NBER WORKING PAPER SERIES

# WEIGHTED-AVERAGE QUANTILE REGRESSION

Denis Chetverikov
Yukun Liu
Aleh Tsyvinski

Working Paper 30014
http://www.nber.org/papers/w30014

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2022

We thank Xiaohong Chen for comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2022 by Denis Chetverikov, Yukun Liu, and Aleh Tsyvinski. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Weighted-Average Quantile Regression
Denis Chetverikov, Yukun Liu, and Aleh Tsyvinski
NBER Working Paper No. 30014
May 2022
JEL No. C01


## ABSTRACT

In this paper, we introduce the weighted-average quantile regression model. We argue that this model is of interest in many applied settings and develop an estimator for parameters of this model. We show that our estimator is $\sqrt{T}$-consistent and asymptotically normal with mean zero under weak conditions, where T is the sample size. We demonstrate the usefulness of our estimator in two empirical settings. First, we study the factor structures of the expected shortfalls of the industry portfolios. Second, we study inequality and social welfare dependence on individual characteristics.



Denis Chetverikov
UCLA, Department of Economics
315 Portola Plaza
Bunche Hall, Room 8283
Los Angeles, CA 90095-1477
chetverikov@econ.ucla.edu

Yukun Liu
Department of Economics
University of Rochester
Rochester, NY 06520-8268
yliu229@simon.rochester.edu

Aleh Tsyvinski
Department of Economics
Yale University
Box 208268
New Haven, CT 06520-8268
and New Economic School
and also NBER
a.tsyvinski@yale.edu

# WEIGHTED-AVERAGE QUANTILE REGRESSION

DENIS CHETVERIKOV, YUKUN LIU, AND ALEH TSYVINSKI

ABSTRACT. In this paper, we introduce the weighted-average quantile regression model. We argue that this model is of interest in many applied settings and develop an estimator for parameters of this model. We show that our estimator is $\sqrt{T}$-consistent and asymptotically normal with mean zero under weak conditions, where $T$ is the sample size. We demonstrate the usefulness of our estimator in two empirical settings. First, we study the factor structures of the expected shortfalls of the industry portfolios. Second, we study inequality and social welfare dependence on individual characteristics.

## 1. INTRODUCTION

Mean and quantile regression models are among the key elements of the econometrics toolbox. However, there is a large set of functionals beyond the mean and quantiles of a distribution that are of interest in applied work. It is therefore important to study other regression models as well. To do so, we consider in this paper a broad class of regression models, which we refer to as the *weighted-average quantile regression:*

$$\int_0^1 q_{Y|X}(u)\psi(u)du = X'\beta, \tag{1}$$

where $Y$ is a dependent variable in $\mathbb{R}$, $X$ is a vector of covariates in $\mathbb{R}^p$, $q_{Y|X}\colon [0,1] \to \mathbb{R}$ is the quantile function of the conditional distribution of $Y$ given $X$, $\psi\colon [0,1] \to \mathbb{R}$ is a signed weighting function, and $\beta$ is the parameter vector in $\mathbb{R}^p$ to be estimated. Such regression models are of interest in a number of applications. First, if $Y$ is the loss of a financial portfolio and $\psi(u) = \mathbb{I}\{u \geq 1-\alpha\}/\alpha$ for, say, $\alpha = 0.1$, we obtain an example of a *risk regression*, namely an expected shortfall regression, which is of interest in finance (e.g., Adrian and Brunnermeier, 2016; Acharya et al., 2017). This regression lets us study how the risk, measured by the expected shortfall, of a financial portfolio comoves with various financial/macro variables. Second, if $Y$ is the wage and $\psi(u) = \mathbb{I}\{u \leq \alpha\}/\alpha$ for, say, $\alpha = 0.2$, we obtain a *lower wage*

*Date*: March 9, 2022.
*Key words and phrases.* Linear regression, quantile regression, double/debiased machine learning, risk measures, expected shortfall, inequality, social welfare.

*regression*, and in the same way, we can also define middle and upper wage regressions. These regressions are similar to the mean regression but apply the mean to separate wage classes: lower, middle, and upper classes. They may be of interest in labor economics as parsimonious alternatives to quantile regression models. Third, if $Y$ is the wage and $\psi(u) = (\mathbb{I}\{u \geq 1-\alpha\} - \mathbb{I}\{u \leq \alpha\})/\alpha$ for, say, $\alpha = 0.1$, we obtain an *inequality regression*. The difference between the high and low income groups captures the inequality of the income distribution (e.g., Angrist et al., 2006; Blundell et al., 2008; Attanasio and Pistaferri, 2016), and therefore the inequality regression may help to identify important determinants of social inequality. Similarly, if $\psi$ is a general decreasing function, we obtain a *social welfare regression*. Finally, if the researcher is concerned about the effect of data contamination, we can define $\psi(u) = \mathbb{I}\{\alpha \leq u \leq 1-\alpha\}/(1-2\alpha)$ for some small $\alpha$ to obtain a *robust regression,* which may be of interest as an alternative to the Huber regression (Huber and Ronchetti, 2009).

We assume that we have a stationary time series dataset $(X_1, Y_1), \ldots, (X_T, Y_T)$ with each $(X_t, Y_t)$ having the same distribution as that of the pair $(X, Y)$, and develop a $\sqrt{T}$-consistent estimator of $\beta$. Our estimator, which we refer to as the weighted-average quantile regression estimator, consists of three steps. First, we use machine learning to estimate the distribution function of the conditional distribution of $Y$ given $X$. Second, we use this distribution function to construct a simple transformation of $Y_t$ for all $t$. Third, we estimate $\beta$ by running OLS of this transformation on $X_t$. We prove that our estimator is asymptotically normal with mean zero and that its asymptotic covariance matrix can be consistently estimated by the Newey-West method on the third step, which is carried out in all commonly used statistical software. Our estimation and inference procedures are thus straightforward to implement.

Importantly, our approach is semi-parametric: we assume that the weighted-average conditional quantile function $x \mapsto \int_0^1 q_{Y|X=x}(u)\psi(u)du$ is linear, to obtain broadly applicable results, but we do not impose any other parametric restrictions. The latter is useful as it minimizes the possibility of misspecification and inconsistent estimation. A parametric alternative to our methods would be to (i) assume that each quantile function $x \mapsto q_{Y|X=x}(u)$ is linear,

$$q_{Y|X}(u) = X'\beta(u), \quad \text{for all } u \in (0,1), \tag{2}$$

(ii) calculate the quantile regression estimator $\widetilde{\beta}(u)$ of each $\beta(u)$, and (iii) given that (2) yields $\beta = \int_0^1 \beta(u)\psi(u)du$, estimate $\beta$ by $\int_0^1 \widetilde{\beta}(u)\psi(u)du$. This alternative

approach, however, can lead to potential misspecification and inconsistent estimation due to extra assumptions (2) and, moreover, requires estimating extreme quantile regressions, which correspond to values of $u$ in (2) that are close to the boundary of the $[0, 1]$ interval, and which are typically difficult to estimate. In contrast, our estimator does not require estimation of extreme quantile regressions. Furthermore, in contrast to classical semiparametric estimation theory, by applying double/debiased machine learning techniques (e.g., Chernozhukov et al., 2018), our estimator is able to handle the case where $X$ is moderate- or large-dimensional, which is particularly useful in a number of applications. It is important to note, however, that our results do not follow from the standard results on double/debiased machine learning techniques because we allow general weighting functions $\psi$ that in particular may feature discontinuities.

We apply our method to two empirical settings: a setting in finance and a setting in labor economics. In the first application, we focus on financial data and study the factor structures of the expected shortfalls of the industry portfolios. We show that the expected shortfalls of the industry portfolios have significant time-varying exposures to the factor models developed in the asset pricing literature. Importantly, the factor structures of the expected shortfalls of the industry portfolios based on the weighted-average quantile regressions can differ significantly from those estimated based on the mean and quantile regressions or based on a parametric estimator. We show that the discrepancies stem from the fact that the quantiles are not linear in the factors in the financial data.

In the second application, we apply the inequality and social welfare regressions to wage data. Using the inequality regression, we study the relationship between wage inequality and individual characteristics that are common in labor economics. We compare the inequality regression results with those based on a parametric estimator and show that the results can differ in important ways. For example, based on the weighted-average quantile regression estimator, the wage inequality is estimated to be negatively related to family size in the recent sample, but the relationship is muted using the parametric estimator. Applying the social welfare regression, we study the dependence of the weighted average wage on individual characteristics, where the weights are exponential with higher weights on the lower income. We call this a social welfare regression as it is consistent with a variety of social welfare functions used in, say, public finance that place higher weights on poorer individuals. We find that the results based on the social welfare regression differ from those from the mean regression. For example, for the early 2000s, the magnitude of the point estimates on

education based on the social welfare regression is only half as large as those using the mean regression.

**Related Literature.** Overall, our results generalize the commonly used mean (OLS) and quantile regression methods to allow for a much larger class of functionals – the weighted-average quantiles, which are of interest in a number of applications as outlined above and as discussed in detail in the next section.

More broadly, we contribute to the literature by providing a general principle for obtaining a double/debiased machine learning estimator of linear regression models of the form $h(F_{Y|X}) = X'\beta$, where $F_{Y|X}$ is the distribution function of the conditional distribution of $Y$ given $X$ and $h$ is a functional of interest, the weighted-average quantile being one example. Specifically, we first calculate the influence function $y \mapsto a(y, F)$ for the functional $F \mapsto h(F)$ and then estimate the vector of parameters $\beta$ by running OLS of an estimated version of $h(F_{Y|X}) + a(Y, F_{Y|X})$ on $X$. We note that although the idea of using influence function adjustments itself is not new, as it can be traced back at least to Bickel (1982) and Schick (1986) and was used recently in Chernozhukov et al. (2016), our key finding is that the adjustment term $a$ in the regression context depends only on the functional $h$ and not on the joint distribution of the pair $(X, Y)$, making the approach broadly available in applied settings. In addition, although our procedure looks similar to that in Firpo et al. (2009), who propose to run OLS of $h(F_Y) + a(Y, F_Y)$ on $X$, where $F_Y$ is the distribution function of the *marginal* distribution of the random variable $Y$, the similarities are superficial: two procedures aim at estimating fundamentally different quantities and also have completely different reasons for the influence function adjustments. The detailed explanations of how our procedure relates to the literature can be found in Section 3, after we fully describe and explain the general principle.

Our paper is also related to the financial literature on estimation of conditional risk measures. Tail risk measures, such as expected shortfall, are an important class of risk measures in finance (e.g., Lettau and Ludvigson, 2013; Jurado et al., 2015; Adrian and Brunnermeier, 2016; Acharya et al., 2017). A few parametric and nonparametric methods for estimating expected shortfall regressions were proposed and analyzed in Scaillet (2005), Cai and Wang (2008), Peracchi and Tanase (2008), Leorato et al. (2012), Kato (2012), and Martins-Filho et al. (2018). However, all of these papers either assume nonparametric expected shortfall, which makes interpretations in applied work difficult, or impose the linear quantile assumption (2), which leads to potential misspecification. Moreover, these papers consider only one risk measure: expected

shortfall, whereas our methods cover a broad class of risk measures; see Section 2 for details. Kato (2012) provides a nice comparison of the existing methods.

Our paper is also related to the semi-parametric methods developed in Chun et al. (2012), who consider the problem of estimating $\beta$ in the model (1) with the integral over $u \in (0,1)$ replaced by the sum over a grid of values of $u$ in $(0,1)$. Clearly, the sum over a fine grid can be used to approximate the integral but the variance of their estimator depends on the inverse of the density of the conditional distribution of $Y$ given $X$ in the tails and generally blows up as we take finer grids, which makes our methods quite different from those developed in Chun et al. (2012).

Moreover, our paper is seemingly related to the methods developed in Rockafellar et al. (2014) and Royset and Rockafellar (2015), who develop super-quantile regression methods. In principle, super-quantile is just another name for the expected shortfall. However, the estimators proposed in these papers do not converge to $\beta$ appearing in (1) when we set $\psi(u) = \mathbb{I}\{u \geq 1-\alpha\}/\alpha$ to obtain the expected shortfall regression. Therefore, from the perspective of our setting, the estimators proposed in these papers are not consistent, even though they do converge to some meaningful quantities, see Rockafellar et al. (2014) for details.

**Outline of the Paper.** The rest of the paper is organized as follows. In the next section, we provide several examples covered by our weighted-average quantile regression framework (1). In Section 3, we derive a general principle for obtaining double/debiased machine learning estimators of linear regression models of the form $h(F_{Y|X}) = X'\beta$ and apply it to the weighted-average quantile regression. In Section 4, we describe our estimation and inference procedures in detail. In Section, 5, we prove consistency and derive the asymptotic distribution for our estimators. In Section 6, we provide results of a small-scale Monte Carlo simulation study confirming good statistical properties of our estimators in finite samples. In Section 7, we apply our procedures in two empirical settings that are of interest in finance and labor economics. In the Online Appendix, we collect all proofs, additional discussions, and extra tables and figures for the empirical applications.

## 2. Examples

In this section, we describe various regression models covered by our general regression framework (1).

### 2.1. **Risk Regression.**

In finance, the concept of risk measures is used to quantify the risk of financial positions, e.g. Föllmer and Schied (2002). Formally, any risk

measure is a functional $\rho$ that is defined on a set of random variables and that has certain desirable properties, so that for a random variable $Y \in \mathbb{R}$ representing a loss of some financial position, the value of $\rho(Y)$ measures the risk associated with $Y$. Most commonly used risk measures belong to the class of *spectral risk measures*. These risk measures have many desirable properties (positive homogeneity, translation invariance, monotonicity, sub-additivity, etc.) and take the following form (Acerbi, 2002):

$$\rho(Y) = \int_0^1 q_Y(u)\psi(u)du, \tag{3}$$

where $q_Y : [0,1] \to \mathbb{R}$ is the quantile function of the random variable $Y$ and $\psi : [0,1] \to \mathbb{R}$ is a an increasing weighting function such that (i) $\psi(u) \geq 0$ for all $u \in (0,1)$ and (ii) $\int_0^1 \psi(u)du = 1$. Here, the function $\psi$ is called the spectrum function associated with the risk measure $\rho$ and different spectrum functions $\psi$ lead to different risk measures. For example, one of the most important spectral risk measures is the expected shortfall, also known as the average value at risk, which corresponds to the spectrum function $\psi(u) = \mathbb{I}\{u \geq 1-\alpha\}/\alpha$, where $\mathbb{I}\{\cdot\}$ denotes the indicator function, and $\alpha \in (0,1)$ is a user-specified parameter, typically taking some small value such as 5% or 10%. Other examples are exponential and polynomial risk measures, which correspond to the spectrum functions $\psi(u) = a\exp(-a(1-u))/(1-\exp(-a))$ with $a > 0$ and $\psi(u) = au^{a-1}$ with $a > 1$, correspondingly, where $a$ is a user-specified parameter, e.g. Leippold (2015). A textbook-level discussion of spectral risk measures can be found in McNeil et al. (2015).

To study how the risk of one random variable, say $Y$, comoves with a vector of other variables, say $X$, we can consider a *risk regression* $\rho(Y|X) = X'\beta$, where $\rho(Y|X) = \int_0^1 q_{Y|X}(u)\psi(u)du$ is the risk measure of the conditional distribution of $Y$ given $X$. Substituting here various functions $\psi$, we obtain various risk regressions, e.g. the expected shortfall and exponential regressions. These regressions are covered by our general framework (1).

In addition, we note also that risk measures, under different names, appear also in behavioral economics, where they are used to rank lotteries, e.g. Kahneman and Tversky (1979) and Yaari (1987), and in actuarial science, where they are used to determine premium principles, e.g. Kaas et al. (2008). Moreover, our methods can be used to estimate an expected shortfall (or any other spectral risk measure) version of CoVar, a concept introduced in Adrian and Brunnermeier (2016) to study systemic risk.

2.2. **Wage Regression.** Quantile regression methods have been used to study conditional wage distributions since Buchinsky (1994) and Chamberlain (1994). If $Y$ is individual's wage and $X$ is a vector of covariates including, for example, education, running a quantile regression of $Y$ on $X$ lets us estimate the effect of education on the conditional distribution of wages for any quantile index $u$ of this distribution. This is useful because the effect of education may vary substantially depending on the quantile index. In practice, however, we may often be interested in the average effect of education for a *group* of quantile indices. For example, we may define the middle-wage class as the set of individuals with quantile indices within the $[20\%, 80\%]$ interval on the conditional wage distribution and we may be interested in the effect of education for this particular set of individuals. In turn, quantile regression methods may not be appropriate for such parameters. Indeed, providing the quantile regression estimate for the average quantile index (50%, in our example) may not give a representative number for the the whole group and providing the quantile regression estimates for each quantile index within the $[20\%, 80\%]$ interval may not be convenient as function-valued estimates are difficult to interpret.[1] Instead, such parameters can be easily estimated by our weighted-average quantile regression methods. Specifically, by setting $\psi(u) = \mathbb{I}\{\alpha \leq u \leq 1-\alpha\}/(1-2\alpha)$ with $\alpha = 0.2$ in (1), we obtain a *middle wage regression*, and the methods developed in our paper can be used to estimate parameters of this regression, yielding in particular the average effect of education on wages for the middle-wage class. Similarly, by setting $\psi(u) = \mathbb{I}\{u \leq \alpha\}/\alpha$ with $\alpha = 0.2$ in (1), we obtain a *lower wage regression*, corresponding to the lower-wage class, and by setting $\psi(u) = \mathbb{I}\{u \geq 1-\alpha\}/\alpha$, again with $\alpha = 0.2$, we obtain an *upper wage regression*, corresponding to the upper-wage class. More generally, since the same techniques can be used with any dependent variable $Y$, we can refer to these types of regression models simply as the lower, middle, and upper regressions. In this case, the expected shortfall regression discussed above becomes an instance of the upper regression.

2.3. **Inequality and Social Welfare Regressions.** Related to our discussion in the previous example, another reason to study conditional wage distributions is that they help us understand the dynamics of the wage inequality over time, e.g. Angrist and Pischke (2008). Again assuming that $Y$ is individual's wage and $X$ is a vector of relevant covariates, we can study wage inequality by our weighted-average quantile

[1]Moreover, averaging quantile regression estimates over quantile indices in the $[20\%, 80\%]$ interval may not be a good idea either, for the reasons explained in the Introduction.

regression methods. Indeed, by setting $\psi(u) = (\mathbb{I}\{u \geq 1-\alpha\} - \mathbb{I}\{u \leq \alpha\})/\alpha$ in (1) for some small $\alpha$, say 0.1, we obtain an *inequality regression*, which allows us to study how the difference between the average wage of 10% individuals with highest wages and the average wage of 10% individuals with lowest wages depend on covariates. Similarly, by considering any decreasing function $\psi$, e.g. polynomial or exponential from Section 2.1 with $u$ replaced by $1-u$, we obtain a *social welfare regression*. Of course, the inequality regression remains meaningful with other dependent variables as well. More broadly, it is straightforward to generalize the weighted-average quantile regression framework to include other inequality measures such as Gini's coefficient, e.g. Cowell (2011).

2.4. **Robust Regression.** Suppose that we are interested in estimating a linear mean regression model

$$\mathrm{E}[Y|X] = X'\beta$$

from a stationary time series $(X_1, Y_1), \ldots, (X_T, Y_T)$, where each $(X_t, Y_t)$ has the same distribution as that of the pair $(X, Y)$. Typically, we would estimate $\beta$ in this model by OLS

$$\widehat{\beta} = \arg\min_{b \in \mathbb{R}^p} \sum_{t=1}^{T} (Y_t - X_t'b)^2.$$

Suppose, however, that for some observations $t$, the values of the dependent variable $Y_t$ are corrupted. These corrupted values may significantly bias the estimator $\widehat{\beta}$ rendering it unreliable. This problem attracted substantial attention in the literature and led to the field called Robust Statistics, which generated many alternatives to OLS, e.g. Least Median of Squares (Rousseeuw, 1984), Least Trimmed Squares (Rousseeuw and Leroy, 1987), and Random Sample Consensus (Fisher and Bolles, 1981); see also recent advances in computer science, e.g. Liu et al. (2018). However, one of the most important methods developed in this field is the Huber estimator (Huber and Ronchetti, 2009), which can be viewed as a modification of the OLS estimator:

$$\widetilde{\beta} = \arg\min_{b \in \mathbb{R}^p} \sum_{t=1}^{T} \rho_c(Y_t - X_t'b),$$

where

$$\rho_c(x) = \begin{cases} x^2, & \text{if } |x| \leq c, \\ 2|x|c, & \text{if } |x| > c, \end{cases}$$

and $c > 0$ is a tuning parameter. Since the derivative of the criterion function $\rho_c$ in the Huber estimator is bounded, this estimator is much more robust with respect to data

corruption in the dependent variable in comparison with the OLS estimator. However, implementing this estimator requires choosing the tuning parameter $c$, which is often unclear in practice: smaller values of $c$ yield more robust but also more biased estimator. We therefore propose to use our weighted-average quantile regression estimator as an alternative. Indeed, suppose that for each observation $t$, the probability of corruption in $Y_t$ does not exceed $\alpha$ for some small user-specified value $\alpha \in (0, 1)$. In this case, we can consider a *robust regression* by setting $w(u) = \mathbb{I}\{\alpha \leq u \leq 1-\alpha\}/(1-2\alpha)$ in (1). Running our estimator based on this regression also requires the choice of the tuning parameter, $\alpha$, but in contrast to the Huber estimator, this choice is rather intuitive: the user simply needs to provide an upper bound on the fraction of corrupted observations.

## 3. Motivation For Estimation Procedure

In this section, we develop a general principle for estimating regression models

$$h(F_{Y|X}) = X'\beta, \tag{4}$$

where $y \mapsto F_{Y|X}(y) = \mathrm{P}(Y \leq y|X)$ is the distribution function of the conditional distribution of $Y$ given $X$ and $h\colon \mathcal{F} \to \mathbb{R}$ is a functional defined on a convex set $\mathcal{F}$ of distribution functions on $\mathbb{R}$ that includes $F_{Y|X}$ almost surely. We then apply the general principle to the weighted-average quantile regression model by substituting $h(F_{Y|X}) = \int_0^1 q_{Y|X}(u)\psi(u)du$. For clarity of the section, we leave technical regularity conditions underlying our derivations for now.

### 3.1. General Principle.

To develop the principle, we need the concept of influence functions. Following Hampel et al. (1986), we say that $h$ is Gateaux differentiable at the distribution function $F \in \mathcal{F}$ if there exists a function $a\colon \mathbb{R} \to \mathbb{R}$ such that for all $G \in \mathcal{F}$, we have

$$\frac{\partial}{\partial t} h((1-t)F + tG)\Big|_{t=0_+} = \lim_{t \downarrow 0} \frac{h((1-t)F + tG) - h(F)}{t} = \int a(y)dG(y). \tag{5}$$

We refer to $a$ as the influence function of $h$ at $F$. To note its dependence on $F$, we will write $a(y, F)$ instead of $a(y)$.

The influence function has two important properties. First, by substituting $G = F$ in (5), we have $\int a(y, F)dF(y) = 0$ and since $F \in \mathcal{F}$ is arbitrary, we obtain

$$\int a(y, F)dF(y) = 0, \quad \text{for all } F \in \mathcal{F}. \tag{6}$$

Second, by substituting $(1-t)F+tG$ instead of $F$ in (6) and taking derivative with respect to $t$ on both sides, we have

$$\frac{\partial}{\partial t}\int a(y,(1-t)F+tG)dF(y)\Big|_{t=0_+}+\int a(y,F)dG(y)-\int a(y,F)dF(y)=0$$

and since $\int a(y,F)dF(y)=0$ and $G$ is arbitrary, we obtain

$$\frac{\partial}{\partial t}\int a(y,(1-t)F+tG)dF(y)\Big|_{t=0_+}=-\int a(y,F)dG(y),\quad \text{for all } G\in\mathcal{F} \tag{7}$$

We will use these two properties below.

Having the concept of influence functions in mind, we propose the following principle: estimate $\beta$ in (4) by running the OLS estimator of $h(\widehat{F}_{Y|X})+a(Y,\widehat{F}_{Y|X})$ on $X$, where $\widehat{F}_{Y|X}$ is a preliminary estimator of $F_{Y|X}$. We claim that this estimator is consistent and robust (in a sense to be made precise below) with respect to the estimation error in $\widehat{F}_{Y|X}$. To see why this is so, observe that, under certain regularity conditions, the probability limit of such an OLS estimator will be

$$\bar{\beta}=(\mathrm{E}[XX'])^{-1}\mathrm{E}\Big[X\Big(h(F_{Y|X})+a(Y,F_{Y|X})\Big)\Big],$$

which can be equivalently written as a set of moment conditions

$$\mathrm{E}\Big[X\Big(h(F_{Y|X})+a(Y,F_{Y|X})-X'\bar{\beta}\Big)\Big]=0_p, \tag{8}$$

where $0_p=(0,\ldots,0)'\in\mathbb{R}^p$. Here, we note that by applying (6) with $F=F_{Y|X}$, we have

$$\int a(y,F_{Y|X})dF_{Y|X}(y)=0.$$

In turn, the left-hand side of this identity is equal to $\mathrm{E}[a(Y,F_{Y|X})|X]$, and so, by the law of iterated expectations,

$$\mathrm{E}[Xa(Y,F_{Y|X})]=\mathrm{E}[X\mathrm{E}[a(Y,F_{Y|X})|X]]=0_p.$$

Substituting this equality into (8) and recalling (4), it follows that the set of moment conditions (8) can be equivalently written as

$$\mathrm{E}[X(X'\beta-X'\bar{\beta})]=0_p.$$

As long as $\mathrm{E}[XX']$ is non-singular, it thus follows that $\bar{\beta}=\beta$ is the unique solution to the set of moment conditions (8). This means that our OLS estimator is consistent.

