LATENT STRUCTURES AND QUANTILES OF THE TREATMENT EFFECT DISTRIBUTION*

ERICH BATTISTIN, CARLOS LAMARCHE and ENRICO RETTORE†

We provide conditions to identify quantiles of the distribution of gains conditional of values of the outcome variable in the base state. The dependence between potential outcomes is modeled through a factor model. We improve on past research by allowing latent factors to affect both the level and the variance of potential outcomes, resulting in a more flexible distribution of gains. We show that the parameter of interest is a functional of features of the factor model that are nonparametrically identified. The analogue principle is used to obtain an estimator, for which we derive asymptotic behavior and finite sample properties via Monte Carlo simulations. We apply our method to evaluate the distributional effects of an Italian labor market policy that combines income support to eligible dismissed employees with wage benefits to employers who hire them. We find that the policy does not seem to have an impact on earnings three years after the program is implemented. The standard factor model specification is rejected in favor of the more flexible and general specification.

Keywords: Factor Models; Quantile Regression; Treatment Effects.

JEL Codes: C21; C23.

---

*This draft: June 5, 2014. The authors would like to thank Roger Koenker, Steve Portnoy, Ken Troske and Jim Ziliak for questions and comments as well as seminar participants at the University of Illinois at Urbana-Champaign, Queen Mary University of London, University of Kentucky, University of St. Gallen and the 14th IZA/CEPR European Summer Symposium.

†Erich Battistin: School of Economics and Finance, Queen Mary, University of London, Mile End Road, London E1 4NS, UK. Email: e.battistin@qmul.ac.uk. Carlos Lamarche: Department of Economics, University of Kentucky, 335A Gatton College of Business and Economics, Lexington, KY 40506, USA. Email: clamarche@uky.edu. Enrico Rettore: Department of Economics and Management, University of Padova, Via del Santo 33, 35123 Padova, Italy. Email: enrico.rettore@unipd.it.

1. Introduction

Countless theoretical and applied work from the early 70's has addressed the evaluation problem, that is the identification of the causal effect of a policy intervention on the outcome of interest. A common feature of many empirical contributions to this literature is the importance of considering heterogeneous returns to participation as opposed to the traditional 'common effect' model. The existence of such heterogeneity notwithstanding, causal inference has been concerned by and large with identification of the *average* effect of the intervention (see Heckman *et al.*, 1999, Heckman and Vytlacil, 2007, and DiNardo and Lee, 2011, for reviews). In this setting, the role played by heterogeneity is typically investigated by comparing average returns for different groups in the population identified by observable characteristics, but ignoring within group variability in returns.

The practice of looking at average effects mostly results from a pragmatic approach to the evaluation problem. Identification of characteristics of the effect *distribution* other than the average requires stronger assumptions. Besides, estimation often poses operational burdens that somewhat limit the widespread application and popularity amongst practitioners. Yet they are crucial to the estimation of a variety of policy parameters, alternative to the average, central for policy making. Abbring and Heckman (2007) thoroughly discuss these parameters, as well as their relevance for drawing informed policy decisions.

This paper derives new conditions for identification of features of the *treatment effect distribution* other than the average, thus allowing to make causal claims that are more general than those from standard average treatment effect analysis. We develop an estimation procedure that yields the quantile function of the individual gain from participation at selected values of the baseline outcome distribution. This is the parameter of interest in what follows, and is needed to understand how widely distributed are benefits amongst targeted beneficiaries. Without additional assumptions, the same policy analysis could not be drawn by considering identification strategies that yield average treatment effects.

The parameter of interest which is central to the discussion in what follows is a functional of the *joint* distribution of potential outcomes under the 'policy on' and 'policy off' scenarios. The *treatment effect distribution* is defined as the distribution of the difference between these two potential outcomes. Starting from the seminal work by Heckman *et al.* (1997), many authors have discussed the theoretical and empirical relevance of understanding the conditions required to learn about this object. Away from an array of special cases, the constant returns assumption being one example, identification of the treatment effect distribution is precluded on a logical ground irrespective of the assignment mechanism at work. With random allocation of units to treatment, data may only unveil the marginal distributions of potential outcomes. Under suitable conditions, such as strong ignorability (Rosenbaum and Rubin, 1983), instrumental variation in the assignment to treatment (Imbens and Rubin, 1997) or difference in differences settings (Athey and Imbens, 2006), the marginal distributions of potential outcomes can be retrieved for specific subgroups of the population even in a non-experimental setting. However, knowledge of the marginal distributions is not sufficient to point identify the joint distribution of potential outcomes. This in turn precludes estimation of many policy parameters away from the average.

A number of papers over the years have circumvented the problem by fueling two alternative streams of empirical and methodological literature. The first approach is by far the most exploited in empirical studies, but in fact moves the goalpost with respect to what we do in this paper. Rather than estimating quantiles of the treatment effect distribution, it provides conditions to identify the quantile treatment effect, which represents the 'horizontal' distance between the distribution under treatment and the distribution in the absence of treatment (Doksum, 1974). In light of what discussed above, methods required to identify averages straightforwardly extend to identification of quantile treatment effects. Abadie *et al.* (2002), Bitler *et al.* (2006), Firpo (2007) and Frandsen *et al.* (2012) are examples and both theoretical and empirical contributions to this literature.

An alternative approach has considered the possibility of constructing bounds on the joint distribution of potential outcomes, very much in the spirit of Fréchet inequalities - which indeed have a long lasting tradition in mathematical statistics - and the seminal work by Manski (1990). Partial identification of any functional of the treatment effect distribution is achieved by imposing restrictions on the *copula* relating the two potential outcomes. Despite its elegance, the contribution of this literature to applied work is still relatively limited. Bounds often prove non-informative for policy parameters, and are derived using conditions whose validity is not uncontroversial in many empirical settings. It turns out that in many instances this approach is used to rule out the 'common effect' model, but seldom can inform policy makers. See Heckman *et al.* (1997), Abbring and Heckman (2007) and Firpo, Pinheiro and

Ridder (2010) for examples and discussions. Rank invariance across policy states imposes restrictions on the *copula* that are sufficient to obtain point identification (see Heckman *et al.*, 1997).

An alternative approach considered in the literature, and the one that we take in what follows, postulates a factor structure that relates potential outcomes (see, for example, Carneiro *et al.*, 2003, Aakvik *et al.*, 2005, Heckman *et al.*, 2006, Cunha and Heckman, 2007, and Heckman *et al.*, 2014). In its bare essentials, the key assumption required for identification is that the dependence across potential outcomes is solely generated by a low dimensional set of factors. Under certain conditions, the availability of additional measurements on top of observed outcomes allows to retrieve non-parametrically the joint distribution of factors and potential outcomes, and from this any functional of interest (see Abbring and Heckman, 2007, for details).

