

DISCUSSION PAPER SERIES

IZA DP No. 14659

**The Use of Quantile Methods in Economic  
History**

Damian Clarke  
Manuel Llorca Jaña  
Daniel Pailañir

AUGUST 2021

## DISCUSSION PAPER SERIES

IZA DP No. 14659

# The Use of Quantile Methods in Economic History

**Damian Clarke**

*University of Chile and IZA*

**Manuel Llorca Jaña**

*Universidad de Valparaíso*

**Daniel Pailañir**

*University of Chile*

AUGUST 2021

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

## ABSTRACT

---

# The Use of Quantile Methods in Economic History\*

Quantile regression and quantile treatment effect methods are powerful econometric tools for considering economic impacts of events or variables of interest beyond the mean. The use of quantile methods allows for an examination of impacts of some independent variable over the entire distribution of continuous dependent variables. Measurement in many quantitative settings in economic history have as a key input continuous outcome variables of interest. Among many other cases, human height and demographics, economic growth, earnings and wages, and crop production are generally recorded as continuous measures, and are collected and studied by economic historians. In this paper we describe and discuss the broad utility of quantile regression for use in research in economic history, review recent quantitative literature in the field, and provide an illustrative example of the use of these methods based on 20,000 records of human height measured across 50-plus years in the 19<sup>th</sup> and 20<sup>th</sup> centuries. We suggest that there is considerably more room in the literature on economic history to convincingly and productively apply quantile regression methods.

**JEL Classification:** N30, B41, C21, C22

**Keywords:** quantile regression, quantile treatment effects, economic history, practitioners

**Corresponding author:**

Damian Clarke  
Department of Economics  
University of Chile  
Diag. Paraguay 257, Santiago  
Región Metropolitana  
Santiago de Chile  
Chile  
E-mail: dclarke@fen.uchile.cl

---

\* We are grateful to Roger Koenker for very useful comments, and we thank Matías Méndez Montenegro for research assistance. We acknowledge the funding of the Chilean National Research Agency (ANID) via Proyecto Anillos "Bienestar y Desigualdad en Chile", 18102020, grant SOC PIA 180001.

# 1 Introduction

Correctly capturing key patterns and relationships in economic history often requires the measurement of entire distributions of variables. For example, consider human demographics. Stunting, or very low heights owing to impaired growth in childhood, is associated with lower cognitive ability in later life, poorer health during adulthood, and reduced labour market earnings (Jayachandran and Pande, 2017; Fogel, 2004; Floud et al., 2011). And while stature is known to be broadly associated with health and other measures of well-being in many populations and time-periods (Costa, 2004; Vogl, 2014; Deaton, 2007), these relationships are not constant over the entire range of heights, nor even monotonic in all populations. Thus, if one wishes to evaluate some historical event as an input to height or other demographic measures, it is necessary to determine how the event has impacted heights across the entire distribution. Similar phenomena are encountered in many applications in economic history, such as economic growth and its impacts on inequality in income distributions, determinants of historical salary distributions, and the study of mortality declines in higher versus lower mortality settings. In each case, the distributional effects of policies or events can be as important as their mean effects in the population under study.

In this article we seek to motivate the importance of considering the distributional impacts of historical events, and specifically survey a broad series of methods which are ideally designed for such analyses. In particular, we lay out a range of methods related to quantile analyses, such as quantile regression, and the estimation of quantile treatment effects. These methods have emerged in a long line of theoretical and computational advances in the econometric literature, first fully described in Koenker and Bassett (1978), and comprehensively discussed in a range of papers or books since (Koenker, 2017; Koenker and Hallock, 2001; Lamarche, 2019; Koenker, 2005). While four decades have passed since the publication of Koenker and Bassett (1978)'s seminal paper on quantile regression, this is still an area of active research, in particular with recent advances considering quantile treatment effects, and identification in broader circumstances.

Despite being well-suited to applications in economic history where dependent variables of interest are often continuously distributed, these are arguably under-utilised in empirical applications. As well as providing an overview of these methods including recent extensions which are likely to be of particular interest to practitioners in economic history, we provide a survey of their usage in economic history, an analysis of their use, and *potential* for their use in papers published over the past 30 years in the field of economic history. In a survey across all principal economic history journals we find around 50

papers that have used quantile regression in some way. However, in a deeper analysis of all the papers published in the journal *Explorations in Economic History*, we find that while around 60% of papers are based on quantitative analyses of continuous dependent variables and thus potentially suited to analysis using quantile regression or related methods, only around 1% of papers actually apply these tools to consider effects beyond the mean.

This article thus seeks to motivate the “use” of quantile analyses in economic history in two ways. The first is to describe how they can be productively used by practitioners in empirical studies to capture relationships of interest which may not be gleaned from simple mean or other average estimators. And the second is to document how they have been used (and arguably under-used) in the literature on economic history up to this point.

In what remains of this paper, we first provide a summary of key methods for distributional analyses in empirical methods in section 2. In section 3 we provide both an overview of papers based on quantile analysis methods in economic history, as well as a review of all papers in a highly cited economic history journal, and their suitability for methods of this type. Finally, in section 4 we document a specific example based on microdata on human height covering around 20,000 individuals exposed to divergent rates of economic growth throughout their life. In closing, we make a number of points related to the computational implementation of these methods.

## 2 Quantile Regression

Generally, in empirical applications one wishes to consider the impact of some group of independent variables  $x_k$  for  $k \in \{1, \dots, K\}$  on a specific dependent variable, denoted  $y$ . Where observations  $i$  refer to units (such as individuals), this is often parameterised using a linear regression model:

$$y_i = \beta_1 + \beta_2 x_1 + \dots + \beta_K x_K + u_i, \tag{1}$$

where  $u_i$  is an unobserved error term. Frequently, and indeed, in the majority of papers in economic history (see Section 3) estimation is implemented using the ordinary least squares (OLS) method. Mathematically, this procedure simply consists of finding the  $K$  parameter estimates of  $\beta$  which minimize the

following problem:

$$\widehat{\beta} = \arg \min_{\beta \in \mathbb{R}^k} \sum_{i=1}^N (y_i - \mathbf{x}'_i \beta)^2,$$

where  $\mathbf{x}_i$  refers to the vector of values of each the independent variables  $x_k$  for unit  $i$ . This optimization returns a vector of parameters capturing the mean impacts of each  $x_k$  on the outcome of interest  $y$ . While mean impacts across the distribution of  $y_i$  may be a logical summary parameter in many cases, it is not the only point with meaningful empirical content. Often, other points of the distribution of  $y$  may be as, or more important, than the mean, particularly in historical processes where extreme outcomes may be of particular interest. This suggests the importance of modelling options allowing for additional heterogeneity.<sup>1</sup> Below we describe a range of quantile estimation techniques which allow for focus across any points of the distribution of outcomes of interest.

## 2.1 The Linear Quantile Regression Model

Quantile regression considers the full distribution of some dependent variable  $y$ . Thus, at a minimum,  $y$  requires a distribution with considerable variation, and is inappropriate in cases where  $y$  is a binary or categorical measure. We will denote as  $\tau$  the quantiles of the distribution, such that (for example)  $\tau = 0.5$  indicates the median of the distribution, and  $\tau = 0.1$  indicates the 10<sup>th</sup> percentile, or the point of the distribution below which 10% of the observations of  $y$  are observed. As originally documented in Koenker and Bassett (1978), we can estimate the parameter vector  $\beta(\tau)$  capturing the impact of each  $x_k$  on the  $\tau^{\text{th}}$  quantile of  $y$  based on the following minimization problem:

$$\widehat{\beta}(\tau) = \arg \min_{\beta \in \mathbb{R}^k} \left[ \sum_{i: y_i \geq \mathbf{x}'_i \beta} \tau |y_i - \mathbf{x}'_i \beta| + \sum_{i: y_i < \mathbf{x}'_i \beta} (1 - \tau) |y_i - \mathbf{x}'_i \beta| \right]. \quad (2)$$

Note that here, estimation is based on absolute deviations of  $\mathbf{x}'_i \beta$  from  $y_i$  rather than quadratic distances in OLS, and indeed, in the case that  $\tau = 0.5$  (the median) this formula collapses to the Least Absolute Deviations estimator. In all cases except for the median, this minimization problem uses the quantity  $\tau$  to ‘tilt’ the estimates towards data which is lower or higher in the distribution of  $y$ , as can be observed in

---

<sup>1</sup>Heterogeneity can be observed for a number of reasons in empirical studies. In this paper we are interested in heterogeneity in the impacts of some independent variable(s) over the full distribution of a dependent variable of interest. A useful alternative example of the interest of heterogeneity in economic history is provided by Bisin and Moro (2020) who discuss heterogeneity in the context of differential take-up of some treatment, and resulting sample-based heterogeneity due to the estimation of Local Average Treatment Effects (LATEs).

the two terms within the parentheses in equation 2: when  $\tau$  is between 0 and 0.5, more weight is given to the right-hand summation for units  $i$  whose  $y_i$  is less than the conditional mean  $x_i'\beta$ , whereas when  $\tau$  is between 0.5 and 1, more weight is assigned to the left-hand summation considering units whose  $y_i$  is greater than the conditional mean.

