



Published in final edited form as:

Biometrika. 2012 ; 99(2): 423–438. doi:10.1093/biomet/ass007.

Multiple imputation in quantile regression

Ying Wei,

Department of Biostatistics, Columbia University, 722 West 168th St., New York, New York 10032, U.S.A

Yanyuan Ma, and

Department of Statistics, Texas A&M University, College Station, Texas 77843-3143, U.S.A

Raymond J. Carroll

Department of Statistics, Texas A&M University, College Station, Texas 77843-3143, U.S.A

Ying Wei: ying.wei@columbia.edu; Yanyuan Ma: ma@stat.tamu.edu; Raymond J. Carroll: carroll@stat.tamu.edu

Summary

We propose a multiple imputation estimator for parameter estimation in a quantile regression model when some covariates are missing at random. The estimation procedure fully utilizes the entire dataset to achieve increased efficiency, and the resulting coefficient estimators are root- n consistent and asymptotically normal. To protect against possible model misspecification, we further propose a shrinkage estimator, which automatically adjusts for possible bias. The finite sample performance of our estimator is investigated in a simulation study. Finally, we apply our methodology to part of the Eating at American's Table Study data, investigating the association between two measures of dietary intake.

Keywords

Missing data; Multiple imputation; Quantile regression; Regression quantile; Shrinkage estimation

1. Introduction

In many regression-type applications, some observations are missing. Ignoring the missing data will undermine study efficiency, and sometimes introduce substantial bias. There is a large literature dealing with missing data; see Little & Rubin (1987) for an early and still fundamental treatment. Quantile regression (Koenker & Bassett, 1978) has been an increasingly important modelling tool, due to its flexibility in exploring how covariates affect the distribution of the response. However, combining quantile regression with missing data is not a well-developed topic. In this paper, we consider a linear quantile regression model, where for $\tau \in (0, 1)$,

© 2012 Biometrika Trust

Supplementary material

Supplementary material available at *Biometrika* online includes additional simulation results with higher rates of missing data and stronger and weaker covariate correlations. Also included are the quantile-quantile plots for the transformed covariates in the data analysis.

$$Q_{\tau}(y) = x^T \beta_{1,\tau} + z^T \beta_{2,\tau}. \quad (1)$$

Here (x, z) are both covariate vectors, but x may be missing, while z is always observed. We assume that z contains the constant 1, so the intercept term is not written out separately. We use n for the total sample size, and assume that n_1 of these n observations are complete, while the remaining n_0 of them have x missing. Thus, observations can be summarized as $\{(y_i, x_i, z_i) : i = 1, \dots, n_1\}$ and $\{(y_j, \cdot, z_j) : j = n_1 + 1, \dots, n\}$. To avoid trivial situations, we assume $0 < \lim_{n \rightarrow \infty} n_0/n_1 = \lambda < \infty$. We make a missing at random assumption that conditional on z , missingness and x are independent. The main interest of this paper is in estimating the regression parameter $\beta_{\tau} = (\beta_{1,\tau}^T, \beta_{2,\tau}^T)^T$ given the assumed missing data mechanism. This research is motivated by the Eating at American's Table Study (Subar et al., 2001), an important study in nutritional epidemiology. In § 5, we describe how this study fits our model framework.

It is not difficult to see that since missingness depends only upon the observed covariates z , using the complete data only yields a consistent estimate of β_{τ} . However, since a part of the data is completely excluded from the analysis, this practice can be highly inefficient. The main goal of this paper is to propose a multiple imputation method to include the incomplete data, so as to improve estimation efficiency. Since additional assumptions on (x, z) are needed to facilitate the imputation procedure, the method risks being inconsistent and we propose a shrinkage estimator to attenuate this risk. The final estimator has an automatic data-driven shrinkage parameter, which guarantees that the resulting estimator is consistent regardless of the correctness of the additional assumptions, and at the same time is more efficient than using the complete data only.

Most existing methods handling missing data are likelihood-based, and hence cannot be applied to quantile regression directly, since there is no likelihood function for quantile regression. Lipsitz et al. (1997) considered an inverse probability approach for longitudinal data with drop-outs. For the same type of data, Yi & He (2009) extended the inverse probability weighted generalized estimating equations proposed by Robins et al. (1995) to correct for the bias from longitudinal drop-out. Our setting is different from those methods, since we are dealing with missing covariates, rather than missing outcomes.

Throughout the paper, we write $Q_{\tau}(y)$ as the τ th quantile of a random variable y . We write $\beta(\tau)$ as the quantile coefficient process for $\tau \in (0, 1)$, and β_{τ} as the quantile coefficient specifically at the τ th quantile. In addition, we use $\|x\|$ to mean Euclidean norm, and write $g'(x)$ as the first derivative of an arbitrary function $g(x)$. If x and y are two random variables, then $E_{(x,y)}\{g(x, y)\}$ stands for the expectation of $g(x, y)$ over the joint distribution of (x, y) .

2. Estimation with multiple imputation

2.1. Method

In this section, we propose a multiple imputation estimator of the quantile coefficient

$\beta_{\tau} = (\beta_{1,\tau}^T, \beta_{2,\tau}^T)^T$ in the linear quantile model (1). The method has the following steps.

Step 1. Perform quantile regression with the complete data only. Run a quantile regression using the complete data only and write the resulting coefficients as $\hat{\beta}_\tau$. That is, for a set of τ values in $(0, 1)$, obtain $\hat{\beta}_\tau = \arg \min_{\beta} \sum_{i=1}^{n_1} \rho_\tau\{y_i - (x_i^T, z_i^T)\beta\}$, where $\rho_\tau(r) = r\{\tau - I(r < 0)\}$ is an asymmetric L_1 loss function. In practice, τ is typically chosen to be evenly spread and sufficiently dense grid points on $(0, 1)$.

Step 2. Impute the missing x based on $f(x | y, z)$. The main challenge is to estimate the conditional density of $f(x | y, z)$. The density $f(x | y, z) \propto f(y | x, z)f(x | z)$, so it can be determined uniquely from the two densities $f(y | x, z)$ and $f(x | z)$.

Step 2a: Estimate the conditional density $f(y | x, z)$. Under the assumption that the linear quantile model (1) holds for all quantile levels τ , we can write the conditional density $f(y | x, z)$ as a function of the quantile coefficient process, that is, $f\{y | x, z; \beta_0(\tau)\} = F'\{y | x, z; \beta_0(\tau)\}$, where $F\{y | x, z; \beta_0(\tau)\} = \inf\{\tau \in (0, 1) : (x^T, z^T)\beta_0(\tau) > y\}$ and $\beta_0(\tau)$ is the true quantile coefficient process. We write the conditional density $f(y | x, z)$ as $f\{y | x, z; \beta_0(\tau)\}$ to indicate its dependence on the quantile coefficient function $\beta_0(\tau)$.

Although the unknown coefficient function $\beta_0(\tau)$ is of infinite dimension, it can be well-approximated by a natural linear spline expanding from a series of estimated $\hat{\beta}_{\tau_k}$ at a fine grid of quantile levels (τ_k) . Specifically, we choose quantile levels $\tau_k = k/(K_n + 1)$ ($k = 1, \dots, K_n$), where K_n is the number of quantile levels. We then define $\hat{\beta}(\tau)$ as a p -dimensional piecewise linear function on $[0, 1]$, which satisfies $\hat{\beta}(\tau_k) = \hat{\beta}_{\tau_k}$ and $\hat{\beta}'(0) = \hat{\beta}'(1) = 0$. Under the conditions in Wei & Carroll (2009), $\hat{\beta}(\tau)$ converges uniformly to the true quantile coefficient process in probability. The quantile function is the inverse distribution function, so the density function can be expressed as the reciprocal of the first derivative of the quantile function at the corresponding quantile level. Consequently, we can approximate the conditional density function by

$$\hat{f}\{y|x, z, \hat{\beta}(\tau)\} = \sum_{k=1}^{K_n} \frac{\tau_{k+1} - \tau_k}{(x^T, z^T)\hat{\beta}_{\tau_{k+1}} - (x^T, z^T)\hat{\beta}_{\tau_k}} I\{(x^T, z^T)\hat{\beta}_{\tau_k} \leq y < (x^T, z^T)\hat{\beta}_{\tau_{k+1}}\}.$$

Here $f\{y | x, z, \hat{\beta}(\tau)\}$ is the previously defined density function that is induced from the estimated conditional quantile function $(x^T, z^T)\hat{\beta}(\tau)$.

Step 2b: Estimate the conditional density $f(x | z)$. The remaining problem is to estimate $f(x | z)$. We model x given z parametrically as $f(x | z, \eta)$. The missing-at-random assumption facilitates the estimation of η based on the complete data. We write the estimate as $\hat{\eta}$, and the estimated conditional density of x given z as $f(x | z, \hat{\eta})$.

Step 2c: Estimate the conditional density $f(x | y, z)$ and impute the missing x accordingly. The estimated conditional density function is $f(\hat{x} | y_j, z_j) \propto f\{y_j | x, z_j, \hat{\beta}(\tau)\} f(x | z_j, \hat{\eta})$. For each $j = n_1 + 1, \dots, n$, we simulate the missing x_j from $f(\hat{x} | y_j, z_j)$ by randomly drawing a $Un(0, 1)$ random variable, and inserting it into the quantile function $F^{-1}(u | y_j, z_j)$, for $u \in (0, 1)$ that is derived from the estimated $f(\hat{x} | y_j, z_j)$. Let u_ℓ be the ℓ th generated $Un(0, 1)$ random

variable. We then define $x_{j(\ell)} = F^{-1}(u_\ell | y_j, z_j)$, the ℓ th imputed x associated with (y_j, z_j) . Consequently, $x_{j(\ell)} \sim f(\hat{x} | y_j, z_j)$.