Further, fix any $\widehat{F}_{Y|X}$ such that $\widehat{F}_{Y|X}\in\mathcal{F}$ almost surely and write $\widetilde{F}=\widehat{F}_{Y|X}-F_{Y|X}$. By applying (5) and (7) with $F=F_{Y|X}$ and $G=\widehat{F}_{Y|X}$ and noting that $(1-t)F+tG=$

$F_{Y|X} + t\widetilde{F}_{Y|X}$ in this case, we have

$$\frac{\partial}{\partial t} h(F_{Y|X} + t\widetilde{F})\Big|_{t=0_+} = \int a(y, F_{Y|X}) d\widehat{F}_{Y|X}(y)$$

and

$$\frac{\partial}{\partial t} \int a(y, F_{Y|X} + t\widetilde{F}) dF_{Y|X}(y)\Big|_{t=0_+} = -\int a(y, F_{Y|X}) d\widehat{F}_{Y|X}(y),$$

respectively. Summing these two identities and observing that

$$\int a(y, F_{Y|X} + t\widetilde{F}) dF_{Y|X}(y) = \mathrm{E}[a(Y, F_{Y|X} + t\widetilde{F})|X],$$

we obtain

$$\frac{\partial}{\partial t} h(F_{Y|X} + t\widetilde{F})\Big|_{t=0_+} + \frac{\partial}{\partial t} \mathrm{E}[a(Y, F_{Y|X} + t\widetilde{F})|X]\Big|_{t=0_+} = 0.$$

The latter in turn implies, via the law of iterated expectations, that

$$\frac{\partial}{\partial t} \mathrm{E}\Big[X\Big(h(F_{Y|X} + t\widetilde{F}) + a(Y, F_{Y|X} + t\widetilde{F}) - X'\bar{\beta}\Big)\Big]\Big|_{t=0_+} = 0,$$

as long as integration and differentiation can be interchanged. This means that our OLS estimator solves a system of equations having the Neyman orthogonality property with respect to $F_{Y|X}$ (Chernozhukov et al., 2018) and is, in this sense, robust with respect to the estimation error in $\widehat{F}_{Y|X}$.

Intuitively, a simple approach to estimate $\beta$ in the model (4) would be to run the OLS estimator of $h(\widehat{F}_{Y|X})$ on $X$. Such an estimator can be shown to be $\sqrt{T}$-consistent and asymptotically normal with mean zero as long as $X$ is low-dimensional, a sufficiently simple estimator $\widehat{F}_{Y|X}$ is used, and its tuning parameters are chosen in a delicate way. When $X$ is moderate- or even large-dimensional, however, we have to rely on machine learning methods to obtain an estimator $\widehat{F}_{Y|X}$. These methods in turn yield heavily biased estimators with relatively slow convergence rates. The estimation error in $\widehat{F}_{Y|X}$ may then propagate into the error of the OLS estimator, leading to estimates of $\beta$ with poor properties. We deal with this problem by adding the influence function $a(Y, \widehat{F}_{Y|X})$ to the functional $h(\widehat{F}_{Y|X})$. This allows us to obtain the OLS estimator of $\beta$ that is robust with respect to the estimation error in $\widehat{F}_{Y|X}$, as explained above.

There are several strands of literature that use influence function adjustments for estimation. First, the so-called one-step estimators, which adjust the plug-in estimators by adding the average value of the estimated influence function, have been used in statistics for a long time as a tool of achieving semiparametric efficiency; see

Bickel (1982) and Schick (1986) for early references and Fisher and Kennedy (2021) for a recent review. Second, Chernozhukov et al. (2016) introduced the idea of adding the influence functions to obtain robust estimators in the setting where the parameter of interest is the expectation of a functional of unknown nonparametric/high-dimensional object that has to be estimated on the first step. We show that in our context the adjustment function $a$ depends only on the functional $h$ and not on the joint distribution of underlying random variables, and thus has a simple form, broadly available in applied settings. We exemplify this last point in the next subsection, where we apply our procedure to the weighted-average quantile regression. Third, Firpo et al. (2009) proposed a procedure that, in its simplest form, consists of running the OLS estimator of $h(\widehat{F}_Y) + a(Y, \widehat{F}_Y)$ on $X$, where $\widehat{F}_Y$ is a preliminary estimator of the distribution function $F_Y$ of $Y$, in order to estimate the impact of $X$ on the functionals of the counterfactual distribution of $Y$ that appears as the distribution of $X$ changes keeping the conditional distribution of $Y$ given $X$ fixed. They use the *marginal* distribution of $Y$, whereas we use the *conditional* distribution of $Y$ given $X$. This seemingly minor change has substantial consequences: two procedures aim at estimating fundamentally different quantities and have different scopes of applicability. Using an example from the Introduction of Firpo et al. (2009), one can say that their procedure estimates how increasing the fraction of unionized workers affects the distribution of wages whereas our procedure estimates how the distribution of wages of unionized workers differs from the distribution of wages of non-unionized workers. In addition, the reasons for adding the influence function in two procedures are completely different. As explained above, we use the influence function adjustment to obtain an OLS estimator that is robust with respect to the estimation error in $\widehat{F}_{Y|X}$, whereas they use the influence function because it directly measures the impact of changing the distribution on the value of the functional, see (5).[2]

More generally, our approach to estimation in this section is an instance of the double/debiased machine learning method (Chernozhukov et al., 2018), which gives estimation procedures based on moment conditions with the Neyman orthogonality property. Our key innovation here is that we demonstrate that although we are interested in an object $h(F_{Y|X})$ that depends on *conditional* distributions, with dependence going through in a potentially complicated way, e.g. via conditional quantile

[2]Also, as long as the intercept is included, the term $h(\widehat{F}_Y)$ plays no role in their procedure: the slope coefficients of the OLS estimator of $h(\widehat{F}_Y) + a(Y, \widehat{F}_Y)$ on $X$ coincide with the slope coefficients of the OLS estimator of $a(Y, \widehat{F}_Y)$ on $X$. In contrast, dropping $h(\widehat{F}_{Y|X})$ in our procedure would lead to a meaningless estimator that would converge in probability to the vector of zeros.

functions, obtaining moment conditions with the Neyman orthogonality property is actually simple: we simply have to add the influence function for the functional $h$, which can be obtained by looking at the values $h(F)$ of the functional $h$ at *unconditional* distributions $F$. In addition, an important issue that arises in almost all of our applications is that the functionals we consider are not twice continuously differentiable (in a Gateaux sense), thus the results of Chernozhukov et al. (2018) can not be applied.

3.2. **Application to Weighted-Average Quantile Regression.** Here, we apply the general principle described in the previous subsection to the weighted-average quantile regression model (1). To do so, we set $h(F) = \int_0^1 q_F(u)\psi(u)du$, where $q_F(u)$ denotes the $u$th quantile of the distribution function $F \in \mathcal{F}$. The influence function for this functional is well-known:

$$a(y, F) = \int_0^1 \frac{u - \mathbb{I}\{y \leq q_F(u)\}}{f(q_F(u))}\psi(u)du, \tag{9}$$

where $f = F'$ is the pdf corresponding to $F$. To see why, suppose first that we are interested in the individual quantile $q_F(u)$ for some $u \in (0, 1)$. Let $F$ and $G$ be two distribution functions with strictly positive derivatives $f$ and $g$, respectively. Then for any $t \in [0, 1]$, we have

$$\int_{-\infty}^{q_{(1-t)F+tG}(u)} ((1-t)f(y) + tg(y))dy = u.$$

Taking derivative of both sides with respect to $t$ at $t = 0_+$, we then obtain

$$f(q_F(u))\frac{\partial}{\partial t}q_{(1-t)F+tG}(u)\Big|_{t=0_+} + \int_{-\infty}^{q_F(u)} (g(y) - f(y))dy = 0.$$

Hence,

$$\begin{aligned}
\frac{\partial}{\partial t}q_{(1-t)F+tG}(u)\Big|_{t=0_+} &= \int_{-\infty}^{q_F(u)} \frac{f(y) - g(y)}{f(q_F(u))}dy \\
&= \frac{u}{f(q_F(u))} - \int_{-\infty}^{+\infty} \frac{\mathbb{I}\{y \leq q_F(u)\}g(y)}{f(q_F(u))}dy \\
&= \int_{-\infty}^{+\infty} \frac{u - \mathbb{I}\{y \leq q_F(u)\}}{f(q_F(u))}g(y)dy.
\end{aligned}$$

Comparing this expression with (5), we thus obtain the influence function for the individual quantile $q_F(u)$:

$$y \mapsto \frac{u - \mathbb{I}\{y \leq q_F(u)\}}{f(q_F(u))}.$$

This expression in turn immediately gives (9) in light of linear additivity of derivatives.

Applying our general principle from the previous subsection, we conclude that in order to estimate $\beta$ in the model (1), we can run OLS of an estimated version of

$$\int_0^1 \left( q_{Y|X}(u) + \frac{u - \mathbb{I}\{Y \leq q_{Y|X}(u)\}}{f_{Y|X}(q_{Y|X}(u))} \right) \psi(u) du \tag{10}$$

on $X$, where $f_{Y|X}$ is the pdf of the conditional distribution of $Y$ given $X$. This, however, is not convenient for two reasons. First, this approach requires estimating extreme quantiles, $q_{Y|X}(u)$ for $u$ close to the boundaries of the interval $[0,1]$, which are typically difficult to estimate. Second, this approach requires estimating the pdf $f_{Y|X}(q_{Y|X}(u))$, which appears in the denominator and which may take small values near the boundaries of the interval $[0,1]$, thus leading to large estimation errors. Note also that simply truncating the interval $[0,1]$ may not lead to good results as in some cases, such as the expected shortfall or inequality regressions, extreme quantiles are of particular importance. To deal with these problems, we rewrite the integral in (10) differently.

First, observe that for all $u \in [0,1]$, we have

$$\int_{-\infty}^{q_{Y|X}(u)} f_{Y|X}(y) dy = u,$$

and so $f_{Y|X}(q_{Y|X}(u)) q'_{Y|X}(u) = 1$ almost surely. Therefore, by applying the change of variables $u \mapsto s(u) = q_{Y|X}(u)$ and recalling that $u = F_{Y|X}(s(u))$ in this case, it follows that the integral in (10) is equal to

$$\begin{aligned} &\int_{-\infty}^{+\infty} \left( s + \frac{F_{Y|X}(s) - \mathbb{I}\{Y \leq s\}}{f_{Y|X}(s)} \right) f_{Y|X}(s) \psi(F_{Y|X}(s)) ds \\ &\quad = \int_{-\infty}^{+\infty} s f_{Y|X}(s) \psi(F_{Y|X}(s)) ds + \int_{-\infty}^{+\infty} (F_{Y|X}(s) - \mathbb{I}\{Y \leq s\}) \psi(F_{Y|X}(s)) ds. \end{aligned} \tag{11}$$

Second, using integration by parts, we can further rewrite the first integral in (11) as

$$\begin{aligned} &\int_{-\infty}^{+\infty} s f_{Y|X}(s) \psi(F_{Y|X}(s)) ds \\ &\quad = \int_{-\infty}^{0} s f_{Y|X}(s) \psi(F_{Y|X}(s)) ds + \int_{0}^{+\infty} s f_{Y|X}(s) \psi(F_{Y|X}(s)) ds \\ &\quad = -\int_{-\infty}^{0} \Psi(F_{Y|X}(s)) ds + \int_{0}^{+\infty} (\bar{\Psi} - \Psi(F_{Y|X}(s))) ds, \end{aligned}$$

where $\Psi\colon [0,1] \to \mathbb{R}$ is the function defined by $\Psi(s) = \int_0^s \psi(u)du$ for all $s \in [0,1]$ and $\bar{\Psi} = \int_0^1 \psi(u)du$. Combining these results, it follows that the integral in (10) is equal to

$$R = -\int_{-\infty}^{0} \Psi(F_{Y|X}(s))ds + \int_0^{+\infty} (\bar{\Psi} - \Psi(F_{Y|X}(s)))ds + \int_{-\infty}^{+\infty} (F_{Y|X}(s) - \mathbb{I}\{Y \leq s\})\psi(F_{Y|X}(s))ds. \tag{12}$$

We thus propose estimating the vector of parameters $\beta$ in the weighted-average quantile regression model (1) by running OLS of an estimated version of $R$ on $X$. In comparison with the integral in (10), the advantage of using $R$ is that it depends only on the distribution function $F_{Y|X}$, which is easy to estimate even in the tails.

## 4. Estimation and Inference

In this section, we provide a detailed discussion of our estimation and inference procedures for the vector of parameters $\beta$ in the weighted-average quantile regression model (1). We assume, throughout the rest of the paper, that we have a strictly stationary time series dataset $(X_1, Y_1), \ldots, (X_T, Y_T)$, where each $(X_t, Y_t)$ has the same distribution as that of the pair $(X, Y)$.

For all $s \in \mathbb{R}$ and $x$ in the support of $X$, denote $F(s|x) = \mathrm{P}(Y \leq s|X = x)$, so that $F(\cdot|X) = F_{Y|X}(\cdot)$ is the distribution function of the conditional distribution of $Y$ given $X$. Also, as in the previous section, denote $\Psi(s) = \int_0^s \psi(u)du$ for all $s \in [0,1]$ and $\bar{\Psi} = \int_0^1 \psi(u)du$. In addition, define $R$ as in (12) and

$$e = \int_{-\infty}^{+\infty} (F(s|X) - \mathbb{I}\{Y \leq s\})\,\psi(F(s|X))ds. \tag{13}$$

By the law of iterated expectations, we then have $\mathrm{E}[e|X] = 0$. In addition, by discussion at the end of the previous section, we also have

$$\int_0^1 q_{Y|X}(u)\psi(u)du = -\int_{-\infty}^{0} \Psi(F(s|X))ds + \int_0^{+\infty} (\bar{\Psi} - \Psi(F(s|X)))ds.$$

Thus, it follows from (1) that

$$R = X'\beta + e, \quad \text{where} \quad \mathrm{E}[e|X] = 0. \tag{14}$$

This equation reinforces our proposal in the previous section to estimate $\beta$ by the OLS method, regressing $R$ on $X$, where the distribution function $F$ appearing in $R$ is replaced by a suitable nonparametric/machine learning estimator, to be discussed

later. For technical reasons, we also rely on sample splitting, so that the function $F$ and the vector $\beta$ are estimated on different subsamples of the whole sample. Formally, we define our estimator of $\beta$ as follows:

(1) split the full sample $(X_1, Y_1), \ldots, (X_T, Y_T)$ into two consequtive subsamples, say, $(X_1, Y_1), \ldots, (X_{T_1}, Y_{T_1})$ and $(X_{T_1+1}, Y_{T_1+1}), \ldots, (X_{T_1+T_2}, Y_{T_1+T_2})$, where $T_1 + T_2 = T$;
(2) use the first subsample, $(X_1, Y_1), \ldots, (X_{T_1}, Y_{T_1})$, to obtain a nonparametric/machine learning estimator $\widehat{F}(s|x)$ of $F(s|x)$ for all $x \in \{X_{T_1+1}, \ldots, X_{T_1+T_2}\}$ and $s \in \mathbb{R}$;
(3) calculate

$$\widehat{R}_t = \int_0^{+\infty} (\bar{\Psi} - \Psi(\widehat{F}(s|X_t)))ds - \int_{-\infty}^{0} \Psi(\widehat{F}(s|X_t))ds + \int_{-\infty}^{+\infty} \left(\widehat{F}(s|X_t) - \mathbb{I}\{Y_t \leq s\}\right) \psi(\widehat{F}(s|X_t))ds, \quad (15)$$

for all $t = T_1 + 1, \ldots, T_1 + T_2$;
(4) calculate the OLS estimator

$$\widehat{\beta} = \left(\sum_{t=T_1+1}^{T_1+T_2} X_t X_t'\right)^{-1} \left(\sum_{t=T_1+1}^{T_1+T_2} X_t \widehat{R}_t\right). \quad (16)$$

In this procedure, $T_1$ and $T_2$ should be chosen to be of the same order, which we assume for the rest of the paper. In our simulations, we choose $T_1 \approx 2T_2$.

By analogy with mean and quantile regression estimators, we refer to $\widehat{\beta}$ as the *weighted-average quantile regression estimator*. By substituting various weighting functions $\psi$ (and the corresponding $\Psi$), we obtain various regression estimators. For instance, if $Y$ is the loss of a financial portfolio, by setting $\psi(u) = \mathbb{I}\{u \geq 1 - \alpha\}/\alpha$, we obtain an expected shortfall regression estimator, as discussed in Section 2.

We will prove in the next section that, under suitable regularity conditions, the estimator $\widehat{\beta}$ is $\sqrt{T}$-consistent and asymptotically normal with mean zero:

$$\sqrt{T_2}(\widehat{\beta} - \beta) = \left(\frac{1}{T_2}\sum_{t=T_1+1}^{T_1+T_2} X_t X_t'\right)^{-1} \left(\frac{1}{\sqrt{T_2}}\sum_{t=T_1+1}^{T_1+T_2} X_t e_t\right) + o_P(1) \to_d N(0, \Sigma), \quad (17)$$

where

$$e_t = \int_{-\infty}^{+\infty} \left(F(s|X_t) - \mathbb{I}\{Y_t \leq s\}\right) \psi(F(s|X_t))ds, \quad \text{for all } t = T_1 + 1, \ldots, T_1 + T_2,$$

and $\Sigma$ is the asymptotic covariance matrix. Moreover, $\Sigma$ can be consistently estimated, for example, by the Newey-West method, where each $e_t$ is replaced by the corresponding residual $\widehat{e}_t = \widehat{R}_t - X_t'\widehat{\beta}$; see the next section for details. This implies that once an estimator of the function $F$ is obtained using data from the first subsample, any standard statistical software can be used to obtain the estimator $\widehat{\beta}$ and corresponding standard errors by simply running the OLS regression of $\widehat{R}_t$ on $X_t$ using data from the second subsample and reporting the Newey-West standard errors.

Next, we discuss estimation of the function $F$. To do so, fix any $s \in \mathbb{R}$ and observe that $F(s|x) = \mathrm{P}(Y \leq s|X = x) = \mathrm{E}[\mathbb{I}\{Y \leq s\}|X = x]$ for all $x \in \mathbb{R}^p$. Therefore, to obtain an estimator of the function $x \mapsto F(s|x)$, we can apply any standard nonparametric/machine learning estimator regressing $\mathbb{I}\{Y_t \leq s\}$ on $X_t$ using data from the first subsample. Applying the estimator separately for each value of $s$, we obtain an estimator $(s, x) \mapsto \widehat{F}(s|x)$ of the function $(s, x) \mapsto F(s|x)$. For example, for our empirical results, we use a version of the random forest method described below.[3]

Further, nonparametric/machine learning estimators will produce numerically identical results for any pair of values of $s$, say $s_1$ and $s_2$, such that there is no $Y_t$ between $s_1$ and $s_2$ since the datasets $\{(X_t, \mathbb{I}\{Y_t \leq s_1\})\}_{t=1}^{T_1}$ and $\{(X_t, \mathbb{I}\{Y_t \leq s_2\})\}_{t=1}^{T_1}$ are identical in this case. This in turn implies that there is no need to apply the nonparametric/machine learning estimator for all values of $s \in \mathbb{R}$, and it suffices to only consider $s \in \{Y_1, \ldots, Y_{T_1}\}$ since the estimator $s \mapsto \widehat{F}(s|x)$ will be piecewise constant and the integrals in (15) will be given by the corresponding sums. More precisely, letting $s_1 \leq \cdots \leq s_{T_1}$ be the sequence of values of $Y_1, \ldots, Y_{T_1}$ arranged in the increasing order and imposing a mild condition that $\widehat{F}(s|X) = 0$ for all $s < \min_{1 \leq t \leq T_1} Y_t$ and $\widehat{F}(s|X) = 1$ for all $s \geq \max_{1 \leq t \leq T_1} Y_t$ a.s., which is satisfied for any reasonable nonparametric/machine learning estimator, it follows that

$$\widehat{R}_t = s_{T_1}\bar{\Psi} + \sum_{j=1}^{T_1-1}(s_{j+1} - s_j)M_t(s_j, s_{j+1}), \quad \text{for all } t = T_1 + 1, \ldots, T_1 + T_2 \tag{18}$$

where we denoted

$$M_t(s_j, s_{j+1}) = -\Psi(\widehat{F}(s_j|X_t)) + (\widehat{F}(s_j|X_t) - \widetilde{\mathbb{I}}\{Y_t \leq s_j, s_{j+1}\})\psi(\widehat{F}(s_j|X_t)) \tag{19}$$

and

$$\widetilde{\mathbb{I}}\{Y_t \leq s_j, s_{j+1}\} = \max\left(\min\left(\frac{s_{j+1} - Y_t}{s_{j+1} - s_j}, 1\right), 0\right)$$

[3]We also tried boosting and $\ell_1$-penalized methods but they did not perform as well as random forests: both methods turned out slower and the latter method suffered from potentially substantial non-linearities in the functions $x \mapsto F(s|x)$.

for all $j = 1, \dots, T_1 - 1$ and $t = T_1 + 1, \dots, T_1 + T_2$.

Moreover, in large samples, where calculating $\widehat{R}_t$ in (18) is computationally costly, the grid $s_1 \leq \cdots \leq s_{T_1}$ used in (18) can be replaced by a much coarser grid $\min_{1 \leq t \leq T_1} Y_t = s_1^* \leq \cdots \leq s_k^* = \max_{1 \leq t \leq T_1} Y_t$, so that

$$\widehat{R}_t = s_k^* + \sum_{j=1}^{k-1} (s_{j+1}^* - s_j^*) M_t(s_j^*, s_{j+1}^*), \quad \text{for all } t = T_1 + 1, \dots, T_1 + T_2,$$

where $k$ is much smaller than $T_1$.

## 5. Asymptotic Theory

In this section, we derive an asymptotic theory for the weighted-average quantile regression estimator $\widehat{\beta}$. To do so, we denote $D_t = (X_t', Y_t)'$ for all $t = 1, \dots, T$ and $D_1^{T_1} = (D_1, \dots, D_{T_1})$. We will assume that the dataset $D_1, \dots, D_T$ is a subset of a strictly stationary time series $\{D_t\}_{t \in \mathbb{Z}}$. Further, for all $j \in \mathbb{N}$, let $\mathcal{I}_{-\infty}^0$ and $\mathcal{I}_j^{+\infty}$ be $\sigma$-algebras generated by $\{D_s\}_{s \leq 0}$ and $\{D_s\}_{s \geq j}$, respectively, and let

$$\beta_j = \mathrm{E}\left[\sup\left\{|\mathrm{P}(B \mid \mathcal{I}_{-\infty}^0) - \mathrm{P}(B)| \colon B \in \mathcal{I}_j^{+\infty}\right\}\right]$$

be the $\beta$-mixing coefficients. In addition, let $\mathcal{X}$ be the support of $X$ and for all $x \in \mathcal{X}$, denote

$$\Delta(x) = \sup_{s \in \mathbb{R}} \left|\widehat{F}(s|x) - F(s|x)\right| + \int_{-\infty}^{+\infty} \left|\widehat{F}(s|x) - F(s|x)\right| ds. \tag{20}$$

Moreover, let $0 < u_0 < 1/2$, $0 < c < \infty$, and $-\infty < s_1 < s_2 < +\infty$ be some constants. We will use the following assumptions.

**Assumption 5.1.** *The strictly stationary time series $\{D_t\}_{t \in \mathbb{Z}}$ has summable $\beta$-mixing coefficients: $\sum_{j=1}^{\infty} \beta_j < \infty$.*

Assumption 5.1 implies that the time series $\{D_t\}_{t \in \mathbb{Z}}$ is absolutely regular. As explained in Chen (2011), many econometric time series models satisfy this assumption. In fact, it is common practice to impose stronger mixing conditions. For example, Chen and Liao (2013) require that $\beta_j \leq \beta_0 j^{-\omega}$ for all $j \geq 1$ and some $\beta_0 > 0$ and $\omega > 2$, which clearly implies that $\sum_{j=1}^{\infty} \beta_j < \infty$. Fan et al. (2016) require that $\phi_j \leq \phi_0 j^{-\omega}$ for all $j \geq 1$ and some $\phi_0 > 0$ and $\omega > 1$, where $\phi_j$'s are $\phi$-mixing coefficients. Since $\phi_j \geq \beta_j$ for all $j \geq 1$, such a condition also implies our Assumption 5.1. Note also that Assumption 5.1 holds trivially if the random vectors $D_t$ are independent across $t$, which means that our results apply for i.i.d. data settings as well.

We refer an interested reader to Fan and Yao (2005) and Bradley (2005) for detailed explanations on various mixing conditions and their plausibility.

**Assumption 5.2.** *(i) Components of the random vector $X$ as well as the random variable $e$ have finite fourth moments:* $\mathrm{E}[\|X\|^4] < \infty$ *and* $\mathrm{E}[e^4] < \infty$. *(ii) In addition, the matrix* $\mathrm{E}[XX']$ *is positive-definite. (iii) Moreover,*

$$\frac{1}{\sqrt{T_2}} \sum_{t=T_1+1}^{T_1+T_2} e_t X_t \to_d N(0, \Omega)$$

*for a positive-definite matrix* $\Omega$.

Assumption 5.2 is standard in time series econometrics. Assumption 5.2(i) is a mild moment condition. Assumption 5.2(ii) is an identification condition. Assumption 5.2(iii) follows from a combination of mixing and moment conditions. For example, since $\beta$-mixing coefficients dominate $\alpha$-mixing coefficients, under Assumption 5.1, Assumption 5.2(iii) holds as long as the random variables $\|e_t X_t\|$ are bounded; see Theorem 2.21(ii) in Fan and Yao (2005). More generally, when the random variables $\|e_t X_t\|$ satisfy $\mathrm{E}[\|e_t X_t\|^\eta] < \infty$ for some $\eta > 2$, Assumption 5.2(iii) holds as long as $\sum_{j=1}^{\infty} \beta_j^{1-2/\eta} < \infty$; see Theorem 2.21(i) in Fan and Yao (2005).

**Assumption 5.3.** *(i) The weighting function $\psi$ has bounded variation. (ii) In addition, $\psi$ is continuously differentiable on $(0, u_0)$ and $(1-u_0, 1)$ with bounded derivative.*

Assumption 5.3(i) means that the function $\psi$ can be decomposed as $\psi = \psi_1 - \psi_2$, where both $\psi_1$ and $\psi_2$ are bounded and increasing functions. Assumption 5.3 is thus satisfied in all our examples from Section 2.

**Assumption 5.4.** *(i) The function $F$ is such that $F(s_1|x) < u_0/2$ and $F(s_2|x) > 1 - u_0/2$ for all $x \in \mathcal{X}$. (ii) In addition, the function $u \mapsto F(u|x)$ is continuously differentiable on $u \in (s_1, s_2)$ with derivative $u \mapsto f(u|x)$ satisfying $f(u|x) \geq c$ for all $u \in (s_1, s_2)$ and $x \in \mathcal{X}$.*

This assumption imposes mild regularity conditions on the conditional distribution of $Y$ given $X$. This assumption can be avoided if the function $\psi$ is smooth.