Equation 2 is the well-known linear quantile regression model which is implemented as standard in many computational languages. Frequently, rather than focusing on a particular quantile of interest, estimates  $\widehat{\beta}(\tau)$  are documented over a range of quantiles in a graphical manner (refer for example to the illustration in section 4 of this paper. What's more, with modern computational tools, standard errors can be calculated quite simply along the range of the distribution of  $y$ , which permits formal hypothesis tests and other inferential procedures at each quantile considered. Formally, following notation from Lamarche (2019), the approximate distribution of the vector  $\widehat{\beta}(\tau)$  can be written as:

$$\widehat{\beta}(\tau) \sim \mathcal{N}\left(\beta, \frac{1}{N}\tau(1-\tau)H^{-1}JH^{-1}\right),$$

where  $\mathcal{N}(\cdot)$  refers to a Gaussian distribution and the matrices  $J = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N x_i x_i'$ , and  $H = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N f_i(x_i'\beta(\tau))x_i x_i'$ , with  $f_i(\cdot)$  being the conditional density function of  $y$ . Note here two key implications: firstly, that this inference holds asymptotically, ie as the sample size grows, and secondly that a key 'ingredient' is the estimate of the density  $f_i(x_i'\beta(\tau))$  at quintile  $\tau$ . If one is willing to assume that error terms are identical and independently distributed (iid), the  $f_i$  term is identical among all units  $i$ , and estimation is simplified, for example using the fitted value of the density at the quantile of interest.<sup>2</sup> However, if heteroscedasticity robust estimates are desired, more complex 'sandwich' estimates are necessary. While these are also implemented as standard in computational routines in languages such as Stata or R, further background on these procedures can be found in Koenker (2005, chapter 3). Alternative options which avoid the estimation of these matrices consist of using bootstrap resampling methods for inference. A review of such resampling methods in quantile regression is provided by He (2018). Finally, note that solutions have also been proposed to resolve cases where clustered inference is desired, allowing for correlated shocks within groups, as well as heteroscedasticity (Hagemann, 2017), once again based on bootstrap resampling procedures.

---

<sup>2</sup>There are also other reasonable approaches generally included as standard in statistical software, such as using a kernel density estimator or residual quantile function to estimate the density at quantile  $\tau$ . These are all available in programs such as Stata with the `qreg` program, or R with the `quantreg` library.

## 2.2 Quantile Treatment Effects

While this quantile regression can be used for  $x_k$  of arbitrary forms and dimensions, a particular case of interest is that of treatment effect models, with some binary ‘treatment’ of interest, which we denote  $D_i$ . Note that while quantile regression requires continuous distributions for the outcome variable of interest, there is no limit on the nature of independent variables  $x_k$ , including binary and categorical measures. The standard average treatment effect aims to identify the mean impact of receiving treatment (versus not receiving treatment) on dependent variable  $y$ :  $\Delta = E(y_i|D = 1) - E(y_i|D = 0)$ . This assumes independence of  $D$  from unobservables. Extending this average treatment effect to its quantile treatment effect (QTE) analogue gives.

$$\Delta(\tau) = Q_y(\tau|D = 1) - Q_y(\tau|D = 0), \quad (3)$$

where  $Q_y(\tau)$  refers to the value of  $y$  at a particular quantile  $\tau$  of the distribution of  $y$ .<sup>3</sup> Such an estimate, beyond the mean is likely to be relevant for a range of historical policies, for example examining whether exposure to a specific type of economic or political system, exposure to a historical environmental shock, or exposure to a historical policy has divergent impacts across the distribution of individual outcomes capturing well-being. All such examples refer to binary ‘treatment’ statuses, and hence are potentially appropriate for QTE methods and their extensions discussed in the following sub-section, provided that outcomes of interest  $y$  are continuous measures. In practice, these estimates can be generated using regression following the procedure laid out in the previous section. We return to discuss computational implementations in the final section of this paper.

This setting can be extended considerably, for example to include covariates, and/or to explicitly consider selection into treatment. This can be done in a number of ways, such as by attempting to correct for selection non-parametrically (Bitler et al., 2006; Firpo, 2007), or extending to panel data settings and using recent methods such as difference-in-differences (Callaway and Li, 2019a). A particular case where QTE methods are extended to deal with endogeneity is laid out in the following sub-section when discussing instrumental variables and Local QTEs. While the inclusion of covariates can complicate the QTE setup (Koenker, 2017), suggestions on how to deal with this have been proposed in Firpo (2007); Frölich and Melly (2010); Callaway and Collins (2018) based on propensity score methods and

---

<sup>3</sup>Formally, this is  $Q_y(\tau) = \inf\{y : F(y) \geq \tau\}$ , where  $F(\cdot)$  refers to the cumulative density function of  $y$ .



re-weighting techniques. For applied work, a particularly useful reference discussing a range of practical estimation methods for quantile treatment effects is the paper of Frölich and Melly (2010).

## 2.3 Quantile Regression Extensions

While standard quantile regression and QTE implementations can provide illustrative results which allow for an understanding of effects beyond the mean, recent theoretical and computation advances in these methods offer considerable additional benefits, including the possibility to address a number of issues in standard quantile regression models, and more explicitly address questions of causality. We discuss the broad scope of these recent advances here, providing key references. Full modeling considerations are available in the papers discussed below, however particularly useful overviews of these methods at more length can also be found in the survey articles or handbook chapters of Koenker (2017) (discussing all below points), or Wei (2018); Lamarche (2019); Melly and Wüthrich (2018); Chernozhukov et al. (2018) with focuses on particular points.

**Models to Correct for Measurement Error** In the case of variables collected over a considerable time-frame from diverse data sources as is often the case in studies in economic history, measurement error or issues of sample selection are valid concerns. Fortunately, a number of methods suggest ways that measurement error can be accounted for within the framework of quantile regression (under assumptions about the nature of these errors). For example, Wei and Carroll (2009) document potential estimation methods when in the presence of measurement error in independent variables, and alternative methods are proposed by Wang et al. (2012). The crux of these models is that if rather than observing a true variable  $x$  we observe some noisy proxy, we must consider the conditional *expectation* of  $x$  based on the noisy proxy, rather than simply using the proxy in the regression. Additionally, even if there is limited error in the measurement of variables, samples may be selected in certain ways, for example if only certain types of records survive following collection of historical data, or if data are based on surviving individuals. Arellano and Bonhomme (2017) propose methods for quantile regression under selection based on an exclusion restriction for estimation, while Blanco et al. (2013) (see also Lamarche (2019)) propose partial identification methods as way ahead in such circumstances. Computational implementations of such selection models are available, for example via the programs of Biewen and Erhardt (2020); Siravegna (2020). Interested readers are pointed to the handbook chapter of Wei (2018) which provides

a deeper discussion of these issues.

**Endogeneity, Instrumental Variables and Causality** In many settings of interest in economic history, independent variables of interest will be endogenous—correlated with unobserved factors which are themselves correlated with outcomes of interest. As is the case in standard econometric models, quantile regressions or QTE models do not allow for causal inference when in the presence of endogeneity. However, there is a rich stream of work which seeks to extend standard linear instrumental variable (IV) models into a quantile framework. Here in particular we discuss the Local Quantile Treatment Effect (LQTE) approach, which, under a number of key assumptions provides consistent (causal) estimates for certain sub-groups of the population across the entire distribution of the outcome variable.

The LQTE framework extends the well-known Local Average Treatment Effect (LATE) framework laid out in Imbens and Angrist (1994). Here, causal estimation using an instrumental variable is based on a monotonicity assumption that requires that the instrumental variable induces exogenous variation in the dependent variable of interest which shifts all individuals weakly in the same direction, for example acting as an as-good-as-randomly assigned incentive or disincentive to uptake an endogenous variable of interest. Here we summarise this and the other identifying assumptions, as well as how an LQTE is estimated, with a binary instrumental variable denoted  $Z$ . A chapter length discussion of these methods, as well as extensions to other circumstances is provided by Melly and Wüthrich (2018). Following the notation of section 2.2, consider an outcome variable of interest  $y$ , a binary treatment variable of interest  $D$ , as well as the instrumental variable  $Z$ . Under 5 assumptions; instrumental independence (or validity), the exclusion restriction, instrumental relevance, monotonicity and the stable unit treatment value assumption (SUTVA) limiting interaction between treatment statuses, Imbens and Angrist (1994) prove that the following Wald estimator gives the LATE:

$$\frac{E(y_i|Z_i = 1) - E(y_i|Z_i = 0)}{E(y_i|D_i = 1) - E(y_i|D_i = 0)}$$

Note that this LATE is a consistent estimate of the impact of treatment ( $D_i$ ) on outcomes ( $y_i$ ) for those individuals whose treatment status would be changed by the instrument (the ‘compliers’).