Step 3. Re-estimate β including the imputed data. We assemble a new objective function including the completely observed data and the ℓ th imputed dataset as

$$S_{n(\ell)}(\beta) = \sum_{i=1}^{n_1} \rho_\tau\{y_i - (x_i^T, z_i^T)\beta\} + \sum_{j=n_1+1}^n \rho_\tau\{y_j - (\tilde{x}_{j(\ell)}^T, z_j^T)\beta\},$$

and define $\hat{\beta}_{*(\ell)} = \operatorname{argmin}_\beta S_{n(\ell)}(\beta)$ as the estimated coefficient using the ℓ th assembled complete data. We repeat this imputation-estimation step m times, and the multiple

imputation estimator is $\tilde{\beta}_\tau = m^{-1} \sum_{\ell=1}^m \hat{\beta}_{*(\ell)}$.

2.2. Large-sample properties of the multiple imputation estimator

In this section, we establish the consistency and asymptotic normality of the multiple imputation estimator $\tilde{\beta}_\tau$. Let $\delta = 0$ when x is missing and $\delta = 1$ otherwise. We first reiterate the assumption on the missingness mechanism.

Assumption 1. For all z , $\operatorname{pr}(\delta = 1 | x, y, z) = \operatorname{pr}(\delta = 1 | z) > 0$.

Assumption 1 ensures that, conditioning on z , the event that x is missing is independent of x and the response y . We then introduce two identifiability conditions.

Assumption 2. There exists a $\beta_{0,\tau} \in \mathbb{R}^p$ such that $\beta_{0,\tau}$ uniquely minimizes the objective function $S_0(\beta) = E_{(y,x,z)}[\rho_\tau\{y - (x^T, z^T)\beta\}]$.

Define $\tilde{S}_0(\beta) = E_{(y,x,z)}[\rho_\tau\{y - (x^T, z^T)\beta\}]$, where, given (y, z) , x follows the conditional distribution $f(\hat{x} | y, z)$. Since \hat{x} is estimated from completely observed data, this expectation is also conditional on the n_1 completely observed data. We then make the following assumptions.

Assumption 3. There exists a compact set $\Omega \in \mathbb{R}^p$, and $\beta_\tau^* \in \Omega$, such that

$$\beta_\tau^* = \operatorname{arg\,min}_\beta \tilde{S}_0(\beta).$$

Assumption 4. The covariate x has bounded support \mathcal{X} . The true conditional density $f(x | z) = f(x | z, \eta = \eta_0)$, where $f(x | z, \eta)$ is a continuous function of η uniformly for (x, z) in a neighbourhood of η_0 and is bounded away from zero and infinity for all (x, z) .

Recall that for any x and z , $(x^T, z^T)\beta_0(\tau)$ defines the conditional quantile function of y given x and z . We further define a functional $h(\tau; x, z) = 1/\{(x^T, z^T)\beta_0'(\tau)\}$, which is the density of y given x and z at the τ th quantile. We call this the conditional quantile density function. Its reciprocal is known as the sparsity function (Welsh, 1988; Koenker & Xiao, 2004). With these definitions, we now introduce the smoothness conditions on $\beta_0(\tau)$.

Assumption 5. The true coefficient functions $\beta_0(\tau)$ are smooth functions on $(0, 1)$, and for any $x \in \mathcal{X}$ and z ,

- i. $0 < h(\tau; x, z) < \infty$, and $\lim_{\tau \rightarrow 0} h(\tau; x, z) = \lim_{\tau \rightarrow 1} h(\tau; x, z) = 0$;
- ii. there exist constants M and $\nu_1, \nu_2 > -1$ such that the first derivative of $h(\cdot)$ satisfies

$$\sup_x |h'(\tau; x, z)| < M\tau^{\nu_1}(1 - \tau)^{\nu_2}. \quad (2)$$

Assumption 5 is similar to Assumption 3 in Wei & Carroll (2009). Assumption 5(i) implies that the conditional density $f(y | x, z)$ is continuous, bounded away from zero and infinity and diminishes to zero as τ converges to 0 and 1, while Assumption 5(ii) is on the tail behaviour of $f(y | x, z)$, since $h'(\tau; x, z)$ determines how smoothly the density function diminishes as the quantile level converges to 0 or 1. Smaller ν_1 and ν_2 indicate heavier tails of the conditional distribution of y given x and z . Assumption 5(ii) covers a wide range of distributions, such as the exponential, Gaussian and the Student t -distributions. Assumption 5, together with Assumptions 2 and 4, ensures the uniform convergence of $\hat{\beta}(\tau)$ over the intervals $[1/(k_n + 1), k_n/(k_n + 1)]$, which in turn ensures consistent estimation of $f(y | x, z)$.

Assumption 6. The matrix $\psi_\tau = (\partial/\partial\beta_{0,\tau})E[\varphi_\tau\{y_i - (x_i^T, z_i^T)\beta_{0,\tau}\}(x_i^T, z_i^T)^T]$ is positive definite, where $\varphi_\tau(r) = \tau - I\{r < 0\}$.

In addition, we also make the definitions

$$\begin{aligned} V_1 &= \text{var}[\varphi_\tau\{y_i - (x_i^T, z_i^T)\beta_{0,\tau}\}(x_i^T, z_i^T)^T], \\ V_0 &= \lim_{n \rightarrow \infty} \text{var}[\varphi_\tau\{y_j - (\tilde{x}_{j(\ell)}^T, z_j^T)\beta_{0,\tau}\}(\tilde{x}_{j(\ell)}^T, z_j^T)^T], \\ U_0 &= \lim_{n \rightarrow \infty} \text{cov}[\varphi_\tau\{y_j - (\tilde{x}_{j(\ell)}^T, z_j^T)\beta_{0,\tau}\}(\tilde{x}_{j(\ell)}^T, z_j^T)^T, \varphi_\tau\{y_j - (\tilde{x}_{j(\ell)}^T, z_j^T)\beta_{0,\tau}\}(\tilde{x}_{j(\ell)}^T, z_j^T)^T]. \end{aligned}$$

With these assumptions and notation, we now present the asymptotic behaviour of $\tilde{\beta}_\tau$. Recall that $0 < \lim_{n \rightarrow \infty} n_0/n_1 = \lambda < \infty$.

Theorem 1. Under Assumptions 1–6, for $K_n \rightarrow \infty$ and $K_n n^{-1} \rightarrow 0$, the multiple imputation estimator $n^{1/2}(\tilde{\beta}_\tau - \beta_{0,\tau}) \rightarrow N(0, \psi_\tau^{-1} \Sigma \psi_\tau^{-1})$ in distribution, where $\Sigma = (\lambda + 1)^{-1} V_1 + (1 + 1/\lambda)^{-1} [m^{-1} V_0 + \{(m - 1)/m\} U_0]$.

The proof of Theorem 1 is provided in Appendix A1, while estimates of ψ_τ and Σ are provided in Appendix A2.

Remark 1. Throughout, we use the phrase complete-data analysis to mean an analysis based only on the completely observed data. The asymptotic variance of the estimator using the completely observed data only is $n_1^{-1} \psi_\tau^{-1} V_1 \psi_\tau^{-1}$. Comparing with the estimation variance $n^{-1} \psi_\tau^{-1} \Sigma \psi_\tau^{-1}$ of the imputed estimator, we see two sources of difference. First, the multiple imputation estimator has an effective sample size n , larger than that for the complete-data analysis, which helps to improve its efficiency. Second, the multiple imputation estimator has additional sources of variability, including the sampling variability from multiple imputation, the inherited variability from using the complete-data estimated parameters and their correlations. Hence, the multiple imputation estimators might be less efficient than the

complete-data estimator. Such phenomena are common for multiple imputation estimators; see Tsiatis (2006, Ch. 14). In practice, one could assess the variabilities of both estimators to decide which to use; see Appendix A2.

3. Shrinkage estimation

The estimator $\hat{\beta}_\tau$ using the complete data only is consistent, but has a potential loss of efficiency. The multiple imputation estimator $\tilde{\beta}_\tau$ is generally more efficient, as will be demonstrated via simulations in § 4. However, imputation may cause bias when the parametric likelihood for x given z is misspecified. There are many ways to balance the two estimators, including test-pretest estimation after testing for the parametric model, but a simple and general strategy that we adopt is a shrinkage estimator, as follows. Let $\theta_\tau = \hat{\beta}_\tau - \tilde{\beta}_\tau$ be the componentwise differences of the multiple imputation and complete-data estimators, respectively, with elements $(\theta_{1,\tau}, \dots, \theta_{p,\tau})^T$. Let V be the covariance matrix of θ_τ with diagonal elements (v_{11}, \dots, v_{pp}) . Then Chen et al. (2009) suggest the estimator

$$\hat{\beta}_\tau^{(s)} = \hat{\beta}_\tau + K(\tilde{\beta}_\tau - \hat{\beta}_\tau), \quad (3)$$

where K is a diagonal matrix with j th diagonal element $= v_{jj} / (v_{jj} + \hat{\theta}_{j,\tau}^2)$. Recall that the asymptotic variances v_{jj} ($j = 1, \dots, p$) are quantities of order n^{-1} . The idea behind this method is that if there is no bias, then $\hat{\theta}_{j,\tau}^2 = O_p(n^{-1})$ and the shrinkage factor K is between 0 and I , so that the multiple imputation estimator and the complete-data estimator both receive weight, although emphasis is on the former. Conversely, if there is a bias, then $\hat{\theta}_{j,\tau}^2 = O(1)$, and the elements of $K \rightarrow 0$, so that the complete-data estimator asymptotically has weight 1.