Despite the transparency of the identification result, the need of flexible estimation strategies often challenges simplicity of the methods proposed, and this has affected their diffusion in empirical research. Leaving aside a range of convenient parametric cases, normality being a leading example, estimation is typically carried out using MCMC methods that make use of mixtures of distributions and involve convolutions of random variables. This approach has the advantage of retrieving the full treatment effect distribution, and thus any policy parameter that can be constructed from this (see the discussion in Abbring and Heckman, 2007). The approach proposed in this paper, on the contrary, has the advantage of being readably implementable using existing software, though it provides conditions for the identification of one specific parameter.

The use of factor models has a long lasting tradition in psychology and other social sciences. Our paper contributes to a growing literature signaling a 'renaissance' of their role in applied micro-econometrics, and in policy evaluation in particular. Recent papers have considered the potential of using latent structures as a tool to gain external validity in regression discontinuity designs (Angrist and Rokkanen, 2013, and Rokkanen, 2013), or to achieve identification of time-varying treatment effects (Cooley Fruehwirth *et al.*, 2011). Bonhomme and Robin (2010), Arcidiacono *et al.* (2011) and Jackson (2013) provide additional examples of empirical research that employs factor representations similar to that we consider in this paper.

The main contributions of this paper can be summarized as follows. First, estimation of treatment effects is carried out by assuming a factor model in which latent factors is allowed to affect the conditional variance of the outcomes, not only its conditional average, a feature that is not considered in the traditional formulation of the model or in previous applications to policy evaluation (see Abbring and Heckman, 2007). The distribution of the causal effect implied by this model is considerably more flexible than that obtained under the traditional factor model. The identification conditions for the parameter of interest are very general in nature, or at least as general as those already presented in other studies estimating economic models in this setup. Second, we derive an estimation strategy that exploits recent results in the literature on quantiles. Monte Carlo simulations show that the proposed estimator outperforms in finite samples that by Chernozhukov and Hansen (2006) even if asymptotically equivalent to their. The methodology developed is then used to evaluate the distributional effects of an Italian labour market policy that combines income support to eligible dismissed employees with wage benefits to employers who hire them.

The remainder of the paper is organised as follows. Section 2 sets up the notation and the parameter of interest. Section **??** discusses the assumptions underlying the factor representation employed. Section 4 derives the main identification result. Section 5 deals with estimation, while asymptotic theory is presented in Section 5.3. The application is in Section 7, and Section 8 concludes. Proofs of theorems are in the Appendix.

2. The Model

2.1. Notation. The notation employed in the potential outcome approach to causal inference is used throughout. Assume that the variables $(Y, D, \mathbf{W}, \mathbf{X})$ are observed for a sample of units randomly drawn from the relevant population, where $Y = Y_0 + D(Y_1 - Y_0)$ is a scalar continuous outcome, $\boldsymbol{W} = (W_1, \ldots, W_K)'$ is a vector of K continuous random variables, D is the *binary* treatment or policy status indicator defining the potential outcomes (Y_1, Y_0) for participation and non-participation, respectively, and $\mathbf{X}$ are control variables exogenous to the model. For example the vector $\mathbf{W}$ may include outcome measurements for periods preceding the policy roll out.

The notation $F_A[a|b]$ indicates the distribution of random variable A calculated at a conditional on random variable B taking value b. A similar notation is employed for the conditional τ-quantile function $Q_A[\tau|b] := F_A^{-1}[\tau|b]$.

2.2. **Factor structure.** The following factor structure is assumed, which we state as conditional on cells defined by $\mathbf{X}$ (the conditioning on these variables is left implicit throughout):

$$W_k = \sum_{l=1}^{r} \delta_{kl}\Theta_l + (1 + \sum_{l=1}^{r} \xi_{kl}\Theta_l)V_k, \quad k = 1, \ldots, K \tag{2.1}$$

$$Y_1 = \sum_{l=1}^{r} \lambda_{1l}\Theta_l + (1 + \sum_{l=1}^{r} \gamma_{1l}\Theta_l)U_1, \tag{2.2}$$

$$Y_0 = \sum_{l=1}^{r} \lambda_{0l}\Theta_l + (1 + \sum_{l=1}^{r} \gamma_{0l}\Theta_l)U_0. \tag{2.3}$$

The Θ_l's are continuous unobserved factors, with $l = 1, \ldots, r$. Their number is known and equal to $r < K$, with $(K - r) \geq r$. The variables V_k, U_0 and U_1 denote continuous uniqueness components. The model features $2(K + 2)r$ unknown parameters, where $(\delta_l, \lambda_{1l}, \lambda_{0l})$ are factor loadings and $(\xi_l, \gamma_{1l}, \gamma_{0l})$ are scale parameters.

We write the model in a more convenient matrix form:

$$\boldsymbol{W} = \boldsymbol{L\Theta} + \boldsymbol{G\Theta} \circ \boldsymbol{V} + \boldsymbol{V}, \tag{2.4}$$

$$Y_1 = \boldsymbol{\lambda}_1'\boldsymbol{\Theta} + (1 + \boldsymbol{\gamma}_1'\boldsymbol{\Theta})U_1, \tag{2.5}$$

$$Y_0 = \boldsymbol{\lambda}_0'\boldsymbol{\Theta} + (1 + \boldsymbol{\gamma}_0'\boldsymbol{\Theta})U_0, \tag{2.6}$$

where $\boldsymbol{V} = (V_1, \ldots, V_K)'$ and $\boldsymbol{\Theta} = (\Theta_1, \ldots, \Theta_r)'$ are $K \times 1$ and $r \times 1$ vectors, respectively, and $\circ$ denotes the Hadamard product. For $d = \{0, 1\}$ there is $\boldsymbol{\lambda}_d = (\lambda_{d1}, \ldots, \lambda_{dr})'$ and $\boldsymbol{\gamma}_d = (\gamma_{d1}, \ldots, \gamma_{dr})'$, and $\boldsymbol{L}$ and $\boldsymbol{G}$ are $K \times r$ matrices defined as follows:

$$\boldsymbol{L} = \begin{bmatrix} \delta_{11} & \delta_{12} & \ldots & \delta_{1r} \\ \delta_{21} & \delta_{22} & \ldots & \delta_{2r} \\ \vdots & \ldots & \ldots & \vdots \\ \delta_{r1} & \delta_{r2} & \ldots & \delta_{rr} \\ \vdots & \ldots & \ldots & \vdots \\ \delta_{K1} & \delta_{K2} & \ldots & \delta_{Kr} \end{bmatrix}, \quad \boldsymbol{G} = \begin{bmatrix} \xi_{11} & \xi_{12} & \ldots & \xi_{1r} \\ \xi_{21} & \xi_{22} & \ldots & \xi_{2r} \\ \vdots & \ldots & \ldots & \vdots \\ \xi_{r1} & \xi_{r2} & \ldots & \xi_{rr} \\ \vdots & \ldots & \ldots & \vdots \\ \xi_{K1} & \xi_{K2} & \ldots & \xi_{Kr} \end{bmatrix}.$$