In order to extend this framework to a quantile framework, distributional results are provided by Imbens and Rubin (1997); Abadie (2002). These give cumulative density functions specifically for the compliers, in this case which we denote  $F_{Y_1}^c$  and  $F_{Y_0}^c$ . Here superscript  $c$  refers to the population of

compliers, and  $F$  refers to the cumulative density of outcomes for cases where treatment is received ( $F_{Y_1}^c$ ) or not received ( $F_{Y_0}^c$ ).<sup>4</sup> Given these definitions based on an IV and compliers, it is very easy to return to the notation from section 2.2, which gives the LQTE estimator. As a clear parallel to equation 3, this is defined as

$$\Delta^c(\tau) = Q_{Y_1}^c(\tau) - Q_{Y_0}^c(\tau), \quad (4)$$

where as previously,  $\tau$  refers to quantiles of the relevant conditional density function above.<sup>5</sup>

It is worth noting that these methods, while more demanding than standard QTE models given the required conditions of instrumental validity, are potentially well suited to applications in economic history where concerns exist relating to endogeneity. For example, Aaronson et al. (2020) estimate the impacts of the number of children born to women – a clearly endogenous variable – on maternal labour supply, using data over more than two centuries, by leveraging the birth of twins as an IV. In this case monotonicity assumptions are very likely met given that twin births should have weakly positive impacts on completed fertility.<sup>6</sup> Very recent work by Valencia Caicedo (2021) provides considerable other examples of the use of IV in economic history (though not discussing quantile methods), describing among other examples IVs based on map borders or other geographical features such as rivers, or slopes of terrain, geographic suitability indexes, or Bartik-type instruments. All such instruments in economic history could be productively introduced into a quantile framework if distributional outcomes are considered.

Finally, prior to turning to applied examples of the use of quantile regression, we note that this particular implementation of IV and endogeneity corrections via LQTE methods is not the only way forward.

---

<sup>4</sup>In extensive form, these can be represented (Abadie, 2002) as:

$$F_{Y_1}^c(y) = \frac{E[\mathbb{1}(Y \leq y)D|Z = 1] - E[\mathbb{1}(Y \leq y)D|Z = 0]}{E(D|Z = 1) - E(D|Z = 0)},$$

and

$$F_{Y_0}^c(y) = \frac{E[\mathbb{1}(Y \leq y)(1 - D)|Z = 1] - E[\mathbb{1}(Y \leq y)(1 - D)|Z = 0]}{E(1 - D|Z = 1) - E(1 - D|Z = 0)}.$$

<sup>5</sup>In the interests of completeness, this is:

$$\begin{aligned} Q_{Y_1}^c(\tau) &= \inf\{y : F_{Y_1}^c(y) \geq \tau\} \\ Q_{Y_0}^c(\tau) &= \inf\{y : F_{Y_0}^c(y) \geq \tau\} \end{aligned}$$

where all notation follows that of section 2.2.

<sup>6</sup>Note however that in general, the assumptions necessary for identification with instruments are not trivial. Bhalotra and Clarke (2019) discuss a number of considerations related to these instruments in a standard linear model. An entirely different take on IV style models which precluded the LQTE approach described here and allows IVs to generate variation locally in endogenous variables is the work of Chesher (2005); Ma and Koenker (2006). This suggests productive ways forward in a quantile framework even if instruments generate shifts at only specific points of endogenous variables of interest.

Regression discontinuity designs can be similarly cast in this “Local” framework, while the instrumental variable quantile regression models (IVQR) of Chernozhukov and Hansen (2005) provides estimation methods which return average Quantile Treatment Effects (rather than LQTEs), but requiring alternative assumptions.<sup>7</sup> These methods are further discussed in Melly and Wüthrich (2018) and Chernozhukov et al. (2018) respectively, and are likely to more productive ways forward when non-binary endogenous variables are considered—in which case LQTE models are less appropriate—or in cases when broader *non-local* QTEs are desired, at the cost of alternative assumptions. A valuable review of these methods is provided in Chernozhukov and Hansen (2013).

**Other Extensions and Methods** A range of other contexts which are potentially of use in quantitative studies of economic history can be productively studied in quantile settings. This includes settings such as longitudinal or panel data and difference-in-difference models, regression discontinuity designs, and non-parametric analyses. We briefly discuss these settings in turn below, pointing interested readers to relevant references.

Extensions of quantile regressions to panel data following individuals over time have been proposed in Koenker (2004) where challenges arise given desires to estimate movements across distributions *within* individuals (or panel units) with potentially few repeated individual level data-points. Koenker (2004) proposes using a penalized estimator to control for relevant individual-level effects, and much additional work has been conducted to take forward these techniques (a complete discussion is provided in Lamarche (2019, pp. 11-14)). Recent work of Gu and Volgushev (2019) has suggested using a grouped fixed effect approach, also based on penalized estimators to group similar individuals in a panel setting. Alternative lines of work, such as Arellano and Bonhomme (2016), suggest viewing (potential individual-level) variation as a problem of unobserved heterogeneity and estimating using an iterative process, super-imposing simulation based estimation procedures on top of standard quantile regression procedures. A particular setting of interest in cases of longitudinal data consists of the estimation of difference-in-differences models where exposure to some treatment varies within a panel over time such that baseline differences between exposed and unexposed individuals can be captured using pre-treatment periods. These models have been extended to a quantile setting, see for example Callaway et al. (2018) for a setting with two time periods and Callaway and Li (2019b) for a broader panel setting.

---

<sup>7</sup>In particular, these methods require a rank preservation assumption, restricting the ranks which individuals can take in terms of the ordering of the outcome variable to be the same across differing potential IV assignments.

Computational software is also available to implement these methods in Callaway (2019)

The regression discontinuity design provides credible identification in cases where some dependent variable of interest is moved discontinuously by some arbitrary cut-off. The use of these methods in economic history has been surveyed by Valencia Caicedo (2021), where examples are often based on distance to geographical features or map boundaries. The regression discontinuity design has been extended to a quantile framework in a very flexible way by Frandsen et al. (2012), which allows for the estimation of quantile treatment effects ‘local’ to a particular discontinuity, and accompanying computational routines are available to implement this method.

Standard non-parametric regression implementations can capture fully flexible relationships between the mean of some variable  $y$  and some independent variable  $x$ . Rather than parametric assumptions and linear functional forms (as imposed in equation 1, the relationship is allowed to vary freely along the support of  $x$ , allowing for  $x$  to have a non-linear impact on the mean of  $y$ . This logic can be extended to a quantile regression, if rather than considering a non-parametric relationship between the mean of  $y$  and  $x$ , a non-parametric relationship between specific *quantiles*  $\tau$  of  $y$  and  $x$  are estimated. This is a particularly flexible way to model heterogeneity which may be well-suited to historical outcomes over which there are few prior assumptions related to the nature of the relationship under study. A review of the methods is available in Koenker (2017, section 3), while computational resources for the implementation of such routines are available in, among others, Koenker (2021); Lipsitz et al. (2017).

### **3 Applications in Economic History**

#### **3.1 Quantile Regression in Economic History Research**

To have some idea about the extensiveness of the use of quantile regression in economic history, we begin by running a search within the main economic history journals, covering the period from 2000 until the present.<sup>8</sup> Namely, we searched within the Journal of Economic History, Explorations in Economic History, Economic History Review, Cliometrica, and the European Review of Economic History. But we also complemented these searches with searches of other economic history journals as discussed below,

---

<sup>8</sup>As we see in the following sub-section, there is even less use of quantile regression prior to 2000. Two notable exceptions are the studies of Conley and Galenson (1998, 1994) which provided early illustrations of the power of quantile regression in historical analyses.

estimating that around 50-55 journal articles have made use of quantile regression as an analytical tool during the last two decades, although only a handful of these are dated pre-2005.

The journal which has most frequently published articles employing this technique is the *Journal of Economic History*, which published 19 articles using quantile regressions for a wide range of topics from 2000, including: Max Weber's hypothesis on the role of Protestantism for economic development (Kersting et al., 2020); the market for paintings in Florence and Italy between 1285 and 1550 (Etro, 2018); the interaction between inequality and financial development in the US during the late nineteenth century (Jaremski and Fishback, 2018); fluctuations in technology during the Great Depression in the US (Watanabe, 2016); the credibility of fixed exchange rates during the classical gold standard era (Mitchener and Weidenmier, 2015); and inequality of wealth in the Ottoman Empire (Coşgel and Ergene, 2012); just to mention those published after 2010. These few examples make clear the broad applicability of quantile methods to questions of interest in economic history, covering issues in micro, macro and financial economics.