Details of implementing the shrinkage estimator are given in Appendix A2. In Appendix A1, we show that the complete-data estimator and the multiple imputation estimator have linear expansions, based on which we outline in Appendix A2 estimation of the joint covariance matrix of $(\hat{\beta}_\tau, \tilde{\beta}_\tau)$. The results enable us to estimate V easily and also mean that the formulae in Chen et al. (2009) are applicable, so that we can construct an estimator of $\text{cov}(\hat{\beta}_\tau^{(s)})$. The general theory for such shrinkage estimators is given by Chen et al. (2009), although constructing the estimate of Σ is nontrivial because of our context.

4. Simulations

Here we investigate the performance of our multiple imputation estimator $\tilde{\beta}_\tau$ and shrinkage estimator $\hat{\beta}_\tau^{(s)}$ based on Monte-Carlo simulations. We first consider two models.

$$y_i = 1 + x_i + z_i + e_{i2}, \quad (4)$$

$$y_i = 1 + x_i + z_i + (0.5x_i + 0.5z_i)e_{i1}, \quad (5)$$

where the errors e_{i1} and e_{i2} are independent and standard normal, and the covariates (x_i, z_i) are jointly normal with mean vector $(4, 4)^T$, variances $(1, 1)^T$ and correlation 0.5. In model

(4), the true intercept at the τ th quantile is $1 + Q_\tau(z)$, where z is a random variable with a standard normal distribution, and both coefficients associated with x_i and z_i equal 1 at every quantile. In model (5), the true intercept equals 1 at every quantile level, but the two slope coefficients vary across the quantiles, both equal to $1 + 0.5Q_\tau(z)$ at quantile level τ . In both models, we further assume that x_i is missing with probability $\text{pr}(x_i \text{ is missing} | z_i) = \max[0, \{(z_i - 3)/10\}^{1/20}]$, which results in approximately 25% missing x_i 's. We then apply the multiple imputation estimation and shrinkage estimation procedures to the simulated data from the two models above. In both settings, the density $f(x | z)$ is estimated by maximum likelihood estimation correctly assuming a joint normal distribution. When the covariates x and z are negative, there is an identifiability issue in model (5) since the distribution of e_{i1} is symmetric around 0. To avoid this trivial situation, we only kept the pairs (x, z) satisfying $x + z > 0$ in model (5). Because the probability of $x + z < 0$ is very small, the resulting true joint probability density function of $(x; z)$ is very close to the joint normal distribution which we used in the imputation procedure. We choose $m = 10$ in the multiple imputation estimation algorithm. The sample size was $n = n_0 + n_1 = 200$. The shrinkage factor is estimated following Appendix A2.

Table 1 displays the means and the standard errors of the estimated quantile coefficients in models (4) and (5) from 500 simulations at $\tau = 0.1, 0.5$ and 0.9 , using the three estimation approaches. The upper half of Table 1 displays the coefficients from model (4), while the bottom half shows those from model (5). All three methods are nearly unbiased. However, as expected from the theory, the variances of the multiple imputation estimators are smaller than the complete-data estimators, especially in the coefficient associated with z_i . Such efficiency improvement is more evident for the heteroscedastic model (2). For example, for estimating the z_i slope at the 0.9th quantile, the relative efficiency of multiple imputation estimation compared with using the complete data only, i.e., the ratio of their variances, is 217%, and that of shrinkage estimation is 149%. To investigate the performance of our methods in various model settings, we also allowed higher missing proportions, and weaker or stronger correlation between the covariates x and z . The resulting estimated coefficients and their standard errors are included in the Supplementary Material. On the basis of those tables, the proposed estimators performed well across various model specifications.

The results in Table 1 are obtained when $f(x | z)$ is estimated from the correct model. To investigate the potential bias that could be induced from misspecified $f(x | z)$, we simulate covariates (x_i, z_i) as $x_i = (0.18u_{i,1} + 0.68u_{i,2}) + 3.14$, and $z_i = (0.68u_{i,1} + 0.18u_{i,2}) + 3.14$, where $u_{i,1}$ and $u_{i,2}$ are two independent χ_1^2 random variables. We choose the constants, 0.18, 0.68 and 3.14, such that (x_i, z_i) have mean 4, variance 1 and correlation of approximately 0.5, as in the earlier simulation. After simulating the nonnormally distributed covariates, we then generate the responses from model (2). For each generated sample, we allow x_i to be missing completely at random with probability 0.25. We apply the same estimation procedures as above, pretending that (x_i, z_i) is jointly normal. Table 2 presents the mean squared errors and standard errors for the resulting estimated coefficients at $\tau = 0.1, 0.5$ and 0.9 . As a comparison, we also re-estimate the coefficients using the imputation method, but use the exact density $f(x | z)$ in the algorithm. On the basis of Table 2, the mean squared errors from the multiple imputation estimators with the exact $f(x | z)$ are the smallest. As

expected, when $f(x | z)$ is misspecified, the mean squared errors are inflated, and the shrinkage estimates have smaller mean squared errors due to the bias correction. Since the complete-data approach only uses part of the data for estimation, its mean squared errors are even larger than the multiple imputation estimator with misspecified $f(x | z)$. Finally, the difference between the multiple imputation estimators using exact and misspecified densities are small relative to their standard errors, indicating that the multiple imputation estimator is also fairly robust against the misspecification of $f(x | z)$.

5. Application

We illustrate the performance of our methods using part of the Eating at American's Table Study (Subar et al., 2001). The dataset consists of 1418 subjects who participated in this study from September 1997 to August 1998. They were required to complete a 24-hour recall on their dietary intakes, and they also completed a dietary history questionnaire. It is commonly thought that the 24-hour recall is an unbiased measure of dietary intake, but is expensive in cohort studies because it must be administered multiple times, and thus costs far more than the dietary history questionnaire. In measurement error modelling of diet and disease, the regression calibration method (Carroll et al., 2006) is to regress the 24-hour recall on the dietary history questionnaire. Since the distributions of nutrition intakes are commonly skewed, quantile regression is a desirable tool for this modelling.

Here we model carbohydrate intake, with y_i being the 24-hour recall for the i th person, x_{i1} the dietary history questionnaire measurement, x_{i2} body mass index, x_{i3} the participant's age, x_{i4} an indicator of Caucasian ethnic status and x_{i5} the gender. The model can be written as

$$y_i = \beta_{0,\tau} + \beta_{1,\tau} x_{i,1} + \beta_{2,\tau} x_{i,2} + \beta_{3,\tau} x_{i,3} + \beta_{4,\tau} x_{i,4} + \beta_{5,\tau} x_{i,5} + e_i. \quad (6)$$

There are 453 randomly selected subjects among the 1418 who do not have measurements of body mass index and did not complete the dietary history questionnaire, because the study was a designed experiment with some participants randomly assigned to complete an alternative questionnaire. Therefore, those covariates are missing completely at random. Here we apply our multiple imputation estimation methodology to obtain the estimate of the β_s , with x as the carbohydrate intake in the dietary history questionnaire and body mass index, and z as gender, ethnicity and age.

In these data, we found that the carbohydrate intake measured in the dietary history questionnaire and body mass index are essentially uncorrelated, with partial correlation 0.0084 conditional on the subject's age and gender. We can thus estimate the conditional density of carbohydrate intakes in the dietary history questionnaire and body mass index separately based on the two Box–Cox transformation models

$$\begin{aligned} \Lambda(x_{i1}, \lambda_1) &= \gamma_{10} + \gamma_{11} x_{i3} + \gamma_{12} x_{i4} + \gamma_{13} x_{i5} + e_{i1}, \quad e_{i1} \sim N(0, \sigma_1^2), \\ \Lambda(x_{i2}, \lambda_2) &= \gamma_{20} + \gamma_{21} x_{i3} + \gamma_{22} x_{i4} + \gamma_{23} x_{i5} + e_{i2}, \quad e_{i2} \sim N(0, \sigma_2^2). \end{aligned}$$

Here $\Lambda(u, \lambda)$ is the Box–Cox transformation function, i.e., $\Lambda(u, \lambda) = \log(u)$ if $\lambda = 0$, and $\Lambda(u, \lambda) = (u^\lambda - 1)/\lambda$ for $\lambda \neq 0$. We used maximum likelihood estimates of the transformation parameters, these being close to 0 and -1 , respectively, which suggests that logarithm and reciprocal transformations are needed for carbohydrate intake in the dietary history questionnaire and body mass index, respectively. In the Supplementary Material, we present the quantile-quantile plot of the residuals from the above two models with their respective best fitted powers, which shows that the transformed variables are approximately normally distributed.