**Assumption 5.5.** *(i) The estimator $\widehat{F}$ is such that*

$$\mathrm{P}\left(\sup_{x \in \mathcal{X}} \Delta(x) > u_0/2\right) \to 0.$$

*(ii) In addition,*

$$\sum_{t=T_1+1}^{T_1+T_2} \mathrm{E}\left[(1+\|X_t\|)^4 \Delta(X_t)^4 \mid D_1^{T_1}\right] = o_P(1). \tag{21}$$

Assumption 5.5 means that the estimator $\widehat{F}$ converges to the function $F$ sufficiently fast. To obtain some intuition about this assumption, note that under Assumptions 5.1 and 5.2(i), it holds as long as $\sup_{x\in\mathcal{X}} \Delta(x) = o_P(T^{-1/4})$, which is plausible for nonparametric/machine learning estimators $\widehat{F}$. Note, however, that (21) does not actually require a bound on the supremum of the function $\Delta$ and instead uses a suitable weighted average value of this function, which is typically easier to bound.

We are now ready to state the main result of this section:

**Theorem 5.1.** *Under Assumptions 5.1–5.5,*

$$\sqrt{T_2}(\widehat{\beta}-\beta) = \left(\frac{1}{T_2}\sum_{t=T_1+1}^{T_1+T_2} X_t X_t'\right)^{-1} \left(\frac{1}{\sqrt{T_2}}\sum_{t=T_1+1}^{T_1+T_2} X_t e_t\right) + o_P(1) \to_d N(0,\Sigma), \tag{22}$$

*where* $\Sigma = (\mathrm{E}[XX'])^{-1}\Omega(\mathrm{E}[XX'])^{-1}$.

**Remark 5.1** (Relation to Double/Debiased Machine Learning)**.** As discussed in Section 3, our approach to estimation of weighted-average quantile regressions is an instance of the double/debiased machine learning method; e.g. Chernozhukov et al. (2018). In particular, we constructed the random variable $R$ in (12) so that (i) $\mathrm{E}[XR] = \mathrm{E}[XX']\beta$, meaning that the least squares projection of $R$ on $X$ correctly identifies $\beta$, and (ii) $\mathrm{E}[XR]$ is first-order insensitive with respect to perturbations of the function $F$, appearing in the definition of $R$. The latter condition, commonly referred to as the Neyman orthogonality, means that if we define

$$\begin{aligned} R(\eta_1,\eta_2) = &\int_0^{+\infty} (1-\Psi(\eta_1(s|X)))ds - \int_{-\infty}^{0} \Psi(\eta_1(s|X))ds \\ &+ \int_{-\infty}^{+\infty} (\eta_1(s|X) - \mathbb{I}\{Y\le s\})\eta_2(s|X)ds, \end{aligned}$$

for all functions $(s,x)\mapsto \eta_1(s|x)$ and $(s,x)\mapsto \eta_2(s|x)$, then $R(F,\psi(F)) = R$ and first-order Gateaux derivatives of the functions $\eta_1 \mapsto \mathrm{E}[XR(\eta_1,\psi(F))]$ and $\eta_2 \mapsto \mathrm{E}[XR(F,\eta_2)]$ at $\eta_1 = F$ and $\eta_2 = \psi(F)$, respectively, both vanish. It is this last condition that allows us to derive asymptotic normality of our estimator $\widehat{\beta}$ under weak conditions on the estimator $\widehat{F}$ of the function $F$, as specified in Assumption 5.5 (in particular, we do not need to impose the common small bias condition). Our

results do not follow from those in Chernozhukov et al. (2018) because the function $\Psi$ is not necessarily continuously differentiable under our assumptions (and in fact has kinks in most examples from Section 2), and so the function $\eta_1 \mapsto \mathrm{E}[XR(\eta_1, \psi(F))]$ does not necessarily have the second-order Gateaux derivative, which is assumed to exist and is required to be suitably bounded in Chernozhukov et al. (2018).[4] Instead, our results employ the smoothing properties of the integrals in the definition of $R$ in (14). ∎

**Remark 5.2** (Cross-Fitting and I.I.D. Setting)**.** Throughout this paper, we are assuming that the observations $D_1, \ldots, D_T$ are coming from a time series under mixing conditions. Of course, this setting covers the case of i.i.d. observations as well. However, we can construct a more efficient estimator in the latter case via cross-fitting, e.g. Chernozhukov et al. (2018). Indeed, let $\widehat{\beta}_1$ be the estimator $\widehat{\beta}$ defined in (16). In addition, let

$$\widehat{\beta}_2 = \left(\sum_{t=1}^{T_1} X_t X_t'\right)^{-1} \left(\sum_{t=1}^{T_1} X_t \widehat{R}_t\right),$$

where $\widehat{R}_t$ is defined by (15) with the estimator $\widehat{F}$ being constructed using data from the second subsample, $D_{T_1+1}, \ldots, D_{T_1+T_2}$. It is then straightforward to show that the estimator $\widehat{\beta} = (\widehat{\beta}_1 + \widehat{\beta}_2)/2$ will satisfy

$$\sqrt{T}(\widehat{\beta} - \beta) = \left(\frac{1}{T_2}\sum_{t=1}^{T} X_t X_t'\right)^{-1} \left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_t e_t\right) + o_P(1) \to_d N(0, \overline{\Sigma}),$$

where

$$\overline{\Sigma} = (\mathrm{E}[XX'])^{-1} \left(\mathrm{E}[e^2 XX']\right) (\mathrm{E}[XX'])^{-1}.$$

For estimation of $\overline{\Sigma}$ and construction of standard errors and confidence intervals, it is then possible to use the conventional Eicker-Huber-White formula. ∎

Next, we consider consistent estimation of the covariance matrix $\Sigma$ appearing in Theorem 5.1. As discussed in the previous section, we focus on the Newey-West estimator:

$$\widehat{\Sigma} = \left(\frac{1}{T_2}\sum_{t=T_1+1}^{T_1+T_2} X_t X_t'\right)^{-1} \left(\widehat{\Omega}_0 + \sum_{j=1}^{m} w(j,m)(\widehat{\Omega}_j + \widehat{\Omega}_j')\right) \left(\frac{1}{T_2}\sum_{t=T_1+1}^{T_1+T_2} X_t X_t'\right)^{-1}, \quad (23)$$

[4]Gateaux differentiability can be retained by assuming that the function $s \mapsto \widehat{F}(s|X)$ is increasing almost surely but this is unattractive because machine learning methods may or may not give increasing estimators and applying monotonization procedure may be computationally costly since the procedure would have to be carried out for each observation separately.

where $m$ is a tuning parameter, $w$ is a weighting function, and

$$\widehat{\Omega}_j = \frac{1}{T_2 - T_1} \sum_{t=T_1+j+1}^{T_1+T_2} \widehat{e}_t \widehat{e}_{t-j} X_t X_{t-j}', \quad \text{for all } j = 0, \ldots, m. \tag{24}$$

The tuning parameter $m$ is often chosen so that $m = m(T) \to \infty$ as $T \to \infty$ and the weighting function is typically defined by $w(j, m) = 1 - j/(m+1)$ for all $j = 1, \ldots, m$.

To prove consistency of the estimator $\widehat{\Sigma}$, we will need the following additional notation:

$$\bar{\Omega} = \bar{\Omega}_0 + \sum_{j=1}^{m} w(j, m)(\bar{\Omega}_j + \bar{\Omega}_j'),$$

where

$$\bar{\Omega}_j = \frac{1}{T_2 - T_1} \sum_{t=T_1+j+1}^{T_1+T_2} e_t e_{t-j} X_t X_{t-j}', \quad \text{for all } j = 0, \ldots, m.$$

We will impose the following assumptions.

**Assumption 5.6.** *(i) The matrix $\bar{\Omega}$ is consistent for $\Omega$: $\bar{\Omega} \to_P \Omega$. (ii) In addition, the weighting function $w$ is such that $0 \leq w(j, m) \leq 1$ for all $j = 1, \ldots, m$. (iii) Moreover, the smoothing parameter $m$ is such that $m = o(T^{1/4})$.*

Assumption 5.6(i) is a high-level condition that is familiar from the literature. Primitive conditions ensuring that this assumption is satisfied can be found in Newey and West (1987). Assumption 5.6(ii) is satisfied if we set $w(j, m) = 1 - j/(m+1)$, for example, which is a typical choice for the weighting function. Assumption 5.6(iii) is a mild growth condition meaning that the tuning parameter $m$ should not grow too fast as $T$ gets large.

In the next result, we prove consistency of the estimator $\widehat{\Sigma}$.

**Theorem 5.2.** *Under Assumptions 5.1–5.6, the estimator $\widehat{\Sigma}$ is consistent:*

$$\widehat{\Sigma} \to_P \Sigma.$$

**Remark 5.3** (Transformations of $X$)**.** Although we focused on the case of linear weighted-average quantile regressions $\int_0^1 q_{Y|X}(u)\psi(u)du = X'\beta$ throughout the paper, inspecting the proofs reveals that our results equally apply to the more general case where we include transformations of $X$, such as interactions and other higher-order polynomial terms, on the right-hand side of the regression: $\int_0^1 q_{Y|X}(u)\psi(u)du = p(X)'\beta$, where $x \mapsto p(x) = (p_1(x), \ldots, p_k(x))'$ is a vector of transformations. In this case, one should simply replace all $X_t$'s in (16), (23), and (24) by the corresponding $p(X_t)$'s. Theorems 5.1 and 5.2 still apply in this case modulo obvious modifications.■

**Remark 5.4** (Weighted-Average Quantile Regression Estimator as Best Linear Predictor)**.** When the weighted-average quantile function $x \mapsto \int_0^1 q_{Y|X=x}(u)\psi(u)du$ is not linear, i.e. there exists no $\beta$ such that $\int_0^1 q_{Y|X}(u)\psi(u)du = X'\beta$, it is straightforward to check that Theorems 5.1 and 5.2 continue to hold with

$$\beta = \arg\min_{b\in\mathbb{R}^p} \mathrm{E}\left[\left(\int_0^1 q_{Y|X}(u)\psi(u)du - X'b\right)^2\right].$$

It thus follows that our weighted-average quantile regression method consistently estimate parameters of the best linear approximation to $x \mapsto \int_0^1 q_{Y|X=x}(u)\psi(u)du$. In this sense, running our estimator makes sense even if it is believed that the linearity assumption of the regression model (1) is satisfied not exactly but only approximately. ■

## 6. Monte Carlo Simulation Study

In this section, we present results of a small-scale Monte Carlo simulation study that sheds light on finite-sample properties of the weighted-average quantile regression estimator.

We consider the following data-generating processes:

$$\begin{aligned} &\text{DGP1:} \quad Y = \varepsilon - X'\bar{\beta}, \\ &\text{DGP2:} \quad Y = (1 + 0.2X^1)\varepsilon - X'\bar{\beta}. \end{aligned}$$

Depending on the experiment, $X = (X^1, \ldots, X^p)'$ and $\bar{\beta} = (\bar{\beta}_1, \ldots, \bar{\beta}_p)'$ are vectors either in $\mathbb{R}^2$ or in $\mathbb{R}^5$, so that $p = 2$ or 5, respectively. in the former case, we set $\bar{\beta}_2 = 0.5$ and vary $\bar{\beta}_1$ over $\{0, 0.3, 0.6, 0.9\}$. In the latter case, we set $\bar{\beta}_2 = 0.5$, $\bar{\beta}_3 = \bar{\beta}_4 = \bar{\beta}_5 = 0$, and again vary $\bar{\beta}_1$ over $\{0, 0.3, 0.6, 0.9\}$. We consider cases with $\varepsilon \sim N(0,1)$ and $\varepsilon \sim t(4)$. In both cases, $\varepsilon$ is independent of $X$. Note that in the latter case, Assumption 5.2 is actually not satisfied, as random variables with the $t(4)$ distribution have finite moments up-to the 4th order but excluding the 4th order itself, and so this case serves as a way to check whether our methods continue to work if our asymptotic theory assumptions are slightly violated. Finally, we set $X = (X^1, \ldots, X^p)'$ so that $X^1 = |\widetilde{X}^1|$ and $X^j = \widetilde{X}^j$ for all $j = 2, \ldots, p$, where $\widetilde{X} = (\widetilde{X}^1, \ldots, \widetilde{X}^p)'$ is a standard normal random vector in $\mathbb{R}^p$. For simplicity, we assume that the data $(X_1, Y_1), \ldots, (X_T, Y_T)$ consists of $T$ i.i.d. realizations of the pair $(X, Y)$. We consider samples of size $T = 1000$ and 2000.

As a machine learning estimator of the function $F$, we use a version of a random forest. Specifically, recall that any random forest estimator takes the weighted-average

form, i.e. an estimator of $\mathrm{E}[V|Z=z]$ based on the data $(Z_1,V_1),\dots,(Z_T,V_T)$ will take the form $\widehat{\mathrm{E}}[V|Z=z]=\sum_{t=1}^T w_t V_t$. We do two changes to this estimator. First, once we have the weights $w_t$ from the random forest, we replace the weighted-average estimator by a local linear estimator:

$$\widehat{\mathrm{E}}[V|Z=z]=z'\arg\min_b \sum_{t=1}^T w_t(V_t-Z_t'b)^2.$$

This helps to improve the accuracy of the estimator, e.g. Friedberg et al. (2021). Second, recall that we need an estimator of $x\mapsto \widehat{F}(s|x)$ for all $s\in\{Y_1,\dots,Y_{T_1}\}$, which is computationally straightforward but costly when $T_1$ is large. We therefore first split the interval $[\min_{1\le t\le T_1} Y_t, \max_{1\le t\le T_1} Y_t]$ into $\log T_1$ equal intervals, calculate random forest weights with $s$ being the center of each interval, and then apply the same weights for all $s$ in the same interval. This substantially reduces computational costs as we now need to calculate only $\log T_1$ random forests rather than $T_1$ of them. Closely related ideas were previously used in Meinshausen (2006) who constructed a quantile random forest by applying a local linear quantile estimator based on weights obtained from a (mean) random forest. The main reason we rely on random forest estimators in this paper is that they are easy to train and allow for computational simplifications as explained here.

Also, for our simulations, we set $T_1=2T_2$, so that the random forest estimator uses twice as many observations as the OLS estimator. This is meaningful because random forest estimator, being a nonparametric estimator, has a much slower rate of convergence than that of OLS. In addition, to choose the number of leaves in each tree of the random forest estimator, we use sample splitting, namely we use $T/2$ observations to build random forest estimators corresponding to different number of leaves and we use remaining $T_1-T/2=T/6$ observations to choose the best random forest estimator according to the mean squared error criterion.

Note that both DGP1 and DGP2 satisfy our linear weighted-average quantile regression model (1). DGP1 corresponds to the homoscedastic case and yields $\beta$ in (1) equal to $-\bar\beta\int_0^1\psi(u)du$. For this data-generating process, we thus have $\beta=\bar\beta$ for the lower, middle, upper as well as exponential and polynomial regression models[5] and $\beta=0_p$ for the inequality regression model. DGP2 corresponds to the heteroscedastic case and yields $\beta$ in (1) such that $\beta_1=0.2\int_0^1 q_\varepsilon(u)\psi(u)du-\bar\beta_1\int_0^1\psi(u)du$ and $(\beta_2,\dots,\beta_p)'=-(\bar\beta_2,\dots,\bar\beta)'\int_0^1\psi(u)du$, where $q_\varepsilon\colon[0,1]\to\mathbb{R}$ is the quantile function

[5]Following Section 2.1, we define the exponential and polynomial regressions by (1) with $\psi(u)=a\exp(-a(1-u))/(1-\exp(-a))$ for $a>0$ and $\psi(u)=au^{a-1}$ for $a>1$, respectively.

of the random variable $\varepsilon$. For this data-generating process, $(\beta_2, \ldots, \beta_p)'$ thus coincides with the same vector under DGP1 but $\beta_1$ can only be calculated numerically.

For each specification of parameter values and each DGP, we repeat the experiment 500 times and estimate the coverage probability for the 90% confidence interval for $\beta_1$ constructed using $t$-statistics. In addition, we estimate the mean absolute error $\mathrm{E}[|\widehat{\beta}_1 - \beta_1|]$. We present results for the coverage probability and the mean absolute error in Tables A.1 and A.2 of the Appendix, respectively. For each case, we give results for 4 regression models: upper, inequality, middle, and exponential regressions, which are denoted in the tables by $\psi$-type 1, 2, 3, and 4, respectively.

Overall, Table A.1 shows that asymptotic theory from the previous section yields good approximation to the finite sample situation. In particular, the empirical coverage probability of 90% confidence intervals is close to the nominal coverage probability. The only exception perhaps is the case of the upper and exponential regressions ($\psi = 1, 4$) with heteroscedastic noise, $T = 1000$, $p = 5$, and the $t(4)$ distribution, in which case the asymptotic confidence intervals undercover the true parameter values. However, the coverage improves substantially as we increase the sample size from $T = 1000$ to $T = 2000$. Table A.2 also shows that the mean absolute error for the case with $p = 5$ is similar to that for the case with $p = 2$, especially when $T = 2000$. This reinforces the conclusion that the asymptotic theory provides a good approximation to the finite-sample situation.

## 7. Empirical Applications

In this section, we apply our weighted-average quantile regression (WAQR) estimator in two empirical settings. In the first one, we focus on financial market data and study the expected shortfall regression. In the second one, we focus on wage data and study the inequality and social welfare regressions.

### 7.1. **Financial Market Data.**

In this subsection, we apply the WAQR estimator in the asset pricing setting. We investigate the factor loadings of the risk measures of the industry returns. Although our method is general, we focus on the expected shortfall (ES), which is one of the most used risk measures in finance (e.g., Gandhi and Lustig, 2014; Adrian and Brunnermeier, 2016; Acharya et al., 2017). In this case, our WAQR can be referred to as the expected shortfall regression estimator.

7.1.1. *Expected Shortfall Regression Estimator.* We investigate the factor structures of the 10% expected shortfalls of the Fama-French 5 industries. We use the Fama-French 5-factor model standard in the literature to capture industries' expected shortfalls (e.g., Fama and French, 2015, 2016). Table 1 reports the factor exposure results for the 10% ES regressions. For comparison, the table also shows the results based on the mean regression and the 10% quantile regression.[6]

The point estimates to the market excess returns based on the 10% ES regression are negative and statistically significant at -1.196, -1.293, -1.400, -1.153, and -1.330 for the consumer, manufacturing, high tech, health, and other industries, respectively. The negative coefficients imply that the 10% ES of the industry returns are lower when the market excess returns are high. The point estimates for the 10% ES regression are slightly larger in magnitude than those of mean regressions and 10% quantile regressions.

The exposures to the other four factors can substantially differ across the mean, quantile, and risk measures. For the size factor, the magnitude of the exposure for the manufacturing industry is similar across the mean, quantile, and risk measures. The magnitudes of the exposures for the other industries are consistently larger for the risk measures relative to those from the mean and quantile regressions. For the value factor, the magnitudes of the exposures for the consumer, high tech, and health industries are similar across the mean, quantile, and risk measures. However, the magnitudes of exposures are markedly higher for the risk measures than for the means and 10% quantiles for the manufacturing and the other industries. For example, for the manufacturing industry, the coefficient estimate to the value factor is -0.514 for the 10% ES regression, while the coefficient estimates are -0.110 and -0.140 for the mean and 10% quantile, respectively. For the profitability factor, the magnitudes of the coefficient estimates for the 10% ES regressions are generally larger to those of the mean and 10% quantile regressions. For example, for the consumer industry, the coefficient estimate to the profitability factor is -0.779 based on the 10% ES regression, while the estimates are -0.197 and -0.195 for the mean and 10% quintile, respectively. For the investment factor, the magnitudes of the coefficient estimates for the 10% ES

[6]The estimates in the mean and 10% quantile regressions are multiplied by -1 to be consistent with the risk regressions.

are consistently higher than those of the mean and 10% quantile regressions for the consumer, the high tech, and the other industries.[7]

In summary, the baseline results show that the 10% ES of the industry returns are highly exposed to the factors that are designed to explain the mean returns. The results highlight the different dynamics between the risk measures and the means or quantiles of returns and the importance of the factors in capturing the variations of the risk measures of the portfolio returns.

We further study the time-varying exposures of the 10% ES of the industry returns to the Fama-French 5 factors. The period used for estimation is the past twenty years and we roll the estimation period every year. The results are summarized in Figure 1. The exposures of the 10% ES to the market factor have a downward spike for the industries around the internet bubble period, implying that the exposures of the 10% ES to the market factor increase during this period.

The exposures of the 10% ES to the other factors vary substantially during the sample period. We discuss several examples. For the health industry, the exposures of the 10% ES to the value factor are consistently positive, suggesting that the risk of the health industry as measured by 10% ES increases when the value premium is high. The magnitude of the coefficient estimate increases substantially to around three from 1990 to early 2000. For the profitability factor, the exposures of the 10% ES of the high tech industry increase drastically in the 200s, but decrease to the pre-2000 level since 2010. For the investment factor, the health industry tends to have negative exposures of its 10% ES, while the high tech industry tends to have positive exposures of their 10% ES. In other words, the risk measured by the 10% ES of the health industry decreases when the investment premium is high, while the risk measured by 10% ES of the high tech industry increases when the investment premium is high. The magnitudes of the 10% ES of the industries all increased during the early 2000s or the burst of the internet bubble period.

The time-series results suggest that exposures of the 10% ES to the factors varied substantially during the 1990s to the early 2000 period, coinciding with the beginning and the subsequent burst of the internet bubble. However, the exposures were relatively stable during the 2008-2009 financial crisis period, although the market experienced drastic volatility during the crisis period.

[7]Table A.3 in the Online Appendix further shows the results of Table 1 and uses the Newey-West procedure to adjust the standard errors. The significance levels are largely unchanged with the adjustment.

7.1.2. *Parametric Estimator.* As discussed above, a (potentially inconsistent) alternative to our WAQR estimator is a parametric estimator. This alternative method estimates exposures of risk measures to a set of covariates by taking the weighted average of the point estimates from the individual quantile regressions. Here, we compare our estimation results with those based on the parametric estimator. The results are documented in Table 2.[8] Table 2 shows the estimation results for the 10% ES using the parametric estimator alongside those from our WAQR estimator.

For the exposures to the market factor, the point estimates of the 10% ES to the market factor from the WAQR estimator are slightly larger than those based on the parametric estimators. However, the coefficient estimates of the 10% ES based on the parametric estimator and the WAQR estimator can differ significantly for the other factors. We provide several examples of the differences. For the high tech industry, the coefficient estimate of the 10% ES to the size factor is 0.119 based on the WAQR estimator but is 0.062 based on the parametric estimator, which is close in magnitude to that of the mean and 10% quantile regressions. For the manufacturing industry, the coefficient estimate of the 10% ES to the value factor is -0.514 based on the WAQR estimator but is only -0.084 based on the parametric estimator. Again, the coefficient estimate based on the parametric estimator is close to those based on the mean and 10% quantile regressions.

Overall, we find that the WAQR and parametric estimators can differ substantially. In particular, when the exposures of the risk measures differ from those of the mean and quantiles, the parametric estimator tends to underestimate the exposures. The magnitudes of the coefficient estimates based on the parametric estimator tend to fall between those based on the WAQR estimator and those from mean and quantile regressions.

Furthermore, we investigate the underlying reasons behind this discrepancy between our WAQR estimator and the simple parametric estimator. Relative to the parametric estimator, an important assumption our WAQR estimator relaxed is the assumption that the quantiles are linear in the covariates. So far, we have shown that the WAQR estimator and the parametric estimator tend to provide different coefficient estimates in financial data. We directly test whether the differences stem from the violation of the linearity assumption the parametric method imposes.

We test whether the 10% quantile and the 5% quantile of the industry returns are significantly exposed to the higher moments and interactions of the Fama-French 5

[8]The individual quantile regression results from 1% to 10% quantiles are reported in Figure A.1.

factors. We document the results in Table A.4 in the Online Appendix. Panel A of the table shows the quantile regression results of regressing the industry returns to the first, second, and third moments of the Fama-French 5 factors. Inconsistent with the assumption that the quantiles are linear with the covariates, we find that the 10% and 5% quantiles of the industry returns are significantly exposed to many of the second and third moments of the Fama-French 5 factors. Panel B of the table reports the quantile regression results of regressing the industry returns to the standalone and interactions of the Fama-French 5 factors. Again, inconsistent with the assumption that the quantiles are linear with the covariates, we show that the 10% and 5% quantiles of the industry returns are significantly exposed to a number of the interaction terms of the Fama-French 5 factors.

Overall, we conclude that the discrepancies of the results between the WAQR estimator and the parametric estimator stem from the fact that the quantiles are not linear in the covariates, and that estimates from the parametric estimator method are not reliable in the financial setting.

7.2. **Wage Data.** In this subsection, we apply our method to study wage inequality and social welfare. We start with the wage inequality.

7.2.1. *Inequality Regression.* We consider several standard individual characteristics in the literature when wage or consumption is studied (e.g., Angrist et al., 2006; Blundell et al., 2008; Attanasio and Pistaferri, 2016), including family size, an indicator variable for no children, age, and education. The sample goes from 2001 to 2018. The sample and variables are discussed in detail in Appendix E. We apply the WAQR estimator for the inequality regression each year using all the independent variables, and show coefficient estimates in Figure 2. For comparison, we also show the estimates based on a parametric estimator which is the differences of the coefficient estimates for the 90% quantile and the 10% quantile regressions.

We discuss the time trends of the coefficient estimates based on the WAQR estimator. The coefficient estimates for the family size decrease over time, going from positive to significantly negative since 2011. That is, the inequality, or the average wage difference between the top and bottom of the distribution, decreases in family size in the latter part of the sample. When the parametric estimator is used, the point estimates slightly decrease over time but stay positive even towards the end of the sample. The WAQR coefficient estimates for the indicator variable of no children increase over time, going from significantly negative to insignificantly positive.

The coefficient estimates for age are relatively stable over time. The point estimates based on the WAQR estimator are relatively similar to those based on the parametric estimator for these two variables.

The point WAQR estimates of education stay significantly positive over time, suggesting that inequality increases with education. When the parametric estimator is used the point estimates are also positive for all years. However, the point estimates based on the parametric estimator are markedly lower than those based on the WAQR estimator before 2010. For example, the magnitude of the point estimate based on the parametric estimator is only half of that based on the WAQR estimator for year 2001. In the latter part of the sample, the point estimates based on the two methods are relatively similar.

7.2.2. *Social Welfare Regression.* We now apply our WAQR estimator to study the relationship between the weighted average wage and the individual characteristics. We assume that the weights are exponential with more weights being placed to the lower income (specifically, we use the exponential weighting function $\psi$ from Section 2.1 with $u$ replaced by $1-u$ and $a=10$).

We report the WAQR estimation results in Figure 3 below. For comparison, we also provide the mean regression (OLS) results. The blue line shows the point estimates based on the WAQR estimator over time, while the red line documents the point estimates for the mean regression (OLS).

We discuss the time trends of the coefficient estimates based on the WAQR. The point estimates for the family size variable increase over time, going from insignificantly negative to positive. For the mean regressions, the point estimates also increase over time but stay negative even towards the end of the sample. The point estimates for the indicator variable of no children decrease over time, and are lower than those based on the mean regression in the latter part of the sample. The point estimates based on the WAQR are similar to those based on the mean regressions for the age variable.