*Explorations in Economic History* has also published articles using quantile regression based methods with some frequency. A scoping review identified six such article (see also the following subsection). Namely, Walker (2000), which explores the degree of economic opportunities in San Francisco compared to other regions around mid-nineteenth century; Dupont (2007), which tests for contagion in bank runs in Kansas during the panic of 1893; Canaday (2008), which deals with the relationship between wealth and wealth accumulation by both blacks and whites in South Carolina between 1910 and 1919, and its determinants; Drelichman and González-Agudo (2014), which reconstructs housing costs for various social groups and traces the effect of exogenous shocks on the rental market for Toledo, Spain, between 1489 and 1600; Álvarez and Ramos-Palencia (2018), which deals with the relationship between human capital and male labour earnings in eighteenth-century Spain; and Callaway and Collins (2018), which measures the union wage premium for several US-cities circa 1950, using unconditional quantile methods.

The *Economic History Review*, in turn, has published five articles where quantile regressions were used: Temin and Voth (2008) analyses the cost and availability of private bank credit between 1702 and 1724; Gazeley and Newell (2011), in turn, estimates urban poverty among working families in the British Isles circa 1904; Brown and Guinnane (2018) deals with the causes of fluctuations in infant mortality rates in Bavaria during the 1820s-1910s; Artunç (2019) examines the composition of firm ownership

and entrepreneurship in Egypt between 1910 and 1949; while Karagedikli and Tunçer (2021) estimates real hedonic house prices and urban wealth inequality for the housing market between 1720 and 1814 in the Ottoman Empire. Note that as above, these articles show both a broad scope of themes, as well as a broad scope of geographic and temporal settings which have been productively analysed with these models.

The *European Review of Economic History* has also published five articles making use of quantile regression: Koepke and Baten (2005), which provides the first anthropometric estimates of the biological standard of living in Europe during the first millennium AD; Dincecco (2009), which performs a statistical analysis of political regimes and sovereign credit risk in Europe from 1750 to 1913; Dribe (2009), who analyses the importance of demand and supply factors in the Swedish fertility transition between 1880 and 1930; Kholodilin (2016), that analyses the housing rental dynamics of Berlin during World War I; and more recently, Jorge-Sotelo (2020) focused on the impact of currency depreciation on international capital flows in Spain between 1928 and 1931 crisis.

*Cliometrica* has published six articles using quantile regressions, the first of these less than a decade ago: Carson (2012), compared body mass index values of late 19th- and early 20th-century amongst African-Americans groups (i.e. blacks versus mulatto); Ogasawara and Kobayashi (2015), which deals with the impact of social workers on reducing infant mortality rates in inter-war Tokyo; Du Plessis et al. (2015), who for the period 1700-1725, estimated hedonic slave price indices and the value of their marginal productivity; González-Val et al. (2017), who analysed the impact of market potential on the structure and growth of some Spanish cities during 1860–1960; Ogasawara and Matsushita (2019), which deals with the treatment effects of piped on diseases in industrializing Japan (1920s-1930s); and finally, Keywood and Baten (2020), who tests the relationship between elite numeracy and elite violence in Europe from 500 to 1900.

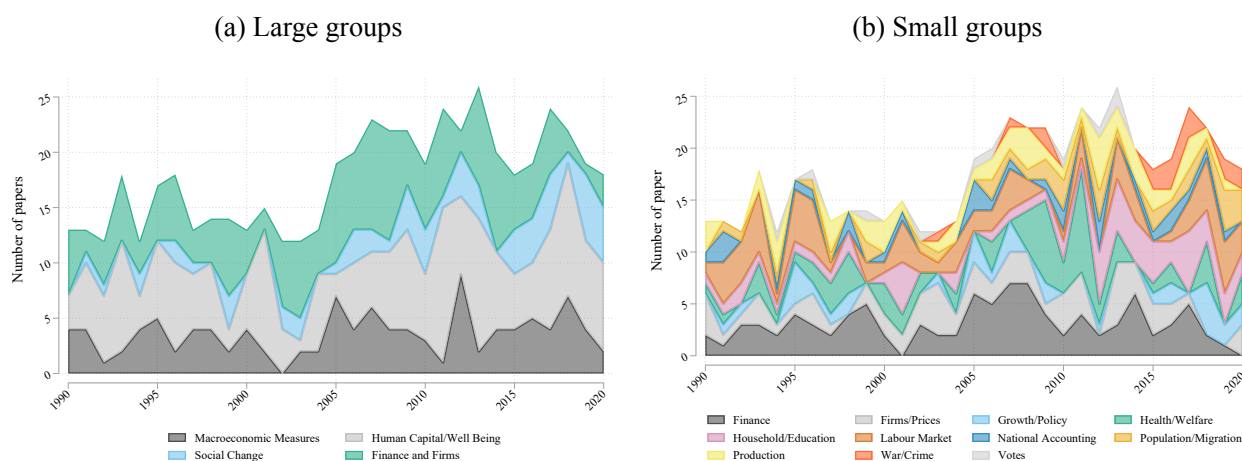
We also run searches in other economic history journals, but where quantile regression was observed to be very infrequently used. For example, in the main Spanish economic history journal, *Revista de Historia Económica – Journal of Iberian and Latin American Economic History*, quantile regressions were used in only four papers; in the *Australian Economic History Review* in only one article; in the *Economic History of Developing Regions*, similarly in only one article; while in the *Scandinavian Economic History Review* it has appeared twice. Finally, in the three main business history journals (*Business History*, *Business History Review* and *Enterprise & Society*), it was used in three articles, twice at *Business*

History and once at *Enterprise & Society*, although this is rather unsurprising since this sub-discipline cultivates a less quantitative approach to history.

### 3.2 The Potential for Quantile Regression Methods in Economic History

Reviewing cases where quantile regression *is* used suggests the broad applicability of these methods across themes and settings in economic history, but does not indicate the additional scope for use in papers which focus principally on mean effects. To see the *potential* for Quantile Regression, we carried out a review based on all papers published over the last 30 years in a specific economic history journal (*Explorations in Economic History*, hereafter EEH). This allows us to gain a more complete picture of trends in published research in a specific highly-cited economic history journal, consider the methods used, and the potential for distributional analyses in economic history. In this review, we read each paper published over the last 3 decades (the first volume of 1990 to the last volume of 2020), and classified each paper according to whether it contained empirical methods, or was theoretical. In the case of empirical papers, we then classified these by method (OLS, Logit, Quantile Regression, and so forth) and finally whether principal dependent and independent variables are continuous, discrete, ordinal, etc.

Figure 1: Trends in the dependent variable in papers published in *Explorations in Economic History*



Notes: Based on our reading of each paper published in EEH we classify each paper as belonging to a particular (single) area, as described in the legend in panel (b). These classes are further aggregated in panel (a), where ‘Macroeconomic Measures’ refers to Growth/Policy, National Accounting and Production; ‘Human Capital/Well Being’ refers to Health/Welfare, Household/Education and Labour Market; ‘Social Change’ refers to Population/Migration, War/Crime and Votes; and Finance and Firms refers to Finance and Firms/Prices. A file summarising all papers in terms of a number of characteristics discussed in this review is provided as an online appendix to this paper.

Prior to considering trends in measures and methods, it is useful to provide descriptive evidence of



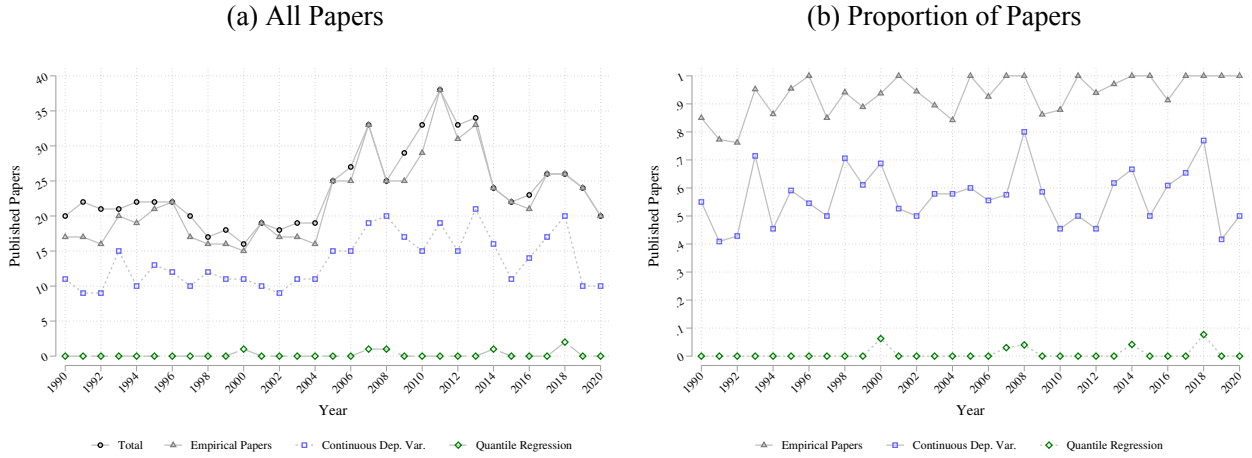
the key themes studied over the past 30 years in this particular journal. In Figure 1, we plot the broad areas which papers study, as classified based on our reading of the literature. These classifications refer to the dependent variable under study in all empirical papers published in *EEH*. The right-hand plot classifies papers into one of four aggregate areas: macroeconomic measures, human capital and well-being, social change (which includes papers studying topics such as conflict, migration, and changing political landscapes), and finance and firms. In general we note the largest of these groups is that related to microeconomic areas such as education, health and labour markets (the shaded light gray area), followed by themes related to finance and firms, with relatively less work focused principally on macroeconomic indicators. In the left-hand panel we provide more disaggregated classes, seeing more noisy patterns, and evidence of broad coverage of topics within economic history.