On the basis of the estimated models, the conditional density of the untransformed carbohydrate intake in the dietary history questionnaire is $f_c(v) = (v\hat{\sigma}_1)^{-1} \phi[\{\log(v) - \hat{\gamma}_{10} - \hat{\gamma}_{11}x_3 - \hat{\gamma}_{12}x_4 - \hat{\gamma}_{13}x_5\}/\hat{\sigma}_1]$, where ϕ is the density function of standard normal. The conditional density of body mass index is $f_b(v) = (v^2\hat{\sigma}_2)^{-1} \phi[\{1/v - \hat{\gamma}_{20} - \hat{\gamma}_{21}x_3 - \hat{\gamma}_{22}x_4 - \hat{\gamma}_{23}x_5\}/\hat{\sigma}_2]$.

Following our multiple imputation algorithm, we estimated model (6) at 50 evenly spaced quantile levels using the completely observed data only in the first step. On the basis of the resulting quantile coefficient process, and the estimated conditional densities $f(x|z)$ using the models above, we imputed the missing carbohydrate intakes and body mass index $m = 10$ times. In Table 3, we listed the multiple imputation estimators at $\tau = 0.1, 0.5$ and 0.9 , as well as their standard errors. To illustrate the improved efficiency from multiple imputation, we calculated the relative efficiency. In addition, we also constructed the shrinkage estimator following (3). The shrinkage factors are estimated following Appendix A2.

Table 3 shows that the multiple imputation estimators are fairly consistent with those using the complete data only, but have much smaller standard errors for the estimates associated with age, ethnicity and gender. Those variables are completely observed when the dietary history questionnaire carbohydrate intakes and body mass index are missing. The multiple imputation estimators make full use of those observations, which improves their efficiency. The shrinkage estimator is generally consistent with the complete-data and multiple imputation estimators; while its standard errors are slightly larger than the multiple imputation estimators, they are still much smaller than those of the complete-data estimators.

6. Discussion

The validity of our multiple imputation method relies on a correct specification of the conditional density $f(x|z)$, which we model parametrically. To further protect against the possible misspecification of $f(x|z)$, a shrinkage estimator was proposed. One could also opt to estimate $f(x|z)$ nonparametrically, which will automatically yield a consistent estimator without an additional shrinkage step. However, nonparametric conditional density estimation is very complex, especially when z is multivariate, and the slow rates of convergence would undermine the usefulness of such an approach.

The missing covariate problem in the quantile regression context is challenging, because the conditional density of y given the covariates is unspecified under a typical quantile

regression setting. Consequently, classical likelihood-based approaches cannot be applied directly. Here, we adopted a joint modelling approach similar to Wei & Carroll (2009) to circumvent this difficulty. However, the proposed method is different from Wei & Carroll (2009) in many aspects. First, the objectives are different. This paper handles missing covariates, while Wei & Carroll (2009) handle mismeasured covariates. Second, the estimation approaches are different. Wei & Carroll (2009) is based on constructing unbiased estimating equations; while this paper uses a multiple imputation approach. Consequently, the estimation algorithms are different; the former involves iterative estimation, while the estimation procedure in this paper does not. Finally, the asymptotic properties are obtained in a very different fashion.

We assumed the conditional quantile functions to be linear at all quantile levels. This assumption holds for location-scale models, i.e., $Y = X^T\beta + X^T\gamma e$, where e is a random error with $Q_\tau(e | X) = 0$. If needed, one can easily relax the linear quantile function to an arbitrary nonlinear or even nonparametric function. The algorithm remains largely unchanged, with the minimal adaptation of setting the linear function to be the new regression function in the check function ρ_τ . Although the method is presented for an independent sample, it can also be extended to longitudinal data using the so-called working independence construction. For a longitudinal sample $(y_{i,j}, x_{i,j}, z_{i,j})$, if the quantiles of $y_{i,j}$ is linear in $(x_{i,j}, z_{i,j})$, then we can estimate the quantile coefficients using a similar algorithm with the longitudinal quantile regression objective function $\sum_i \sum_j \rho_\tau(y_{i,j} - x_{i,j}^T\beta - z_{i,j}^T\gamma)$. The estimation of the conditional density $f(x | z)$ also needs to be adapted for the longitudinal data. The resulting estimators would still be consistent, but the limiting distribution would need to be derived separately.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Wei's research was supported by the National Science Foundation (DMS-0906568) and a career award from NIEHS Center for Environmental Health in Northern Manhattan (ES009089). Ma's research was supported by a grant from the National Science Foundation (DMS-0906341). Carroll's research was supported by a grant from the National Cancer Institute (CA57030).

Appendix

A1. Technical arguments

Recall that $\tilde{x}_{j(\ell)}$ is the ℓ th imputed x associated with (y_j, z_j) , based on the estimated density $\hat{f}(x | y_j, z_j)$. We define a partial objective function with the imputed proportion of the data

$$\tilde{S}_{n_0}^{(\ell)}(\beta) = \sum_{j=n_1+1}^n \rho_\tau\{y_j - (\tilde{x}_{j(\ell)}^T, z_j^T)\beta\},$$

and define its minimizer $\hat{\beta}_{0,(\ell)} = \arg \min_{\beta} \tilde{S}_{n_0}^{(\ell)}(\beta)$.

We say that $\hat{\beta}_{0,(\ell)}$ is the estimated coefficient using the ℓ th imputed portion of the data only. In later steps, we show that the multiple imputation estimator $\hat{\beta}_{\tau}$ can be written as a linear combination of $\hat{\beta}_{\tau}$ and $\hat{\beta}_{0,(\ell)}$ s. Hence, to find the asymptotic distribution of $\hat{\beta}_{\tau}$, a key step is to find the asymptotic distribution of $\hat{\beta}_{0,(\ell)}$ as $n = n_0 + n_1 \rightarrow \infty$, and $0 < \lim_n n_0/n_1 = \lambda < \infty$. To do that, we first show that

$$\sup_{\beta \in \Omega} |\tilde{S}_0(\beta) - S_0(\beta)| \rightarrow 0 \quad (A1)$$

in probability as $n_1 \rightarrow \infty$. Here $\tilde{S}_0(\beta)$ and $S_0(\beta)$ are the two expected objective functions defined before Assumptions 2 and 3.

Recall that $f\{\hat{y} | x, z, \hat{\beta}(\tau)\}$ is the estimated conditional density of y given x and z using the complete data only. We first decompose the difference between the estimated density $f\{\hat{y} | x, z, \hat{\beta}(\tau)\}$ and its true value as

$$\begin{aligned} & \sup_x |f\{\hat{y} | x, z, \hat{\beta}(\tau)\} \\ & \quad - f\{y | x, z, \beta_0(\tau)\} \\ & = \sup_x \left| \sum_{k=1}^{K_n} [f\{\hat{y} | x, z, \hat{\beta}(\tau)\} - f\{y | x, z, \beta_0(\tau)\}] I\{(x^T, z^T) \hat{\beta}_{\tau_k} \leq y < (x^T, z^T) \hat{\beta}_{\tau_{k+1}}\} - f\{y | x, z, \beta_0(\tau)\} I\{y < (x^T, z^T) \hat{\beta}_{\tau_1}\} \right. \\ & \quad \left. - f\{y | x, z, \beta_0(\tau)\} I\{y > (x^T, z^T) \hat{\beta}_{\tau_{K_n}}\} \right| \leq \sup_x A_1 \\ & + \sup_x h(\tau_1, x, z) \\ & + \sup_x h(\tau_{K_n}, x, z), \end{aligned}$$

where

$$\sup_x A_1 = \sup_x \sum_{k=1}^{K_n} |f\{\hat{y} | x, z, \hat{\beta}(\tau)\} - f\{y | x, z, \beta_0(\tau)\}| I\{(x^T, z^T) \hat{\beta}_{\tau_k} \leq y < (x^T, z^T) \hat{\beta}_{\tau_{k+1}}\}.$$

Following the definition of $f\{\hat{y} | x, z, \hat{\beta}(\tau)\}$, and since for any given value of y , it can only be contained in one of those subintervals $\{(x^T, z^T) \hat{\beta}_{\tau_k}, (x^T, z^T) \hat{\beta}_{\tau_{k+1}}\}$, we have

$$\sup_x A_1 \leq \sup_x \max_k \left| \frac{\tau_{k+1} - \tau_k}{(x^T, z^T)(\hat{\beta}_{\tau_{k+1}} - \hat{\beta}_{\tau_k})} - f\{y | x, z, \beta_0(\tau)\} \right| I\{(x^T, z^T) \hat{\beta}_{\tau_k} \leq y < (x^T, z^T) \hat{\beta}_{\tau_{k+1}}\}.$$

Following the uniform convergence of $\hat{\beta}_{\tau}$, readily available from the result in Wei & Carroll (2009) by considering in their context a special case where the measurement error variance is zero, the convergence $(x^T, z^T)(\hat{\beta}_{\tau_k} - \beta_{0,\tau_k}) = o_p(K_n^{1/2} n_1^{-1/2})$ holds uniformly for any k . Consequently, we can rewrite the upper bound as

$$\begin{aligned} \sup_x A_1 &\leq \sup_x \max_k \left| \frac{\tau_{k+1} - \tau_k}{(x^T, z^T)(\beta_{0,\tau_{k+1}} - \beta_{0,\tau_k}) + o_p(K_n^{1/2} n_1^{-1/2})} \right. \\ &\quad \left. - f\{y|x, z, \beta_0(\tau)\} \right| \\ &\times I\{(x^T, z^T)\hat{\beta}_{\tau_k} \leq y < (x^T, z^T)\hat{\beta}_{\tau_{k+1}}\} = \sup_x \max_k \left| \frac{\tau_{k+1} - \tau_k}{(x^T, z^T)(\beta_{0,\tau_{k+1}} - \beta_{0,\tau_k})} - f\{y|x, z, \beta_0(\tau)\} \right. \\ &\quad \left. + o_p(K_n^{1/2} n_1^{-1/2}) \right. \\ &\quad \left. \times I\{(x^T, z^T)\hat{\beta}_{\tau_k} \leq y < (x^T, z^T)\hat{\beta}_{\tau_{k+1}}\} \right|. \end{aligned}$$

