widehat{\theta})^\top \nabla_{\theta} g(w_i, \widehat{\theta})^\top \widehat{x}_i^\top \widehat{\beta}(\tau)^\top\| (2nc_n)^{-1} \\
 &\quad \times \sum_{i=1}^n [\int_{c_n - \delta c_n[\|\widehat{x}_i\| + \max_j D_j(w_j)]}^{c_n + \delta c_n[\|\widehat{x}_i\| + \max_j D_j(w_j)]} f_i(\lambda | x_i) d\lambda + \int_{-\delta c_n[\|\widehat{x}_i\| + \max_j D_j(w_j)]}^{\delta c_n[\|\widehat{x}_i\| + \max_j D_j(w_j)]} f_i(\lambda | x_i) d\lambda] \} \\
 &\leq E\{ \|\widehat{x}_i r(\widehat{\theta})^\top \nabla_{\theta} g(w_i, \widehat{\theta})^\top \widehat{x}_i^\top \widehat{\beta}(\tau)^\top\| (2nc_n)^{-1} 4\delta c_n \sum_{i=1}^n [\|\widehat{x}_i\| + \max_j D_j(w_j)] A_f(x_i) \} \\
 &\leq 2C \cdot \delta = O(\delta)
 \end{aligned}$$

where  $C < \infty$  is some bound for  $E[\|\widehat{x}_i r(\widehat{\theta})^\top \nabla_{\theta} g(w_i, \widehat{\theta})^\top \widehat{x}_i^\top \widehat{\beta}(\tau)^\top\| \cdot [\|\widehat{x}_i\| + \max_j D_j(w_j)] A_f(x_i)]$  by assumptions **A2**, **A4** and **A5**.  $\square$

**Proof of Theorem 4.** The proof of Theorem 4 is simple; it follows from observing that for any fixed  $\tau$ , by Theorem 2

$$\sqrt{n}(\widehat{\beta}(\tau) - \beta_0(\tau)) \xrightarrow{d} N(0, \Omega(\tau)),$$

and under the null hypothesis,

$$\sqrt{n}(R\widehat{\beta}(\tau) - r) \xrightarrow{d} N(0, R\Omega(\tau)R^\top).$$

Since by Theorem 3,  $\widehat{\Omega}(\tau)$  is a consistent estimator of  $\Omega(\tau)$ , by the Slutsky's theorem,

$$\mathcal{W}_T = T(R\widehat{\beta}(\tau) - r)^\top [R\widehat{\Omega}(\tau)R^\top]^{-1} (R\widehat{\beta}(\tau) - r) \xrightarrow{d} \chi_s^2.$$

$\square$

#### Appendix A.2. OLS with Generated Regressors

In order to compare our quantile regression with generated regressor framework with the OLS with generated regressor (OLS-GR), we include the detailed assumptions and propositions for the OLS case in this section.

$$y = x^\top \beta_0 + u \quad (\text{A15})$$

where  $E(u|x) = 0$ ,  $\beta_0$  is a  $k \times 1$  vector,  $x = h(w, \delta)$ ,  $\delta$  is a  $p \times 1$  vector and  $h(\cdot)$  is known but  $\delta$  is unknown and  $E(u|w) = 0$ .

Let  $\widehat{\delta}$  be a  $\sqrt{n}$ -consistent estimator of  $\delta$ , and we obtain the generated regressors as  $\widehat{x}_i = h(w_i, \widehat{\delta})$ . Let  $\widehat{\beta}$  be the OLS-GR estimator from the equation

$$y_i = \widehat{x}_i^\top \beta + e_i,$$

where  $\widehat{x}_i = h(w_i, \widehat{\delta})$ . We impose the following regularity conditions.

**Conditions:**

**C1.**  $n \geq k$ ,  $\text{rank}(X) = k$ .

**C2.**  $E(u|x) = 0$ .

**C3.** The observations  $\{(y_i, \mathbf{x}_i) : i = 1, 2, \dots, n\}$  are i.i.d. across  $i$  with  $E(y_i^{4+\epsilon}) < \infty$  and  $E(\|\mathbf{x}_i\|^{4+\epsilon}) < \infty$  for some  $\epsilon > 0$ ,  $\text{plim}_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n (\mathbf{x}_i \mathbf{x}_i^\top) = H$  is a nonsingular matrix, and  $E(u|\mathbf{w}) = 0$ .