The point estimates for the education variable stay significantly positive over time based on the WAQR, suggesting that the average wage of the low income group increases with education. When OLS is used, the point estimates are also positive for all years. However, the point estimates based on the WAQR are significantly lower than those based on the mean regression before 2010. The point estimates converge towards the latter part of the sample.

## 8. Conclusion

We introduce the weighted-average quantile regression that significantly generalizes the commonly used mean and quantile regressions. We develop estimators of such regressions that are straightforward to apply in a variety of empirical settings. In the examples of risk, inequality, and social welfare regressions, the weighted-average quantile regression estimators yield results that are different from those based on both mean and quantile regression methods.

## References

Acerbi, C. (2002). Spectral measures of risk: A coherent representation of subjective risk aversion. *Journal of Banking and Finance* **26** 1505-1518.

Acharya, V., Pedersen, L., Philippon, T., and Richardson, M. (2017). Measuring systemic risk. *Review of Financial Studies* **20** 2-47.

Adrian, T. and Brunnermeier, M. (2016). CoVar. *American Economic Review* **106** 1705-1741.

Angrist, J., Chernozhukov, V., and Fernandez-Val, I. (2006). Quantile regression under misspecification, with an application to the US wage structutre. *Econometrica* **74** 539-563.

Angrist, J. and Pischke, J. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press.

Artzner, P., Delbaen, F., Eber, J., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance* **9** 203-228.

Attanasio, O. and Pistaferri, L. (2016). Consumption inequality. *Journal of Economic Perspectives* **30** 3-28.

Bickel, P. (1982). On adaptive estimation. *Annals of Statistics* **10** 647-671.

Blundell, R., Pistaferri, L., and Preston, I. (2008). Consumption inequality and partial insurance. *American Economic Review* **98** 1887-1921.

Bradley, R. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys* **2** 107-144.

Buchinsky, M. (1994). Changes in the US wage structure 1963-1987: Application of quantile regression. *Econometrica* **62** 405-458.

Cai, Z. and Wang, X. (2008). Nonparametric methods for estimating conditional VaR and expected shortfall. *Journal of Econometrics* **147** 120-130.

Chamberlain, G. (1994). Quantile regression, censoring, and the structure of wages. *In Advances in Econometrics, Sixth World Congress* **1** 171-209.

Chen, X. (2011). Penalized sieve estimation and inference of semi-nonparametric dynamic models: a selective review. *Cowles Foundation Discussion Paper*.

Chen, X. and Liao, Z. (2013). Asymptotic properties of penalized M estimators with time series observations. *Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis* 97-120.

Chernozhukov, V. (2005). Extremal quantile regression. *Annals of Statistics* **2** 806-839.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *Econometrics Journal* **21** 1-68.

Chernozhukov, V., Escanciano, J., Ichimura, H., Newey, W., and Robins, J. (2018). Locally robust semiparametric estimation. *Econometrics Journal* **21** 1-68.

Chernozhukov, V. and Fernandez-Val, I. (2011). Inference for extremal conditional quantile models, with an application to market and birthweight risks. *Review of Economic Studies* **11** 559-589.

Chernozhukov, V., Fernandez-Val, I., and Kaji, T. (2018). Extremal quantile regression: an overview. *Handbook of Quantile Regression*.

Chernozhukov, V. and Umantsev, L. (2001). Conditional value-at-risk: aspects of modeling and estimation. *Journal of Econometrics* **147** 120-130.

Chun, S., Shapiro, A., and Uryasev, S. (2012). Conditional value-at-risk and average value-at-risk: estimation and asymptotics. *Operations Research* **60** 739-756.

Cotter, J. and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance* **9** 203-228.

Cowell, F. (2011). *Measuring Inequality, 3rd edition.* Oxford University Press.

Denneberg, D. (1990). Premium calculation: why standard deviation should be replaced by absolute deviation. *ASTIN Bulletin* **20** 181-190.

Fama, E. and French, K. (2015). A five-factor asset pricing model. *Journal of Financial Economics* **116** 1-22.

Fama, E. and French, K. (2016). Dissecting anomalies with a five-factor model. *Review of Financial Studies* **29** 69-103.

Fan, J., Han, F., Liu, H., and Vickers, B. (2016). Robust inference of risks of large portfolios. *Journal of Econometrics* **194** 298-308.

Fan, J. and Yao, Q. (2005). *Nonlinear Time Series: Nonparametric and Parametric Methods.* Springer Series in Statistics.

Firpo, S., Fortin, N., and Lemieux, T. (2009). Unconditional Quantile Regressions. *Econometrica* **77** 953-973.

Fisher, M. and Bolles, R. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM* **24** 381-395.

Fisher, A. and Kennedy, E. (2021). Visually communicating and teaching intuiting for influence functions. *American Statistician* **75** 162-172.

Föllmer, H. and Schied, A. (2002). *Stochastic Finance: An Introduction in Discrete Time.* de Gruyter Studies in Mathematics.

Friedberg, R., Tibshirani, J., Athey, S., and Wager, S. (2021). Local linear forest. *Journal of Computational and Graphical Statistics* **30** 503-517.

Gandhi, P. and Lustig, H. (2014). Size anomalies in U.S. bank stock returns. *The Journal of Finance* **70** 733-768.

Gzyl, H. and Mayoral, S. (2006). On a relationship between distorted and spectral risk measures. *Working paper.*

Hampel, F., Ronchetti, E., Rousseeeuw, P., and Stahel, W. (1986). *Robust Statistics: The approach based on influence functions.* John Wiley and Sons.

Huber, P. and Ronchetti, E. (2009). *Robust Statistics, Second Edition.* John Wiley and Sons.

Jurado, K., Ludvigson, S., and Ng, S. (2015). *Measuring uncertainty. The American Economic Review* **105** 1177-1215.

Kaas, R., Goovaerts, M., Dhaene, J., and Denuit, M. (2002). *Modern Actuarial Risk Theory.* Springer.

Kahneman, D. and Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica* **47** 263-292.

Kato, K. (2012). Weighted Nadaraya-Watson estimation of conditional expected shortfall. *Journal of Financial Econometrics* **10** 265-291.

Leippold, M. (2015). Value-at-risk and other risk measures. *Investment Risk Management* 283-303.

Leorato, S., Peracchi, F., and Tanase, A. (2012). Asymptotically efficient estimation of the conditional expected shortfall. *Computational Statistics and Data Analysis* **56** 768-784.

Lettau, M. and Ludvigson, S. (2013). Shocks and crashes. *National Bureau of Economic Research Twenty-eighth Macroeconomics Annual* 293-354.

Liu, J., Cosman, P., and Rao, B. (2018). Robust linear regression via $\ell_0$ regularization. *IEEE Transactions on Signal Processing* **66**.

Martins-Filho, C., Yao, F., and Torero, M. (2018). Nonparametric estimation of conditional value-at-risk and expected shortfall based on extreme value theory. *Econometric Theory* **34** 23-67.

McNeil, A., Frey, R., and Embrechts, P. (2015). *Quantitative Risk Management.* Princeton Series in Finance.

Meinshausen, N. (2006). Quantile regression forests. *Journal of Machine Learning Research* **7** 983-999.

Newey, W. and West, K. (1987). A simple, positive-definite, heteroscedasticity and autocorrelation consistent covariance matrix. *Econometrica* **55** 703-708.

Peracchi, F. and Tanase, A. (2008). On estimating the conditional expected shortfall. *Applied stochastic models in business and industry* **24** 471-493.

Prelec, D. (1998). The probability weighting function. *Econometrica* **66** 497-527.

Rockafellar, R., Royset, J., and Miranda, S. (2014). Superquantile regression with applications to buffered reliability, uncertainty quantification, and conditional value-at-risk. *European Journal of Operational Research* **234** 140-154.

Royset, J. and Rockafellar, R. (20145. Measures of residual risk with connections to regression, risk tracing, surrogate models, and ambiguity. *SIAM Journal of Optimization* **25** 1179-1208.

Rousseeuw, P. (1984). Least median of squares regression. *Journal of American Statistical Association* **79** 871-880.

Rousseeuw, P. and Leroy, A. (1987). *Robust Regression and Outlier Detection.* John Wiley and Sons.

Scaillet, O. (2005). Nonparametric estimation of conditional expected shortfall. *Insurance and Risk Management Journal* **74** 639-660.

Schick, A. (1986). On asymptotically efficient estimation in semiparametric models. *Annals of Statistics* **14** 1139-1151.

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: cumulative representation of uncertainty. *Journal of Risk and Uncertainty* **5** 297-323.

Vershynin, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science.* Cambridge Series in Statistical and Probabilistic Mathematics.

Wang, S. (1995). Insurance pricing and increased limits ratemaking by proportional hazards transforms. *Insurance: Mathematics and Economics* **17** 43-54.

Wang, S. (1996). Premium calculation by transforming the layer premium density. *ASTIN Bulletin* **26** 71-92.

Wang, W. and Xu, H. (2021). Preference robust distortion risk measure and its application. *Working Paper.*

Yaari, M. (1987). The dual theory of choice under risk. *Econometrica* **55** 95-115.

TABLE 1. Industry Factor Loadings

This table shows the industry factor loadings for the Fama-French 5 industries. The "Mean" rows reports the results for mean regressions. The "10% VAR" rows report the results for 10% quantile regressions. The "10% ES" rows report results for the 10% ES risk regressions. The estimates in the mean and 10% quantile regressions are multiplied by -1 to be consistent with the risk regressions. The Fama-French 5-factor model is used. Standard errors are reported in parentheses.

| Cnsmr | MKTRF | | SMB | | HML | | RMW | | CMA | | Constant | | $R^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Mean | -0.869 | (0.004) | -0.059 | (0.008) | 0.165 | (0.007) | -0.278 | (0.012) | -0.197 | (0.015) | -0.003 | (0.005) | 0.904 |
| 10% VAR | -0.873 | (0.008) | -0.074 | (0.016) | 0.177 | (0.014) | -0.294 | (0.023) | -0.195 | (0.030) | 0.374 | (0.009) | |
| 10% ES | -1.196 | (0.022) | -0.122 | (0.043) | 0.214 | (0.039) | -0.779 | (0.061) | -0.250 | (0.081) | 0.909 | (0.025) | 0.388 |
| Manuf | MKTRF | | SMB | | HML | | RMW | | CMA | | Constant | | $R^2$ |
| Mean | -1.026 | (0.006) | -0.075 | (0.011) | -0.110 | (0.010) | -0.440 | (0.016) | -0.182 | (0.021) | 0.008 | (0.007) | 0.879 |
| 10% VAR | -0.997 | (0.012) | -0.101 | (0.023) | -0.140 | (0.021) | -0.406 | (0.033) | -0.099 | (0.043) | 0.527 | (0.013) | |
| 10% ES | -1.293 | (0.029) | -0.066 | (0.057) | -0.514 | (0.051) | -0.832 | (0.081) | 0.122 | (0.106) | 1.188 | (0.033) | 0.345 |
| Hitec | MKTRF | | SMB | | HML | | RMW | | CMA | | Constant | | $R^2$ |
| Mean | -1.074 | (0.005) | 0.068 | (0.010) | 0.316 | (0.009) | 0.348 | (0.014) | -0.045 | (0.018) | -0.008 | (0.006) | 0.922 |
| 10% VAR | -1.075 | (0.011) | 0.056 | (0.021) | 0.339 | (0.019) | 0.370 | (0.030) | -0.049 | (0.040) | 0.421 | (0.012) | |
| 10% ES | -1.400 | (0.027) | 0.119 | (0.052) | 0.297 | (0.046) | 1.125 | (0.074) | -0.280 | (0.097) | 1.082 | (0.030) | 0.445 |
| Hlth | MKTRF | | SMB | | HML | | RMW | | CMA | | Constant | | $R^2$ |
| Mean | -0.816 | (0.007) | 0.021 | (0.013) | 0.319 | (0.012) | 0.084 | (0.019) | -0.124 | (0.025) | -0.003 | (0.008) | 0.765 |
| 10% VAR | -0.803 | (0.013) | 0.016 | (0.025) | 0.327 | (0.022) | 0.114 | (0.035) | -0.098 | (0.046) | 0.592 | (0.014) | |
| 10% ES | -1.153 | (0.026) | 0.096 | (0.051) | 0.251 | (0.045) | 0.190 | (0.072) | -0.045 | (0.094) | 1.207 | (0.029) | 0.322 |
| Other | MKTRF | | SMB | | HML | | RMW | | CMA | | Constant | | $R^2$ |
| Mean | -1.060 | (0.004) | 0.005 | (0.007) | -0.632 | (0.007) | 0.153 | (0.011) | 0.311 | (0.014) | 0.002 | (0.004) | 0.962 |
| 10% VAR | -1.059 | (0.007) | 0.012 | (0.013) | -0.604 | (0.012) | 0.175 | (0.018) | 0.308 | (0.024) | 0.338 | (0.007) | |
| 10% ES | -1.330 | (0.033) | 0.349 | (0.064) | -1.685 | (0.058) | 0.278 | (0.092) | 0.765 | (0.120) | 1.043 | (0.037) | 0.427 |

TABLE 2. Compare with Parametric Estimator

This table shows the industry factor loadings for the 10% ES regression for the Fama-French 5 industries based on the parametric estimator and the WAQR estimator. The "10% Parametric" rows report results using the parametric estimator. The "10% ES" rows report results using the WAQR estimator. The Fama-French 5-factor model is used. Standard errors are reported in parentheses.

| Cnsmr | MKTRF | | SMB | | HML | | RMW | | CMA | | Constant | | $R^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 10% Parametric | -0.862 | | -0.094 | | 0.179 | | -0.294 | | -0.181 | | 0.548 | | |
| 10% ES | -1.196 | (0.022) | -0.122 | (0.043) | 0.214 | (0.039) | -0.779 | (0.061) | -0.250 | (0.081) | 0.909 | (0.025) | 0.388 |
| Manuf | MKTRF | | SMB | | HML | | RMW | | CMA | | Constant | | $R^2$ |
| 10% Parametric | -1.016 | | -0.111 | | -0.084 | | -0.440 | | -0.038 | | 0.782 | | |
| 10% ES | -1.293 | (0.029) | -0.066 | (0.057) | -0.514 | (0.051) | -0.832 | (0.081) | 0.122 | (0.106) | 1.188 | (0.033) | 0.345 |
| Hitec | MKTRF | | SMB | | HML | | RMW | | CMA | | Constant | | $R^2$ |
| 10% Parametric | -1.077 | | 0.062 | | 0.360 | | 0.440 | | -0.070 | | 0.639 | | |
| 10% ES | -1.400 | (0.027) | 0.119 | (0.052) | 0.297 | (0.046) | 1.125 | (0.074) | -0.280 | (0.097) | 1.082 | (0.030) | 0.445 |
| Hlth | MKTRF | | SMB | | HML | | RMW | | CMA | | Constant | | $R^2$ |
| 10% Parametric | -0.813 | | 0.009 | | 0.342 | | 0.110 | | -0.128 | | 0.900 | | |
| 10% ES | -1.153 | (0.026) | 0.096 | (0.051) | 0.251 | (0.045) | 0.190 | (0.072) | -0.045 | (0.094) | 1.207 | (0.029) | 0.322 |
| Other | MKTRF | | SMB | | HML | | RMW | | CMA | | Constant | | $R^2$ |
| 10% Parametric | -1.050 | | 0.000 | | -0.609 | | 0.182 | | 0.284 | | 0.486 | | |
| 10% ES | -1.330 | (0.033) | 0.349 | (0.064) | -1.685 | (0.058) | 0.278 | (0.092) | 0.765 | (0.120) | 1.043 | (0.037) | 0.427 |

FIGURE 1. Rolling 20 Year Coefficient

This figure shows the WAQR estimates for the time-varying exposures of the 10% ES of the industry returns to the Fama-French 5 factors. The industries include the Fama-French 5 industries: consumer, manufacturing, high tech, health, and other. The period used for estimation is the past 20 years and the estimation period is rolled over every year.

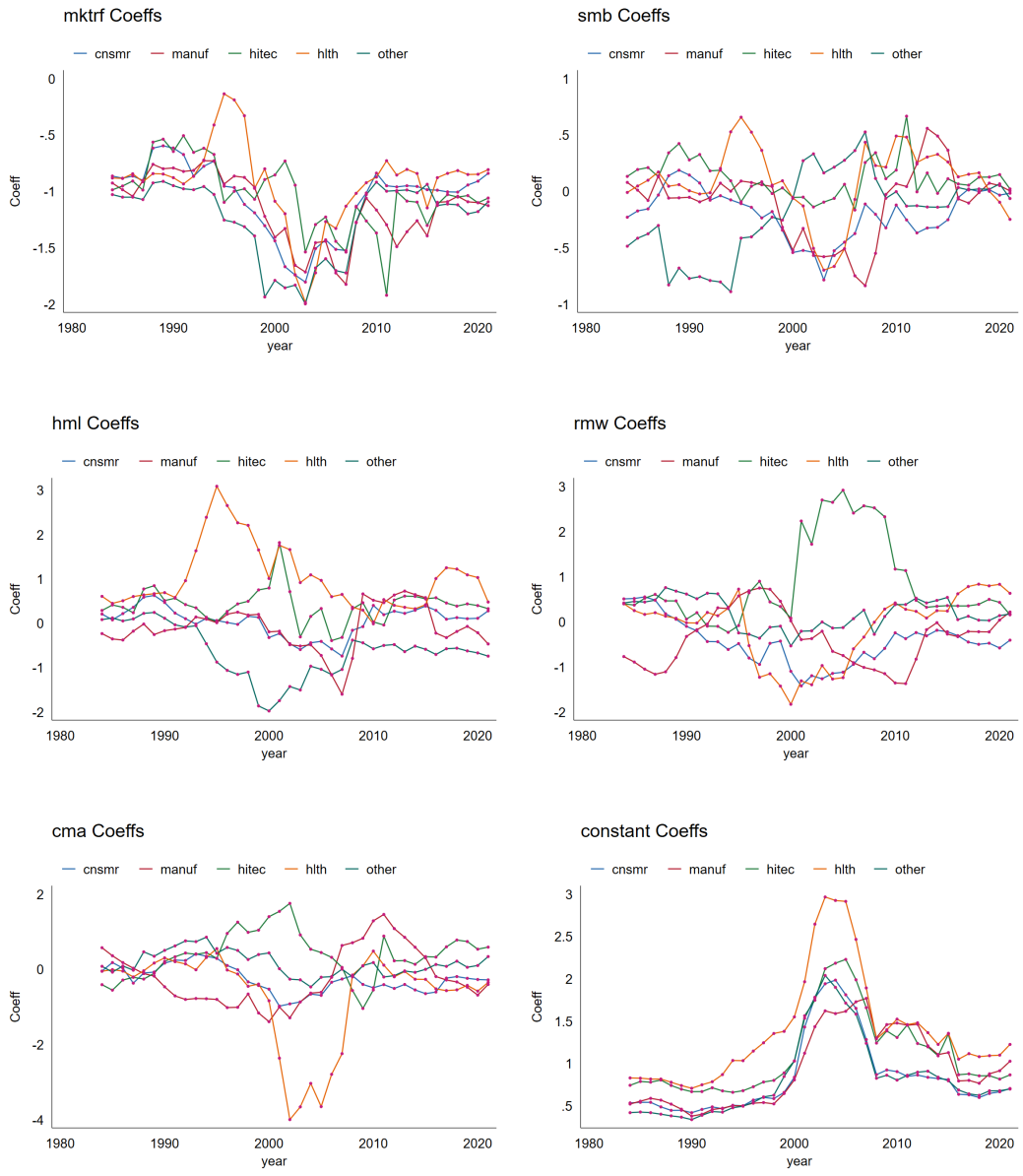


FIGURE 2. Inequality Regression Coefficients

This figure shows the time-varying WAQR coefficient estimates of the difference between the average logged weekly wages of the top and bottom 10% to several standard individual characteristics, including family size, an indicator variable of no children, age, and education. Each year, one inequality regression is conducted with all the independent variables. The blue lines represent coefficient estimates based on the inequality regression and red lines represent coefficient estimates based on a native method. The 95% confidence intervals for the point estimates based on the inequality regressions are plotted in dash lines.

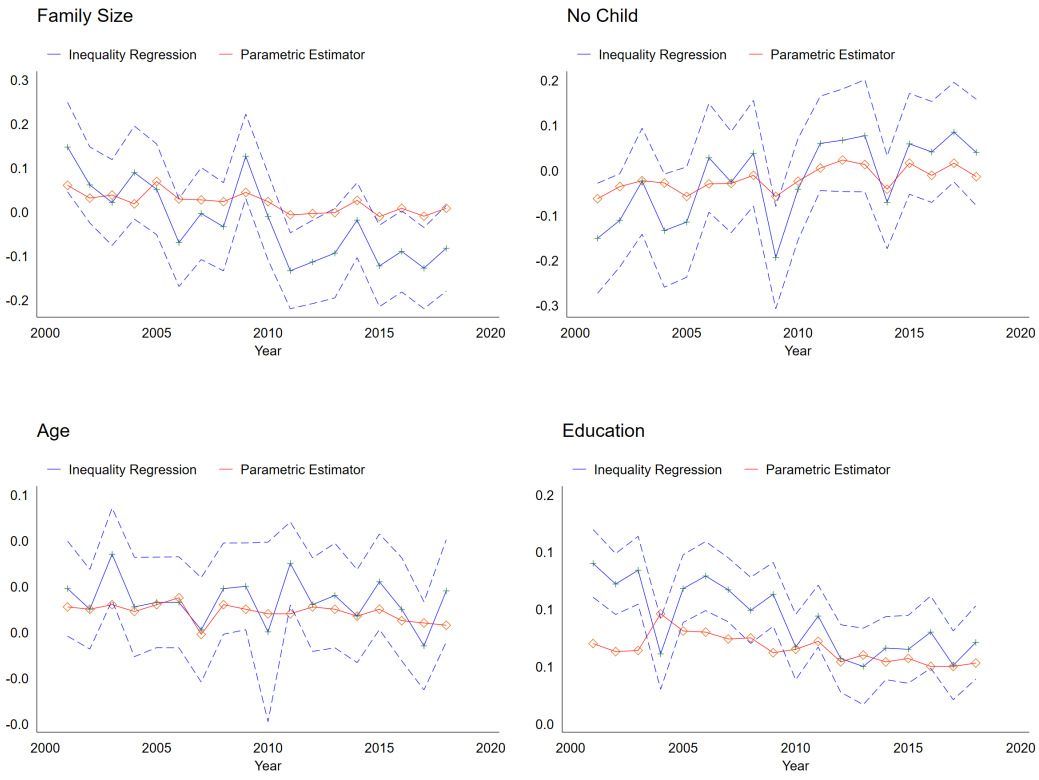
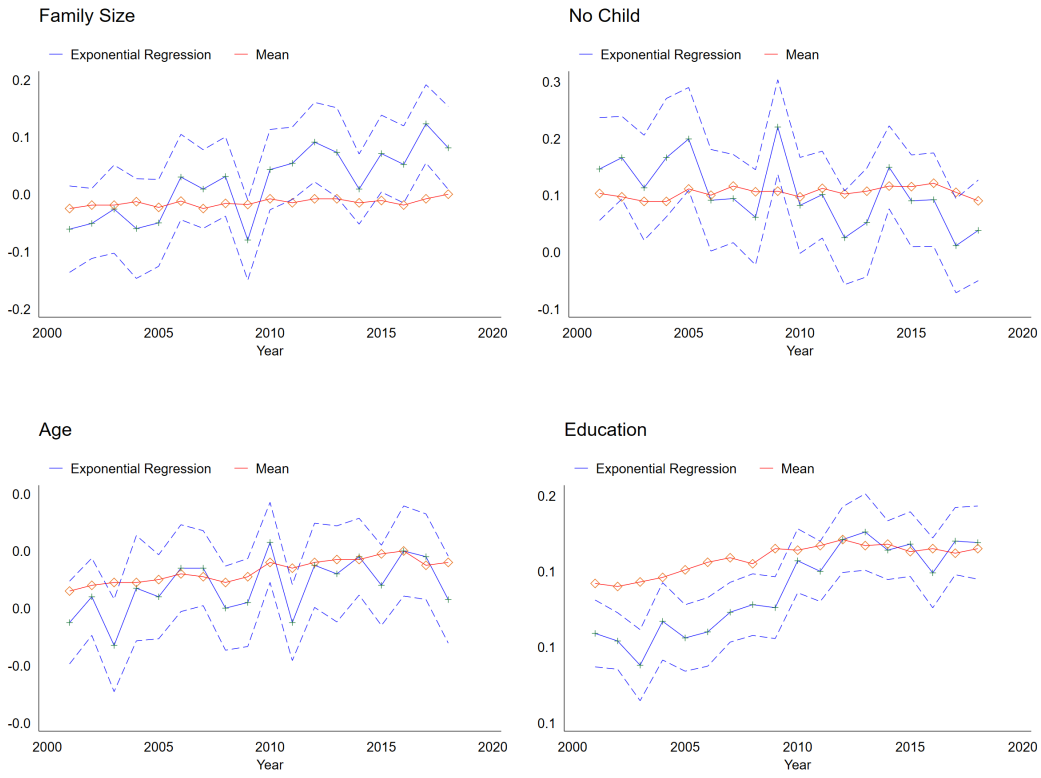


FIGURE 3. Social Welfare Regression Coefficients

This figure shows the time-varying WAQR coefficient estimates for the social welfare (exponential) regression, with dependent variable being wage and the vector of independent variables including individual characteristics (family size, an indicator variable of no children, age, and education). Each year, one social welfare (exponential) regression is conducted with all the independent variables. The blue lines represent coefficient estimates based on the social welfare (exponential) regression and red lines represent coefficient estimates based on the mean regression. The 95% confidence intervals for the point estimates based on the social welfare (exponential) regressions are plotted in dash lines.



# For Online Publication

## Appendix A. Proof of Theorem 5.1

The proof of the theorem is long. We therefore start with a sequence of useful lemmas. Throughout this section, we will assume, without loss of generality, that $D = (X', Y)'$ is independent of $\{D_t\}_{t\in\mathbb{Z}}$.