By the mean value theorem, there exists a $\tau^* \in (\tau_k, \tau_{k+1})$ such that $(\tau_{k+1} - \tau_k)/\{(x^T, z^T)(\beta_{0,\tau_{k+1}} - \beta_{0,\tau_k})\} = h(\tau^*, x, z)$. On the other hand, let τ_y be the quantile level of y with respect to true quantile function $(x^T, z^T)\beta_0(\tau)$ for $y \in [(x^T, z^T)\hat{\beta}_{\tau_k}, (x^T, z^T)\hat{\beta}_{\tau_{k+1}}]$, then $f\{y|x, z, \beta_0(\tau)\} = h(\tau_y, x, z)$ by definition. Since the true quantile function $(x^T, z^T)\beta_0(\tau)$ is a continuous function that satisfies the Lipschitz condition, the quantile level of $(x^T, z^T)\hat{\beta}_{\tau_{k+1}}$ with respect to the true quantile function is $\tau_{k+1} + o_p(K_n^{1/2} n_1^{-1/2})$. Moreover, due to the uniform convergence of $\hat{\beta}(\tau)$, the quantile level of $(x^T, z^T)\hat{\beta}_{\tau_k}$ is $\tau_k + o_p(K_n^{1/2} n_1^{-1/2})$, for any k . Therefore, together with the monotonicity of quantile function, we have $\tau_k + o_p(K_n^{1/2} n_1^{-1/2}) < \tau_y < \tau_{k+1} + o_p(K_n^{1/2} n_1^{-1/2})$. Following these arguments, we have

$$\begin{aligned} \sup_x A_1 &= \sup_x \max_k |h(\tau^*, x, z) \\ &\quad - h(\tau_y, x, z) \\ &\quad + o_p(n_1^{-1/2})| I\{\tau_k \\ &\quad + o_p(n_1^{-1/2}) \leq \tau^*, \tau_y \leq \tau_{k+1} \\ &\quad + o_p(K_n^{1/2} n_1^{-1/2})\} \leq \sup_x \max_k [|h'(\tau_k, x, z)| \{O(K_n^{-1}) + o_p(K_n^{1/2} n_1^{-1/2})\} + o_p(K_n^{1/2} n_1^{-1/2})] \\ &= O(K_n^{-\nu_1 \Lambda \nu_2 - 1}) \\ &\quad + o_p(K_n^{1/2} n_1^{-1/2}) = o_p(1). \end{aligned}$$

The last step follows from Assumption 5(i) and the fact that $K_n/n \rightarrow 0$. Consequently, for any given values of y and z , as $n_1 \rightarrow \infty$ and $K_n \rightarrow \infty$, we have

$$\sup_x \|\hat{f}\{y|x, z, \hat{\beta}(\tau)\} - f\{y|x, z, \beta_0(\tau)\}\| \leq O_p(K_n^{-\nu_1 \Lambda \nu_2 - 1}) + o_p(n_1^{-1/2}) + \sup_x h(\tau_1, x, z) + \sup_x h(\tau_{K_n}, x, z) = o_p(1). \quad (A2)$$

Let $D_0(y, z) = \int_x f\{y|x, z, \beta_0(\tau)\} f(x) dx$, and $D_{n_1}(y, z) = \int_x \hat{f}\{y|x, z, \hat{\beta}(\tau)\} f(x) dx$. Since $f(x|z)$ is an integrable function, the convergence (A2) also implies that

$$|D_{n_1}(y, z) - D_0(y, z)| = \int_x |\hat{f}\{y|x, z, \hat{\beta}(\tau)\} - f\{y|x, z, \beta_0(\tau)\}| f(x|z) dx = o_p(1). \quad (A3)$$

It follows that, for any y and z ,

$$\begin{aligned}
& \sup_x |\hat{f}(x|y, z) - f(x|y, z)| \\
&= \sup_x \left| \frac{\hat{f}\{y|x, z, \hat{\beta}(\tau)\}f(x)}{D_{n_1}(y, z)} - \frac{f\{y|x, z, \beta_0(\tau)\}f(x|z)}{D_0(y, z)} \right| = \sup_x \left| \frac{\hat{f}\{y|x, z, \hat{\beta}(\tau)\}f(x)}{D_0(y, z)} \right. \\
&\quad \left. - \frac{f\{y|x, z, \beta_0(\tau)\}f(x|z)}{D_0(y, z)} \right| + O_p\{|D_{n_1}(y, z) - D_0(y, z)|\} = \sup_x \left| \frac{[\hat{f}\{y|x, z, \hat{\beta}(\tau)\} - f\{y|x, z, \beta_0(\tau)\}]f(x|z)}{D_0(y, z)} \right| + O_p\{|D_{n_1}(y, z) - D_0(y, z)|\} = o_p(1).
\end{aligned} \tag{A4}$$

The last step is implied by (A2), (A3), together with the facts that $D_0(y, z) > 0$ for any (y, z) , and the density $f(x)$ is bounded away from infinity under Assumption 4. Moreover, the distance between the two objective functions can be written as

$$\begin{aligned}
& \sup_{\beta \in \Omega} |\tilde{S}_0(\beta) - S_0(\beta)| \\
&= \sup_{\beta \in \Omega} |E_{(y, \tilde{x}, z)}[\rho_\tau\{y - (\tilde{x}^T, z^T)\beta\}] - E_{(y, \tilde{x}, z)}[\rho_\tau\{y - (x^T, z^T)\beta\}]| \\
&= \int_{(y, x, z)} \sup_{\beta \in \Omega} \rho_\tau\{y - (x^T, z^T)\beta\} f(y, z) |\hat{f}(x|y, z) - f(x|y, z)| d(y, x, z) \stackrel{\Delta}{=} \int_{(y, x, z)} g(y, x, z) |\hat{f}(x|y, z) - f(x|y, z)| d(y, x, z),
\end{aligned}$$

where $g(y, x, z) = \sup_{\beta \in \Omega} \rho_\tau\{y - (x^T, z^T)\beta\} f(y, z)$. Since x has bounded support, and Ω is a compact set, under the assumptions that $E(y) < \infty$ and $E(z) < \infty$, the function $g(y, x, z)$ is integrable, i.e.,

$$\int_{(y, x, z)} g(y, x, z) d(y, x, z) < \infty. \tag{A5}$$

On the other hand, due to the uniform convergence of $\hat{\beta}(\tau)$, there exists a constant C_1 , such that for large enough n_1 , $f\{y|x, z, \hat{\beta}(\tau)\} \leq h(\tau, x, z) + C_1$. Following Assumption 5(i), the quantile density function $h(\tau, x, z)$ is bounded for any τ, x and z , it follows that $f\{y|x, z, \hat{\beta}(\tau)\}$ is bounded for any (y, x, z) . Moreover, since $f(x)$ is bounded with bounded support, $D_{n_1}(y, z)$ is also bounded. Consequently, the estimated density $\hat{f}(x|y, z)$ is bounded for any (y, x, z) . Following the dominated convergence theorem, the convergence (A4), the integrability (A5) and the boundedness of $\hat{f}(x|y, z)$ together imply the convergence $\sup_{\beta \in \Omega} |\tilde{S}_0(\beta) - S_0(\beta)| = o_p(1)$ as n_1 and $K_n \rightarrow \infty$.

Since $S_0(\beta)$ is a continuous function, and uniquely minimized in $\beta_{0,\tau}$, following the arguments in Amemiya (1985, pp. 106–8), the convergence (A1) suffices for $\|\beta_\tau^* - \beta_{0,\tau}\| = o_p(1)$, where β_τ^* is the minimizer of $S_0(\beta)$. Recall that

$\tilde{S}_{n_0}^{(\ell)}(\beta) = \sum_{j=n_1+1}^n \rho_\tau\{y_j - (\tilde{x}_{j(\ell)}^T, z_j^T)\beta\}$ is the objective function which is minimized at $\hat{\beta}_{0(\ell)}$. Of course,

$$n_0^{-1} E\{\tilde{S}_{n_0}^{(\ell)}(\beta)\} = n_0^{-1} \sum_{j=n_1+1}^n E_{(y_j, \tilde{x}_{j(\ell)}, z_j)}[\rho_\tau\{y_j - (\tilde{x}_{j(\ell)}^T, z_j^T)\beta\}] = \tilde{S}_0(\beta).$$

Then, following standard arguments for M -estimation (van der Vaart, 1998, 44–7), the estimator $\hat{\beta}_{0(\ell)}$ converges to β_τ^* in probability, conditioning on the completely observed data. Therefore,

$$\|\hat{\beta}_{0(\ell)} - \beta_{0,\tau}\| \leq \|\hat{\beta}_{0(\ell)} - \beta_\tau^*\| + \|\beta_\tau^* - \beta_{0,\tau}\| = o_p(1) \quad (\text{A6})$$

as $n_0 + n_1 \rightarrow \infty$. Thus, we have shown the consistency of $\hat{\beta}_{0(\ell)}$.