**C4.**  $\sqrt{n}(\hat{\delta} - \delta) = n^{-1/2} \sum_{i=1}^n \mathbf{r}_i(\delta) + o_p(1)$ , where  $E[\mathbf{r}_i(\delta)] = 0$ .

**C5.**  $\text{plim}_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n (\boldsymbol{\beta} \otimes \mathbf{x}_i^\top)^\top \nabla_\delta h(\mathbf{w}_i, \delta)^\top = G$ , which is bounded.

We first state the consistency of the OLS-GR estimator.

**Proposition A1.** In model (A15), under the assumptions C1–C3, the OLS-GR estimator is consistent, i.e.,  $\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}_0$ .

**Proof of Proposition A1.** Recall that

$$\hat{\boldsymbol{\beta}} = (\hat{\mathbf{X}}^\top \hat{\mathbf{X}})^{-1} \hat{\mathbf{X}}^\top \mathbf{y} = \left( \sum_{i=1}^n \hat{\mathbf{x}}_i \hat{\mathbf{x}}_i^\top \right)^{-1} \left[ \sum_{i=1}^n \hat{\mathbf{x}}_i y_i \right].$$

Thus,

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = \left( \frac{1}{n} \sum_{i=1}^n \hat{\mathbf{x}}_i \hat{\mathbf{x}}_i^\top \right)^{-1} \left\{ \frac{1}{n} \sum_{i=1}^n \hat{\mathbf{x}}_i [(x_i - \hat{x}_i)^\top \boldsymbol{\beta}_0 + u_i] \right\} = \hat{H}^{-1} \left\{ \frac{1}{n} \sum_{i=1}^n \hat{\mathbf{x}}_i [(x_i - \hat{x}_i)^\top \boldsymbol{\beta}_0 + u_i] \right\},$$

where  $\hat{H} = \frac{1}{n} \sum_{i=1}^n \hat{\mathbf{x}}_i \hat{\mathbf{x}}_i^\top$ . Using the mean value expansion, we have

$$\frac{1}{n} \sum_{i=1}^n \hat{\mathbf{x}}_i u_i = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i u_i + \left[ \frac{1}{n} \sum_{i=1}^n \nabla_\delta h(\mathbf{w}_i, \delta)^\top u_i \right] (\hat{\delta} - \delta) + o_p(1).$$

Since  $E(u|\mathbf{w}) = 0$ ,  $E[\nabla_\delta h(\mathbf{w}_i, \delta)^\top u_i] = 0$ . It implies that  $\frac{1}{n} \sum_{i=1}^n \nabla_\delta h(\mathbf{w}_i, \delta)^\top u_i = o_p(1)$ . Since  $\hat{\delta} - \delta = o_p(1)$  and  $E(\mathbf{x}_i u_i) = 0$ , it follows that

$$\frac{1}{n} \sum_{i=1}^n \hat{\mathbf{x}}_i u_i = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i u_i + o_p(1) = o_p(1).$$

Or,

$$\frac{1}{n} \sum_{i=1}^n \hat{\mathbf{x}}_i (x_i - \hat{x}_i)^\top \boldsymbol{\beta}_0 = - \left[ \frac{1}{n} \sum_{i=1}^n (\boldsymbol{\beta}_0 \otimes \mathbf{x}_i^\top)^\top \nabla_\delta h(\mathbf{w}_i, \delta)^\top \right] (\hat{\delta} - \delta) + o_p(1) = -G(\hat{\delta} - \delta) + o_p(1),$$

where  $G = E[(\boldsymbol{\beta}_0 \otimes \mathbf{x}_i^\top)^\top \nabla_\delta h(\mathbf{w}_i, \delta)^\top]$ , using  $\hat{\mathbf{x}}_i (x_i - \hat{x}_i)^\top \boldsymbol{\beta}_0 = (\boldsymbol{\beta}_0 \otimes \hat{\mathbf{x}}_i^\top)^\top (\mathbf{x}_i - \hat{\mathbf{x}}_i)$ . Since  $\hat{\delta} - \delta = o_p(1)$  and  $G$  is bounded, then

$$\frac{1}{n} \sum_{i=1}^n \hat{\mathbf{x}}_i (x_i - \hat{x}_i)^\top \boldsymbol{\beta}_0 = o_p(1).$$

Therefore,  $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = o_p(1)$ .  $\square$

Next, we state the asymptotic normality of the OLS-GR estimator.