**Lemma A.1.** *Under Assumptions 5.3, 5.4, and 5.5(i), we have*

$$\int_{-\infty}^{+\infty} \left|\psi(\widehat{F}(s|x)) - \psi(F(s|x))\right| ds \leq C\Delta(x) \tag{25}$$

*for all $x \in \mathcal{X}$ with probability approaching one, where $C > 0$ is some constant.*

*Proof.* By Assumption 5.3(i), the function $\psi$ can be decomposed as $\psi = \psi_1 - \psi_2$, where both $\psi_1 \colon [0,1] \to \mathbb{R}$ and $\psi_2 \colon [0,1] \to \mathbb{R}$ are bounded and increasing functions. Moreover, by Assumption 5.3(ii), we can choose $\psi_1$ and $\psi_2$ such that they are both continuously differentiable on $(0, u_0)$ and $(1-u_0, 1)$ with bounded derivatives. Moreover, by suitable shifting these functions, we can assume, without loss of generality, that they are both non-negative. We will show how to prove that (25) holds with $\psi$ replaced by $\psi_1$. We will then note that the same argument applies in the case of $\psi_2$ as well, and so (25) will follow from the triangle inequality.

For all $x \in \mathcal{X}$, denote

$$\Delta_1(x) = \sup_{s\in\mathbb{R}} \left|\widehat{F}(s|x) - F(s|x)\right| \text{ and } \Delta_2(x) = \int_{-\infty}^{+\infty} \left|\widehat{F}(s|x) - F(s|x)\right| ds$$

so that $\Delta(x) = \Delta_1(x) + \Delta_2(x)$. In addition, extend the function $\psi_1$ from $[0,1]$ to $\mathbb{R}$ by setting $\psi_1(u) = \psi_1(0)$ for all $u < 0$ and $\psi_1(u) = \psi_1(1)$ for all $u > 1$. Defined this way, the function $\psi_1$ is bounded, increasing, and non-negative on $\mathbb{R}$.

Now, since $\psi_1$ is increasing, we have for all $x \in \mathcal{X}$ and $s \in \mathbb{R}$ that

$$\psi_1(\widehat{F}(s|x)) - \psi_1(F(s|x)) \leq \psi_1(F(s|x) + \Delta_1(x)) - \psi_1(F(s|x))$$

and

$$\psi_1(F(s|x)) - \psi_1(\widehat{F}(s|x)) \leq \psi_1(F(s|x)) - \psi_1(F(s|x) - \Delta_1(x)).$$

Hence,

$$\begin{aligned}&\left|\psi_1(\widehat{F}(s|x)) - \psi_1(F(s|x))\right| \\ &\quad\leq \max\Big(\psi_1(F(s|x) + \Delta_1(x)) - \psi_1(F(s|x)), \psi_1(F(s|x)) - \psi_1(F(s|x) - \Delta_1(x))\Big)\end{aligned}$$

$$\leq \psi_1(F(s|x) + \Delta_1(x)) - \psi_1(F(s|x) - \Delta_1(x)).$$

Therefore,

$$\begin{aligned}
&\int_{s_1}^{s_2} |\psi_1(\widehat{F}(s|x)) - \psi_1(F(s|x))| ds \\
&\quad \leq \int_{s_1}^{s_2} (\psi_1(F(s|x) + \Delta_1(x)) - \psi_1(F(s|x) - \Delta_1(x)))\, ds \\
&\quad \leq \frac{1}{c} \int_{F(s_1|x)}^{F(s_2|x)} (\psi_1(z + \Delta_1(x)) - \psi_1(z - \Delta_1(x))) dz \\
&\quad = \frac{1}{c} \int_{F(s_2|x)-\Delta_1(x)}^{F(s_2|x)+\Delta_1(x)} \psi_1(z) dz - \frac{1}{c} \int_{F(s_1|x)-\Delta_1(x)}^{F(s_1|x)+\Delta_1(x)} \psi_1(z) dz \leq \frac{2\Delta_1(x)\psi_1(1)}{c}, \qquad (26)
\end{aligned}$$

where the second inequality follows from Assumption 5.4(ii) by carrying out the change of variables $s \mapsto F(s|x) = z$, and the third from the fact that $\psi_1$ is non-negative.

Moreover, by Assumptions 5.4(i) and 5.5(i), we have

$$\sup_{s \leq s_1} \widehat{F}(s|x) \leq \sup_{s \leq s_1} F(s|x) + \Delta_1(x) \leq F(s_1|x) + \Delta_1(x) < u_0/2 + u_0/2 \leq u_0$$

and

$$\inf_{s \geq s_2} \widehat{F}(s|x) \geq \inf_{s \geq s_2} F(s|x) - \Delta_1(x) \geq F(s_2|x) - \Delta_1(x) > 1 - u_0/2 - u_0/2 = 1 - u_0$$

with probability approaching one uniformly over $x \in \mathcal{X}$. Therefore, by Assumptions 5.3(iii) and 5.4(i), for some constant $C_\psi > 0$,

$$\int_{-\infty}^{s_1} \left|\psi_1(\widehat{F}(s|x)) - \psi_1(F(s|x))\right| ds \leq C_\psi \int_{-\infty}^{s_1} \left|\widehat{F}(s|x) - F(s|x)\right| ds \leq C_\psi \Delta_2(x) \qquad (27)$$

and

$$\int_{s_2}^{\infty} \left|\psi_1(\widehat{F}(s|x)) - \psi_1(F(s|x))\right| ds \leq C_\psi \int_{s_2}^{\infty} \left|\widehat{F}(s|x) - F(s|x)\right| ds \leq C_\psi \Delta_2(x) \qquad (28)$$

with probability approaching one uniformly over $x \in \mathcal{X}$. Combining (26), (27), and (28) gives (25) with $\psi$ replaced by $\psi_1$. In addition, we can prove by the same argument that (25) holds with $\psi$ replaced by $\psi_2$ as well. The asserted claim now follows from the triangle inequality. ∎

**Lemma A.2.** *Consider a sequence of functions $\{f_T\}_{T\geq 2}$ such that for all $T \geq 2$, the function $f_T$ is mapping $\mathcal{D} \times \mathcal{D}^{T_1}$ into $\mathbb{R}$, where $\mathcal{D}$ is the support of $D$. Then*

$$\operatorname{Var}\left(\sum_{t=T_1+1}^{T_1+T_2} f_T(D_t, D_1^{T_1}) \mid D_1^{T_1}\right) = o_P(T)$$

*as long as*

$$\sum_{t=T_1+1}^{T_1+T_2} \mathrm{E}\left[|f_T(D_t, D_1^{T_1})|^4 \mid D_1^{T_1}\right] = o_P(1) \tag{29}$$

*and Assumption 5.1 is satisfied.*

*Proof.* For brevity of notations, for all $t = T_1+1, \ldots, T_1+T_2$, we will write $f_t$ instead of $f_T(D_t, D_1^{T_1})$ throughout the proof. Then it follows from (29) that there exists $\gamma_T \to 0$ as $T \to \infty$ such that

$$\mathrm{P}\left(\sum_{t=T_1+1}^{T_1+T_2} \mathrm{E}\left[|f_t|^4 \mid D_1^{T_1}\right] \leq \gamma_T^2\right) \geq 1 - \gamma_T. \tag{30}$$

Next, for all $t = T_1+1, \ldots, T_1+T_2$, denote $f_{t,+} = f_t \mathbb{I}\{f_t \geq 0\}$ and $f_{t,-} = -f_t \mathbb{I}\{f_t < 0\}$, so that $f_t = f_{t,+} - f_{t,-}$. Then

$$f_t = \int_0^{+\infty} (\mathbb{I}\{f_{t,+} > s\} - \mathbb{I}\{f_{t,-} > s\}) ds,$$

and so

$$\begin{aligned}
\operatorname{Var}\left(\sum_{t=T_1+1}^{T_1+T_2} f_t \mid D_1^{T_1}\right) &= \sum_{t_1,t_2=T_1+1}^{T_1+T_2} \operatorname{Cov}(f_{t_1}, f_{t_2} \mid D_1^{T_1}) \\
&= \int_0^{+\infty}\int_0^{+\infty} \sum_{t_1,t_2=T_1+1}^{T_1+T_2} \operatorname{Cov}\Big(\mathbb{I}\{f_{t_1,+} > s_1\} - \mathbb{I}\{f_{t_1,-} > s_1\}, \\
&\qquad \mathbb{I}\{f_{t_2,+} > s_2\} - \mathbb{I}\{f_{t_2,-} > s_2\} \mid D_1^{T_1}\Big) ds_1 ds_2.
\end{aligned} \tag{31}$$

Denoting the integrand here by $R(s_1, s_2)$, we now derive three different bounds on it.

First, observe that for any $t_1 < t_2$ and any random variables $Z_1$ and $Z_2$ such that $Z_1$ depends only on $D_{t_1}$ and $D_1^{T_1}$ and $Z_2$ depends only on $D_{t_2}$ and $D_1^{T_1}$, we have

$$\begin{aligned}
\operatorname{Cov}(Z_1, Z_2 \mid D_1^{T_1}) &= \mathrm{E}[Z_1 Z_2 \mid D_1^{T_1}] - \mathrm{E}[Z_1 \mid D_1^{T_1}]\mathrm{E}[Z_2 \mid D_1^{T_1}] \\
&= \mathrm{E}\left[Z_1(\mathrm{E}[Z_2 \mid D_{t_1}, D_1^{T_1}] - \mathrm{E}[Z_2 \mid D_1^{T_1}]) \mid D_1^{T_1}\right],
\end{aligned}$$

and so if $|Z_1| \leq 1$ a.s., then

$$|\mathrm{Cov}(Z_1, Z_2 \mid D_1^{T_1})| \leq \mathrm{E}[|\mathrm{E}[Z_2 \mid D_{t_1}, D_1^{T_1}] - \mathrm{E}[Z_2 \mid D_1^{T_1}]| \mid D_1^{T_1}].$$

Substituting here $Z_1 = \mathbb{I}\{f_{t_1,+} > s_1\} - \mathbb{I}\{f_{t_1,-} > s_1\}$ and $Z_2 = \mathbb{I}\{f_{t_2,+} > s_2\} - \mathbb{I}\{f_{t_2,-} > s_2\}$, we obtain

$$\begin{aligned}
&\Big|\mathrm{Cov}\Big(\mathbb{I}\{f_{t_1,+} > s_1\} - \mathbb{I}\{f_{t_1,-} > s_1\}, \mathbb{I}\{f_{t_2,+} > s_2\} - \mathbb{I}\{f_{t_2,-} > s_2\} \mid D_1^{T_1}\Big)\Big| \\
&\quad \leq \mathrm{E}\left[\left|\mathrm{P}(f_T(D_{t_2}, D_1^{T_1}) > s_2 \mid D_{t_1}, D_1^{T_1}) - \mathrm{P}(f_T(D_{t_2}, D_1^{T_1}) > s_2 \mid D_1^{T_1})\right| \mid D_1^{T_1}\right] \\
&\qquad + \mathrm{E}\left[\left|\mathrm{P}(f_T(D_{t_2}, D_1^{T_1}) < -s_2 \mid D_{t_1}, D_1^{T_1}) - \mathrm{P}(f_T(D_{t_2}, D_1^{T_1}) < -s_2 \mid D_1^{T_1})\right| \mid D_1^{T_1}\right] \\
&\quad \leq 2\mathrm{E}\left[\sup_B \left|\mathrm{P}(D_{t_2} \in B \mid D_{t_1}, D_1^{T_1}) - \mathrm{P}(D_{t_2} \in B \mid D_1^{T_1})\right| \mid D_1^{T_1}\right],
\end{aligned}$$

where the penultimate inequality follows from the definition of the $\beta$-mixing coefficients. Therefore,

$$\begin{aligned}
\mathrm{E}\left[\sup_{s_1,s_2 \in (0,\infty)} |R(s_1, s_2)|\right] &\leq \sum_{t=T_1+1}^{T_1+T_2} 1 + 2 \sum_{t_1=T_1+1}^{T_1+T_2-1} \sum_{t_2=t_1+1}^{T_1+T_2} 2(\beta_{t_2-t_1} + \beta_{t_2-T_1}) \\
&\leq (T_2 - T_1)\left(1 + 8\sum_{t=1}^{\infty} \beta_t\right)
\end{aligned}$$

by the definition of the $\beta$-mixing coefficients. Hence, by Markov's inequality,

$$|R(s_1, s_2)| \leq \frac{T_2 - T_1}{\gamma_T}\left(1 + 8\sum_{t=1}^{\infty} \beta_t\right) \tag{32}$$

with probability at least $1 - \gamma_T$ uniformly over $s_1, s_2 \in (0, \infty)$, which is our first bound on $R(s_1, s_2)$.

Next, observe that for any random variables $Z_1$ and $Z_2$ such that $|Z_1| \leq 1$ a.s., we have

$$\begin{aligned}
|\mathrm{Cov}(Z_1, Z_2 \mid D_1^{T_1})| &= |\mathrm{E}[(Z_1 - \mathrm{E}[Z_1 \mid D_1^{T_1}])(Z_2 - \mathrm{E}[Z_2 \mid D_1^{T_1}]) \mid D_1^{T_1}]| \\
&= |\mathrm{E}[Z_1(Z_2 - \mathrm{E}[Z_2 \mid D_1^{T_1}]) \mid D_1^{T_1}]| \leq \mathrm{E}[|Z_2 - \mathrm{E}[Z_2 \mid D_1^{T_1}]| \mid D_1^{T_1}] \\
&\leq \mathrm{E}[|Z_2| \mid D_1^{T_1}] + |\mathrm{E}[Z_2 \mid D_1^{T_1}]| \leq 2\mathrm{E}[|Z_2| \mid D_1^{T_1}].
\end{aligned}$$

Substituting here $Z_1 = \mathbb{I}\{f_{t_1,+} > s_1\} - \mathbb{I}\{f_{t_1,-} > s_1\}$ and $Z_2 = \mathbb{I}\{f_{t_2,+} > s_2\} - \mathbb{I}\{f_{t_2,-} > s_2\}$ again, we obtain

$$\begin{aligned}
&\Big|\mathrm{Cov}\Big(\mathbb{I}\{f_{t_1,+} > s_1\} - \mathbb{I}\{f_{t_1,-} > s_1\}, \mathbb{I}\{f_{t_2,+} > s_2\} - \mathbb{I}\{f_{t_2,-} > s_2\} \mid D_1^{T_1}\Big)\Big| \\
&\qquad \leq 2\Big(\mathrm{P}(f_{t_2} > s_2 \mid D_1^{T_1}) + \mathrm{P}(f_{t_2} < -s_2 \mid D_1^{T_1})\Big) \leq 2\mathrm{P}(|f_{t_2}| > s_2 \mid D_1^{T_1}).
\end{aligned}$$

Therefore,

$$|R(s_1,s_2)| \le 2 \sum_{t_1,t_2=T_1+1}^{T_1+T_2} \mathrm{P}(|f_{t_2}| > s_2 \mid D_1^{T_1}) \le 2(T_2-T_1) \sum_{t=T_1+1}^{T_1+T_2} \mathrm{P}(|f_t| > s_2 \mid D_1^{T_1}),$$

and so, by Markov's inequality and (30),

$$|R(s_1,s_2)| \le \frac{2(T_2-T_1)}{s_2^4} \sum_{t=T_1+1}^{T_1+T_2} \mathrm{E}[|f_t|^4 \mid D_1^{T_1}] \le \frac{2\gamma_T^2(T_2-T_1)}{s_2^4}$$

with probability at least $1-\gamma_T$ uniformly over $s_1, s_2 \in (0,\infty)$, which is our second bound on $R(s_1,s_2)$. In addition, by the same argument, with interchanged $Z_1$ and $Z_2$,

$$|R(s_1,s_2)| \le 2(T_2-T_1) \sum_{t=T_1+1}^{T_1+T_2} \mathrm{P}(|f_t| > s_1 \mid D_1^{T_1}) \le \frac{2\gamma_T^2(T_2-T_1)}{s_1^4},$$

with probability at least $1-\gamma_T$ uniformly over $s_1, s_2 \in (0,\infty)$, which is our third bound on $R(s_1,s_2)$.

Now, denoting the right-hand side of (32) by $\bar{R}$ and combining all three bounds, we have

$$|R(s_1,s_2)| \le \int_0^{\bar{R}} \mathbb{I}\left\{u \le \frac{2\gamma_T^2(T_2-T_1)}{s_1^4}\right\} \mathbb{I}\left\{u \le \frac{2\gamma_T^2(T_2-T_1)}{s_2^4}\right\} du$$

with probability at least $1-3\gamma_T$. Substituting this bound into (31), we obtain

$$\begin{aligned}
\mathrm{Var}\left(\sum_{t=T_1+1}^{T_1+T_2} f_t \mid D_1^{T_1}\right) &\le \int_0^{+\infty}\int_0^{+\infty} |R(s_1,s_2)| ds_1 ds_2 \\
&\le \int_0^{+\infty}\int_0^{+\infty}\int_0^{\bar{R}} \mathbb{I}\left\{u \le \frac{2\gamma_T^2(T_2-T_1)}{s_1^4}\right\} \mathbb{I}\left\{u \le \frac{2\gamma_T^2(T_2-T_1)}{s_2^4}\right\} du ds_1 ds_2 \\
&= \int_0^{\bar{R}} \left(\frac{2\gamma_T^2(T_2-T_1)}{u}\right)^{1/4} \left(\frac{2\gamma_T^2(T_2-T_1)}{u}\right)^{1/4} du \\
&= 2\gamma_T \sqrt{2(T_2-T_1)}\sqrt{\bar{R}} \le 2\sqrt{2\gamma_T\left(1+\sum_{t=1}^{\infty}\beta_t\right)(T_2-T_1)}
\end{aligned}$$

with probability at least $1-3\gamma_T$. Since $\gamma_T \to 0$ and $\sum_{t=1}^{\infty}\beta_t < \infty$ by Assumption 5.1, the asserted claim follows. ∎

**Lemma A.3.** *Consider a sequence of functions $\{f_T\}_{T\ge 2}$ such that for all $T \ge 2$, the function $f_T$ is mapping $\mathcal{D} \times \mathcal{D}^{T_1}$ into $\mathbb{R}$, where $\mathcal{D}$ is the support of $D$. Let $\{A_T\}_{T\ge 2}$ be a sequence of positive numbers. Also, let $D = (X', Y)'$ be independent of $\{D_t\}_{t\in\mathbb{Z}}$.*

*Finally, let $\delta > 0$ be some number. Then*

$$\sum_{t=T_1+1}^{T_1+T_2} \left( \mathrm{E}\left[ f_T(D_t, D_1^{T_1}) \mid D_1^{T_1} \right] - \mathrm{E}\left[ f_T(D, D_1^{T_1}) \mid D_1^{T_1} \right] \right) = o_P(A_T)$$

*as long as*

$$\sum_{t=T_1+1}^{T_1+T_2} \left( \mathrm{E}\left[ |f_T(D_t, D_1^{T_1})|^{1+\delta} \mid D_1^{T_1} \right] + \mathrm{E}\left[ |f_T(D, D_1^{T_1})|^{1+\delta} \mid D_1^{T_1} \right] \right) = o_P(A_T^{1+\delta}) \tag{33}$$

*and Assumption 5.1 is satisfied.*

*Proof.* For brevity of notations, for all $t = T_1 + 1, \ldots, T_2$, we will write $f_t$ instead of $f_T(D_t, D_1^{T_1})$ throughout the proof. In addition, we will write $\tilde{f}$ instead of $f_T(D, D_1^{T_1})$. Then it follows from (33) that there exists $\gamma_T \to 0$ as $T \to \infty$ such that

$$\mathrm{P}\left( \sum_{t=T_1+1}^{T_1+T_2} \left( \mathrm{E}\left[ |f_t|^{1+\delta} \mid D_1^{T_1} \right] + \mathrm{E}\left[ |\tilde{f}|^{1+\delta} \mid D_1^{T_1} \right] \right) \leq (\gamma_T A_T)^{1+\delta} \right) \geq 1 - \gamma_T. \tag{34}$$

Next, for all $t = T_1 + 1, \ldots, T_2$, denote $f_{t,+} = f_t \mathbb{I}\{f_t \geq 0\}$ and $f_{t,-} = -f_t \mathbb{I}\{f_t < 0\}$, so that

$$f_t = f_{t,+} - f_{t,-} = \int_0^{+\infty} (\mathbb{I}\{f_{t,+} > s\} - \mathbb{I}\{f_{t,-} > s\}) ds. \tag{35}$$

Similarly, denote $\tilde{f}_+ = \tilde{f}\mathbb{I}\{\tilde{f} \geq 0\}$ and $\tilde{f}_- = -\tilde{f}\mathbb{I}\{\tilde{f} < 0\}$, so that

$$\tilde{f} = \tilde{f}_+ - \tilde{f}_- = \int_0^{+\infty} (\mathbb{I}\{\tilde{f}_+ > s\} - \mathbb{I}\{\tilde{f}_- > s\}) ds. \tag{36}$$

Further, for all $s > 0$, denote

$$R_1(s) = \sum_{t=T_1+1}^{T_1+T_2} \left| \mathrm{P}(f_{t,+} > s \mid D_1^{T_1}) - \mathrm{P}(\tilde{f}_+ > s \mid D_1^{T_1}) \right|$$

and

$$R_2(s) = \sum_{t=T_1+1}^{T_1+T_2} \left| \mathrm{P}(f_{t,-} > s \mid D_1^{T_1}) - \mathrm{P}(\tilde{f}_- > s \mid D_1^{T_1}) \right|.$$

Then it follows from (35), (36), and the triangle inequality that

$$\left| \sum_{t=T_1+1}^{T_1+T_2} (\mathrm{E}[f_t \mid D_1^{T_1}] - \mathrm{E}[\tilde{f} \mid D_1^{T_1}]) \right| \leq \int_0^{+\infty} R_1(s) ds + \int_0^{+\infty} R_2(s) ds. \tag{37}$$

We will bound $\int_0^{+\infty} R_1(s) ds$ and note that $\int_0^{+\infty} R_2(s) ds$ can be bounded by the same argument.

Observe that

$$\begin{aligned} R_1(s) &\leq \sum_{t=T_1+1}^{T_1+T_2} \sup_B \left|\mathrm{P}(D_t \in B \mid D_1^{T_1}) - \mathrm{P}(D \in B \mid D_1^{T_1})\right| \\ &= \sum_{t=T_1+1}^{T_1+T_2} \sup_B \left|\mathrm{P}(D_t \in B \mid D_1^{T_1}) - \mathrm{P}(D \in B)\right| \\ &= \sum_{t=T_1+1}^{T_1+T_2} \sup_B \left|\mathrm{P}(D_t \in B \mid D_1^{T_1}) - \mathrm{P}(D_t \in B)\right|, \end{aligned}$$

and so

$$\mathrm{E}\left[\sup_{s\in(0,\infty)} R_1(s)\right] \leq \sum_{t=1}^{\infty} \beta_t.$$

Therefore, $R_1(s) \leq \sum_{t=1}^{\infty} \beta_t/\gamma_T$ with probability at least $1-\gamma_T$ uniformly over $s \in (0,\infty)$ by Markov's inequality. In addition,

$$\begin{aligned} R_1(s) &\leq \sum_{t=T_1+1}^{T_1+T_2} \left(\mathrm{P}(f_{t,+} > s \mid D_1^{T_1}) + \mathrm{P}(\tilde{f}_+ > s \mid D_1^{T_1})\right) \\ &\leq \frac{1}{s^{1+\delta}} \sum_{t=T_1+1}^{T_1+T_2} \left(\mathrm{E}[|f_{t,+}|^{1+\delta} \mid D_1^{T_1}] + \mathrm{E}[|\tilde{f}_+|^{1+\delta} \mid D_1^{T_1}]\right) \leq \frac{(\gamma_T A_T)^{1+\delta}}{s^{1+\delta}} \end{aligned}$$

with probability at least $1-\gamma_T$ uniformly over $s \in (0,\infty)$ by Markov's inequality and (34). Hence, for any $s_0 > 0$, we have

$$\int_0^{\infty} R_1(s)ds = \int_0^{s_0} R_1(s)ds + \int_{s_0}^{\infty} R_1(s)ds \leq \frac{s_0}{\gamma_T} \sum_{t=1}^{\infty} \beta_t + \frac{(\gamma_T A_T)^{1+\delta}}{\delta s_0^{\delta}}$$

with probability at least $1-2\gamma_T$. Therefore, setting $s_0 = A_T(\gamma_T^{2+\delta}/\delta \sum_{t=1}^{\infty} \beta_t)^{1/(1+\delta)}$, it follows that

$$\int_0^{\infty} R_1(s)ds \leq 2A_t \left(\sum_{t=1}^{\infty} \beta_t\right)^{\delta/(1+\delta)} \left(\frac{\gamma_T}{\delta}\right)^{1/(1+\delta)}$$

with probability at least $1-2\gamma_T$. Hence, given that $\sum_{t=1}^{\infty} \beta_t < \infty$ by Assumption 5.1, it follows that $\int_0^{\infty} R_1(s)ds = o_P(A_T)$ and, by the same argument, $\int_0^{\infty} R_2(s)ds = o_P(A_T)$. Substituting these bounds into (37), we obtain the asserted claim. ∎

**Lemma A.4.** *Under Assumptions 5.1, 5.2, and 5.5,*

$$\mathrm{E}\left[\|X\|^2 \Delta(X)^2 \mid D_1^{T_1}\right] = o_P(1).$$

*Proof.* By Jensen's inequality,

$$\mathrm{E}\left[\|X\|^2\Delta(X)^2 \mid D_1^{T_1}\right] \leq \sqrt{\mathrm{E}[\|X\|^4 \mid D_1^{T_1}]}\sqrt{\mathrm{E}[\Delta(X)^4 \mid D_1^{T_1}]}.$$

Therefore, given that $\mathrm{E}[\|X\|^4 \mid D_1^{T_1}] = O_P(1)$ by Assumption 5.2(i) and Markov's inequality, it suffices to prove that $\mathrm{E}[\Delta(X)^4 \mid D_1^{T_1}] = o_P(1)$. To do so, observe that by Assumption 5.5(ii), there exists $\gamma_T \to 0$ as $T \to \infty$ such that

$$\mathrm{P}\left(\sum_{t=T_1+1}^{T_1+T_2} \mathrm{E}[\Delta(X_t)^4 \mid D_1^{T_1}] \leq \gamma_T\right) \geq 1 - \gamma_T.$$

Hence,

$$\begin{aligned}\sum_{t=T_1+1}^{T_1+T_2} \mathrm{P}(\Delta(X_t) > \gamma_T^{1/8} \mid D_1^{T_1}) &= \sum_{t=T_1+1}^{T_1+T_2} \mathrm{P}(\Delta(X_t)^4 > \sqrt{\gamma_T} \mid D_1^{T_1}) \\ &\leq \frac{1}{\sqrt{\gamma_T}} \sum_{t=T_1+1}^{T_1+T_2} \mathrm{E}[\Delta(X_t)^4 \mid D_1^{T_1}] \leq \sqrt{\gamma_T}\end{aligned}$$

with probability at least $1 - \gamma_T$. Also,

$$\sum_{t=T_1+1}^{T_1+T_2} \mathrm{E}\left[\left|\mathrm{P}(\Delta(X_t) > \gamma_T^{1/8} \mid D_1^{T_1}) - \mathrm{P}(\Delta(X) > \gamma_T^{1/8} \mid D_1^{T_1})\right|\right] \leq \sum_{t=1}^{\infty} \beta_t,$$

and so, by Markov's inequality,

$$\frac{1}{T_2 - T_1} \sum_{t=T_1+1}^{T_1+T_2} \left|\mathrm{P}(\Delta(X_t) > \gamma_T^{1/8} \mid D_1^{T_1}) - \mathrm{P}(\Delta(X) > \gamma_T^{1/8} \mid D_1^{T_1})\right| \leq \frac{1}{\sqrt{T_2 - T_1}} \sum_{t=1}^{\infty} \beta_t$$

with probability at least $1 - 1/\sqrt{T_2 - T_1}$. Therefore, by the union bound,

$$\mathrm{P}(\Delta(X) > \gamma_T^{1/8} \mid D_1^{T_1}) \leq \frac{1}{\sqrt{T_2 - T_1}} \sum_{t=1}^{\infty} \beta_t + \frac{\sqrt{\gamma_T}}{T_2 - T_1}$$

with probability at least $1 - \gamma_T - 1/\sqrt{T_2 - T_1}$. Combining this bound with Assumption 5.5(i) shows that $\mathrm{E}[\Delta(X)^4 \mid D_1^{T_1}] = o_P(1)$ and completes the proof of the lemma. ∎

We are now ready to proof Theorem 5.1:

*Proof of Theorem 5.1.* Observe that

$$\frac{1}{T_2} \sum_{t=T_1+1}^{T_1+T_2} X_t X_t' \to_P \mathrm{E}[XX'] \tag{38}$$

by Assumptions 5.1 and 5.2(i) and Proposition 2.8 in Fan and Yao (2005) since $\beta$-mixing coefficients dominate $\alpha$-mixing coefficients. Combining this result with Assumptions 5.2(ii,iii) and using the continuous mapping theorem and the Slutsky lemma gives the second convergence result in (22).