We now use a Taylor expansion to derive the asymptotic normality of $\hat{\beta}_{0(\ell)}$. Define the directional derivative function of $\tilde{S}_{n_0}^{(\ell)}(\beta)$ as

$$\tilde{S}'_{n_0}(\beta) = \sum_{j=n_1+1}^n \varphi_\tau\{y_j - (\tilde{x}_{j(\ell)}^T, z_j^T)\beta\} (\tilde{x}_{j(\ell)}^T, z_j^T)^T.$$

Arguments similar to those used in proving He & Shao (1996, Lemma 4.6) yield the uniform convergence result

$$\sup_{\|\beta - \beta_{0,\tau}\| < \delta_n} n_0^{-1/2} \|\tilde{S}'_{n_0}(\beta) - \tilde{S}'_{n_0}(\beta_{0,\tau}) - E\{\tilde{S}'_{n_0}(\beta)\} + E\{\tilde{S}'_{n_0}(\beta_{0,\tau})\}\| = o_p(1), \quad (\text{A7})$$

for any descending sequence δ_n . Combining (A6) and (A7), we have

$$n_0^{-1/2} \|\tilde{S}'_{n_0}(\hat{\beta}_{0(\ell)}) - \tilde{S}'_{n_0}(\beta_{0,\tau}) - E\{\tilde{S}'_{n_0}(\hat{\beta}_{0(\ell)})\} + E\{\tilde{S}'_{n_0}(\beta_{0,\tau})\}\| = o_p(1). \quad (\text{A8})$$

Since $\tilde{S}'_{n_0}(\hat{\beta}_{0(\ell)}) \approx 0$, we Taylor expand $E\{\tilde{S}'_{n_0}(\hat{\beta}_{0(\ell)})\}$ in (A8) around $\beta_{0,\tau}$, so that

$$\begin{aligned} 0 &\approx n_0^{-1/2} \sum_{j=n_1+1}^n \varphi_\tau\{y_j \\ &\quad - (\tilde{x}_{j(\ell)}^T, z_j^T)\hat{\beta}_{0(\ell)}\} (\tilde{x}_{j(\ell)}^T, z_j^T)^T = n_0^{-1/2} \sum_{j=n_1+1}^n \varphi_\tau\{y_j \\ &\quad - (\tilde{x}_{j(\ell)}^T, z_j^T)\beta_{0,\tau}\} (\tilde{x}_{j(\ell)}^T, z_j^T)^T + n_0^{-1} \frac{\partial \sum_{j=n_1+1}^n E\{\varphi_\tau\{y_j - (\tilde{x}_{j(\ell)}^T, z_j^T)\beta_{0,\tau}\}\} (\tilde{x}_{j(\ell)}^T, z_j^T)^T}{\partial \beta_{0,\tau}^T} n_0^{1/2} (\hat{\beta}_{0(\ell)} \\ &\quad - \beta_{0,\tau}) + o_p(1), \end{aligned}$$

and thus $\hat{\beta}_{0(\ell)}$ has Bahadur representation

$$n_0^{1/2}(\hat{\beta}_{0(\ell)} - \beta_{0,\tau}) = -(\Psi_{n_0,\tau}/n_0)^{-1} n_0^{-1/2} \sum_{i=1}^{n_0} \varphi_\tau\{y_j - (\tilde{x}_{j(\ell)}^T, z_j^T)\beta_{0,\tau}\}(\tilde{x}_{j(\ell)}^T, z_j^T)^T + o_p(1), \quad (\text{A9})$$

where $\psi_{n_0,\tau} = (\partial/\partial\beta_{0,\tau}) \sum_{j=n_1+1}^n E[\varphi_\tau\{y_j - (\tilde{x}_{j(\ell)}^T, z_j^T)\beta_{0,\tau}\}(\tilde{x}_{j(\ell)}^T, z_j^T)^T]$. Since the conditional density of $f(\tilde{x} | y_j, z_j)$ converges to the true density $f(x | y_j, z_j)$ as $n_1 \rightarrow \infty$ for any x , the joint distribution of $(y_j, \tilde{x}_{j(\ell)}, z_j)$ converges to the joint distribution of (y_i, x_i, z_i) as $n_1 \rightarrow \infty$. Consequently, using Assumption 1 and the dominated convergence theorem, we have that $n_0^{-1}\psi_{n_0,\tau}$ converges to ψ_τ in probability as $n_1 \rightarrow \infty$, and $\text{var}[\varphi_\tau\{y_j - (\tilde{x}_{j(\ell)}^T, z_j^T)\beta_{0,\tau}\}(\tilde{x}_{j(\ell)}^T, z_j^T)^T]$ converges to V_0 as $n_1 \rightarrow \infty$. It follows that $\hat{\beta}_{0(\ell)}$ is asymptotically normally distributed with mean $\beta_{0,\tau}$ and covariance matrix $\psi_\tau^{-1}V_0\psi_\tau^{-1}$. This finishes our analysis of $\hat{\beta}_{0(\ell)}$.

We now define

$$\hat{\beta}_{*(\ell)} = \underset{\tau}{\text{argmin}} \left[\sum_{i=1}^{n_1} \rho_\tau\{y_i - (x_i^T, z_i^T)\beta\} + \sum_{j=n_1+1}^n \rho_\tau\{y_j - (\tilde{x}_{j(\ell)}^T, z_j^T)\beta\} \right]$$

as the estimated coefficient using the ℓ th assembled complete data. Following similar lines in proving (A9) by treating the observed x_i as an imputed value using the true density function $f(x | z, \eta_0)$, we have

$$n^{1/2}(\hat{\beta}_{*(\ell)} - \beta_{0,\tau}) = -n(\Psi_{n_1,\tau} + \Psi_{n_0,\tau})^{-1} \left[n^{-1/2} \sum_{i=1}^{n_1} \varphi_\tau\{y_i - (x_i^T, z_i^T)\beta_{0,\tau}\}(x_i^T, z_i^T)^T + n^{-1/2} \sum_{j=n_1+1}^n \varphi_\tau\{y_j - (\tilde{x}_{j(\ell)}^T, z_j^T)\beta_{0,\tau}\}(\tilde{x}_{j(\ell)}^T, z_j^T)^T \right] + o_p(1), \quad (\text{A10})$$

where $\psi_{n_1,\tau} = (\partial/\partial\beta_{0,\tau}) \sum_{i=1}^{n_1} E[\varphi_\tau\{y_i - (x_i^T, z_i^T)\beta_{0,\tau}\}(x_i^T, z_i^T)^T]$. Using the law of large numbers, the matrix $\psi_{n_1,\tau}$ converges to ψ_τ in probability. On the other hand, recall that $\hat{\beta}_\tau$ is the estimated coefficient based on n_1 complete data only. For any τ , $\hat{\beta}_\tau$ has the Bahadur representation (Koenker, 2005, Equation (4.4)),

$$n_1^{1/2}(\hat{\beta}_\tau - \beta_{0,\tau}) = -(\Psi_{n_1,\tau}/n_1)^{-1} n_1^{-1/2} \sum_{i=1}^{n_1} \psi_\tau\{y_i - (x_i^T, z_i^T)\beta_{0,\tau}\}(x_i^T, z_i^T)^T + o_p(n_1^{-1/2}). \quad (\text{A11})$$

Combining (A9)–(A11), and using $n^{1/2}(\tilde{\beta} - \beta_{0,\tau}) = m^{-1} \sum_{\ell=1}^m n^{1/2}(\hat{\beta}_{*(\ell)} - \beta_{0,\tau})$, we obtain

$$\begin{aligned}
& n^{1/2}(\tilde{\beta} - \beta_{0,\tau}) \\
&= - \{(\lambda+1)^{-1}n_1^{-1}\Psi_{n_1,\tau} + (1+1/\lambda)^{-1}n_0^{-1}\Psi_{n_0,\tau}\}^{-1} \\
&\quad \times \left((\lambda+1)^{-1/2}n_1^{-1/2} \sum_{i=1}^{n_1} \varphi_\tau\{y_i - (x_i^T, z_i^T)\beta_{0,\tau}\} (x_i^T, z_i^T)^T \right. \\
&\quad \left. + \left[(1+1/\lambda)^{-1/2}m^{-1} \sum_{\ell=1}^m n_0^{-1/2} \sum_{j=n_1+1}^n \varphi_\tau\{y_j - (\tilde{x}_{j(\ell)}^T, z_j^T)\beta_{0,\tau}\} (\tilde{x}_{j(\ell)}^T, z_j^T)^T \right] \right) \\
&= - \{(\lambda+1)^{-1}n_1^{-1}\Psi_{n_1,\tau} + (1+1/\lambda)^{-1}n_0^{-1}\Psi_{n_0,\tau}\}^{-1} \tag{A12} \\
&\quad \times \left((\lambda+1)^{-1/2}(n_1^{-1}\Psi_{n_1,\tau})\{n_1^{1/2}(\hat{\beta}_\tau - \beta_{0,\tau})\} \right. \\
&\quad \left. + \left[(1+1/\lambda)^{-1/2}m^{-1} \sum_{\ell=1}^m (\Psi_{n_0,\tau}/n_0)\{n_0^{1/2}(\hat{\beta}_{0(\ell)} - \beta_{0,\tau})\} \right] \right) \\
&= - \{(\lambda+1)^{-1}n_1^{-1}\Psi_{n_1,\tau} + (1+1/\lambda)^{-1}n_0^{-1}\Psi_{n_0,\tau}\}^{-1} \\
&\quad \times \left\{ (\lambda+1)^{-1/2}\mathcal{U}_n + (1+1/\lambda)^{-1/2}m^{-1} \sum_{\ell=1}^m \mathcal{V}_{n(\ell)} \right\},
\end{aligned}$$

where

$$\mathcal{U}_n = n_1^{-1/2} \sum_{i=1}^{n_1} \varphi_\tau\{y_i - (x_i^T, z_i^T)\beta_{0,\tau}\} (x_i^T, z_i^T)^T, \quad \mathcal{V}_{n(\ell)} = n_0^{-1/2} \sum_{j=n_1+1}^n \varphi_\tau\{y_j - (\tilde{x}_{j(\ell)}^T, z_j^T)\beta_{0,\tau}\} (\tilde{x}_{j(\ell)}^T, z_j^T)^T.$$