Consider the following partition of the matrix $\boldsymbol{L} = [\boldsymbol{L}_{(1)}, \boldsymbol{L}_{(2)}]$, where $\boldsymbol{L}_{(1)}$ is $r \times r$ and $\boldsymbol{L}_{(2)}$ is $(K - r) \times r$. Similarly, $\boldsymbol{G} = [\boldsymbol{G}_{(1)}, \boldsymbol{G}_{(2)}]$, where $\boldsymbol{G}_{(1)}$ is $r \times r$ and $\boldsymbol{G}_{(2)}$ is $(K - r) \times r$. We maintain the following normalization:

$$\boldsymbol{L}_{(1)} = \boldsymbol{I}_r, \qquad \boldsymbol{G}_{(1)} = \boldsymbol{0}_r, \tag{2.7}$$

which in turn implies:

$$\boldsymbol{\Theta} = \boldsymbol{W}_{(1)} - \boldsymbol{V}_{(1)}, \tag{2.8}$$

where $\boldsymbol{W}_{(1)} = (W_1, ..., W_r)'$ and $\boldsymbol{V}_{(1)} = (V_1, ..., V_r)'$. Naturally this implies that $\boldsymbol{W}_{(2)} = (W_{r+1}, ..., W_K)'$. The normalization $\boldsymbol{L}_{(1)} = \boldsymbol{I}_r$ is standard, and sets the scale of the unobserved factors. The normalization $\boldsymbol{G}_{(1)} = \mathbf{0}_r$ imposes homoskedasticity for measurements $\boldsymbol{W}_{(1)}$. This, together with Assumption 1, implies that the r components of $\boldsymbol{W}_{(1)}$ are error ridden measurements of the latent factors $\boldsymbol{\Theta}$. The latter normalization is required to derive our identification result.

Components of the vector $(U_0, \mathbf{V}')$ are assumed mutually independent and independent of the latent factors $\boldsymbol{\Theta}$. The same assumption is made for the vector $(U_1, \mathbf{V}')$. Means of factors and uniqueness are normalized to zero. The conditions embedded in the model can be summarized as follows.

Assumption 1. (*Factor model*).

- **(i)** The factor model is defined by equations (2.4), (2.5), (2.6) and the normalization (2.7).
- **(ii)** Conditional on $\mathbf{X}$ there is for $d = \{0, 1\}$:

$$F_{U_d \boldsymbol{V} \boldsymbol{\Theta}}(u, \mathbf{v}, \boldsymbol{\theta}) = F_{U_d}(u) \left[\prod_{k=1}^{K} F_{V_k}(v_k) \right] F_{\boldsymbol{\Theta}}(\boldsymbol{\theta}).$$

- **(iii)** The random variables $(U_1, U_0, \mathbf{V}, \boldsymbol{\Theta})$ have continuous distributions with zero mean, variance matrix $\boldsymbol{\Sigma}_{2+K+r} < \infty$, and finite fourth moments.

Assumptions 1.ii and 1.iii are standard in the literature on factor models. Instead Assumption 1.i marks an important departure, as the variance of measurements $(Y_0, Y_1, \mathbf{W})$ conditional on $\boldsymbol{\Theta}$ is allowed to vary with the latent factors $\boldsymbol{\Theta}$.

The standard factor model is obtained by setting scale parameters to zero. In the standard setting, an application of Kotlarski's (1967) theorem yields non-parametric identification of $F_{U_d \boldsymbol{V} \boldsymbol{\Theta}}(u, \mathbf{v}, \boldsymbol{\theta})$, $d = \{0, 1\}$. This result paves the way for identification of $F_{Y_0 Y_1}(y_0, y_1)$, or conditional versions of it, if combined with the assumptions discussed further below (see Abbring and Heckman, 2007). The Kotlarski's (1967) theorem does not apply to equations (2.2) and (2.3), as the uniqueness is allowed to depend on latent factors through heteroskedasticity. We can however retrieve features of the model that are sufficient to define a parameter of interest that is informative on the distribution of gains.

2.3. **Policy assignment mechanism.** It is assumed throughout that the treatment status is randomly allocated to units of the relevant population.

Assumption 2. (*Policy assignment mechanism*). Conditional on $\mathbf{X}$ the random variables $(U_1, U_0, \mathbf{V}, \boldsymbol{\Theta})$ are independent of policy status:

$$F_{U_1 U_0 \mathbf{V} \boldsymbol{\Theta}}(u_1, u_0, \mathbf{v}, \boldsymbol{\theta}|d) = F_{U_1 U_0 \mathbf{V} \boldsymbol{\Theta}}(u_1, u_0, \mathbf{v}, \boldsymbol{\theta}).$$

We maintain randomization (or conditional versions of it, such as in regression discontinuity designs) as leading example. The discussion will clarify that identification is driven by the outcome equations together with independence assumptions that come with the factor structure, rather than from the assignment mechanism. Variants to Assumption 2 may be considered to allow for non-random selection into treatment (see Abbring and Heckman, 2007). Further modifications can be allowed by assuming that units have two periods of outcome data in one counterfactual state or the other (Cunha and Heckman, 2008).[1]

3. Parameter of interest

3.1. **Definition.** The contrast of potential outcomes for alternative populations of individuals defines a variety of *causal parameters*. The treatment or policy effect is defined as $\Delta \equiv Y_1 - Y_0$, namely the difference that results from contrasting realizations of the outcome under the two (mutually exclusive) scenarios for the policy status. Knowledge of the distribution of Δ allows to answer policy questions regarding, for instance, how widely treatment gains are distributed across recipients, or to study the effect on recipients for specific values of the base state distribution. Even when individuals are randomized into/out of treatment, identification of these parameters requires additional assumptions to retrieve the joint distribution of Y_0 and Y_1, and thus that of Δ, from the two marginals. The unrestricted set of joint distributions consistent with the marginals can be exploited to partially-identify the distribution of Δ via classical probability inequalities. However, the resulting identification set is generally uninformative (see the discussion in Heckman *et al.*, 1997). Measures based on the marginal distributions of Y_0 and Y_1 are useful to document the heterogeneity of the treatment across individuals investigating quantile treatment effects (Bitler *et al.*, 2006).

[1] Enrico to check that it is truly the case.