**Proposition A2.** In model (A15), under the assumptions C4–C5 and the conditions in Proposition 1, the OLS-GR is asymptotically normal, i.e.,

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N(0, H^{-1} M H^{-1}),$$

where  $M = \text{Var}[\mathbf{x}_i u_i - \text{Gr}_i(\delta)]$ .



**Proof of Proposition A2.** Note that from

$$\hat{\beta} = (\hat{X}^\top \hat{X})^{-1} \hat{X}^\top y = \left( \sum_{i=1}^n \hat{x}_i \hat{x}_i^\top \right)^{-1} \left[ \sum_{i=1}^n \hat{x}_i y_i \right],$$

By rearranging we obtain

$$\begin{aligned} \sqrt{n}(\hat{\beta} - \beta_0) &= \left( \frac{1}{n} \sum_{i=1}^n \hat{x}_i \hat{x}_i^\top \right)^{-1} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{x}_i [(x_i - \hat{x}_i)^\top \beta_0 + u_i] \right\} \\ &= \hat{H}^{-1} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{x}_i [(x_i - \hat{x}_i)^\top \beta_0 + u_i] \right\}, \end{aligned}$$

where  $\hat{H} = \frac{1}{n} \sum_{i=1}^n \hat{x}_i \hat{x}_i^\top$ . Using the mean value expansion, we have

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{x}_i u_i = \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i u_i + \left[ \frac{1}{n} \sum_{i=1}^n \nabla_\delta h(w_i, \delta)^\top u_i \right] \sqrt{n}(\hat{\delta} - \delta) + o_p(1).$$

Since  $E(u|w) = 0$ ,  $E[\nabla_\delta h(w_i, \delta)^\top u_i] = 0$ . It implies that  $\frac{1}{n} \sum_{i=1}^n \nabla_\delta h(w_i, \delta)^\top u_i = o_p(1)$ . Since  $\sqrt{n}(\hat{\delta} - \delta) = O_p(1)$ , it follows that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{x}_i u_i = \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i u_i + o_p(1).$$

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{x}_i (x_i - \hat{x}_i)^\top \beta_0 = - \left[ \frac{1}{n} \sum_{i=1}^n (\beta_0 \otimes x_i^\top)^\top \nabla_\delta h(w_i, \delta)^\top \right] \sqrt{n}(\hat{\delta} - \delta) + o_p(1) = -G \sqrt{n}(\hat{\delta} - \delta) + o_p(1),$$

where  $G = E[(\beta_0 \otimes x_i^\top)^\top \nabla_\delta h(w_i, \delta)^\top]$ , using  $\hat{x}_i (x_i - \hat{x}_i)^\top \beta_0 = (\beta_0 \otimes \hat{x}_i^\top)^\top (x_i - \hat{x}_i)$ . By Assumption C4,

$$\sqrt{n}(\hat{\delta} - \delta) = n^{-1/2} \sum_{i=1}^n r_i(\delta) + o_p(1).$$

Hence,  $\sqrt{n}(\hat{\beta} - \beta_0) = H^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n [x_i u_i - G r_i(\delta)] + o_p(1)$ .

By the Central Limit Theorem, we obtain

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, H^{-1} M H^{-1}),$$

where  $M = \text{Var}[x_i u_i - G r_i(\delta)]$ .  $\square$

## Notes

- Examples of generated regressors include models of interest involving expectations of future variables, such as expected prices or sales or inflation that have been generated as the predictions of some dynamic model (Engle (1982)). “Unanticipated” components of aggregate money growth in macroeconomic models (Barro (1977); Barro (1978)).
- In particular, Xiao and Koenker (2009) develop QR with GR in the context of GARCH models. Chen et al. (2015) propose a quantile factor model. Lee (2007) applied a control function approach to generate instruments and resolve the endogeneity, and Ma and Koenker (2006) develop QR for recursive structural equation models. Chernozhukov et al. (2015) suggest QR with censoring and endogeneity. Arellano and Bonhomme (2016) discuss the correction of the QR estimates for nonrandom sample selection. Chernozhukov and Hansen (2005 2006) develop a model of quantile treatment effects. Akerberg et al. (2014) suggest two-step GMM where the moment functions can be seen as the score of QR.
- R codes are provided for all methods, simulations, and applications.