It follows immediately from the central limit theorem that $\mathcal{U}_n \rightarrow N(0, V_1)$ in distribution. On the other hand, conditioning on the complete data, $\mathcal{V}_{n(\ell)}$ converges to $N(0, V_n)$ in

distribution, where $V_n = \text{var}[\varphi_\tau\{y_j - (\tilde{x}_{j(\ell)}^T, z_j^T)\beta_{0,\tau}\} (\tilde{x}_{j(\ell)}^T, z_j^T)^T]$. Since V_n converges to V_0 with the increase of the total sample size, $\mathcal{V}_{n(\ell)}$ converges to $N(0, V_0)$ in distribution as n goes to infinity by Slutsky's theorem. Because $N(0, V_0)$ does not depend on the complete data, this is also the limit of the marginal distribution of $\mathcal{V}_{n(\ell)}$. Moreover, it is easy to show that $E(\mathcal{U}_n \mathcal{V}_{n(\ell)}) \rightarrow 0$ and $\text{cov}(\mathcal{V}_{n(\ell)}, \mathcal{V}_{n(\ell)}) \rightarrow U_0$. It follows that

$n^{1/2}(\tilde{\beta} - \beta_{0,\tau}) \rightarrow N(0, \psi_\tau^{-1}\Sigma\psi_\tau^{-1})$, where $\Sigma = (\lambda+1)^{-1}V_1 + (1+1/\lambda)^{-1}[m^{-1}V_0 + \{(m-1)/m\}U_0]$, as claimed.

A2. Implementing the shrinkage estimator

Define $\hat{\mathcal{B}} = (\hat{\beta}_\tau^T, \tilde{\beta}_\tau^T)^T$. Let Γ be the estimated covariance matrix of $\hat{\mathcal{B}}$, which is derived on a case-by-case basis. Then $\hat{\theta}_\tau = (\theta_{1,\tau}, \dots, \theta_{p,\tau})^T = \hat{\beta}_\tau - \tilde{\beta}_\tau$. Let V be the estimated covariance matrix of $\hat{\theta}_\tau$, with diagonal elements (v_1, \dots, v_p) . Define

$$K = \text{diag} \left(\frac{\hat{v}_1}{\hat{v}_1 + \hat{\psi}_1^2}, \dots, \frac{\hat{v}_p}{\hat{v}_p + \hat{\psi}_p^2} \right);$$

and define $G = (K, I_p - K)$. Then the shrinkage estimator is $\hat{\beta}_s(\tau) = G\hat{\mathcal{B}}$. Its estimated covariance matrix is $\widehat{\text{cov}}\{\hat{\beta}_s(\tau)\} = G\hat{\Gamma}G^T$.

We estimate the covariance matrixes of $\tilde{\beta}_\tau$ and $\hat{\beta}_\tau$ based on their Bahadur representations (A11) and (A12), respectively. That requires the estimation of the variance component matrices, $\psi_{n_1,\tau}$, $\psi_{n_0,\tau}$, the variances of \mathcal{U}_n and $\mathcal{V}_{n(\ell)}$ and the covariance of $\mathcal{V}_{n(\ell)}$ and $\mathcal{V}_{n(\ell')}$. In what follows, we provide sample estimation of those variance component matrices. First,

$\hat{\psi}_{n_1,\tau} = n_1^{-1} \sum_{i=1}^{n_1} -\hat{f}_i(\tau)(x_i^T, z_j^T)^T (x_i^T, z_j^T)$, where

$$\hat{f}_i(\tau) = \frac{2h_\tau}{(x_i^T, z_i^T)\{\hat{\beta}_{(\tau+h_\tau)} - \hat{\beta}_{(\tau-h_\tau)}\}}.$$

Here h_τ is the bandwidth chosen by the method of Hall & Sheather (1988). Compared with the density estimator that we used in the estimation procedure, here we incorporated a bandwidth selection h_τ to improve the stability of $\hat{f}_i(\tau)$. Of course,

$\psi_{n_0,\tau} = \lim_{n_0 \rightarrow \infty} n_0^{-1} \sum_j \partial E\{\varphi_\tau(y_j - (\tilde{x}_{j(\ell)}^T, z_j^T)\beta_{0,\tau})(\tilde{x}_{j(\ell)}^T, z_j^T)^T\} / \partial \beta^T$. Following similar lines, we approximate this last term by

$$\hat{\Psi}_{n_0,\tau} \approx n_0^{-1} \sum_{j=n_1+1}^n \frac{1}{m} \sum_{\ell=1}^m -\hat{f}_{j(\ell)}(\tau)(\tilde{x}_{j(\ell)}^T, z_j^T)^T (\tilde{x}_{j(\ell)}^T, z_j^T),$$

where the estimated density function is

$$\hat{f}_{j(\ell)}(\tau) = \frac{2h}{(\tilde{x}_{j(\ell)}^T, z_j^T)\{\hat{\beta}_{(\tau+h_\tau)} - \hat{\beta}_{(\tau-h_\tau)}\}}.$$

Following the linear expansions of $\tilde{\beta}_\tau$ and $\hat{\beta}_\tau$, we first estimate $\text{var}(\mathcal{U}_n)$ and $\text{var}(\mathcal{V}_{n(\ell)})$, and $\text{cov}(\mathcal{V}_{n(\ell)}, \mathcal{V}_{n(\ell)})$ using sample variances, i.e., we define the estimator

$$\hat{V}_1 = \widehat{\text{var}}(\mathcal{U}_n) = n_1^{-1} \sum_{i=1}^{n_1} \varphi_\tau^2\{y_i - (x_i^T, z_i^T)\hat{\beta}_\tau\} (x_i^T, z_i^T)^T (x_i^T, z_i^T).$$

Let Q_ℓ be sample covariance matrix of $[\varphi_\tau\{y_j - (\tilde{x}_{j(\ell)}^T, z_j^T)\beta_\tau\}(\tilde{x}_{j(\ell)}^T, z_j^T)^T]_{j=n_1+1}^n$. The variance component matrix $\text{var}(\mathcal{V}_{n(\ell)})$, for any ℓ , can be estimated by

$$\hat{V}_0 = \widehat{\text{var}}(\mathcal{V}_{n(\ell)}) = m^{-1} \sum_{\ell=1}^m Q_\ell.$$

For any $\ell \neq \ell'$, we define $Q(\ell, \ell')$ as the sample covariance matrix between

$$[\varphi_\tau\{y_j - (\tilde{x}_{j(\ell)}^T, z_j^T)\hat{\beta}_\tau\}(\tilde{x}_{j(\ell)}^T, bz_j^T)^T]_{j=n_1+1}^n \text{ and } [\varphi_\tau\{y_j - (\tilde{x}_{j(\ell')}^T, z_j^T)\hat{\beta}_\tau\}(\tilde{x}_{j(\ell')}^T, bz_j^T)^T]_{j=n_1+1}^n.$$

We define

$$\hat{U} = \text{cov}(\mathcal{Y}_{n(\ell)}, \mathcal{Y}_{n(\ell')}) = \{m(m-1)\}^{-1} \sum_{\ell} \sum_{\ell' \neq \ell} Q(\ell, \ell'),$$

for any (ℓ, ℓ') . With the considerations above, we have

$$\hat{M} = (\hat{\lambda} + 1)^{-1} \hat{\Psi}_{n_1, \tau} + (1 + 1/\hat{\lambda})^{-1} \hat{\Psi}_{n_0, \tau}, \quad \hat{\Sigma} = (\hat{\lambda} + 1)^{-1} \hat{V}_1 + (1 + 1/\hat{\lambda})^{-1} [m^{-1} \hat{V}_0 + (m-1)/m \hat{U}_0],$$

where $\hat{\lambda} = n_0/n_1$. Consequently, the estimated covariance matrix of $\tilde{\beta}_\tau$ is

$$\widehat{\text{var}}(\tilde{\beta}_\tau) = n^{-1} \hat{M}^{-1} \hat{\Sigma} \hat{M}^{-1}, \text{ and the estimated covariance matrix of } \hat{\beta}_\tau \text{ is}$$