To prove the first convergence result in (22), denote

$$r_{t1} = \int_{-\infty}^{+\infty} (\Psi(F(s|X_t)) - \Psi(\widehat{F}(s|X_t)))ds + \int_{-\infty}^{+\infty} (\widehat{F}(s|X_t) - F(s|X_t))\psi(\widehat{F}(s|X_t))ds$$

and

$$r_{t2} = \int_{-\infty}^{+\infty} (F(s|X_t) - \mathbb{I}\{Y_t \leq s\})(\psi(\widehat{F}(s|X_t)) - \psi(F(s|X_t)))ds$$

for all $t = T_1 + 1, \ldots, T_1 + T_2$. Then

$$\begin{aligned}\sqrt{T_2}(\widehat{\beta} - \beta) = &\left(\frac{1}{T_2}\sum_{t=T_1+1}^{T_1+T_2} X_t X_t'\right)^{-1} \left(\frac{1}{\sqrt{T_2}}\sum_{t=T_1+1}^{T_1+T_2} X_t e_t\right) \\ &+ \left(\frac{1}{T_2}\sum_{t=T_1+1}^{T_1+T_2} X_t X_t'\right)^{-1} \left(\frac{1}{\sqrt{T_2}}\sum_{t=T_1+1}^{T_1+T_2} X_t (r_{t1} + r_{t2})\right).\end{aligned}$$

Therefore, given that $T_2^{-1}\sum_{t=T_1+1}^{T_1+T_2} X_t X_t'$ converges to a positive-definite matrix by (38) and Assumption 5.2(ii), we only need to prove that

$$E_1 = \frac{1}{\sqrt{T_2}}\sum_{t=T_1+1}^{T_1+T_2} X_t r_{t1} = o_P(1) \quad \text{and} \quad E_2 = \frac{1}{\sqrt{T_2}}\sum_{t=T_1+1}^{T_1+T_2} X_t r_{t2} = o_P(1). \tag{39}$$

We do so in turn. In addition, as in the proof of Lemma A.1, we can decompose the function $\psi$ as $\psi = \psi_1 - \psi_2$, where the functions $\psi_1$ and $\psi_2$ are both bounded, increasing, and non-negative. Therefore, given that both $r_{t1}$ and $r_{t2}$ are linear in $\psi$ (and $\Psi$), it suffices to prove (39) assuming that the function $\psi$ is itself bounded, increasing, and non-negative. This is what we do below. (Note also that the new function $\psi$ still satisfies Assumption 5.3, and so Lemma A.1 is still applicable.)

We start with $E_1$. Since $\Psi(s) = \int_0^s \psi(u)du$ and $\psi$ is increasing, the function $\Psi$ is convex, and so

$$\Psi(F(s|X_t)) - \Psi(\widehat{F}(s|X_t)) = \Psi'(\widetilde{F}(s|X_t))(F(s|X_t) - \widehat{F}(s|X_t)),$$

where $\widetilde{F}(s|X_t)$ belongs to the interval connecting $F(s|X_t)$ and $\widehat{F}(s|X_t)$, and $\Psi'(\widetilde{F}(s|X_t))$ is an element of the sub-differential of $\Psi(\widetilde{F}(s|X_t))$. Hence,

$$r_{t1} = \int_{-\infty}^{+\infty} (\widehat{F}(s|X_t) - F(s|X_t))(\psi(\widehat{F}(s|X_t)) - \Psi'(\widetilde{F}(s|X_t)))ds,$$

and so, for some constant $C > 0$,

$$|r_{t1}| \leq \int_{-\infty}^{+\infty} |\widehat{F}(s|X_t) - F(s|X_t)| \times |\psi(\widehat{F}(s|X_t)) - \Psi'(\widetilde{F}(s|X_t))| ds$$
$$\leq \int_{-\infty}^{+\infty} |\widehat{F}(s|X_t) - F(s|X_t)| \times |\psi(\widehat{F}(s|X_t)) - \psi(F(s|X_t))| ds \leq C\Delta(X_t)^2,$$

with probability approaching one uniformly over $t = T_1 + 1, \ldots, T_1 + T_2$, where the second inequality follows from the facts that the function $\psi$ is increasing and that $\Psi'(\widetilde{F}(s|X_t)) \in [\psi(\widetilde{F}(s|X_t) - 0), \psi(\widetilde{F}(s|X_t) + 0)]$ and the third from Lemma A.1. Therefore,

$$\|E_1\| \leq \frac{1}{\sqrt{T_2}} \sum_{t=T_1+1}^{T_1+T_2} \|X_t\| \times |r_{t1}| \leq \frac{C}{\sqrt{T_2}} \sum_{t=T_1+1}^{T_1+T_2} \|X_t\| \Delta(X_t)^2 \tag{40}$$

with probability approaching one. In addition,

$$\frac{1}{\sqrt{T_2}} \sum_{t=T_1+1}^{T_1+T_2} \mathrm{E}[\|X_t\| \Delta(X_t)^2 \mid D_1^{T_1}] \leq \sqrt{\sum_{t=T_1+1}^{T_1+T_2} \mathrm{E}[\|X_t\|^2 \Delta(X_t)^4 \mid D_1^{T_1}]} = o_P(1) \tag{41}$$

by the Cauchy-Schwarz inequality and Assumption 5.5(ii). Combining (40) and (41) with Markov's inequality gives $E_1 = o_P(1)$.

Next, we consider $E_2$. Observe that

$$\mathrm{E}\left[X \int_{-\infty}^{+\infty} (F(s|X) - \mathbb{I}\{Y \leq s\})(\psi(\widehat{F}(s|X)) - \psi(F(s|X)))ds \mid D_1^{T_1}\right] = 0.$$

Also, for some constant $C > 0$,

$$\mathrm{E}\left[\left(\|X\| \int_{-\infty}^{+\infty} (F(s|X) - \mathbb{I}\{Y \leq s\})(\psi(\widehat{F}(s|X)) - \psi(F(s|X)))ds\right)^2 \mid D_1^{T_1}\right]$$
$$\leq C\mathrm{E}\left[\|X\|^2 \Delta(X)^2 \mid D_1^{T_1}\right] + o_P(1) = o_P(1)$$

by Lemmas A.1 and A.4. In addition,

$$\sum_{t=T_1+1}^{T_1+T_2} \mathrm{E}\left[\left(\|X_t\| \int_{-\infty}^{+\infty} (F(s|X_t) - \mathbb{I}\{Y_t \leq s\})(\psi(\widehat{F}(s|X_t)) - \psi(F(s|X_t)))ds\right)^2 \mid D_1^{T_1}\right]$$
$$\leq C \sum_{t=T_1+1}^{T_1+T_2} \mathrm{E}\left[\|X_t\|^2 \Delta(X_t)^2 \mid D_1^{T_1}\right] + o_P(1)$$
$$\leq \sqrt{T_2 - T_1} \sqrt{\sum_{t=T_1+1}^{T_1+T_2} \mathrm{E}\left[\|X_t\|^4 \Delta(X_t)^4 \mid D_1^{T_1}\right]} + o_P(1) = o_P(T)$$

by Lemma A.1, the Cauchy-Schwarz inequality, and Assumption 5.5(ii). Hence,

$$\|\mathrm{E}[E_2 \mid D_1^{T_1}]\| = \left\| \mathrm{E}\left[ \frac{1}{\sqrt{T_2}} \sum_{t=T_1+1}^{T_1+T_2} X_t r_{t2} \mid D_1^{T_1} \right] \right\| = o_P(1) \tag{42}$$

by Lemma A.3. Further,

$$\sum_{t=T_1+1}^{T_1+T_2} \mathrm{E}\left[ (\|X_t\| \times |r_{t2}|)^4 \mid D_1^{T_1} \right] \leq C^2 \sum_{t=T_1+1}^{T_1+T_2} \mathrm{E}\left[ \|X_t\|^4 \Delta(X_t)^4 \mid D_1^{T_1} \right] + o_P(1) = o_P(1)$$

by Lemma A.1 and Assumption 5.5(ii). Hence, by Lemma A.2,

$$\|\mathrm{Var}(E_2 \mid D_1^{T_1})\| = o_P(1). \tag{43}$$

Combining (42) and (43) gives $E_2 = o_P(1)$ and completes the proof of the theorem. ∎

## Appendix B. Proof of Theorem 5.2

Denote $w(0,m) = 1/2$, so that $\bar{\Omega} = \sum_{j=0}^{m} w(j,m)(\bar{\Omega}_j + \bar{\Omega}_j')$. Also, denote

$$\widehat{\Omega} = \sum_{j=0}^{m} w(j,m)(\widehat{\Omega}_j' + \widehat{\Omega}_j'), \text{ so that } \widehat{\Sigma} = \left( \frac{1}{T_2} \sum_{t=T_1+1}^{T_1+T_2} X_t X_t' \right)^{-1} \widehat{\Omega} \left( \frac{1}{T_2} \sum_{t=T_1+1}^{T_1+T_2} X_t X_t' \right)^{-1}.$$

Then, recalling (38) from the proof of Theorem 5.1 and observing that $\bar{\Omega} \to_P \Omega$ by Assumption 5.6(i), it follows that $\widehat{\Sigma} \to_P \Sigma$ as long as $\widehat{\Omega} - \bar{\Omega} \to_P 0$. Thus, it suffices to prove that $\sum_{j=0}^{m} w(j,m)(\widehat{\Omega}_j - \bar{\Omega}_j) \to_P 0$. To do so, observe that for all $j = 0, \ldots, m$ and $t = T_1 + j + 1, \ldots, T_1 + T_2$, we have

$$\widehat{e}_t \widehat{e}_{t-j} - e_t e_{t-j} = e_{t-j}(\widehat{e}_t - e_t) + \widehat{e}_t(\widehat{e}_{t-j} - e_{t-j}),$$

and so, denoting

$$\mathcal{S}_1 = \sum_{j=0}^{m} \frac{w(j,m)}{T_2 - T_1} \sum_{t=T_1+j+1}^{T_1+T_2} e_{t-j}(\widehat{e}_t - e_t) X_t X_{t-j}'$$

and

$$\mathcal{S}_2 = \sum_{j=0}^{m} \frac{w(j,m)}{T_2 - T_1} \sum_{t=T_1+j+1}^{T_1+T_2} \widehat{e}_t(\widehat{e}_{t-j} - e_{t-j}) X_t X_{t-j}',$$

we have $\sum_{j=0}^{m} w(j,m)(\widehat{\Omega}_j - \bar{\Omega}_j) = \mathcal{S}_1 + \mathcal{S}_2$. We will prove that $\mathcal{S}_1 \to_P 0$ and note that $\mathcal{S}_2 \to_P 0$ by a similar argument.

By the Cauchy-Schwarz inequality and Assumption 5.6(ii),

$$\|\mathcal{S}_1\| \leq \sum_{j=0}^{m} \frac{w(j,m)}{T_2 - T_1} \sqrt{\sum_{t=T_1+j+1}^{T_1+T_2} \|e_{t-j} X_{t-j}\|^2} \sqrt{\sum_{t=T_1+j+1}^{T_1+T_2} \|(\widehat{e}_t - e_t) X_t\|^2}$$
$$\leq \frac{m}{T_2 - T_1} \sqrt{\sum_{t=T_1+1}^{T_1+T_2} \|e_t X_t\|^2} \sqrt{\sum_{t=T_1+1}^{T_1+T_2} \|(\widehat{e}_t - e_t) X_t\|^2}.$$

Here, given that $\mathrm{E}[\|eX\|^2] \leq \sqrt{\mathrm{E}[e^4]\mathrm{E}[\|X\|^4]} < \infty$ by Assumption 5.2(i),

$$\frac{1}{T_2 - T_1} \sum_{t=T_1+1}^{T_1+T_2} \|e_t X_t\|^2 = O_P(1)$$

by Assumptions 5.1 and Proposition 2.8 in Fan and Yao (2005). Also,

$$\frac{1}{T_2 - T_1} \sum_{t=T_1+1}^{T_1+T_2} \|(\widehat{e}_t - e_t) X_t\|^2$$
$$\leq \frac{2}{T_2 - T_1} \sum_{t=T_1+1}^{T_1+T_2} |r_{t1} + r_{t2}|^2 \|X_t\|^2 + \frac{2}{T_2 - T_1} \sum_{t=T_1+1}^{T_1+T_2} \|\widehat{\beta} - \beta\|^2 \|X_t\|^4$$

for $r_{t1}$ and $r_{t2}$ defined in the proof of Theorem 5.1. Moreover,

$$\frac{1}{T_2 - T_1} \sum_{t=T_1+1}^{T_1+T_2} \|X_t\|^4 = O_P(1)$$

by Assumptions 5.1 and 5.2(i) and Proposition 2.8 in Fan and Yao (2005). Therefore,

$$\frac{1}{T_2 - T_1} \sum_{t=T_1+1}^{T_1+T_2} \|\widehat{\beta} - \beta\|^2 \|X_t\|^4 = O_P\left(\frac{1}{T}\right)$$

by Theorem 5.1. In addition, as in the proof of Theorem 5.1, for some constant $C > 0$, we have $|r_{t1} + r_{t2}| \leq C(\Delta(X_t)^2 + \Delta(X_t))$ with probability approaching one uniformly over $t = T_1 + 1, \ldots, T_1 + T_2$. Hence,

$$\frac{1}{T_2 - T_1} \sum_{t=T_1+1}^{T_1+T_2} |r_{t1} + r_{t2}|^2 \|X_t\|^2 \leq \frac{2C^2}{T_2 - T_1} \sum_{t=T_1+1}^{T_1+T_2} (\Delta(X_t)^4 + \Delta(X_t)^2) \|X_t\|^2 = o_P\left(\frac{1}{\sqrt{T}}\right)$$

by Assumption 5.5(ii). We therefore conclude that $\|\mathcal{S}_1\| = o_P(m/T^{1/4}) = o_P(1)$ by Assumption 5.6(iii). Thus, given that $\|\mathcal{S}_2\| = o_P(1)$ by a similar argument, it follows that $\sum_{j=0}^{m} w(j,m)(\widehat{\Omega}_j - \bar{\Omega}_j) \to_P 0$, which completes the proof of the theorem.

## Appendix C. Weighted-Average Quantile Regression Estimators versus Parametric Estimators

In this section, we compare our weighted-average quantile regression estimators with parametric estimators outlined in the Introduction. Recall that given a weighting function $\psi$, we define the parametric estimator by

$$\widetilde{\beta} = \int_0^1 \widetilde{\beta}(u)\psi(u)du,$$

where each $\widetilde{\beta}(u)$ is the classical (linear) $u$-quantile regression estimator of $Y$ on $X$.

This parametric estimator is rather intuitive and is simple to implement. However, the key advantage of our weighted-average quantile regression estimator $\widehat{\beta}$ over the parametric estimator $\widetilde{\beta}$ is that our estimator is much more robust with respect to possible misspecification. In particular, our estimator requires fewer parametric assumptions for consistency. Indeed, we claim that consistency of the parametric estimator $\widetilde{\beta}$ can only be guaranteed under a *continuum* of constraints, namely $q_{Y|X}(u) = X'\beta(u)$ for all $u \in (0,1)$, whereas consistency of our estimator $\widehat{\beta}$, as discussed in the previous section, requires only one constraint: $\int_0^1 q_{Y|X}(u)\psi(u)du = X'\beta$.

To prove this claim, suppose that $\int_0^1 q_{Y|X}(u)\psi(u)du = X'\beta$ and recall that the classical $u$-quantile regression estimator

$$\widetilde{\beta}(u) = \arg\min_{b\in\mathbb{R}^d} \left( \frac{u}{T}\sum_{t=1}^{T}(Y_t - X_t'b)_+ + \frac{1-u}{T}\sum_{t=1}^{T}(Y_t - X_t'b)_- \right) \tag{44}$$

converges in probability to

$$\bar{\beta}(u) = \arg\min_{b\in\mathbb{R}^d} \Big( u\mathrm{E}[(Y - X'b)_+] + (1-u)\mathrm{E}[(Y - X'b)_-] \Big), \tag{45}$$

where for any random variable $Z$, we use $Z_+ = Z\mathbb{I}\{Z \geq 0\}$ and $Z_- = Z\mathbb{I}\{Z < 0\}$ to denote its positive and negative parts. Whenever $q_{Y|X}(u)$ is linear in $X$, i.e. $q_{Y|X}(u) = X'\beta(u)$ for some $\beta(u)$ almost surely, it is a standard exercise to show that $\bar{\beta}(u) = \beta(u)$ by taking the first-order conditions of (45), meaning that $\widetilde{\beta}(u) \to_P \beta(u)$, and so

$$\widetilde{\beta} = \int_0^1 \widetilde{\beta}(u)\psi(u)du \to_P \int_0^1 \beta(u)\psi(u)du = \beta,$$

where the last equality follows from substituting $q_{Y|X}(u) = X'\beta(u)$ into the regression model $\int_0^1 q_{Y|X}(u)\psi(u)du = X'\beta$. On the other hand, whenever $q_{Y|X}(u)$ is not linear in $X$, we still have $\widetilde{\beta}(u) \to_P \bar{\beta}(u)$, so that $\widetilde{\beta} = \int_0^1 \widetilde{\beta}(u)\psi(u)du \to_P \int_0^1 \bar{\beta}(u)\psi(u)du$,

but in general $\int_0^1 \bar{\beta}(u)\psi(u)du \neq \beta$ in this case. Indeed, consider the following data-generating process:

$$Y = X\beta + X^2\gamma(U),$$

where $X \sim U[0,2]$ and $U \sim U[0,1]$ are independent random variables, $\beta$ is any constant, and $\gamma(u) = 4u - 3$ for all $u \in [0,1]$. Suppose that $\psi(u) = 2\mathbb{I}\{u > 1/2\}$ for all $u \in (0,1)$. It is then easy to check that $\int_0^1 q_{Y|X}(u)\psi(u)du = X'\beta$ but, as we show below, $2\int_{1/2}^1 \bar{\beta}(u)du = \beta + 19/6 - 8/\sqrt{6}$. Therefore, the parametric estimator $\widetilde{\beta}$ is not consistent in this case, whereas our estimator $\widehat{\beta}$ is. Of course, the problem for the parametric estimator here is that $q_{Y|X}(u) = X\beta(u) + X^2\gamma(u)$ is not linear in $X$.

In addition, another advantage of our estimator $\widehat{\beta}$ over the parametric estimator $\widetilde{\beta}$ is that the latter requires estimating $u$-quantile regressions for values of $u$ that are close to the boundaries of the interval $[0,1]$. This is problematic because such quantile regression estimators may have a slow rate of convergence, undermining the properties of the estimator $\widetilde{\beta}$. In principle, one could consider a truncated version of $\widetilde{\beta}$, namely

$$\widetilde{\beta}^{\varepsilon} = \int_{\varepsilon}^{1-\varepsilon} \widetilde{\beta}(u)\psi(u)du$$

for some $\varepsilon = \varepsilon_T \to 0$ as $T \to \infty$ but in this case, one has to find a data-driven method to choose $\varepsilon$, and we are not aware of such methods. In contrast, although our estimator $\widehat{\beta}$ requires estimating the function $F$ via nonparametric/machine learning methods, which also rely on tuning parameters, there is a variety of methods in the literature, such as sample splitting and cross-validation, to choose these tuning parameters.

We now prove that $2\int_{1/2}^1 \bar{\beta}(u)du = 2\int_{1/2}^2 \beta(u)du + 19/6 - 8/\sqrt{6}$. This calculation demonstrates that the parametric estimator described above is not consistent. Fix $u \in (0,1)$ and $b \in \mathbb{R}$ and denote $\tilde{b} = b - \beta$. First, consider the case $b \geq \beta$. In this case, we have

$$\begin{aligned}
\mathrm{E}[(Y - Xb)_+ \mid X] &= \mathrm{E}[(X^2\gamma(U) - X\tilde{b})_+ \mid X] = \int_0^\infty \mathrm{P}(X^2\gamma(U) - X\tilde{b} > s \mid X)ds \\
&= \int_0^\infty \mathrm{P}\left(U > \frac{1}{4}\left(3 + \frac{\tilde{b}}{X} + \frac{s}{X^2}\right) \mid X\right) ds = \begin{cases} 0 & \text{if } X \leq \tilde{b}, \\ \frac{(X-\tilde{b})^2}{8} & \text{if } X > \tilde{b}, \end{cases}
\end{aligned}$$

and

$$\mathrm{E}[(Y - Xb)_- \mid X] = \int_0^\infty \mathrm{P}\left(U < \frac{1}{4}\left(3 + \frac{\tilde{b}}{X} - \frac{s}{X^2}\right) \mid X\right) = \begin{cases} X^2 + X\tilde{b} & \text{if } X \leq \tilde{b}, \\ \frac{(3X+\tilde{b})^2}{8} & \text{if } X > \tilde{b}. \end{cases}$$

Therefore, for $\beta \le b < \beta + 2$,

$$\frac{d}{db}\mathrm{E}[(Y - Xb)_+] = \mathrm{E}\left[\frac{d}{db}\mathrm{E}[(Y - Xb)_+ \mid X]\right] = \mathrm{E}\left[\frac{\tilde{b} - X}{4}\mathbb{I}\{X > \tilde{b}\}\right] = -\frac{\tilde{b}^2}{16} + \frac{\tilde{b}}{4} - \frac{1}{4}$$

and

$$\begin{aligned}\frac{d}{db}\mathrm{E}[(Y - Xb)_-] &= \mathrm{E}\left[\frac{d}{db}\mathrm{E}[(Y - Xb)_- \mid X]\right] \\ &= \mathrm{E}\left[\frac{3X + \tilde{b}}{4}\mathbb{I}\{X > \tilde{b}\} + X\mathbb{I}\{X \le \tilde{b}\}\right] = -\frac{\tilde{b}^2}{16} + \frac{\tilde{b}}{4} + \frac{3}{4},\end{aligned}$$

whereas for $b \ge \beta + 2$,

$$\frac{d}{db}\mathrm{E}[(Y - Xb)_+] = 0 \text{ and } \frac{d}{db}\mathrm{E}[(Y - Xb)_-] = \mathrm{E}[X] = 1.$$

Next, consider the case $b < \beta$. In this case, we have

$$\mathrm{E}[(Y - Xb)_+ \mid X] = \begin{cases} -X^2 - X\tilde{b} & \text{if } X \le -\tilde{b}/3, \\ \frac{(X - \tilde{b})^2}{8} & \text{if } X > -\tilde{b}/3. \end{cases}$$

and

$$\mathrm{E}[(Y - Xb)_- \mid X] = \begin{cases} 0 & \text{if } X \le -\tilde{b}/3, \\ \frac{(3X + \tilde{b})^2}{8} & \text{if } X > -\tilde{b}/3. \end{cases}$$

Therefore, for $\beta - 6 < b < \beta$,

$$\begin{aligned}\frac{d}{db}\mathrm{E}[(Y - Xb)_+] &= \mathrm{E}\left[\frac{d}{db}\mathrm{E}[(Y - Xb)_+ \mid X]\right] \\ &= \mathrm{E}\left[\frac{\tilde{b} - X}{4}\mathbb{I}\{X > -\tilde{b}/3\} - X\mathbb{I}\{X \le -\tilde{b}/3\}\right] = \frac{\tilde{b}^2}{48} + \frac{\tilde{b}}{4} - \frac{1}{4}\end{aligned}$$

and

$$\frac{d}{db}\mathrm{E}[(Y - Xb)_-] = \mathrm{E}\left[\frac{d}{db}\mathrm{E}[(Y - Xb)_- \mid X]\right] = \mathrm{E}\left[\frac{3X + \tilde{b}}{4}\mathbb{I}\{X > -\tilde{b}/3\}\right] = \frac{\tilde{b}^2}{48} + \frac{\tilde{b}}{4} + \frac{3}{4},$$

whereas for $b \le \beta - 6$,

$$\frac{d}{db}\mathrm{E}[(Y - Xb)_+] = -\mathrm{E}[X] = -1 \text{ and } \frac{d}{db}\mathrm{E}[(Y - Xb)_-] = 0.$$

Hence,

$$\frac{d}{db}\Big\{u\mathrm{E}[(Y-Xb)_+]+(1-u)\mathrm{E}[(Y-Xb)_-]\Big\}=\begin{cases}-1 & \text{if } b\leq\beta-6,\\ \frac{\tilde{b}^2}{48}+\frac{\tilde{b}}{4}+\frac{3}{4}-u & \text{if } \beta-6<b<\beta,\\ -\frac{\tilde{b}^2}{16}+\frac{\tilde{b}}{4}+\frac{3}{4}-u & \text{if } \beta\leq b<\beta+2,\\ +1 & \text{if } b\geq\beta+2.\end{cases}$$

Thus, by the first-order conditions, the solution to the optimization problem in (45) is

$$\bar{\beta}(u)=\begin{cases}\beta-6+4\sqrt{3u} & \text{if } u<3/4,\\ \beta+2-4\sqrt{1-u} & \text{if } u\geq 3/4.\end{cases}$$

We conclude that

$$2\int_{1/2}^{1}\bar{\beta}(u)du=\beta+2\int_{1/2}^{3/4}(-6+4\sqrt{3u})du+2\int_{3/4}^{1}(2-4\sqrt{1-u})du=\beta+\frac{19}{6}-\frac{8}{\sqrt{6}}\neq\beta.$$

This means that the parametric estimator described above is not consistent.