The parameter of interest in what follows is the quantile function of the distribution of gains (QDG) conditional on $\boldsymbol{\Theta} = \boldsymbol{\theta}$ and $U_0 = u_0$ (within cells defined by $\mathbf{X}$):

$$QDG[\tau; \boldsymbol{\theta}, u_0] = Q_{\Delta}(\tau|\boldsymbol{\theta}, u_0).$$

We will consider a collection of QDGs defined at quantiles $Q_{U_0}(\tau_0)$ of the uniqueness U_0. In light of equation (2.6), the conditioning on $\boldsymbol{\Theta}$ and U_0 makes such collection of QDGs equivalent to learning about the distribution of gains at different values of the base state (or policy off) distribution Y_0. This is one of the parameters discussed in the seminal paper by Heckman *et al.* (1997).[2]

3.2. **Representation.** Assumption 1 and Assumption 2 are needed to learn about the factor model. The following Assumption 3.i has been used to achieve identification of the distribution of causal effects, and is standard in the literature on factor models and distribution of gains (see, for example, Carneiro *et al.*, 2003, Aakvik *et al.*, 2005, and Abbring and Heckman, 2007). The assumption plays a key role in the derivation of QDG. As it will be clear in the next section, Assumption 3.i can be replaced by Assumption 3.ii conveying a similar identification result. Rather than assuming independence between uniqueness U_0 and U_1, the latter assumption restricts the conditional distribution of $U_1 - U_0$ given U_0.

Assumption 3. (*Distribution of uniqueness*). Conditional on $\mathbf{X}$:

(i) the random variables (U_1, U_0) are independent:

$$F_{U_0 U_1}(u_0, u_1) = F_{U_0}(u_0) F_{U_1}(u_1);$$

or alternatively,

(ii) the participation effect on uniqueness $U_1 - U_0$ is independent of the baseline value U_0: $F_{U_1 - U_0}(\eta|u_0) = F_{U_1 - U_0}(\eta)$.

The independence Assumption 3.i, together with equations (2.5) and (2.6), allows to write the fundamental representation for the parameter of interest. There is:

$$\Delta = (\boldsymbol{\lambda}_1' - \boldsymbol{\lambda}_0')\boldsymbol{\Theta} + (U_1 \boldsymbol{\gamma}_1' - U_0 \boldsymbol{\gamma}_0')\boldsymbol{\Theta} + U_1 - U_0,$$

which under Assumption 3.i yields:

$$QDG[\tau_1; \boldsymbol{\theta}, Q_{U_0}(\tau_0)] = (\boldsymbol{\lambda}_1[\tau_1]' - \boldsymbol{\lambda}_0[\tau_0]')\boldsymbol{\theta} + Q_{U_1}(\tau_1) - Q_{U_0}(\tau_0), \tag{3.1}$$

[2]Add graphical interpretation? Maybe yes, as the audience may like this.

where here the QDG is calculated at $U_0 = Q_{U_0}[\tau_0]$ and:

$$\begin{aligned} \boldsymbol{\lambda}_1[\tau_1] &\equiv \boldsymbol{\lambda}_1 + \boldsymbol{\gamma}_1 Q_{U_1}[\tau_1], \\ \boldsymbol{\lambda}_0[\tau_0] &\equiv \boldsymbol{\lambda}_0 + \boldsymbol{\gamma}_0 Q_{U_0}[\tau_0]. \end{aligned}$$

The quantity in (3.1) is the quantile function of the distribution of gains conditional on having $\boldsymbol{\Theta} = \boldsymbol{\theta}$ and $U_0 = Q_{U_0}(\tau_0)$. It defines the following correspondence:

$$QDG[\tau_1; \boldsymbol{\theta}, Q_{U_0}(\tau_0)] = \mathcal{H}_{\lambda_0[\tau_0],\lambda_1[\tau_1]}\{Q_{U_0}[\tau_0], Q_{U_1}[\tau_1]\}, \tag{3.2}$$

where $\mathcal{H}$ is a (known) point identifying functional. For known values of the loadings $\boldsymbol{\lambda}_1[\tau_1]$ and $\boldsymbol{\lambda}_0[\tau_0]$, this establishes a correspondence between the unobserved term on the left hand side of the equation and the two quantile functions $Q_{U_0}[\tau_0]$ and $Q_{U_1}[\tau_1]$ on the right hand side. Knowledge of factor loadings and these quantile functions is sufficient to retrieve the parameter of interest. This sets the stage for analogue estimation presented below.

Assumption 1 improves on the previous literature by adding flexibility to the relationship between Δ and $\boldsymbol{\Theta}$. In the traditional formulation of the factor model there is $\boldsymbol{\gamma}_0 = \boldsymbol{\gamma}_1 = \mathbf{0}$, so that latent factors affect the distribution of gains only through a location shift constant across quantiles of the treatment effect distribution. This is a feature that may be implausible in many empirical applications. Our setting instead implies:

$$\frac{\partial}{\partial \Theta_l} QDG[\tau_1; \boldsymbol{\theta}, Q_{U_0}(\tau_0)] = \lambda_{1l}[\tau_1] - \lambda_{0l}[\tau_0], \qquad l = 1, \ldots, r \tag{3.3}$$

where λ_{1l} and λ_{0l} are the l-th components of vectors $\boldsymbol{\lambda}_1[\tau_1]$ and $\boldsymbol{\lambda}_0[\tau_0]$, respectively, and the derivative is allowed to vary across quantiles through its dependence on $\boldsymbol{\gamma}_0$ and $\boldsymbol{\gamma}_1$. It will be clear in the next section that the identification of (3.3) requires less assumptions than those needed to identify (3.1). This result follows upon noting that identification of the derivative does not require direct knowledge of $Q_{U_0}[\tau_0]$ and $Q_{U_1}[\tau_1]$. Knowledge of the first derivative yields a formal test for the validity of the standard factor model, as in the latter case the value of (3.3) should be constant across quantiles.

Remark 1. [3] Under Assumptions 2 and 3, if the latent factors Θ_l's are observable, it is possible to obtain the parameter of interest $QDG[\tau_1; \boldsymbol{\theta}, Q_{U_0}(\tau_0)]$ as the difference between two conditional quantile functions: $Q_{Y_1}(\tau_1|\boldsymbol{\theta}) - Q_{Y_0}(\tau_0|\boldsymbol{\theta})$. Furthermore, if

[3] EB/ER propose to change into this: Note that, under Assumption 2 and Assumption 3, the expression in (3.1) coincides with the difference between two conditional quantile functions:

$U_d \sim F$ and $\tau_d = \tau$ for $d = \{0, 1\}$, under randomization, we obtain a conditional on unobservables version of the Lehmann's quantile treatment effect (QTE), $Q_Y(\tau|D = 1, \boldsymbol{\theta}) - Q_Y(\tau|D = 0, \boldsymbol{\theta})$. It should be noted that our model does not impose these restrictive assumptions. More importantly, our parameter of interest departs from Lehmann's quantile treatment effect since it measures gains at selected base state values y_0 which are determined by conditioning on $\boldsymbol{\theta}$ and $Q_{U_0}(\tau_0)$.

4. Identification

We first consider identification of factor loadings $\boldsymbol{\lambda}_d$ and scale parameters $\boldsymbol{\gamma}_d$ for $d = \{0, 1\}$. We then present the identification result to the case of the quantile of the uniqueness, employing instrumental quantile regressions.