$\widehat{\text{var}}(\hat{\beta}_\tau) = n^{-1} (1 + \hat{\lambda}) \psi_{n_1, \tau}^{-1} V_1 \psi_{n_1, \tau}^{-1}$. Since \mathcal{U}_n and $\mathcal{Y}_{n(\ell)}$ are asymptotically independent and have means zero, we have that $E(\mathcal{U}_n \mathcal{Y}_{n(\ell)}) = o(1)$. We can estimate the covariance between $\tilde{\beta}_\tau$ and $\hat{\beta}_\tau$ by $\widehat{\text{cov}}(\tilde{\beta}_\tau, \hat{\beta}_\tau) = n^{-1} \hat{M}^{-1} \hat{V}_1 \hat{\psi}_{n_1, \tau}^{-1}$. Assembling these components together, we obtain

$$\hat{\Gamma} = n^{-1} \begin{pmatrix} \hat{M}^{-1} \hat{\Sigma} \hat{M}^{-1} & \hat{M}^{-1} \hat{V}_1 \hat{\psi}_{n_1, \tau}^{-1} \\ (\hat{M}^{-1} \hat{V}_1 \hat{\psi}_{n_1, \tau}^{-1})^T & (1 + \hat{\lambda}) \hat{\psi}_{n_1, \tau}^{-1} \hat{V}_1 \hat{\psi}_{n_1, \tau}^{-1} \end{pmatrix}.$$

References

- Amemiya, T. *Advanced Econometrics*. Boston: Harvard University Press; 1985.
- Carroll, RJ.; Ruppert, D.; Stefanski, LA.; Crainiceanu, CM. *Measurement Error in Nonlinear Models: A Modern Perspective*. 2nd ed.. London: Chapman and Hall CRC Press; 2006.
- Chen YH, Chatterjee N, Carroll RJ. Shrinkage estimators for robust and efficient inference in haplotype-based case-control studies. *J. Am. Statist. Assoc.* 2009; 104:220–233.
- Hall P, Sheather S. On the distribution of a studentized quantile. *J. R. Statist. Soc. B.* 1988; 50:381–391.
- He X, Shao QM. A general Bahadur representation of M -estimators and its application to linear regression with nonstochastic designs. *Ann. Statist.* 1996; 24:2608–2630.
- Koenker, R. *Quantile Regression*. Cambridge: Cambridge University Press; 2005.
- Koenker R, Bassett GJ. Regression quantiles. *Econometrica.* 1978; 46:33–50.
- Koenker R, Xiao ZJ. Unit root quantile autoregression inference. *J. Am. Statist. Assoc.* 2004; 99:775–787.
- Lipsitz SR, Fitzmaurice GM, Molenberghs G, Zhao LP. Quantile regression methods for longitudinal data with drop-outs: application to CD4 cell counts of patients infected with the human immunodeficiency virus. *Appl. Statist.* 1997; 46:463–476.
- Little, RJA.; Rubin, DB. *Statistical Analysis with Missing Data*. New York: Wiley; 1987.
- Robins JM, Rotnitzky A, Zhao LP. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *J. Am. Statist. Assoc.* 1995; 90:106–121.
- Subar AF, Thompson FE, Kipnis V, Midthune D, Hurwitz P, McNutt S, Mcintosh A, Rosenfeld S. Comparative validation of the Block, Willett, and National Cancer Institute Food Frequency

Questionnaires: the Eating at American's Table Study. *Am. J. Epidemiol.* 2001; 154:1089–1099. [PubMed: 11744511]

Tsiatis, AA. *Semiparametric Theory and Missing Data*. New York: Springer; 2006.

Van der Vaart, AW. *Asymptotic Statistics*. Cambridge: Cambridge University Press; 1998.

Wei Y, Carroll RJ. Quantile regression with measurement error. *J. Am. Statist. Assoc.* 2009; 104:1129–1143.

Welsh AH. Asymptotically efficient estimation of the sparsity function at a point. *Statist. Prob. Lett.* 1988; 6:427–432.

Yi GY, He W. Median regression models for longitudinal data with dropouts. *Biometrics.* 2009; 65:618–625. [PubMed: 18759840]

Table 1

Means and standard errors of the estimated coefficients at quantile levels 0.1, 0.5 and 0.9 from 500 simulations in models (4) and (5)

Intercept	True	$\tau = 0.1$			$\tau = 0.5$			$\tau = 0.9$			
		Mean	SE	True	Mean	SE	True	Mean	SE	True	
Model (4)											
	$\hat{\beta}$	-0.28	0.29	1.00	1.00	0.21	2.28	2.28	0.28	2.28	0.28
	$\tilde{\beta}$	-0.28	0.23	1.00	1.03	0.17	2.28	2.29	0.24	2.29	0.24
	$\beta^{(5)}$	-0.28	0.25	1.00	1.01	0.19	2.28	2.29	0.25	2.29	0.25
x	$\hat{\beta}$	1.00	0.16	1.00	1.00	0.12	1.00	1.00	0.16	1.00	0.16
	$\tilde{\beta}$	1.00	0.16	1.00	0.98	0.11	1.00	0.96	0.16	1.00	0.16
	$\beta^{(5)}$	1.00	0.16	1.00	0.99	0.12	1.00	0.99	0.16	1.00	0.16
z	$\hat{\beta}$	1.00	0.20	1.00	1.00	0.14	1.00	1.00	0.19	1.00	0.19
	$\tilde{\beta}$	1.00	0.18	1.00	1.01	0.12	1.00	1.03	0.16	1.00	0.16
	$\beta^{(5)}$	1.00	0.19	1.00	1.01	0.13	1.00	1.01	0.17	1.00	0.17
Model (5)											
	$\hat{\beta}$	1.00	3.88	1.00	0.61	2.68	1.00	1.00	3.16	1.00	3.16
	$\tilde{\beta}$	1.00	2.04	1.00	0.84	1.59	1.00	1.56	2.11	1.00	2.11
	$\beta^{(5)}$	1.00	3.22	1.00	0.69	2.24	1.00	1.37	2.64	1.00	2.64
x	$\hat{\beta}$	0.36	0.62	1.00	1.01	0.51	1.64	1.68	0.62	1.64	0.62
	$\tilde{\beta}$	0.36	0.52	1.00	0.92	0.46	1.64	1.45	0.53	1.64	0.53
	$\beta^{(5)}$	0.36	0.58	1.00	0.98	0.49	1.64	1.62	0.60	1.64	0.60
z	$\hat{\beta}$	0.36	0.92	1.00	1.08	0.64	1.64	1.59	0.77	1.64	0.77
	$\tilde{\beta}$	0.36	0.52	1.00	1.11	0.41	1.64	1.68	0.52	1.64	0.52
	$\beta^{(5)}$	0.36	0.78	1.00	1.09	0.54	1.64	1.60	0.64	1.64	0.64

$\hat{\beta}$, the estimated coefficient using the completely observed data only; $\tilde{\beta}$, the multiple imputation estimator with 10 imputations; $\beta^{(5)}$, the shrinkage estimator; True, the true coefficients; SE, standard errors.

Table 2

Mean squared errors of the estimated coefficients at quantile levels 0.1, 0.5 and 0.9 from 500 simulations in model (2) when $f(x|z)$ is misspecified

	$\tau=0.1$		$\tau=0.5$		$\tau=0.9$	
	MSE	SE	MSE	SE	MSE	SE
$\hat{\beta}$	1.58	0.09	0.76	0.05	1.33	0.08
$\tilde{\beta}$	1.49	0.08	0.70	0.04	1.14	0.06
$\hat{\beta}^{(s)}$	1.36	0.08	0.72	0.04	1.07	0.07
β^*	1.31	0.07	0.68	0.04	1.02	0.06

$\hat{\beta}$, the estimated coefficient using the completely observed data only; $\tilde{\beta}$, the multiple imputation estimator with 10 imputations; $\hat{\beta}^{(s)}$, the shrinkage estimator; β^* , the multiple imputation estimator using the exact $f(x|z)$; SE, the standard error of the mean squared error; MSE, mean squared errors.

Table 3

Estimated coefficients in the Eating at American's Table Study

Covariates	τ	Raw		Multiple imputation			Shrinkage		
		$\hat{\beta}$	$\hat{s.e.}$	$\hat{\beta}$	$\hat{s.e.}$	$\hat{\tau}\hat{e}(\%)$	$\hat{\beta}^{(s)}$	$\hat{s.e.}$	$\hat{\tau}\hat{e}(\%)$
Carbohydrate intake	0.1	0.08	0.06	0.04	0.06	99	0.06	0.06	104
	0.5	0.27	0.04	0.24	0.03	109	0.27	0.03	102
	0.9	0.60	0.07	0.48	0.07	112	0.59	0.07	101
Body mass index	0.1	-0.94	0.88	-0.84	0.91	94	-0.85	0.90	96
	0.5	-1.68	0.54	-1.63	0.54	99	-1.63	0.54	100
	0.9	-0.70	1.20	-0.35	1.21	98	-0.51	1.19	101
Age	0.1	-0.53	0.36	-0.39	0.35	108	-0.42	0.33	117
	0.5	-0.86	0.28	-1.00	0.24	136	-0.95	0.25	132
	0.9	-1.38	0.62	-1.71	0.51	147	-1.54	0.55	126
Caucasian	0.1	5.95	14.43	14.95	12.57	132	11.61	12.02	144
	0.5	4.67	10.87	6.22	8.38	168	6.16	8.37	169
	0.9	-38.45	41.02	-1.39	25.57	257	-27.91	35.76	132
Gender	0.1	-47.34	11.03	-38.34	10.25	116	-43.20	10.11	119
	0.5	-73.48	8.27	-66.90	7.05	137	-70.77	7.57	119
	0.9	-108.07	15.58	-114.92	12.96	145	-113.41	13.59	131

$\hat{s.e.}$ standard errors following the estimation method described in Appendix A2; $\hat{\tau}\hat{e}$ relative efficiency, which is defined as the ratio between the estimated variance of the complete-data estimator and that of the multiple imputation/shrinkage estimates.