While it is illustrative to observe trends in the published record in economic history, here we are most interested in considering how the types of questions and measures used in these papers relate to their potential for using quantile methods. At minimum, this implies that dependent variables need to be continuous, given that quantile methods examine impacts of some independent variable across the distribution of the dependent variable.

Figure 2 plots the number and share of papers published in *EEH*, classifying them by whether they are empirical or not, and in the case of empirical papers, whether the dependent variable is continuous, and hence suitable for quantile analyses. Finally, we plot the number and share of papers that effectively do undertake quantile analyses. In Figure 2 panel (a) we can see how the amount of total articles published in *EEH* was around 20 annual publications, with an increase to around 30 from 2005 onwards. We can see how the number of empirical papers follows a similar trend to that of total articles, and corresponds to 93.1% on average of the number of total articles. In recent years, this value is even higher, with 100%, or close to 100% of articles containing at least some considerable empirical portion, pointing to the growing frequency of empirical evidence in studies in economic history.

While this is illustrative of the importance of quantitative analyses in understanding historical processes in economics, it does not necessarily imply a suitability for analysis with quantile methods. However, if we observe the articles that are based on continuous dependent variables and hence amenable to quantile-based methods, we observe that they correspond to about 56.5% of the quantity of total articles (over the past 30 years). Despite these two stylised facts: (i) that there is considerable quantitative evidence brought to bear in economic history papers, and (ii) that many of these analyses are based on

Figure 2: Trends in Publications in an Economic History Journal



Notes: The total number of papers (panel A) and the proportion of all papers (panel B) refer to all papers in the journal EEH, based on the issue in which they are published. This data was collected based on a reading of each paper and classification of the methods and ‘principal’ dependent variables.

variables which have continuous distributions and hence potentially interesting distributional features, we observe little use of quantile methods. Among the 738 papers examined, 692 of which were empirical, only 6 studies used quantile regression (as described in the previous sub-section). Unless there is a particular interest in the mean, these facts taken together point to a considerable margin for the adoption of other estimation methods such as quantile regression or QTE methods within economic history.

Table 1: Classification of Dependent and Independent variables

Type	Dependent		Independent	
	Number	Proportion	Number	Proportion
Continuous	418	76.70	207	37.91
Binary	58	10.64	47	8.61
Discrete	23	4.22	12	2.20
Ordinal	3	0.55	0	0
Multiple	10	1.83	115	21.06
Continuous-Binary	19	3.49	77	14.10
Continuous-Discrete	9	1.65	21	3.85
Continuous-Dummy	0	0	25	4.58
Continuous-Ordinal	1	0.18	0	0
Dummy-Dummies	0	0	30	5.49
Discrete-Binary	4	0.73	12	2.20

Notes: Multiple classification is article with more than two types of variables. The classification that presents 2 types of variables, correspond to papers that present more than one regression in which its dependent variables are continuous plus one or more variables of another type, and both regressions are relevant in the research.

Finally, before providing a simple illustrative example of patterns observable in quantile regressions

in a particular study of historical economic processes, we note how these continuous dependent variables are situated within analyses. In Table 1 we provide a full break down of all the classes of dependent variables observed across all studies from EEH. 418 papers had a single continuous outcome variable of interest, while another 29 had a continuous outcome variable as well as other non-continuous measures. Of these 447 studies, 201 had independent variables which were also continuous, and so more suited to quantile regression methods, while 33 had a binary independent variable of interest, and so are potentially well suited to QTE methods. Many other studies have various types of independent variables, and as such either QTE or quantile regression methods may be appropriate.

## **4 An Illustrative Example Based on 19<sup>th</sup> Century Demographics and Economic Growth**

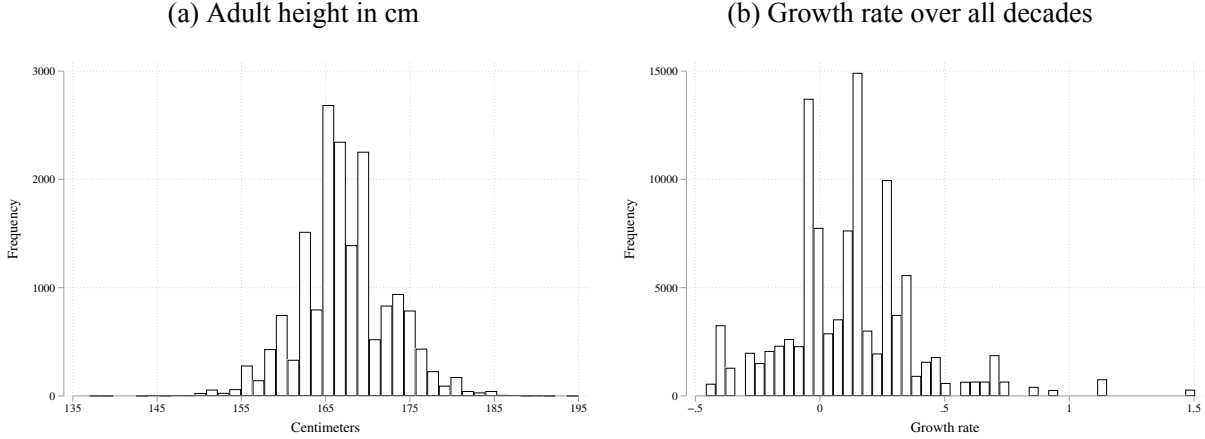
As a brief illustration of the use of quantile regression, and how it can shed light on patterns across the distribution of outcomes which are hidden by standard analyses, we consider the empirical setting described in Llorca-Jaña et al. (2019, 2021). As laid out there, we gathered information on all the 36,371 records of Military Personnel born in the 20<sup>th</sup> century and 3,283 record of Military Personnel born in the 19<sup>th</sup> century in Chile.<sup>9</sup> These records relate mainly to soldiers or low-ranking officers (the Army's Historical Archive). Full information on this process and these data are available in Llorca-Jaña et al. (2019, 2021). Of those individuals, we generate a final database of the Chilean-born individuals aged between 17-55, which here we cross with rates of economic growth in the province in which the individual was born. These rates of economic growth are calculated from historical evidence collected by Badia-Miró (2008), which is the best sub-national evidence of economic conditions in Chile, available covering the periods of 1890-1950. From these data sources, we are able to combine a large micro-level sample of human height as well as measures of economic growth by decade and province of birth. Finally, we have a database of 17.293 individuals. The reduction of the database is explained given the

---

<sup>9</sup>There are many historical precedents to the study of height in economic history, although less work extending to quantile analyses. Among many other references Wachter and Trussell (1982) discuss historical measurement of height, though work goes back much further, for example Quetelet's discussion of classifying populations.

data availability of growth rates (1890-1940).<sup>10</sup> Descriptive plots of these data on height and exposure to economic growth are provided in Figure 3, suggesting considerable variation in observed heights, and also in changes in economic conditions within provinces over time.

Figure 3: Distributions of individual height and sub-national rates of historical growth



Notes: Histogram: Panel (a) distribution of adult male height in Chile, 1890s–1940s (in centimetres, 17,293 observations); Panel (b) distribution of growth rate in Chile, 1890s–1940s.  
Source of data: Panel (a) Llorca-Jaña et al. (2019, 2021), Panel (b) Badia-Miró (2008).