4.1. **Factor loadings.** By substituting (2.8) into (2.5) and (2.6) there is:

$$Y_d = \boldsymbol{\lambda}_d' \boldsymbol{W}_{(1)} - \boldsymbol{\lambda}_d' \boldsymbol{V}_{(1)} + (1 + \boldsymbol{\gamma}_d' \boldsymbol{W}_{(1)} - \boldsymbol{\gamma}_d' \boldsymbol{V}_{(1)}) U_d, \tag{4.1}$$

which defines the (feasible) regression of Y on $\boldsymbol{W}_{(1)} = (W_1, ..., W_r)'$. Let:

$$\dot{Y} \equiv Y - \boldsymbol{\lambda}_1' D \boldsymbol{W}_{(1)} - \boldsymbol{\lambda}_0' (1 - D) \boldsymbol{W}_{(1)},$$

and recall that $\boldsymbol{W}_{(2)} = (W_{r+1}, ..., W_K)'$ and that σ_d^2 denotes the variance of the uniqueness U_d as in Assumption 1.iii.

Theorem 1. *Under Assumption 1 and Assumption 2, there is for* $d = \{0, 1\}$*:*

$$\begin{aligned} E_Y(Y|\boldsymbol{w}_{(2)}, d) &= \boldsymbol{\lambda}_d' E_{\boldsymbol{W}_{(1)}}(\boldsymbol{W}_{(1)}|\boldsymbol{w}_{(2)}, d), \\ E_{\dot{Y}}(\dot{Y}^2|\boldsymbol{w}_{(2)}, d) &= \sigma_d^2 [2\boldsymbol{\gamma}_d' E_{\boldsymbol{W}_{(1)}}(\boldsymbol{W}_{(1)}|\boldsymbol{w}_{(2)}, d) + \boldsymbol{\gamma}_d' E_{\boldsymbol{W}_{(1)}}(\boldsymbol{W}_{(1)} \boldsymbol{W}_{(1)}'|\boldsymbol{w}_{(2)}, d) \boldsymbol{\gamma}_d] + \kappa_d. \end{aligned}$$

The first moment condition in Theorem 1 implies that a consistent estimate of $\boldsymbol{\lambda}_1$ can be retrieved by estimating with 2SLS the regression of Y on $\boldsymbol{W}_{(1)}$ for the 'policy on' group using functions of $\boldsymbol{W}_{(2)}$ as instruments. The same approach applied to the 'policy off' group yields a consistent estimate of $\boldsymbol{\lambda}_0$. Knowledge of these quantities allows to retrieve $\dot{Y}$. Separately for the 'policy on' and 'policy off' groups, $\dot{Y}^2$ is regressed on the elements of $\boldsymbol{W}_{(1)}$, their squares and interactions using 2SLS, with functions of $\boldsymbol{W}_{(2)}$ as instruments. The elements of $\boldsymbol{\gamma}_d$ are identified by taking

$Q_{Y_1}(\tau_1|\boldsymbol{\theta}) - Q_{Y_0}(\tau_0|\boldsymbol{\theta})$. The latter difference for $\tau_1 = \tau_0$ identifies a version of the Lehmann's quantile treatment effect (QTE) which is conditional on $\boldsymbol{\theta}$. Our parameter of interest if evaluated at $\tau_1 = \tau_0$ coincides with this conditional QTE.

the appropriate ratio between the coefficients on $E_{\boldsymbol{W}_{(1)}}(\boldsymbol{W}_{(1)}\boldsymbol{W}'_{(1)}|\boldsymbol{w}_{(2)}, d)$ and the coefficients on $E_{\boldsymbol{W}_{(1)}}(W_1|\boldsymbol{w}_{(2)}, d)$. Note that the degree of over-identification of these parameters increases with r.

4.2. **Quantile functions of uniqueness.** The identification result is contained in Theorem 2 and Corollaries 1 and 2, for which conditions are discussed below. For $d = \{0, 1\}$, define:

$$Y_d = \boldsymbol{\lambda}'_d\boldsymbol{\theta} + (1 + \boldsymbol{\gamma}'_d\boldsymbol{\theta})U_d := q_d(\boldsymbol{\theta}, U_d),$$

so that the corresponding quantile function can be written as:

$$Q_{Y_d}(\tau_d|\boldsymbol{\theta}) = \boldsymbol{\lambda}'_d\boldsymbol{\theta} + (1 + \boldsymbol{\gamma}'_d\boldsymbol{\theta})Q_{U_d}(\tau_d) = q_d(\boldsymbol{\theta}, Q_{U_d}(\tau_d)).$$

Assumption 4. (*Monotonicity*) Conditional on $\boldsymbol{\theta}$ and $\boldsymbol{W}_{(1)}$, the conditional quantile functions $q_d(\boldsymbol{\theta}, Q_{U_d}(\tau_d))$ and $q_d(\boldsymbol{W}_{(1)}, \tau_d)$ are strictly increasing in τ_d.

EB: How is $q_d(\boldsymbol{W}_{(1)}, \tau_d)$ defined? It seems to be a quantile function conditional on $\boldsymbol{W}_{(1)}$, but this has not been defined - I think. Check also whether the continuity assumption in Assumption 1.iii implies Assumption 4.

Assumption[4] 4 is similar to A1 in Chernozhukov and Hansen (2005) and imposes strictly increasing functions in τ_d. As in Chernozhukov and Hansen (2005), we also employ a conventional conditional independence restriction which in this case is implied by the factor model and Assumptions 1 and 2. In contrast with their work, we do not use a Skorohod representation for the uniqueness, and more importantly, we do not require Assumption A4 on rank invariance, $U_0 \sim U_1$. This assumption imposes a restriction on U_d to do not vary with potential treatment states, implying for instance that subjects with high wages with training remain strong earners without training. Using the Skorohod representation, this essentially implies that the rank $U_0 = \tau$ remains $U_1 = \tau$, in contrast with our case where $U_d = \tau_d$ for $d = \{0, 1\}$. It is important to note that in our simplest conceivable model, assuming that $\lambda_{dl} = \gamma_{dl} = 1$ for all d and l and $r = 1$, the term $\Theta + U_d$ implies that the rank of the worker might not vary across subjects with $\Theta = \theta$. But the model allows for changes since U_d is not assumed to be identically distributed across d. Although our model does not require rank invariance or rank similarity, it imposes independence Assumptions 2 and 3, as previously stated.

[4]Discussion on rank invariance to be moved earlier in the paper.

The following theorem on latent structures and conditional quantile functions is the main result of this section:

Theorem 2. *Under Assumptions 1, 2, (3.i), and 4, for* $d = \{0,1\}$ *and* $\tau_d \in (0,1)$, $P(Y_d \leq q_d(\boldsymbol{\theta}, \tau_d)|\boldsymbol{\theta}) = P(Y_d \leq q_d(\boldsymbol{W}_{(1)}, \tau_d)|\boldsymbol{W}_{(2)}) = \tau_d$.