## Appendix D. Additional Tables & Figures

Table A.1. Results of Monte Carlo simulation study for the coverage probability of 90% confidence intervals.

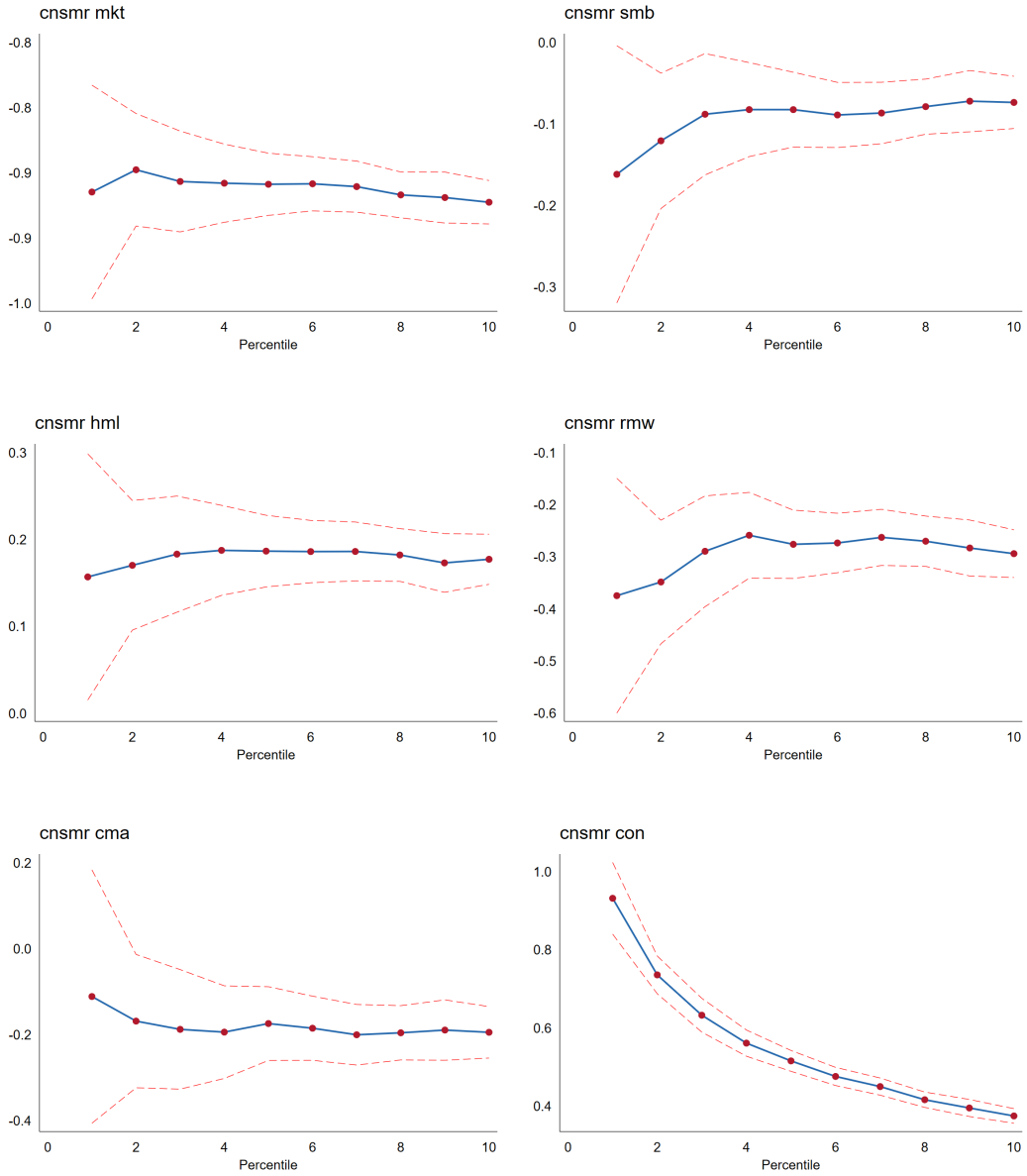
| Panel A: Homoscedastic Noise | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $e \sim N(0,1)$ | | | | $e \sim t(4)$ | | | |
| $\psi$-type | $\beta_1$ | $p=2$ | | $p=5$ | | $p=2$ | | $p=5$ | |
| | | $T=1000$ | $T=2000$ | $T=1000$ | $T=2000$ | $T=1000$ | $T=2000$ | $T=1000$ | $T=2000$ |
| 1 | .0 | 0.912 | 0.918 | 0.894 | 0.874 | 0.896 | 0.898 | 0.874 | 0.874 |
| | .3 | 0.926 | 0.9 | 0.918 | 0.872 | 0.892 | 0.888 | 0.86 | 0.882 |
| | .6 | 0.906 | 0.89 | 0.906 | 0.888 | 0.916 | 0.898 | 0.866 | 0.878 |
| | .9 | 0.912 | 0.884 | 0.886 | 0.876 | 0.886 | 0.892 | 0.868 | 0.886 |
| 2 | .0 | 0.882 | 0.908 | 0.922 | 0.888 | 0.91 | 0.888 | 0.89 | 0.882 |
| | .3 | 0.89 | 0.908 | 0.896 | 0.898 | 0.908 | 0.902 | 0.876 | 0.9 |
| | .6 | 0.9 | 0.906 | 0.878 | 0.866 | 0.89 | 0.89 | 0.874 | 0.898 |
| | .9 | 0.892 | 0.892 | 0.87 | 0.89 | 0.902 | 0.884 | 0.854 | 0.886 |
| 3 | .0 | 0.884 | 0.874 | 0.892 | 0.904 | 0.874 | 0.896 | 0.888 | 0.898 |
| | .3 | 0.878 | 0.878 | 0.884 | 0.88 | 0.872 | 0.912 | 0.878 | 0.896 |
| | .6 | 0.89 | 0.886 | 0.876 | 0.89 | 0.878 | 0.898 | 0.882 | 0.908 |
| | .9 | 0.88 | 0.874 | 0.886 | 0.886 | 0.884 | 0.896 | 0.87 | 0.894 |
| 4 | .0 | 0.92 | 0.906 | 0.912 | 0.884 | 0.906 | 0.896 | 0.866 | 0.89 |
| | .3 | 0.92 | 0.914 | 0.914 | 0.876 | 0.902 | 0.888 | 0.86 | 0.882 |
| | .6 | 0.918 | 0.914 | 0.912 | 0.874 | 0.914 | 0.882 | 0.848 | 0.896 |
| | .9 | 0.928 | 0.906 | 0.9 | 0.876 | 0.898 | 0.894 | 0.872 | 0.892 |
| Panel B: Heteroscedastic noise | | | | | | | | | |
| | | $e \sim N(0,1)$ | | | | $e \sim t(4)$ | | | |
| $\psi$-type | $\beta_1$ | $p=2$ | | $p=5$ | | $p=2$ | | $p=5$ | |
| | | $T=1000$ | $T=2000$ | $T=1000$ | $T=2000$ | $T=1000$ | $T=2000$ | $T=1000$ | $T=2000$ |
| 1 | .0 | 0.916 | 0.906 | 0.872 | 0.862 | 0.884 | 0.888 | 0.846 | 0.88 |
| | .3 | 0.906 | 0.906 | 0.916 | 0.848 | 0.886 | 0.88 | 0.862 | 0.868 |
| | .6 | 0.902 | 0.908 | 0.91 | 0.886 | 0.88 | 0.89 | 0.844 | 0.872 |
| | .9 | 0.902 | 0.902 | 0.912 | 0.88 | 0.876 | 0.886 | 0.842 | 0.862 |
| 2 | .0 | 0.882 | 0.896 | 0.894 | 0.884 | 0.888 | 0.87 | 0.872 | 0.888 |
| | .3 | 0.886 | 0.884 | 0.906 | 0.876 | 0.884 | 0.878 | 0.88 | 0.894 |
| | .6 | 0.886 | 0.91 | 0.894 | 0.88 | 0.882 | 0.892 | 0.854 | 0.89 |
| | .9 | 0.866 | 0.884 | 0.898 | 0.86 | 0.858 | 0.866 | 0.87 | 0.896 |
| 3 | .0 | 0.89 | 0.87 | 0.884 | 0.882 | 0.886 | 0.918 | 0.878 | 0.9 |
| | .3 | 0.884 | 0.876 | 0.88 | 0.87 | 0.88 | 0.906 | 0.868 | 0.898 |
| | .6 | 0.886 | 0.88 | 0.878 | 0.884 | 0.892 | 0.914 | 0.876 | 0.908 |
| | .9 | 0.888 | 0.868 | 0.886 | 0.868 | 0.886 | 0.906 | 0.88 | 0.902 |
| 4 | .0 | 0.898 | 0.916 | 0.882 | 0.876 | 0.876 | 0.894 | 0.846 | 0.878 |
| | .3 | 0.926 | 0.912 | 0.9 | 0.86 | 0.878 | 0.892 | 0.84 | 0.884 |
| | .6 | 0.914 | 0.912 | 0.918 | 0.864 | 0.88 | 0.892 | 0.842 | 0.878 |
| | .9 | 0.91 | 0.898 | 0.91 | 0.876 | 0.876 | 0.874 | 0.844 | 0.882 |

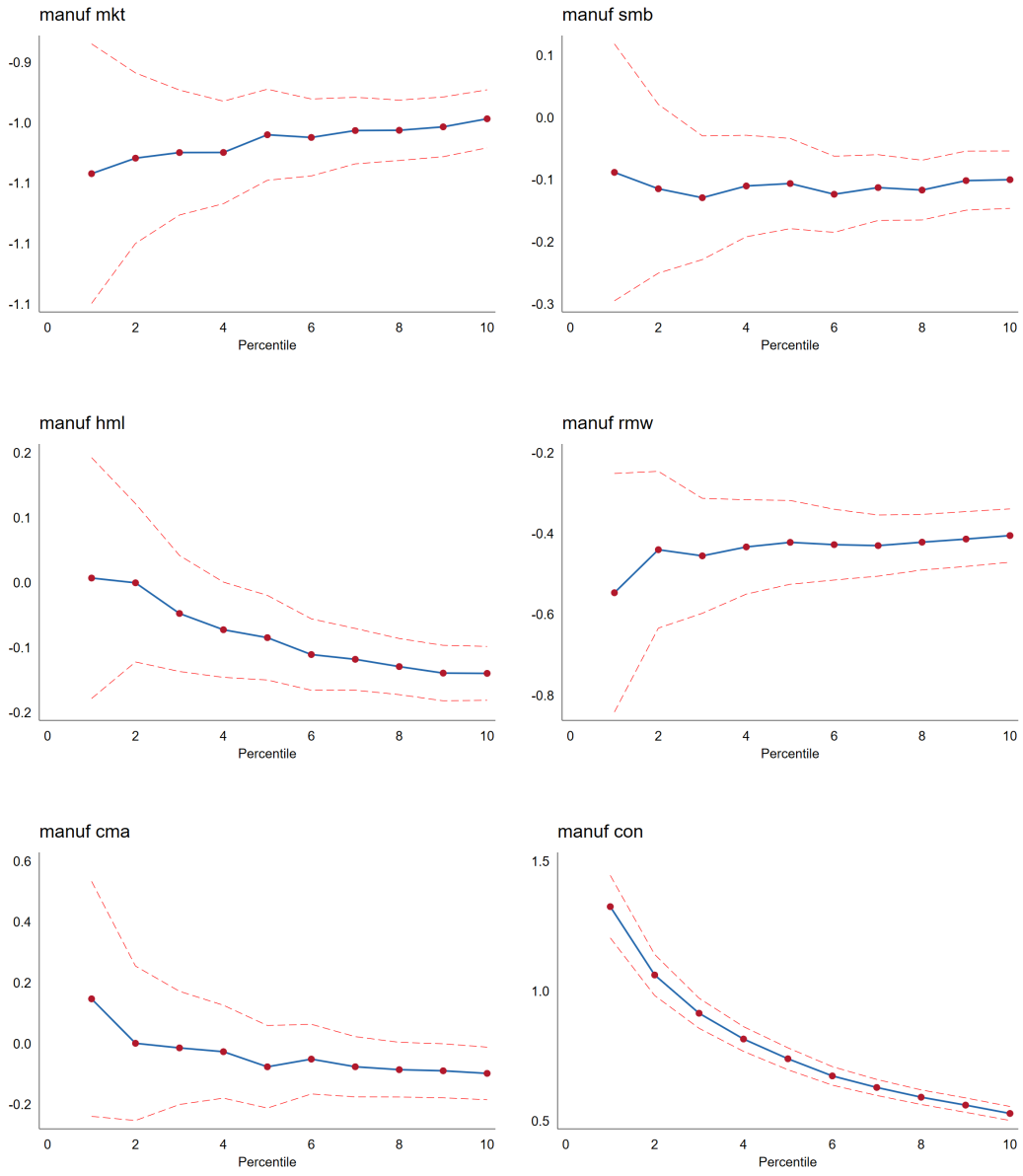
TABLE A.2. Results of Monte Carlo simulation study for the mean absolute error.

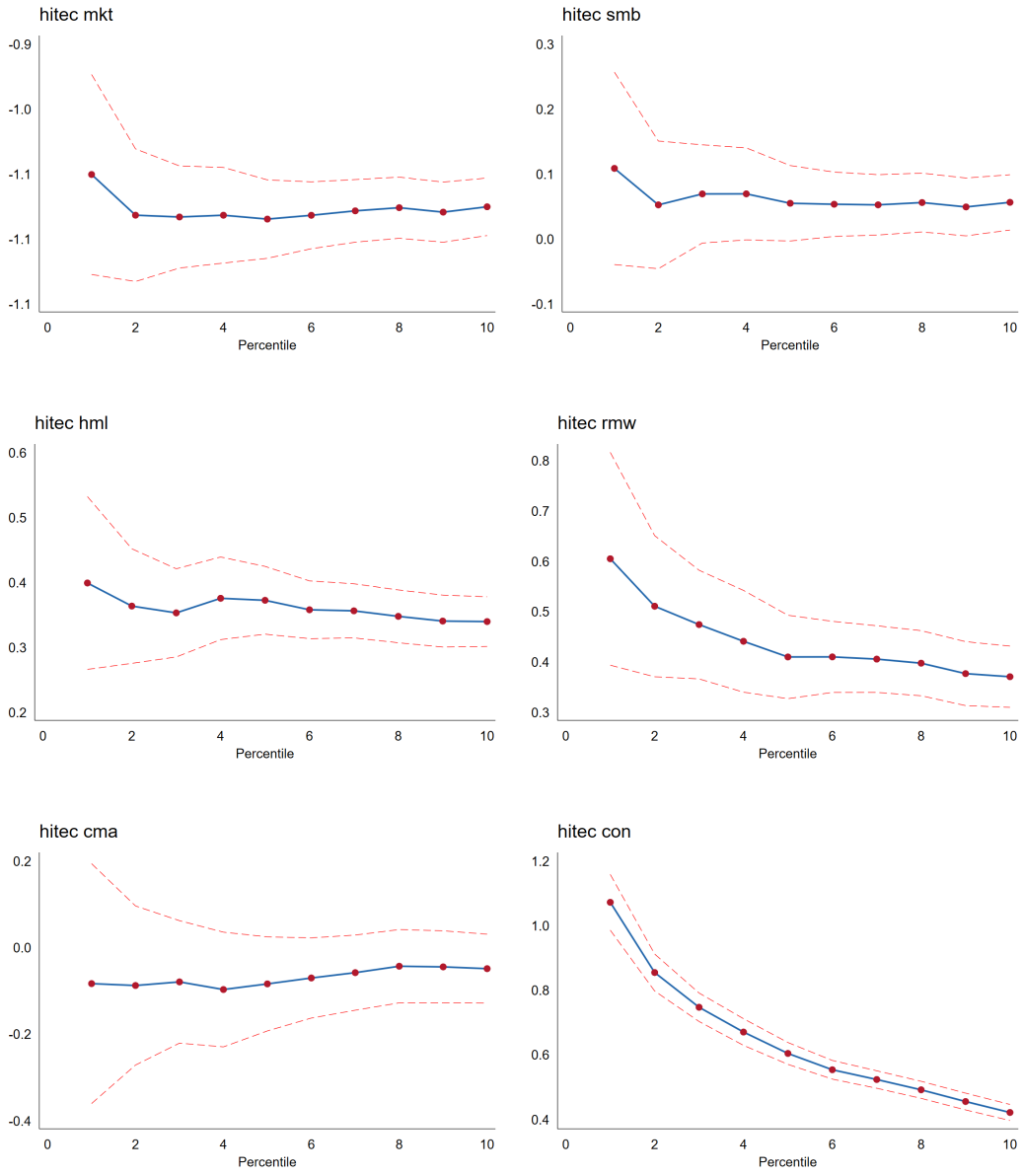
| Panel A: DGP1, Homoscedastic Noise | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $\psi$-type | $\beta_1$ | $e \sim N(0,1)$ | | | | $e \sim t(4)$ | | | |
| | | $p=2$ | | $p=5$ | | $p=2$ | | $p=5$ | |
| | | $T=1000$ | $T=2000$ | $T=1000$ | $T=2000$ | $T=1000$ | $T=2000$ | $T=1000$ | $T=2000$ |
| 1 | .0 | 0.158 | 0.107 | 0.171 | 0.12 | 0.373 | 0.255 | 0.381 | 0.254 |
| | .3 | 0.152 | 0.109 | 0.166 | 0.123 | 0.368 | 0.252 | 0.379 | 0.257 |
| | .6 | 0.151 | 0.113 | 0.171 | 0.123 | 0.369 | 0.253 | 0.376 | 0.259 |
| | .9 | 0.152 | 0.115 | 0.175 | 0.123 | 0.368 | 0.243 | 0.375 | 0.255 |
| 2 | .0 | 0.214 | 0.152 | 0.215 | 0.154 | 0.478 | 0.339 | 0.485 | 0.316 |
| | .3 | 0.212 | 0.147 | 0.224 | 0.153 | 0.463 | 0.332 | 0.472 | 0.319 |
| | .6 | 0.21 | 0.147 | 0.219 | 0.158 | 0.464 | 0.333 | 0.465 | 0.318 |
| | .9 | 0.2 | 0.15 | 0.229 | 0.159 | 0.467 | 0.321 | 0.483 | 0.319 |
| 3 | .0 | 0.076 | 0.057 | 0.074 | 0.054 | 0.086 | 0.058 | 0.088 | 0.058 |
| | .3 | 0.078 | 0.058 | 0.074 | 0.056 | 0.086 | 0.058 | 0.089 | 0.058 |
| | .6 | 0.077 | 0.058 | 0.075 | 0.056 | 0.087 | 0.059 | 0.088 | 0.057 |
| | .9 | 0.079 | 0.06 | 0.077 | 0.056 | 0.089 | 0.06 | 0.093 | 0.059 |
| 4 | .0 | 0.14 | 0.097 | 0.153 | 0.104 | 0.326 | 0.226 | 0.336 | 0.218 |
| | .3 | 0.138 | 0.097 | 0.152 | 0.108 | 0.324 | 0.218 | 0.336 | 0.221 |
| | .6 | 0.136 | 0.099 | 0.156 | 0.109 | 0.327 | 0.219 | 0.335 | 0.222 |
| | .9 | 0.131 | 0.1 | 0.152 | 0.107 | 0.323 | 0.215 | 0.331 | 0.224 |
| Panel B: DGP2, Heteroscedastic noise | | | | | | | | | |
| $\psi$-type | $\beta_1$ | $e \sim N(0,1)$ | | | | $e \sim t(4)$ | | | |
| | | $p=2$ | | $p=5$ | | $p=2$ | | $p=5$ | |
| | | $T=1000$ | $T=2000$ | $T=1000$ | $T=2000$ | $T=1000$ | $T=2000$ | $T=1000$ | $T=2000$ |
| 1 | .0 | 0.201 | 0.143 | 0.232 | 0.159 | 0.479 | 0.335 | 0.497 | 0.32 |
| | .3 | 0.204 | 0.147 | 0.219 | 0.16 | 0.487 | 0.336 | 0.494 | 0.327 |
| | .6 | 0.206 | 0.152 | 0.234 | 0.168 | 0.483 | 0.332 | 0.494 | 0.331 |
| | .9 | 0.203 | 0.146 | 0.222 | 0.161 | 0.48 | 0.324 | 0.482 | 0.329 |
| 2 | .0 | 0.278 | 0.194 | 0.278 | 0.194 | 0.596 | 0.435 | 0.634 | 0.407 |
| | .3 | 0.277 | 0.194 | 0.279 | 0.199 | 0.603 | 0.43 | 0.624 | 0.399 |
| | .6 | 0.272 | 0.191 | 0.279 | 0.208 | 0.601 | 0.426 | 0.622 | 0.401 |
| | .9 | 0.267 | 0.191 | 0.278 | 0.216 | 0.604 | 0.427 | 0.602 | 0.406 |
| 3 | .0 | 0.098 | 0.074 | 0.095 | 0.071 | 0.112 | 0.075 | 0.113 | 0.073 |
| | .3 | 0.097 | 0.074 | 0.096 | 0.072 | 0.108 | 0.075 | 0.111 | 0.073 |
| | .6 | 0.098 | 0.074 | 0.095 | 0.071 | 0.111 | 0.075 | 0.111 | 0.073 |
| | .9 | 0.099 | 0.076 | 0.095 | 0.072 | 0.111 | 0.076 | 0.112 | 0.073 |
| 4 | .0 | 0.183 | 0.127 | 0.2 | 0.138 | 0.425 | 0.294 | 0.429 | 0.278 |
| | .3 | 0.184 | 0.128 | 0.202 | 0.144 | 0.429 | 0.291 | 0.432 | 0.282 |
| | .6 | 0.187 | 0.132 | 0.205 | 0.149 | 0.42 | 0.291 | 0.439 | 0.286 |
| | .9 | 0.182 | 0.131 | 0.204 | 0.15 | 0.427 | 0.291 | 0.434 | 0.284 |

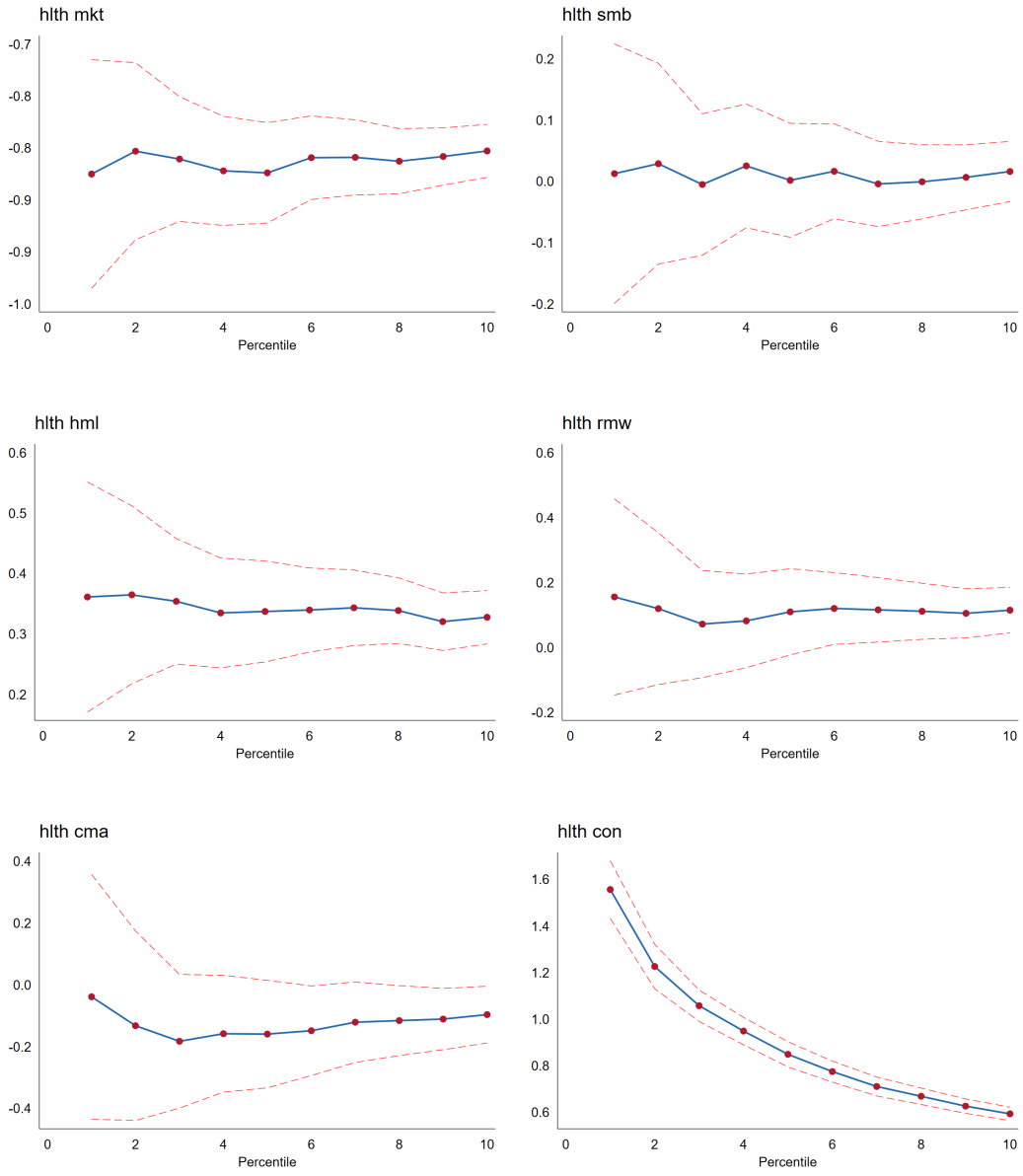
FIGURE A.1. Coefficient by Percentiles

This figure plots the coefficient estimates and the 95% confidence intervals for the 1% to 10% quantile regressions. The dependent variables are the excess returns of the Fama-French 5 industries. The dependent variables are the Fama-French 5 factors. The estimates are multiplied by -1 to be consistent with the risk regressions.