- <sup>4</sup> We note that  $\mathbf{x}_i = (x_{i1}, \dots, x_{iq}, x_{iq+1}, \dots, x_{ik})$  where  $x_{ij} = g_j(\mathbf{w}_{ji}, \boldsymbol{\theta}_j)$  for  $j = 1, \dots, q$ . We also note that for any  $\boldsymbol{\theta}_j \in \Theta_j$  and for all  $i$ ,  $E[\|\mathbf{x}_i\|^3 + \max_j D_j(\mathbf{w}_j)\|\mathbf{x}_i\|^2]A_f(\mathbf{x}_i) < A < \infty$  and  $E[D_j(\mathbf{w}_j)\{\|\mathbf{x}_i\|^2 + \max_{j'} D_{j'}(\mathbf{w}_{j'})\|\mathbf{x}_i\}]A_f(\mathbf{x}_i) < B_j < \infty$ . Additionally,  $E[\hat{\mathbf{x}}_i \mathbf{r}(\hat{\boldsymbol{\theta}})^\top \nabla_{\boldsymbol{\theta}} g(\mathbf{w}_i, \hat{\boldsymbol{\theta}})^\top \hat{\mathbf{x}}_i^\top \hat{\boldsymbol{\beta}}(\tau)^\top \cdot [\|\hat{\mathbf{x}}_i\| + \max_j D_j(\mathbf{w}_j)]A_f(\mathbf{x}_i)] < C < \infty$  by applying A2, A4 and A5.
- <sup>5</sup> These data have been previously used by Barigozzi and Moneta (2016).
- <sup>6</sup> The way to aggregate consumption follows Barigozzi and Moneta (2016).
- <sup>7</sup> RPI is obtained from UK Office for National Statistics (<http://www.ons.gov.uk/> (accessed on January 2019)). To deflate the data, we divided the nominal total expenditure by the aggregate price index. Additionally, the nominal budget share, as a ratio of nominal level of expenditure over nominal total budget, was multiplied by a ratio of the total price index over a price index for the particular expenditure.