The result of Theorem 2 is important as it shows that the quantile of the uniqueness associated with a generalization of the traditional factor model can be identified. The result might be seen as an extension of Chernozhukov and Hansen (2005) to a model with latent variables and offers a foundation for estimation of a fairly general factor model based on the restrictions previously stated. Note that because of this, for instance, the standard factor model can be tested against the data, as in this case the quantile-specific factor loadings, hence the derivative (3.3), should be constant across quantiles.

The following result shows that, once $\boldsymbol{\lambda}_d$ and $\boldsymbol{\gamma}_d$ are identified (Theorem 1), it is possible to identify the quantile of the uniqueness considering the normalization adopted in (2.7) and the conditions of Theorem 2.

Corollary 1. *Under the conditions of Theorem 1 and Theorem 2, for* $\tau_d \in (0,1)$ *and* $d = \{0,1\}$*, (i) the quantiles of the uniqueness* U_d *are identified, and (ii) the quantile specific loading* $\boldsymbol{\lambda}_d(\tau_d)$ *is identified.*

Corollary 1 contains two important results. First, it shows that the quantiles of the uniqueness U_d and $\boldsymbol{\lambda}_d(\tau_d)$ are identified and, consequently for a given $\boldsymbol{\theta}$, the quantile of the distribution of gains (QDG) introduced in equation (3.2) is also identified. The second result is important as it shows that derivative of the QDG as in equation (3.3) can be identified.

Lastly, the following result shows that, if Assumption 3.i is replaced with Assumption 3.ii, knowledge of loadings and scale parameters in the factor model together with that of quantile functions $Q_{U_0}(\tau_0)$ and $Q_{U_1}(\tau_1)$ is sufficient to write an identifying correspondence for the QDG as in equation (3.2).

Corollary 2. *Under Assumption 1, Assumption 2, Assumption 3.ii, and Assumption 4, for* $\tau_d \in (0,1)$*, the following identifying correspondence is defined*

$$QDG[\tau_1; \boldsymbol{\theta}, Q_{U_0}(\tau_0)] = \mathcal{G}, \tag{4.2}$$

with $\mathcal{G}$ *known point identifying functional.*

4.3. **Quantile function of the latent factors.** It follows that identification and estimation of how a marginal change to $\boldsymbol{\theta}$ affects QDGs as in equation (3.3) does not require knowledge of $F_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ but the QDGs defined in equations (3.2) and (4.2) do. In this paper, we propose a solution to this problem based on approximations with small uniqueness variance (Chesher 1991). Following the normalization adopted in (2.7) and Assumption 1.iii, consider $W_l = \Theta_l + \sigma_{V_l} V_l$ for $l = 1, \ldots, r$. In what follows, we suppress the index l for simplicity. There is:

$$\begin{aligned} f_W[w] &= E_V[f_\Theta(w - \sigma_V V)] \\ &= f_\Theta(w) - \sigma_V f_\Theta^{(1)}(w) E_V[V] + \frac{\sigma_V^2}{2} f_\Theta^{(2)}(w) E_V[V^2] + o(\sigma_V^2), \end{aligned}$$

where the last expression follows from a Taylor approximation around $\sigma_V = 0$ and $f^{(j)}$ indicates the j-th partial derivative. Integration yields:

$$F_W[w] = F_\Theta(w) - \sigma f_\Theta(w) E_V[V] + \frac{\sigma_V^2}{2} f_\Theta^{(1)}(w) E_V[V^2] + o(\sigma_V^2).$$

Under 1.iii, it follows that:

$$F_W[w] \simeq F_\Theta(w) + \frac{\sigma^2}{2} f_\Theta^{(1)}(w),$$

where $A \simeq B$ indicates $A = B + o(\sigma_V^2)$. Defining $Q_W(\tau)$ such that $\tau = F_W[Q_W(\tau)]$ and adopting the convention that $w = Q_W(\tau)$, it follows that:

$$F_\Theta(Q_W(\tau)) + \frac{\sigma^2}{2} f_\Theta^{(1)}(Q_W(\tau)) \simeq \tau,$$

which also implies,

$$Q_\Theta(\tau) \simeq F_\Theta^{-1}\{F_\Theta(Q_W(\tau)) + \frac{\sigma^2}{2} f_\Theta^{(1)}(Q_W(\tau))\}.$$

By expanding the last expression around $\sigma^2 = 0$ and using differentiation of inverse functions, we obtain,

$$Q_\Theta(\tau) \simeq Q_W(\tau) + \frac{\sigma^2}{2} \frac{f_\Theta^{(1)}\{Q_W(\tau)\}}{f_\Theta\{Q_W(\tau)\}}.$$

Finally, by Chesher (2001, Appendix B), we approximate the quantile of the factor Θ in terms of the inverse of the cumulative distribution function and density function of the observable pre-intervention measure W:

$$Q_\Theta(\tau) \simeq Q_W(\tau) + \frac{\sigma^2}{2}\frac{f_W^{(1)}\{Q_W(\tau)\}}{f_W\{Q_W(\tau)\}} \tag{4.3}$$

$$= Q_W(\tau) + \frac{\sigma^2}{2}\left[\frac{\partial}{\partial w}\log(f_W\{w\})\right]_{w=Q_W(\tau)}. \tag{4.4}$$

By adopting the convention that $\theta = Q_\Theta(\tau)$, it is possible to identify the QDGs using equation (4.3). The last expression (4.4) turns out to be convenient to estimate the parameter of interest because avoids the natural challenges of estimating $f_W^{(1)}(\cdot)$.

It is important to emphasize that conditioning on $\boldsymbol{\theta}$ has a natural economic interpretation and allows the policy evaluator to learn about the causal relationship between two potential outcomes. At the core of this paper, there is the idea that an evaluator interested in the causal effect at different values of the base state distribution Y_0 can learn about gains by conditioning on values $\boldsymbol{\theta}$ and $Q_{U_0}(\tau_0)$. We note that there are alternatives to the approximation with small uniqueness variance but we focus on one approach for practical purposes (see, e.g., Hall and Lahiri 2008).

Remark 2. Identification has interesting connections with the literature on measurement error. Schennach (2008) discusses conditions to identify $Q_{Y|D\Theta}[\tau|d,\theta]$, interpreting W_1 as an error ridden measurement of the latent variable Θ. As in our context there is $Q_{Y|D\Theta}[\tau|d,0] = Q_{U_d}[\tau]$, the object she is after coincides with that entering the right end side of (3.2). As we do in this paper, Schennach (2008) imposes no specific functional form on the distribution of the dependent variable or Θ. Her approach does not employ any factor representation. However, results in Schennach (2008) do not apply to the case discussed here because her condition (6) is not met. This is because in our model $(\Theta - E\Theta|W_2)$ depends on W_2 as long as ξ_2 in (2.1) is not zero. Wei and Carroll (2009, see equation (2)) consider a problem similar to that arising in our context. Their solution requires knowledge of the distribution of $\Theta|W_2$ or, alternatively, the availability of auxiliary information allowing to estimate the distribution of the uniqueness V_1 (they consider the case in which V_1 is a zero-mean Gaussian random variable, so that only its variance is unknown). Their strategy cannot be applied to our context, as we don't make use of information coming from auxiliary data.