To document these methods, we consider a particular model seeking to determine the effect of economic growth during an individual’s formative years on their adult height. Relationships between height and economic development have been discussed in the past, including over long periods, such as Peracchi (2008)’s study of Italian height from the 1730s-1980s. Here, we are interested in a model of the following type:

$$\text{Height}_{ipt} = \beta_0 + \beta_1 \text{Growth}_{pt}^0 + \beta_2 \text{Growth}_{pt}^6 + \beta_3 \text{Growth}_{pt}^{12} + \beta_4 \text{Growth}_{pt}^{18} + \mu_p + \lambda_t + \varepsilon_{pt}, \quad (5)$$

where the height of an individual  $i$  born in province  $p$  and year  $t$  is regressed on the growth rate in that province when the individual is born ( $\text{Growth}^0$ ), when they are aged 6 years ( $\text{Growth}^6$ ), when they are aged 12 years ( $\text{Growth}^{12}$ ) and when they are aged 18 years ( $\text{Growth}^{18}$ ). It is important to note that given

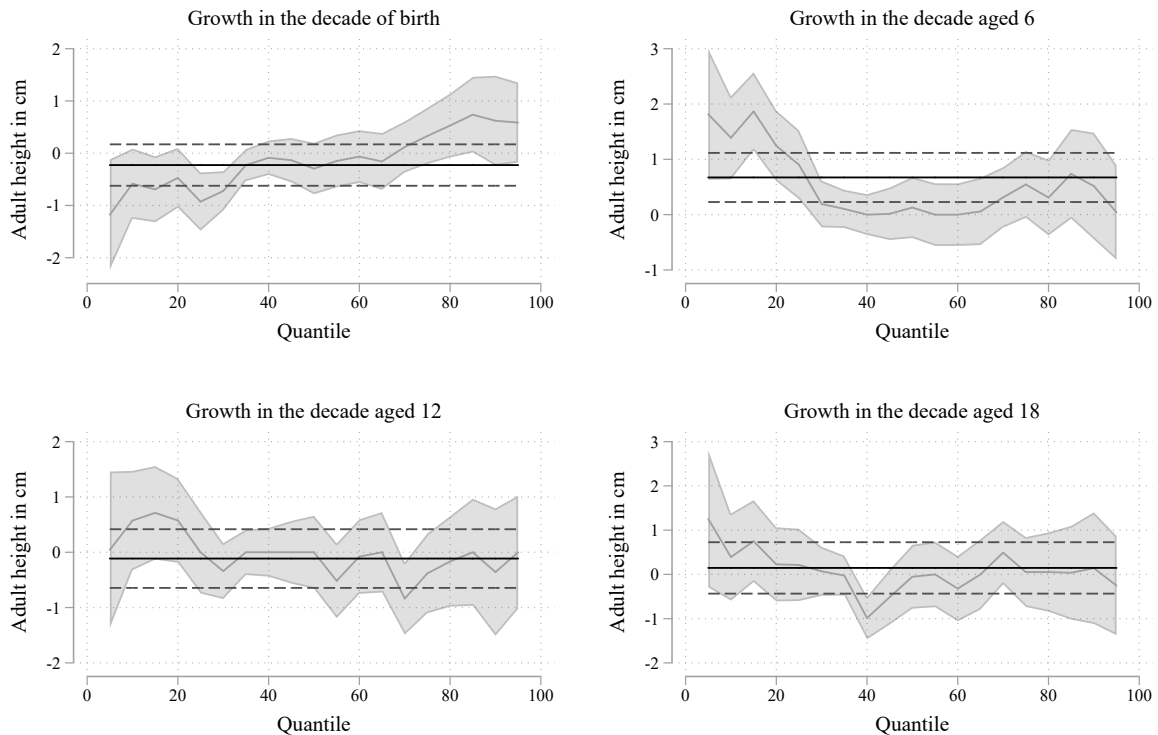
<sup>10</sup>We note that the sample reduction from the initial 36,371 digitized records owes to the availability of historical measures of growth. We calculate growth records in each decade as:

$$\frac{\text{GDP}_{pc_{p,t+1}} - \text{GDP}_{pc_{p,t}}}{\text{GDP}_{pc_{p,t}}},$$

where  $p$  indexes provinces and  $t$  indicates decades, and as such this will not be defined in the decade of 1950, given that comparable records are not available for 1960. We thus limit the final estimation sample to all individuals born in the 60 years between the 1890s and 1940s.

challenges in collecting data on economic patterns around 200 years in the past, these measures of growth are the best available estimates, but should be considered as noisy measure of the economic conditions during an individual’s growing years.<sup>11</sup> This model includes province and decade fixed effects ( $\mu_p$  and  $\lambda_t$  respectively), capturing idiosyncratic regional or temporal factors, such that measures of growth are not simply proxying regional or time-specific factors that correlated with height.

Figure 4: Quantile Regression



Notes: Each sub-plot presents the OLS and quantile regression estimates of the effect of local rates of economic growth at particular points of an individual’s life on their adult height. Solid black lines and dashed lines refer to OLS point estimates and 95% CIs, while gray solid lines and shaded areas refer to quantile regression point estimates and 95% CIs.

To consider what can be gained from estimating quantile regression in this particular setting of interest in economic history, we present estimates from equation 5 by both standard OLS, and using quantile regression following equation 2 at quantiles 5, 10, . . . , 95. These results are displayed in Figure 4, where we plot estimates on each of the parameters  $\beta_1, \beta_2, \beta_3$  and  $\beta_4$  on average (OLS, indicated by the solid horizontal black line), and across the distribution (quantile regression, indicated by the solid gray line).

<sup>11</sup>We also note that we consistently use measures related to an individual’s province of birth, given that from Military data we know where they are born. However, in the case of individuals moving between provinces in the country, these measures are noisy proxies of exposure to local economic conditions.

95% confidence intervals are indicated by dashed lines (OLS) or shaded intervals (quantile regression).

The top-left panel documents a null effect of exposure to growth in the decade of birth on adult height when considering average effects, although marginally negative effects lower in the distribution of adult height, and marginally positive effects higher in the distribution. Where results are most striking is in the top right-hand panel. While exposure to growth in the decade in which a child is aged 6 years is observed to significantly increase adult height (OLS estimates suggest a 1% higher rate of growth is associated with nearly 1cm in additional height), this effect is very heterogeneous, with large impacts at the lower end of the distribution of height, and null impacts among taller individuals. Here the value of considering quantile regression is clear. These results point to a particular value of economic growth in human condition, which – at least in this early life period where children are growing significantly – is most relevant among those who are shorter. This is significant, as adult height is a well known marker of health, suggesting that economic growth during this period of childhood in this context has done the most to pick up individuals who have the worst health stocks. While some similar patterns may even persist when children are older (bottom panels), these effects are observed to be most notable in the early years of life when children are particularly sensitive to their conditions. While these results are descriptive, and presented in part to illustrate the utility of quantile regression in research in economic history, they are nonetheless able to point to key distributional factors which are of relevance in understanding human demographics and sensitive periods of human capital accumulation, as well as the value of historical periods of growth – beyond just population averages.

## **5 Conclusions and Ways Ahead**

This paper seeks to provide a preview of a range of empirical methods which are relevant to consider impacts of some independent variable(s) of interest, across the entire distribution of a continuous dependent variable of interest. We specifically seek to describe these methods, and motivate their adoption more widely in literature in economic history where considerable work is often spent to collect rich (continuous) outcome measures of interest. These methods can thus contribute to fully taking advantage of such data collection processes or existing data repositories which have been collected based on considerable efforts in collating, systematising, or digitising historical records.

We discuss both standard quantile regression methods originally laid out by Koenker and Bassett

(1978) and also document how these have been fruitfully applied in a more recent “treatment effects” literature, with the application of Quantile Treatment Effects. We discuss a number of other extensions which may be of particular interest for researchers in economic history such as quantile regression with measurement error, additional ways to loosen parametric assumptions, and potential solutions to endogeneity in these models. This is a very large and ever-growing literature in econometrics, and so here, while aiming to provide a broad overview of the field, we do not claim to comprehensively survey the entire field nor the full depth of all models. Fortunately, there are a number of full textbook or handbook references such as work of Koenker (2004); Koenker et al. (2017), to which we point interested readers in cases where a more comprehensive econometric base of these models is desired.

While we argue that these models are well suited to research in economic history, we suggest that there is scope for considerably more work in this line. Fortunately, these methods are accompanied by a range of computational tools which mean that their application can be viewed as part of a quite standard toolbox for interested practitioners. In closing, it is worth pointing to the functionality of these packages, which are significant and generally open source contributions of the methods discussed in this paper, allowing for these methods to be adopted at relatively low cost.

Computational languages widely used in economics and economic history such as R, Stata, Julia, Python, MATLAB and so forth generally all have a standard implementation of quantile regression allowing for simple implementation of these models. For example, (Koenker, 2021) provides the R `quantreg` package which contains (among *many* other things) a standard quantile regression interpretation as `rq`, while Stata’s `qreg` provides a stable option for both estimation and various inference procedures. However, many extensions to these commands’ ‘standard’ procedures are available, including packages to extend analyses to quantile treatment effects such as `ivqte` in Stata (Frölich and Melly, 2010) or the `QTE` package in R (Callaway, 2019). Each of these QTE libraries extends in numerous ways to cases where treatment assignment is endogenous, and, in the case of QTE to difference-in-differences models (Callaway, 2019). Within the ‘universe of `quantreg` in R there are many other extensions, including a wide range of inference procedures, non-linear models, LASSO models, and plotting functions. Given the highly applied nature of these methods, and the considerable recent extensions in the field, many new papers are also accompanied by computational code – most frequently in R or Stata – including the examples discussed in the paper such as recent advances in selection models (Siravegna, 2020; Biewen and Erhardt, 2020), and quantile analyses in regression discontinuity designs (Frandsen

et al., 2012). All in all, these models present an extremely flexible, accessible and extendable series of analytical routes to researchers in economic history, and should be viewed as a key component of analyses, allowing for a much richer consideration of the distributional effects of historical phenomena in economic processes.