manuf mkt
-0.9
-1.0
-1.1
-1.1
-1.1
0
2
4
6
8
10
Percentile
manuf smb
0.1
0.0
-0.1
-0.2
-0.3
0
2
4
6
8
10
Percentile
manuf hml
0.2
0.1
0.0
-0.1
-0.2
0
2
4
6
8
10
Percentile
manuf rmw
-0.2
-0.4
-0.6
-0.8
0
2
4
6
8
10
Percentile
manuf cma
0.6
0.4
0.2
0.0
-0.2
0
2
4
6
8
10
Percentile
manuf con
1.5
1.0
0.5
0
2
4
6
8
10
Percentile


hitec mkt
-0.9
-1.0
-1.1
-1.1
-1.1
0
2
4
6
8
10
Percentile
hitec smb
0.3
0.2
0.1
0.0
-0.1
0
2
4
6
8
10
Percentile
hitec hml
0.6
0.5
0.4
0.3
0.2
0
2
4
6
8
10
Percentile
hitec rmw
0.8
0.7
0.6
0.5
0.4
0.3
0
2
4
6
8
10
Percentile
hitec cma
0.2
0.0
-0.2
-0.4
0
2
4
6
8
10
Percentile
hitec con
1.2
1.0
0.8
0.6
0.4
0
2
4
6
8
10
Percentile


hlth mkt
-0.7
-0.8
-0.8
-0.9
-0.9
-1.0
0
2
4
6
8
10
Percentile
hlth smb
0.2
0.1
0.0
-0.1
-0.2
0
2
4
6
8
10
Percentile
hlth hml
0.6
0.5
0.4
0.3
0.2
0
2
4
6
8
10
Percentile
hlth rmw
0.6
0.4
0.2
0.0
-0.2
0
2
4
6
8
10
Percentile
hlth cma
0.4
0.2
0.0
-0.2
-0.4
0
2
4
6
8
10
Percentile
hlth con
1.6
1.4
1.2
1.0
0.8
0.6
0
2
4
6
8
10
Percentile

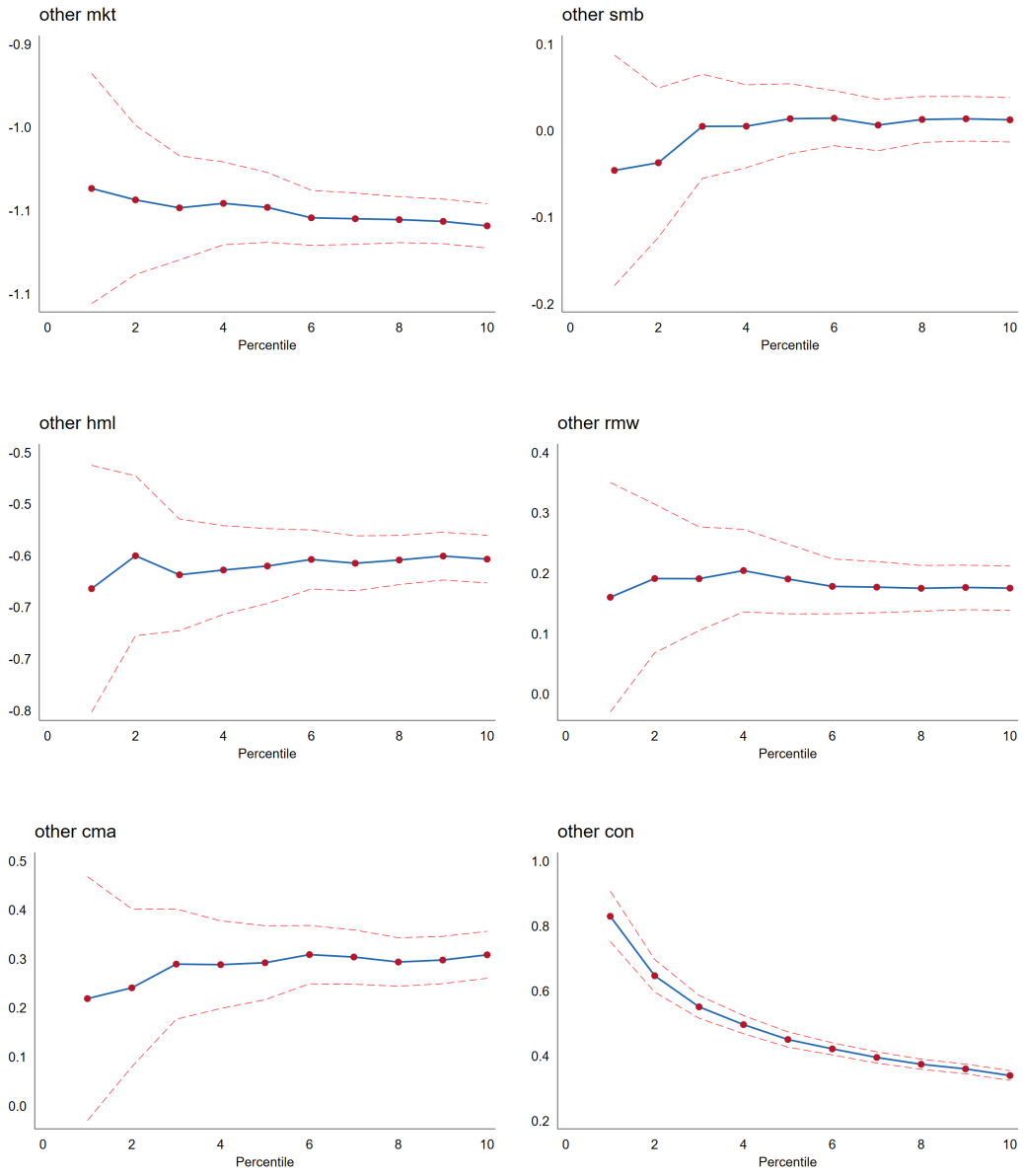

other mkt
-0.9
-1.0
-1.1
-1.1
0
2
4
6
8
10
Percentile
other smb
0.1
0.0
-0.1
-0.2
0
2
4
6
8
10
Percentile
other hml
-0.5
-0.5
-0.6
-0.7
-0.7
-0.8
0
2
4
6
8
10
Percentile
other rmw
0.4
0.3
0.2
0.1
0.0
0
2
4
6
8
10
Percentile
other cma
0.5
0.4
0.3
0.2
0.1
0.0
0
2
4
6
8
10
Percentile
other con
1.0
0.8
0.6
0.4
0.2
0
2
4
6
8
10
Percentile

TABLE A.3. Industry Factor Loadings – Newey-West Adjusted

This table shows the WAQR estimates of industry factor loadings for the Fama-French 5 industries for the 10% ES regression. The standard errors are reported for both the normal and Newey-West adjusted method. Standard errors are reported in parentheses.

| | MKTRF | | SMB | | HML | | RMW | | CMA | | Constant | | Adjustment |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Cnsmr | -1.196 | (0.022) | -0.122 | (0.043) | 0.214 | (0.039) | -0.779 | (0.061) | -0.250 | (0.081) | 0.909 | (0.025) | No Adjust |
| | -1.196 | (0.050) | -0.122 | (0.093) | 0.214 | (0.112) | -0.779 | (0.134) | -0.250 | (0.185) | 0.909 | (0.029) | Newey-West |
| Manuf | -1.293 | (0.029) | -0.066 | (0.057) | -0.514 | (0.051) | -0.832 | (0.081) | 0.122 | (0.106) | 1.188 | (0.033) | No Adjust |
| | -1.293 | (0.058) | -0.066 | (0.125) | -0.514 | (0.131) | -0.832 | (0.122) | 0.122 | (0.205) | 1.188 | (0.037) | Newey-West |
| Hitec | -1.400 | (0.027) | 0.119 | (0.052) | 0.297 | (0.046) | 1.125 | (0.074) | -0.280 | (0.097) | 1.082 | (0.030) | No Adjust |
| | -1.400 | (0.058) | 0.119 | (0.099) | 0.297 | (0.117) | 1.125 | (0.199) | -0.280 | (0.210) | 1.082 | (0.033) | Newey-West |
| Hlth | -1.153 | (0.026) | 0.096 | (0.051) | 0.251 | (0.045) | 0.190 | (0.072) | -0.045 | (0.094) | 1.207 | (0.029) | No Adjust |
| | -1.153 | (0.056) | 0.096 | (0.087) | 0.251 | (0.095) | 0.190 | (0.105) | -0.045 | (0.168) | 1.207 | (0.033) | Newey-West |
| Other | -1.330 | (0.033) | 0.349 | (0.064) | -1.685 | (0.058) | 0.278 | (0.092) | 0.765 | (0.120) | 1.043 | (0.037) | No Adjust |
| | -1.330 | (0.068) | 0.349 | (0.129) | -1.685 | (0.197) | 0.278 | (0.144) | 0.765 | (0.224) | 1.043 | (0.041) | Newey-West |

TABLE A.4. Quantile Regression – Higher Order

This table shows the quantile regression results of regressing the industry returns to the higher moments and interaction terms of the Fama-French 5 factors. Standard errors are reported in parentheses.

| Panel A | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
|---|---|---|---|---|---|---|---|---|---|---|
| Ind | cnsmr | cnsmr | manuf | manuf | hitec | hitec | hlth | hlth | other | other |
| Quantile | 0.1 | 0.05 | 0.1 | 0.05 | 0.1 | 0.05 | 0.1 | 0.05 | 0.1 | 0.05 |
| $mktrf$ | 0.88 | 0.89 | 0.99 | 0.99 | 1.08 | 1.10 | 0.82 | 0.83 | 1.07 | 1.07 |
| | (0.01) | (0.01) | (0.01) | (0.02) | (0.01) | (0.01) | (0.01) | (0.02) | (0.01) | (0.01) |
| $smb$ | 0.04 | 0.08 | 0.14 | 0.13 | -0.06 | -0.10 | -0.06 | -0.09 | -0.03 | -0.03 |
| | (0.02) | (0.02) | (0.02) | (0.03) | (0.02) | (0.02) | (0.03) | (0.04) | (0.01) | (0.02) |
| $hml$ | -0.17 | -0.14 | 0.22 | 0.19 | -0.34 | -0.33 | -0.43 | -0.49 | 0.54 | 0.56 |
| | (0.02) | (0.02) | (0.02) | (0.03) | (0.02) | (0.02) | (0.03) | (0.04) | (0.01) | (0.02) |
| $rmw$ | 0.31 | 0.26 | 0.39 | 0.44 | -0.33 | -0.34 | -0.31 | -0.41 | -0.20 | -0.24 |
| | (0.03) | (0.04) | (0.04) | (0.05) | (0.03) | (0.04) | (0.04) | (0.07) | (0.02) | (0.03) |
| $cma$ | 0.26 | 0.20 | -0.05 | -0.09 | 0.09 | 0.09 | 0.08 | 0.12 | -0.19 | -0.19 |
| | (0.03) | (0.04) | (0.05) | (0.07) | (0.03) | (0.05) | (0.05) | (0.08) | (0.03) | (0.04) |
| $mktrf^2$ | -0.01 | -0.02 | -0.02 | -0.02 | -0.00 | -0.00 | -0.02 | -0.01 | -0.01 | -0.01 |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |
| $smb^2$ | -0.05 | -0.06 | -0.06 | -0.02 | -0.00 | -0.02 | -0.10 | -0.14 | -0.02 | -0.02 |
| | (0.01) | (0.01) | (0.01) | (0.02) | (0.01) | (0.01) | (0.02) | (0.03) | (0.01) | (0.01) |
| $hml^2$ | -0.00 | 0.01 | -0.06 | -0.10 | -0.02 | -0.02 | -0.05 | -0.08 | -0.02 | -0.03 |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.00) | (0.01) |
| $rmw^2$ | -0.15 | -0.24 | -0.03 | -0.15 | -0.30 | -0.31 | -0.17 | -0.24 | -0.14 | -0.22 |
| | (0.02) | (0.03) | (0.03) | (0.05) | (0.02) | (0.03) | (0.04) | (0.06) | (0.02) | (0.03) |
| $cma^2$ | -0.17 | -0.20 | -0.41 | -0.54 | -0.45 | -0.57 | -0.26 | -0.42 | -0.13 | -0.09 |
| | (0.04) | (0.05) | (0.05) | (0.07) | (0.04) | (0.05) | (0.05) | (0.09) | (0.03) | (0.04) |
| $mktrf^3$ | -0.00 | -0.00 | 0.00 | 0.00 | -0.00 | -0.00 | -0.00 | -0.00 | -0.00 | -0.00 |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |
| $smb^3$ | 0.01 | -0.00 | -0.02 | -0.02 | 0.00 | 0.00 | 0.02 | 0.03 | 0.00 | 0.00 |
| | (0.00) | (0.00) | (0.00) | (0.01) | (0.00) | (0.00) | (0.00) | (0.01) | (0.00) | (0.00) |
| $hml^3$ | -0.00 | -0.00 | -0.02 | -0.02 | 0.00 | 0.00 | 0.02 | 0.03 | 0.01 | 0.01 |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |
| $rmw^3$ | 0.03 | 0.08 | -0.01 | -0.05 | -0.06 | -0.09 | 0.19 | 0.23 | 0.08 | 0.11 |
| | (0.01) | (0.02) | (0.02) | (0.03) | (0.01) | (0.02) | (0.02) | (0.03) | (0.01) | (0.01) |
| $cma^3$ | -0.07 | 0.10 | 0.02 | 0.09 | -0.20 | -0.25 | 0.13 | -0.05 | -0.08 | -0.03 |
| | (0.02) | (0.03) | (0.03) | (0.04) | (0.02) | (0.03) | (0.03) | (0.06) | (0.02) | (0.02) |
| Cons | -0.30 | -0.40 | -0.40 | -0.55 | -0.30 | -0.42 | -0.46 | -0.62 | -0.27 | -0.36 |
| | (0.01) | (0.01) | (0.01) | (0.02) | (0.01) | (0.01) | (0.02) | (0.03) | (0.01) | (0.01) |

| Panel B | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
|---|---|---|---|---|---|---|---|---|---|---|
| Ind | cnsmr | cnsmr | manuf | manuf | hitec | hitec | hlth | hlth | other | other |
| Quantile | 0.1 | 0.05 | 0.1 | 0.05 | 0.1 | 0.05 | 0.1 | 0.05 | 0.1 | 0.05 |
| $mktrf$ | 0.87 | 0.87 | 1.02 | 1.03 | 1.07 | 1.08 | 0.80 | 0.79 | 1.06 | 1.04 |
| | (0.01) | (0.01) | (0.01) | (0.02) | (0.01) | (0.02) | (0.01) | (0.02) | (0.01) | (0.01) |
| $smb$ | 0.06 | 0.09 | 0.10 | 0.07 | -0.07 | -0.08 | -0.02 | -0.05 | 0.00 | -0.01 |
| | (0.02) | (0.02) | (0.02) | (0.03) | (0.02) | (0.03) | (0.02) | (0.04) | (0.01) | (0.02) |
| $hml$ | -0.19 | -0.18 | 0.14 | 0.11 | -0.33 | -0.32 | -0.32 | -0.32 | 0.60 | 0.62 |
| | (0.01) | (0.02) | (0.02) | (0.03) | (0.02) | (0.03) | (0.02) | (0.04) | (0.01) | (0.02) |
| $rmw$ | 0.30 | 0.30 | 0.41 | 0.45 | -0.40 | -0.47 | -0.14 | -0.19 | -0.16 | -0.17 |
| | (0.02) | (0.03) | (0.04) | (0.05) | (0.03) | (0.04) | (0.03) | (0.06) | (0.02) | (0.03) |
| $cma$ | 0.21 | 0.19 | 0.03 | -0.02 | 0.07 | 0.05 | 0.07 | 0.09 | -0.30 | -0.30 |
| | (0.03) | (0.04) | (0.05) | (0.06) | (0.04) | (0.05) | (0.04) | (0.08) | (0.02) | (0.04) |
| $mktrf \times smb$ | 0.03 | 0.02 | 0.06 | 0.09 | 0.03 | 0.04 | 0.04 | 0.05 | 0.02 | 0.01 |
| | (0.01) | (0.01) | (0.01) | (0.02) | (0.01) | (0.01) | (0.01) | (0.02) | (0.01) | (0.01) |
| $mktrf \times hml$ | -0.01 | -0.02 | -0.06 | -0.06 | -0.01 | -0.00 | -0.03 | -0.04 | -0.01 | -0.01 |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.00) | (0.01) |
| $mktrf \times rmw$ | 0.02 | 0.00 | -0.01 | -0.02 | 0.06 | 0.06 | 0.08 | 0.08 | -0.00 | -0.02 |
| | (0.01) | (0.02) | (0.02) | (0.03) | (0.02) | (0.02) | (0.02) | (0.03) | (0.01) | (0.02) |
| $mktrf \times cma$ | 0.01 | -0.00 | 0.14 | 0.13 | 0.03 | 0.02 | 0.03 | 0.01 | 0.04 | 0.05 |
| | (0.02) | (0.02) | (0.03) | (0.03) | (0.02) | (0.03) | (0.02) | (0.04) | (0.01) | (0.02) |
| $smb \times hml$ | -0.03 | -0.03 | -0.10 | -0.13 | -0.05 | -0.05 | -0.03 | -0.06 | -0.06 | -0.05 |
| | (0.01) | (0.02) | (0.02) | (0.02) | (0.01) | (0.02) | (0.02) | (0.03) | (0.01) | (0.02) |
| $smb \times rmw$ | -0.01 | -0.03 | 0.04 | 0.13 | 0.08 | 0.08 | -0.02 | -0.06 | 0.03 | 0.01 |
| | (0.03) | (0.04) | (0.04) | (0.05) | (0.03) | (0.05) | (0.04) | (0.07) | (0.02) | (0.03) |
| $smb \times cma$ | 0.05 | 0.03 | 0.21 | 0.20 | -0.06 | -0.04 | 0.10 | 0.13 | 0.04 | 0.04 |
| | (0.03) | (0.05) | (0.05) | (0.07) | (0.04) | (0.06) | (0.05) | (0.09) | (0.03) | (0.04) |
| $hml \times rmw$ | 0.02 | 0.03 | 0.09 | 0.15 | 0.08 | 0.09 | -0.05 | -0.00 | 0.01 | 0.04 |
| | (0.02) | (0.03) | (0.03) | (0.04) | (0.02) | (0.04) | (0.03) | (0.05) | (0.02) | (0.03) |
| $hml \times cma$ | -0.05 | -0.08 | -0.22 | -0.30 | -0.15 | -0.15 | -0.13 | -0.21 | -0.11 | -0.09 |
| | (0.03) | (0.04) | (0.04) | (0.05) | (0.03) | (0.05) | (0.04) | (0.07) | (0.02) | (0.03) |
| $rmw \times cma$ | -0.04 | -0.07 | 0.22 | 0.20 | 0.08 | -0.04 | -0.10 | -0.16 | 0.04 | -0.03 |
| | (0.05) | (0.07) | (0.07) | (0.09) | (0.06) | (0.08) | (0.07) | (0.12) | (0.04) | (0.06) |
| Cons | -0.36 | -0.50 | -0.50 | -0.69 | -0.39 | -0.56 | -0.57 | -0.80 | -0.32 | -0.43 |
| | (0.01) | (0.01) | (0.02) | (0.02) | (0.01) | (0.02) | (0.01) | (0.03) | (0.01) | (0.01) |

TABLE A.5. Inequality Regression

This table reports results of the inequality regression for each year from 2001 to 2018. Standard errors are reported in parentheses.

| | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 |
|---|---|---|---|---|---|---|---|---|---|
| Fam Size | 0.147 | 0.061 | 0.021 | 0.089 | 0.051 | -0.070 | -0.004 | -0.034 | 0.126 |
| | (0.052) | (0.044) | (0.049) | (0.054) | (0.053) | (0.051) | (0.048) | (0.051) | (0.049) |
| No Child | -0.151 | -0.111 | -0.024 | -0.134 | -0.115 | 0.028 | -0.026 | 0.038 | -0.194 |
| | (0.062) | (0.053) | (0.060) | (0.064) | (0.063) | (0.061) | (0.057) | (0.061) | (0.058) |
| Age | 0.019 | 0.010 | 0.034 | 0.011 | 0.013 | 0.013 | 0.001 | 0.019 | 0.020 |
| | (0.010) | (0.009) | (0.010) | (0.011) | (0.010) | (0.010) | (0.010) | (0.010) | (0.010) |
| Edu | 0.140 | 0.122 | 0.134 | 0.061 | 0.118 | 0.129 | 0.117 | 0.099 | 0.113 |
| | (0.015) | (0.014) | (0.015) | (0.016) | (0.015) | (0.015) | (0.014) | (0.015) | (0.014) |
| Constant | 0.055 | 0.705 | -0.419 | 1.138 | 0.778 | 0.886 | 1.365 | 0.784 | 0.377 |
| | (0.480) | (0.429) | (0.482) | (0.504) | (0.489) | (0.498) | (0.467) | (0.480) | (0.463) |
| | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 |
| Fam Size | -0.011 | -0.134 | -0.114 | -0.094 | -0.019 | -0.123 | -0.090 | -0.128 | -0.083 |
| | (0.048) | (0.044) | (0.048) | (0.052) | (0.044) | (0.047) | (0.047) | (0.047) | (0.049) |
| No Child | -0.042 | 0.060 | 0.067 | 0.077 | -0.071 | 0.059 | 0.041 | 0.085 | 0.040 |
| | (0.058) | (0.053) | (0.058) | (0.063) | (0.053) | (0.057) | (0.057) | (0.056) | (0.060) |
| Age | 0.000 | 0.030 | 0.012 | 0.016 | 0.007 | 0.022 | 0.010 | -0.006 | 0.018 |
| | (0.010) | (0.009) | (0.010) | (0.011) | (0.010) | (0.011) | (0.011) | (0.010) | (0.011) |
| Edu | 0.067 | 0.094 | 0.057 | 0.050 | 0.066 | 0.065 | 0.080 | 0.051 | 0.071 |
| | (0.015) | (0.014) | (0.015) | (0.017) | (0.014) | (0.015) | (0.016) | (0.015) | (0.016) |
| Constant | 1.912 | 0.618 | 1.696 | 1.623 | 1.748 | 1.274 | 1.704 | 2.643 | 1.366 |
| | (0.469) | (0.442) | (0.490) | (0.542) | (0.459) | (0.499) | (0.520) | (0.496) | (0.528) |

TABLE A.6. Social Welfare Regression

This table reports results of the social welfare (exponential) regression for each year from 2001 to 2018. Standard errors are reported in parentheses.

| | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 |
|---|---|---|---|---|---|---|---|---|---|
| Fam Size | -0.061 | -0.051 | -0.026 | -0.060 | -0.050 | 0.030 | 0.009 | 0.031 | -0.080 |
| | (0.038) | (0.031) | (0.039) | (0.045) | (0.039) | (0.038) | (0.033) | (0.036) | (0.036) |
| No Child | 0.146 | 0.166 | 0.113 | 0.166 | 0.199 | 0.091 | 0.094 | 0.061 | 0.220 |
| | (0.046) | (0.037) | (0.047) | (0.053) | (0.046) | (0.046) | (0.040) | (0.043) | (0.043) |
| Age | -0.005 | 0.004 | -0.013 | 0.007 | 0.004 | 0.014 | 0.014 | 0.000 | 0.002 |
| | (0.008) | (0.006) | (0.008) | (0.009) | (0.008) | (0.008) | (0.007) | (0.007) | (0.007) |
| Edu | 0.109 | 0.104 | 0.088 | 0.117 | 0.106 | 0.110 | 0.123 | 0.128 | 0.126 |
| | (0.011) | (0.010) | (0.012) | (0.013) | (0.011) | (0.012) | (0.010) | (0.010) | (0.010) |
| Constant | 5.179 | 4.836 | 5.675 | 4.613 | 4.738 | 4.162 | 4.222 | 4.750 | 4.828 |
| | (0.356) | (0.302) | (0.380) | (0.418) | (0.361) | (0.375) | (0.327) | (0.339) | (0.339) |

| | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 |
|---|---|---|---|---|---|---|---|---|---|
| Fam Size | 0.043 | 0.054 | 0.091 | 0.073 | 0.009 | 0.071 | 0.052 | 0.123 | 0.081 |
| | (0.036) | (0.032) | (0.035) | (0.040) | (0.031) | (0.034) | (0.035) | (0.035) | (0.037) |
| No Child | 0.082 | 0.101 | 0.025 | 0.052 | 0.149 | 0.090 | 0.092 | 0.011 | 0.038 |
| | (0.043) | (0.039) | (0.042) | (0.049) | (0.037) | (0.041) | (0.042) | (0.042) | (0.045) |
| Age | 0.023 | -0.005 | 0.015 | 0.012 | 0.018 | 0.008 | 0.020 | 0.018 | 0.003 |
| | (0.007) | (0.007) | (0.007) | (0.009) | (0.007) | (0.008) | (0.008) | (0.008) | (0.008) |
| Edu | 0.157 | 0.150 | 0.171 | 0.176 | 0.164 | 0.168 | 0.149 | 0.170 | 0.169 |
| | (0.011) | (0.010) | (0.011) | (0.013) | (0.010) | (0.011) | (0.012) | (0.011) | (0.012) |
| Constant | 3.370 | 4.651 | 3.533 | 3.598 | 3.622 | 3.890 | 3.516 | 3.370 | 4.178 |
| | (0.347) | (0.325) | (0.359) | (0.415) | (0.326) | (0.360) | (0.382) | (0.368) | (0.396) |

## Appendix E. Data

**Financial Market Data.** The Fama-French industry daily returns are obtained from Kenneth French's website. In our main specifications, we use the Fama-French 5 industry definition. In the Appendix, we also provide results based on the Fama-French 30-industry definition. The factor model data, including the Fama-French 3-factor and the Fama-French 5-factor, are also obtained from Kenneth French's website. The industry returns and the factor model returns span from 1963 to 2021.

**Wage Data.** The data are drawn from the IPUMS website. We apply filters similar to Angrist et al. (2006). The sample for the calculations consists of US-born black and white men with age 40-49 with at least 5 years of eduation, with positive wages and hours worked. The data span from 2001 to 2018. For each year, we use a 30,000 random sample. The logged wage variable is the average logged weekly wage and is calculated as the log of the annual income from work divided by weeks worked.

(D. Chetverikov) Department of Economics, UCLA, Bunche Hall, 8283, 315 Portola Plaza, Los Angeles, CA 90095, USA.

*E-mail address*: chetverikov@econ.ucla.edu

(Y. Liu) Simon Business School, University of Rochester, 3-147 Carol Simon Hall, Rochester, NY 14611, USA.

*E-mail address*: yliu229@simon.rochester.edu

(A. Tsyvinski) Department of Economics, Yale University, 28 Hillhouse Ave, New Haven, CT 06511, USA.

*E-mail address*: a.tsyvinski@yale.edu