5. Estimation

This section proposes a series of estimators and investigates their large sample properties. We estimate conditional mean models and conditional quantile functions considering independently and identically distributed (i.i.d.) samples $\{(Y_i, D_i, \boldsymbol{X}_i', \boldsymbol{W}_i') : i = 1, ..., n_d\}$. We assume that n_1 individuals are in the treatment group and n_0 are in the control group.

5.1. **Factor loadings.** We begin by rewriting (4.1) as

$$\boldsymbol{Y}_d = \boldsymbol{W}_1 \boldsymbol{\lambda}_d + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}_d, \tag{5.1}$$

where $\boldsymbol{Y}_d = (Y_{1d}, \ldots, Y_{n_d d})'$ is an n_d-dimensional vector, $\boldsymbol{W}_1 = [\boldsymbol{W}_{(1)1}', \ldots, \boldsymbol{W}_{(1)n_d}']$ is a $n_d \times r$ matrix, $\boldsymbol{X}$ is a $n_d \times p$ matrix of exogenous variables, and $\boldsymbol{\epsilon}_d$ is a $n_d \times 1$ error term. We write:

$$\boldsymbol{y}_d(\boldsymbol{\lambda}_d) = \boldsymbol{Y}_d - \boldsymbol{W}_1 \boldsymbol{\lambda}_d = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{W}_2 \boldsymbol{\eta} + \boldsymbol{\varepsilon}_d, \tag{5.2}$$

where $\boldsymbol{W}_2$ is a matrix of instruments of dimension $n_d \times (K - r)$. Letting $\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$, it is possible to obtain $\hat{\boldsymbol{\eta}}(\boldsymbol{\lambda}_d) = (\boldsymbol{W}_2'\boldsymbol{M}\boldsymbol{W}_2)^{-1}\boldsymbol{W}_2'\boldsymbol{M}\boldsymbol{y}_d(\boldsymbol{\lambda}_d)$ and then:

$$\hat{\boldsymbol{\lambda}}_d = \underset{\lambda_d \in \Lambda_d}{\operatorname{argmin}} \left\{ \hat{\boldsymbol{\eta}}(\boldsymbol{\lambda}_d)' \boldsymbol{W}_2' \boldsymbol{M} \boldsymbol{W}_2 \hat{\boldsymbol{\eta}}(\boldsymbol{\lambda}_d) \right\}. \tag{5.3}$$

We now define the products,

$$Z_{1s} := W_{1k} W_{1l}, \ \text{ for } \ s = 1, \ldots, R; \ k = 1, \ldots, r; \ l = 1, \ldots, r,$$

where $R = r(r+1)/2$ and $\boldsymbol{Z}_1 = (Z_{11}, \ldots, Z_{1R})'$. Following Theorem 1, we define

$$\dot{\boldsymbol{Y}}_d = (\boldsymbol{Y}_d - \boldsymbol{W}_1 \boldsymbol{\lambda}_d)^2 = \boldsymbol{W}_1^2 \boldsymbol{a}_d + \boldsymbol{W}_1 \boldsymbol{b}_d + \boldsymbol{Z}_1 \boldsymbol{c}_d + \boldsymbol{X}\boldsymbol{\beta} + \dot{\boldsymbol{\epsilon}}_d, \tag{5.4}$$

where $\boldsymbol{W}_1^2$ is the $n_d \times r$ matrix whose elements are squares of the elements of $\boldsymbol{W}_1$. For a given $(\boldsymbol{a}_d', \boldsymbol{b}_d', \boldsymbol{c}_d')$ the previous equation can be written as,

$$\dot{\boldsymbol{y}}_d(\boldsymbol{a}_d, \boldsymbol{b}_d, \boldsymbol{c}_d) = \dot{\boldsymbol{Y}}_d - \boldsymbol{W}_1^2 \boldsymbol{a}_d - \boldsymbol{W}_1 \boldsymbol{b}_d - \boldsymbol{Z}_1 \boldsymbol{c}_d = \boldsymbol{X}\boldsymbol{\beta} + \dot{\boldsymbol{W}}_2 \dot{\boldsymbol{\eta}} + \dot{\boldsymbol{\varepsilon}}_d, \tag{5.5}$$

where $\dot{\boldsymbol{W}}_2$ is a vector of instruments that include transformations of $\boldsymbol{W}_2$. Similarly as before we obtain the estimator, $\tilde{\boldsymbol{\eta}}(\boldsymbol{a}_d, \boldsymbol{b}_d, \boldsymbol{c}_d) = (\dot{\boldsymbol{W}}_2'\boldsymbol{M}\dot{\boldsymbol{W}}_2)^{-1}\dot{\boldsymbol{W}}_2'\boldsymbol{M}\boldsymbol{y}_d(\boldsymbol{a}_d, \boldsymbol{b}_d, \boldsymbol{c}_d)$. It can be shown that:

$$(\hat{\boldsymbol{a}}_d', \hat{\boldsymbol{b}}_d', \hat{\boldsymbol{c}}_c') = \underset{(a,b,c) \in \mathcal{A} \times \mathcal{B} \times \mathcal{C}}{\operatorname{argmin}} \left\{ \tilde{\boldsymbol{\eta}}(\boldsymbol{a}_d, \boldsymbol{b}_d, \boldsymbol{c}_d)' \dot{\boldsymbol{W}}_2' \boldsymbol{M} \dot{\boldsymbol{W}}_2 \tilde{\boldsymbol{\eta}}(\boldsymbol{a}_d, \boldsymbol{b}_d, \boldsymbol{c}_d) \right\}. \tag{5.6}$$

Following Theorem 1 the estimator for the scale parameter is defined as follows for $l, s = 1, \dots, r$:

$$\hat{\gamma}_{dl} = \frac{\hat{c}_{dls}}{\hat{b}_{ds}}.$$

Asymptotically valid inference relies on the delta method provided that n_d is sufficiently large.