## References

- D. Aaronson, R. Dehejia, A. Jordan, C. Pop-Eleches, C. Samii, and K. Schulze. The Effect of Fertility on Mothers' Labor Supply over the Last Two Centuries. *The Economic Journal*, 131(633):1–32, 08 2020.
- A. Abadie. Bootstrap tests for distributional treatment effects in instrumental variable models. *Journal of the American Statistical Association*, 97:284–292, 2002.
- B. Álvarez and F. Ramos-Palencia. Housing and the cost of living in early modern Toledo. *Explorations in Economic History*, 67:105–137, 2018.
- M. Arellano and S. Bonhomme. Nonlinear panel data estimation via quantile regressions. *The Econometrics Journal*, 19(3):C61–C94, 2016. doi: <https://doi.org/10.1111/ectj.12062>.
- M. Arellano and S. Bonhomme. Quantile selection models with an application to understanding changes in wage inequality. *Econometrica*, 85(1):1–28, 2017.
- C. Artunç. Religious minorities and firm ownership in early twentieth century Egypt. *Economic History Review*, 72(3):979–1007, 2019.
- M. Badia-Miró. *La localización de la actividad económica en Chile, 1890-1973. Su impacto de largo plazo*. PhD thesis, Universitat de Barcelona. Departament d'Història i Institucions Econòmiques, November 2008.
- S. Bhalotra and D. Clarke. Twin Birth and Maternal Condition. *The Review of Economics and Statistics*, 101(5):853–864, December 2019.
- M. Biewen and P. Erhardt. ARHOMME: Stata module to estimate Arellano and Bonhomme quantile selection model, 2020. URL <https://EconPapers.repec.org/RePEc:boc:bocode:s458890>.
- A. Bisin and A. Moro. LATE for History. Working Paper 28113, National Bureau of Economic Research, November 2020. URL <http://www.nber.org/papers/w28113>.
- M. P. Bitler, J. Gelbach, and H. Hoynes. What mean impacts miss: distributional effects of welfare reform experiments. *American Economic Review*, 96(4):988–1012, 2006.
- G. Blanco, C. A. Flores, and A. Flores-Lagunes. Bounds on average and quantile treatment effects of Job Corps training on wages. *Journal of Human Resources*, 48(3):659–701, 2013.
- J. C. Brown and T. W. Guinnane. Infant mortality decline in rural and urban Bavaria: fertility, economic transformation, infant care, and inequality in Bavaria and Munich, 1825–1910. *Economic History Review*, 71(3), 2018.
- B. Callaway. *qte: Quantile Treatment Effects*, 2019. R package version 1.3.0 – Available at <https://cran.r-project.org/web/packages/qte/>.
- B. Callaway and W. J. Collins. Unions, workers, and wages at the peak of the American labor movement. *Explorations in Economic History*, 68:95–118, 2018.
- B. Callaway and T. Li. Quantile Treatment Effects in Difference in Differences Models with Panel Data. *Quantitative Economics*, 10(4):1579–1618, 2019a.

- B. Callaway and T. Li. Quantile treatment effects in difference in differences models with panel data. *Quantitative Economics*, 10(4):1579–1618, 2019b. doi: <https://doi.org/10.3982/QE935>. URL <https://onlinelibrary.wiley.com/doi/abs/10.3982/QE935>.
- B. Callaway, T. Li, and T. Oka. Quantile treatment effects in difference in differences models under dependence restrictions and with only two time periods. *Journal of Econometrics*, 206(2):395–413, 2018. ISSN 0304-4076. doi: <https://doi.org/10.1016/j.jeconom.2018.06.008>. URL <https://www.sciencedirect.com/science/article/pii/S0304407618301027>. Special issue on Advances in Econometric Theory: Essays in honor of Takeshi Amemiya.
- N. Canaday. The accumulation of property by southern blacks and whites: Individual-level evidence from a South Carolina cotton county, 1910–1919. *Explorations in Economic History*, 45(1):51–75, 2008.
- S. Carson. A quantile approach to the demographic, residential, and socioeconomic effects on 19th-century African-American body mass index values. *Cliometrica*, 6:193–209, 2012. doi: 10.1007/s11698-011-0069-0.
- V. Chernozhukov and C. Hansen. An iv model of quantile treatment effects. *Econometrica*, 73(1): 245–261, 2005. doi: <https://doi.org/10.1111/j.1468-0262.2005.00570.x>.
- V. Chernozhukov and C. Hansen. Quantile models with endogeneity. *Annual Review of Economics*, 5 (1):57–81, 2013. doi: 10.1146/annurev-economics-080511-110952.
- V. Chernozhukov, C. Hansen, and K. Wüthrich. Instrumental variable quantile regression. In R. Koenker, V. Chernozhukov, X. He, and L. Peng, editors, *Handbook of Quantile Regression*, Handbooks of Modern Statistical Methods, pages 119–143. Chapman & Hall/CRC, 2018.
- A. Chesher. Nonparametric identification under discrete variation. *Econometrica*, 73(5):1525–1550, 2005. doi: <https://doi.org/10.1111/j.1468-0262.2005.00628.x>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-0262.2005.00628.x>.
- T. G. Conley and D. W. Galenson. Quantile regression analysis of censored wealth data. *Historical Methods*, 27(4):149–165, 1994.
- T. G. Conley and D. W. Galenson. Nativity and Wealth in Mid-Nineteenth-Century Cities. *The Journal of Economic History*, 58(2):468–493, 1998. doi: 10.1017/S0022050700020581.
- D. L. Costa. The Measure of Man and Older Age Mortality: Evidence from the Gould Sample. *The Journal of Economic History*, 64(1):1–23, March 2004. URL [https://ideas.repec.org/a/cup/jechis/v64y2004i01p1-23\\_00.html](https://ideas.repec.org/a/cup/jechis/v64y2004i01p1-23_00.html).
- M. Coşgel and B. Ergene. Inequality of Wealth in the Ottoman Empire: War, Weather, and Long-Term Trends in Eighteenth-Century Kastamonu. *The Journal of Economic History*, 72(2):308–331, 2012. doi: 10.1017/S0022050712000046.
- A. Deaton. Height, health, and development. *Proceedings of the National Academy of Sciences*, 104 (33):13232–13237, 2007. ISSN 0027-8424. doi: 10.1073/pnas.0611500104. URL <https://www.pnas.org/content/104/33/13232>.

- M. Dincecco. Political regimes and sovereign credit risk in Europe, 1750–1913. *European Review of Economic History*, 13(1):31–63, 2009. doi: 10.1017/S136149160900238X.
- M. Drelichman and D. González-Agudo. Housing and the cost of living in early modern Toledo. *Explorations in Economic History*, 54:27–47, 2014.
- M. Dribe. Demand and supply factors in the fertility transition: A county-level analysis of age-specific marital fertility in Sweden, 1880–1930. *European Review of Economic History*, 13(1):65–94, 2009. doi: 10.1017/S1361491608002372.
- S. Du Plessis, Jansen, A., and D. Von Fintel. Slave prices and productivity at the Cape of Good Hope from 1700 to 1725: Did everyone win from the trade? *Cliometrica*, 9:289–330, 2015. doi: 10.1007/s11698-014-0116-8.
- B. Dupont. Bank runs, information and contagion in the panic of 1893. *Explorations in Economic History*, 44(3):411–431, 2007.
- F. Etro. The Economics of Renaissance Art. *The Journal of Economic History*, 78(2):500–538, 2018. doi: 10.1017/S0022050718000244.
- S. Firpo. Efficient semiparametric estimation of quantile treatment effects. *Econometrica*, 75(1):259–276, 2007.
- R. Floud, R. Fogel, B. Harris, and S. Hong. *The Changing Body: Health, Nutrition, and Human Development in the Western World since 1700*. Cambridge University Press, 2011.
- R. W. Fogel. *The Escape from Hunger and Premature Death, 1700–2100: Europe, America, and the Third World*. Cambridge Studies in Population, Economy and Society in Past Time. Cambridge University Press, 2004. doi: 10.1017/CBO9780511817649.
- B. R. Frandsen, M. Frölich, and B. Melly. Quantile treatment effects in the regression discontinuity design. *Journal of Econometrics*, 168(2):382–395, 2012. ISSN 0304-4076. doi: <https://doi.org/10.1016/j.jeconom.2012.02.004>. URL <https://www.sciencedirect.com/science/article/pii/S0304407612000607>.
- M. Frölich and B. Melly. Estimation of quantile treatment effects with Stata. *Stata Journal*, 10(3):423–457(35), 2010.
- I. Gazeley and A. Newell. Poverty in Edwardian Britain. *Economic History Review*, 64(1):52–71, 2011.
- R. González-Val, D. Tirado-Fabregat, and E. Viladecans-Marsal. Market potential and city growth: Spain 1860–1960. *Cliometrica*, 11:31–61, 2017. doi: 10.1007/s11698-015-0139-9.
- J. Gu and S. Volgushev. Panel data quantile regression with grouped fixed effects. *Journal of Econometrics*, 213(1):68–91, 2019. doi: <https://doi.org/10.1016/j.jeconom.2019.04.006>. Annals: In Honor of Roger Koenker.
- A. Hagemann. Cluster-robust bootstrap inference in quantile regression models. *Journal of the American Statistical Association*, 112(517):446–456, 2017. doi: 10.1080/01621459.2016.1148610.