## References

- Ackerberg, Daniel, Xiaohong Chen, Jinyong Hahn, and Zhipeng Liao. 2014. Asymptotic efficiency of semiparametric two-step gmm. *Review of Economic Studies* 288: 919–43. [CrossRef]
- Arellano, Manuel, and Stéphane Bonhomme. 2018. Sample selection in quantile regression: A survey. In *Handbook of Quantile Regression*. Edited by Koenker, Roger, Victor Chernozhukov, Xuming He and Limin Peng. Boca Raton: CRC/Chapman-Hall.
- Bai, Jushan, and Serena Ng. 2002. Determining the number of factors in approximate factor models. *Econometrica* 70: 191–221. [CrossRef]
- Barigozzi, Matteo, and Alessio Moneta. 2016. Identifying the independent sources of consumption variation. *Journal of Applied Econometrics* 31: 420–49. [CrossRef]
- Barro, Robert J. 1977. Unanticipated money growth and unemployment in the united states. *The American Economic Review* 67: 101–15.
- Barro, Robert J. 1978. Unanticipated money, output, and the price level in the united states. *Journal of Political Economy* 86: 549–80. [CrossRef]
- Blundell, Richard, Xiaohong Chen, and Dennis Kristensen. 2007. Semi-nonparametric iv estimation of shape-invariant engel curves. *Econometrica* 75: 1613–69. [CrossRef]
- Buchinsky, Moshe, and Jinyong Hahn. 1998. An alternative estimator for the censored quantile regression model. *Econometrica* 66: 653–71. [CrossRef]
- Chai, Andreas, and Alessio Moneta. 2010. Retrospectives: Engel curves. *Journal of Economic Perspectives* 24: 225–40. [CrossRef]
- Chen, Liang, Juan Jose Dolado, and Jesus Gonzalo. 2015. Quantile factor models. *arXiv* arXiv:1911.02173.
- Chernozhukov, Victor, Ivan Fernandez-Val, and A. E. Kowalski. 2015. Quantile regression with censoring and endogeneity. *Journal of Econometrics* 186: 201–21. [CrossRef]
- Chernozhukov, Victor, and Christian Hansen. 2005. An iv model of quantile treatment effects. *Econometrica* 73: 245–61. [CrossRef]
- Chernozhukov, Victor, and Christian Hansen. 2006. Instrumental quantile regression inference for structural and treatment effects models. *Journal of Econometrics* 132: 491–525. [CrossRef]
- Engel, Ernst. 1857. Die productions-und consumtionsverhältnisse des königreichs sachsen. *Zeitschrift des Statistischen Bureaus des Königlich Sächsischen Ministeriums des Innern* 8: 1–54.
- Engle, Robert F. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica* 50: 987–1007. [CrossRef]
- Firpo, Sergio, Antonio F. Galvao, and Suyong Song. 2017. Measurement errors in quantile regression models. *Journal of Econometrics* 198: 146–64. [CrossRef]
- Hahn, Jinyong, and Geert Ridder. 2013. Asymptotic variance of semiparametric estimators with generated regressors. *Econometrica* 81: 315–40.
- Hendricks, Wallace, and Roger Koenker. 1991. Hierarchical spline models for conditional quantiles and the demand for electricity. *Journal of the American Statistical Association* 87: 58–68. [CrossRef]
- Koenker, Roger, and Gilbert Bassett. 1978. Regression quantiles. *Econometrica* 46: 33–50. [CrossRef]
- Koenker, Roger, and Gilbert Bassett. 1982a. Robust tests for heteroscedasticity based on regression quantiles. *Econometrica* 50: 43–61. [CrossRef]
- Koenker, Roger, and Gilbert Bassett. 1982b. Tests of linear hypotheses and  $l_1$  estimation. *Econometrica* 50: 1577–84. [CrossRef]
- Koenker, Roger, and José A. F. Machado. 1999. Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association* 94: 1296–310. [CrossRef]
- Lee, Sokbae. 2007. Endogeneity in quantile regression models: A control function approach. *Journal of Econometrics* 141: 1131–58. [CrossRef]
- Lewbel, Arthur. 1997. Consumer demand systems and household equivalence scales. *Handbook of Applied Econometrics* 2: 167–201.
- Lewbel, Arthur. 2008. Engel curves. *The New Palgrave Dictionary of Economics* 2: 1–4.
- Ma, Lingjie, and Roger Koenker. 2006. Quantile regression methods for recursive structural equation models. *Journal of Econometrics* 134: 471–506. [CrossRef]
- Mammen, Enno, Christoph Rothe, and Melanie Schienle. 2012. Nonparametric regression with nonparametrically generated covariates. *Annals of Statistics* 40: 1132–70. [CrossRef]

- Murphy, Kevin M., and Robert H. Topel. 2002. Estimation and inference in two-step econometric models. *Journal of Business & Economic Statistics* 20: 88–97.
- Pagan, Adrian. 1984. Econometric issues in the analysis of regressions with generated regressors. *International Economic Review* 25: 221–47. [[CrossRef](#)]
- Powell, James L. 1991. Estimation of monotonic regression models under quantile regressions. In *Nonparametric and Semiparametric Models in Econometrics*. Edited by William A. Barnett, James Powell and George E. Tauchen. Cambridge: Cambridge University Press.
- Wang, Huixia Judy, Leonard A. Stefanski, and Zhongyi Zhu. 2012. Corrected-loss estimation for quantile regression with covariate measurement errors. *Biometrika* 99: 405–21. [[CrossRef](#)]
- Wei, Ying, and Raymond J. Carroll. 2009. Quantile regression with measurement error. *Journal of the American Statistical Association* 104: 1129–43. [[CrossRef](#)] [[PubMed](#)]
- Working, Holbrook. 1943. Statistical laws of family expenditure. *Journal of the American Statistical Association* 38: 43–56. [[CrossRef](#)]
- Xiao, Zhijie, and Roger Koenker. 2009. Conditional quantile estimation for generalized autoregressive conditional heteroscedasticity models. *Journal of the American Statistical Association* 104: 1696–712. [[CrossRef](#)]