- X. He. Resampling Methods. In R. Koenker, V. Chernozhukov, X. He, and L. Peng, editors, *Handbook of Quantile Regression*, Handbooks of Modern Statistical Methods, pages 7–19. Chapman & Hall/CRC, 2018.
- G. W. Imbens and J. D. Angrist. Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62(2):467–475, March 1994.
- G. W. Imbens and D. B. Rubin. Estimating Outcome Distributions for Compliers in Instrumental Variables Models. *Review of Economic Studies*, 64(4):555–574, 1997.
- M. Jaremski and P. Fishback. Did Inequality in Farm Sizes Lead to Suppression of Banking and Credit in the Late Nineteenth Century? *The Journal of Economic History*, 78(1):155–195, 2018. doi: 10.1017/S0022050718000062.
- S. Jayachandran and R. Pande. Why Are Indian Children So Short? The Role of Birth Order and Son Preference. *American Economic Review*, 107(9):2600–2629, September 2017. doi: 10.1257/aer.20151282. URL <https://www.aeaweb.org/articles?id=10.1257/aer.20151282>.
- E. Jorge-Sotelo. The limits to lender of last resort interventions in emerging economies: evidence from the Gold Standard and the Great Depression in Spain. *European Review of Economic History*, 24(1): 98–133, 2020. doi: 10.1093/ereh/hey030.
- G. Karagedikli and A. C. Tunçer. House prices in the Ottoman Empire: evidence from eighteenth-century Edirne. *Economic History Review*, 74(1):710–745, 2021.
- F. Kersting, I. Wohnsiedler, and N. Wolf. Weber Revisited: The Protestant Ethic and the Spirit of Nationalism. *The Journal of Economic History*, 80(3):710–745, 2020. doi: 10.1017/S0022050720000364.
- T. Keywood and J. Baten. Elite violence and elite numeracy in Europe from 500 to 1900 CE: roots of the divergence. *Cliometrica*, 2020. doi: 10.1007/s11698-020-00206-1.
- K. A. Kholodilin. Demand and supply factors in the fertility transition: A county-level analysis of age-specific marital fertility in Sweden, 1880–1930. *European Review of Economic History*, 20(3): 322–344, 2016. doi: 10.1093/ereh/hew008.
- R. Koenker. Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, 91(1):74–89, 2004. ISSN 0047-259X. doi: <https://doi.org/10.1016/j.jmva.2004.05.006>. URL <https://www.sciencedirect.com/science/article/pii/S0047259X04001113>. Special Issue on Semiparametric and Nonparametric Mixed Models.
- R. Koenker. *Quantile Regression*. Econometric Society Monographs. Cambridge University Press, 2005. doi: 10.1017/CBO9780511754098.
- R. Koenker. Quantile regression: 40 years on. *Annual Review of Economics*, 9(1):155–176, 2017. doi: 10.1146/annurev-economics-063016-103651.
- R. Koenker. *quantreg: Quantile Regression*, 2021. R package version 5.86 – Available at <https://CRAN.R-project.org/package=quantreg>.
- R. Koenker and G. Bassett. Regression quantiles. *Econometrica*, 46(1):33–50, 1978.

- R. Koenker and K. F. Hallock. Quantile regression. *The Journal of Economic Perspectives*, 15(4): 143–156, 2001. ISSN 08953309. URL <http://www.jstor.org/stable/2696522>.
- R. Koenker, V. Chernozhukov, X. He, and L. Peng. *“Handbook of Quantile Regression”*. CRC Press, 2017. doi: 10.1201/9781315120256.
- N. Koepke and J. Baten. The biological standard of living in Europe during the last two millennia. *European Review of Economic History*, 9(1):61–95, 2005. doi: 10.1017/S1361491604001388.
- C. Lamarche. Quantile regression methods. In K. F. Zimmermann, editor, *Handbook of Labor, Human Resources and Population Economics*, pages 1–15. Springer International Publishing, Cham, 2019. ISBN 978-3-319-57365-6. doi: 10.1007/978-3-319-57365-6\_42-1. URL [https://doi.org/10.1007/978-3-319-57365-6\\_42-1](https://doi.org/10.1007/978-3-319-57365-6_42-1).
- M. Lipsitz, A. Belloni, V. Chernozhukov, and I. Fernandez-Val. Quantreg.nonpar: an r package for performing nonparametric series quantile regression. CeMMAP working papers CWP29/17, Centre for Microdata Methods and Practice, Institute for Fiscal Studies, 2017. URL <https://EconPapers.repec.org/RePEc:ifs:cemmap:29/17>.
- M. Llorca-Jaña, J. Navarrete-Montalvo, R. Araya-Valenzuela, and F. Droller. The physical stature of men in 19th-century chile: Another case of stagnation during an export boom. *Revista de Historia Económica / Journal of Iberian and Latin American Economic History*, 37(2):239–270, 2019. doi: 10.1017/S0212610918000198.
- M. Llorca-Jaña, J. Navarrete-Montalvo, R. Araya-Valenzuela, F. Droller, M. Allende, and J. Rivas. Height in twentieth-century chilean men: growth with divergence. *Cliometrica*, 15(1):135–166, 2021. doi: <https://doi.org/10.1007/s11698-020-00205-2>.
- L. Ma and R. Koenker. Quantile regression methods for recursive structural equation models. *Journal of Econometrics*, 134(2):471–506, 2006. URL <https://EconPapers.repec.org/RePEc:eee:econom:v:134:y:2006:i:2:p:471-506>.
- B. Melly and K. Wüthrich. Local Quantile Treatment Effects. In R. Koenker, V. Chernozhukov, X. He, and L. Peng, editors, *Handbook of Quantile Regression*, Handbooks of Modern Statistical Methods, pages 145–164. Chapman & Hall/CRC, 2018.
- K. J. Mitchener and M. D. Weidenmier. Was the Classical Gold Standard Credible on the Periphery? Evidence from Currency Risk. *The Journal of Economic History*, 75(2), 2015.
- K. Ogasawara and G. Kobayashi. The impact of social workers on infant mortality in inter-war Tokyo: Bayesian dynamic panel quantile regression with endogenous variables. *Cliometrica*, 9:97–130, 2015. doi: 10.1007/s11698-014-0110-1.
- K. Ogasawara and Y. Matsushita. Heterogeneous treatment effects of safe water on infectious disease: Do meteorological factors matter? *Cliometrica*, 13:55–82, 2019. doi: 10.1007/s11698-017-0169-6.
- F. Peracchi. Height and economic development in italy, 1730-1980. *American Economic Review*, 98(2): 475–81, May 2008. doi: 10.1257/aer.98.2.475. URL <https://www.aeaweb.org/articles?id=10.1257/aer.98.2.475>.

- M. Siravegna. Implementing quantile selection models in Stata. 2020 Stata Conference 2, Stata Users Group, Aug. 2020. URL <https://ideas.repec.org/p/boc/scon20/2.html>.
- P. Temin and H. Voth. Private borrowing during the financial revolution: Hoare's Bank and its customers, 1702–24. *Economic History Review*, 61(3):541–564, 2008.
- F. Valencia Caicedo. Historical Econometrics: Instrumental Variables and Regression Discontinuity Designs. In A. Bisin and G. Federico, editors, *The Handbook of Historical Economics*, Elsevier, pages 179–211. Academic Press, 2021.
- T. S. Vogl. Height, skills, and labor market outcomes in Mexico. *Journal of Development Economics*, 107:84–96, 2014. ISSN 0304-3878. doi: <https://doi.org/10.1016/j.jdeveco.2013.11.007>. URL <https://www.sciencedirect.com/science/article/pii/S0304387813001697>.
- K. W. Wachter and J. Trussell. Estimating historical heights. *Journal of the American Statistical Association*, 77(378):279–293, 1982. doi: 10.1080/01621459.1982.10477796.
- T. R. Walker. Economic Opportunity on the Urban Frontier: Wealth and Nativity in Early San Francisco. *Explorations in Economic History*, 37(3):258–277, 2000.
- H. J. Wang, L. A. Stefanski, and Z. Zhu. Corrected-loss estimation for quantile regression with covariate measurement errors. *Biometrika*, 99:405–421, 2012.
- S. Watanabe. Technology Shocks and the Great Depression. *The Journal of Economic History*, 76(3): 909–933, 2016. doi: 10.1017/S0022050716000772.
- Y. Wei. Quantile regression with measurement error. In R. Koenker, V. Chernozhukov, X. He, and L. Peng, editors, *Handbook of Quantile Regression*, Handbooks of Modern Statistical Methods, pages 119–139. Chapman & Hall/CRC, 2018.
- Y. Wei and R. Carroll. Quantile regression with measurement error. *Journal of the American Statistical Association*, 104:1129–1143, 2009.