5.2. **Quantiles of Uniqueness.** It is convenient to define:

$$C_{id} = \rho_{\tau_d}(Y_{id} - \boldsymbol{\lambda}_d' \boldsymbol{W}_{1i} - \alpha_d \boldsymbol{\gamma}_d' \boldsymbol{W}_{1i} - \boldsymbol{X}_i' \boldsymbol{\beta} - \boldsymbol{W}_{2i}' \boldsymbol{\eta}) \tag{5.7}$$

where $\rho_{\tau_d}(u) = u(\tau_d - I(u \leq 0))$ is the standard quantile loss function (see, e.g., Koenker 2005). The term $\boldsymbol{W}_{2i} = g(\tau; W_{(2),r+1}, ..., W_{(2),K-r}, \boldsymbol{X}_i')'$ is a vector of transformations of instruments as introduced by Chernozhukov and Hansen (2006) [**E2: not sure this specification is needed**]. In practice, it is possible to estimate $g(\cdot)$ by a least squares projection of the endogenous variables $\boldsymbol{W}_{1i}$ on the instruments $\boldsymbol{W}_{2i}$ and the exogenous variables $\boldsymbol{X}_i$. In the simulations and empirical example, we consider the case of $r = (K - r)$ with $r > 0$, although the vector $g(\cdot)$ may include more elements than the vector of endogenous variables.

It is important to note that the generalization of the factor model proposed in this paper leads to a location-scale shift model, and consequently, [**E2: not sure to understand the consequence here**] the unknown parameter α_d in (5.7) is equal to the quantile of the uniqueness $Q_{U_d}(\tau_d)$. Thus, once we obtain consistent estimates of the factor loadings and scale parameters in a model with endogenous variables, we can adopt an inverse quantile regression approach which is similar in spirit [**E2: if this is a contribution, perhaps it would be important to spell out the nature of it**] to Chernozhukov and Hansen (2006, 2008).

We proceed in two steps. First, we minimize C_{id} over $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ as functions of τ_d, $\boldsymbol{\lambda}_d$, $\boldsymbol{\gamma}_d$ and α_d,

$$\hat{\boldsymbol{\vartheta}}(\tau_d, \hat{\boldsymbol{\lambda}}_d, \hat{\boldsymbol{\gamma}}_d, \alpha_d) = \begin{pmatrix} \hat{\boldsymbol{\beta}}(\tau_d, \hat{\boldsymbol{\lambda}}_d, \hat{\boldsymbol{\gamma}}_d, \alpha_d) \\ \hat{\boldsymbol{\eta}}(\tau_d, \hat{\boldsymbol{\lambda}}_d, \hat{\boldsymbol{\gamma}}_d, \alpha_d) \end{pmatrix} = \underset{\beta, \eta \in \mathcal{B} \times \mathcal{G}}{\operatorname{argmin}} \sum_{i=1}^{n_d} C_{id}(\tau_d, \hat{\boldsymbol{\lambda}}_d, \hat{\boldsymbol{\gamma}}_d, \alpha_d). \tag{5.8}$$

Then we estimate the coefficient on the endogenous variable by finding the value of α which minimizes the following objective function:

$$\hat{\alpha}(\tau_d) = \hat{Q}_{U_d}(\tau_d) = \underset{\alpha \in \mathcal{A}}{\operatorname{argmin}} \left\{ \hat{\boldsymbol{\eta}}(\tau_d, \boldsymbol{\lambda}_d, \gamma_d, \alpha_d)' \hat{\boldsymbol{A}}(\tau_d) \hat{\boldsymbol{\eta}}(\tau_d, \boldsymbol{\lambda}_d, \gamma_d, \alpha_d) \right\}, \tag{5.9}$$

for a positive definite matrix $\boldsymbol{A}$.

The method has several advantages. It is well known that the implementation of inverse quantile regression approaches requires a low-dimensional search over coefficients associated with the endogenous variables. Note that while r can be potentially large in empirical applications, the search over α_d is simply done over the real line. Moreover, in contrast with other approaches, the proposed approach has the additional advantage to estimate directly quantiles of the uniqueness, avoiding issues of solving for α_d based on $\hat{\boldsymbol{\lambda}}_d(\tau_d)$, $\hat{\boldsymbol{\lambda}}_d$ and $\hat{\boldsymbol{\gamma}}_d$ with unknown weights and potentially small scale parameter estimates.

Lastly, it is useful to compare the approach with a control variate approach (Chesher 2003 and Ma and Koenker 2006). A control variate type estimator for the model defined in (4.1) can be defined as follows:

$$(5.10)\quad (\tilde{\boldsymbol{\lambda}}_d(\tau_d)', \tilde{\boldsymbol{\delta}}(\tau_d)', \tilde{\boldsymbol{\beta}}(\tau_d)') = \underset{\lambda,\delta,\beta\in\Lambda\times\Delta\times\mathcal{B}}{\operatorname{argmin}} \sum_{i=1}^{n_d} \rho_{\tau_d}\left(Y_{id} - \boldsymbol{\lambda}_d'\boldsymbol{W}_{1i} - \boldsymbol{\delta}'\boldsymbol{V}_{1i} - \boldsymbol{X}_i'\boldsymbol{\beta}\right).$$

If the vector $\boldsymbol{V}_{1i}$ is observed, it is natural to directly estimate the quantile of the uniqueness as $\tilde{\beta}_0(\tau_d)$ which estimates $\beta_0(\tau_d) = \beta_0 + Q_{U_d}(\tau_d)$ or directly $Q_{U_d}(\tau_d)$ in the model defined in equations (2.2) and (2.3). However, estimation of $\boldsymbol{V}_{1i}$ based on residuals from a linear or non-linear model might not lead to estimation of the uniqueness even in the case that $\beta_0 = 0$ and the loadings $\boldsymbol{\lambda}_d$ and scale parameters $\boldsymbol{\gamma}_d$ are not estimated.

5.3. **Theory of estimation and basic inference.** This section states additional conditions and a series of results to facilitate the estimation of the standard error of the quantile of the uniqueness. Consider the following assumptions:

Assumption 5. The variables Y_{id} are independent across i with conditional distribution functions F_{id}, differentiable conditional densities, $0 < f_{id} < \infty$, with bounded derivatives f_{id}' at the conditional quantiles for $i = 1, ..., n_d$.

Assumption 6. There exists positive constants Δ_0 and Δ_1 such that $E(|\epsilon_d^4|^{1+\Delta_0}) < \Delta_1$ and $E(|W_{1j}W_{1k}W_{1l}W_{1m}|^{1+\Delta_0}) < \Delta_1$ for $j, k, l, m = 1, \ldots, r$. In addition, we have that $\inf_i \xi_{\min}(E[W_{2i}'\boldsymbol{M}W_{2i}]) > 0$ and $\inf_i \xi_{\min}(E[\tilde{W}_{2i}'\boldsymbol{M}\tilde{W}_{2i}]) > 0$ where $\xi_{\min}$ is the smallest eigenvalue.

Assumption 7. For all $\tau_d \in (0, 1)$, $(\alpha(\tau_d), \boldsymbol{\beta}(\tau_d)) \in \operatorname{int} \mathcal{A}\times\mathcal{B}$, where $\mathcal{A}\times\mathcal{B}$ is compact and convex.

Assumption 8. Let $\mathbf{\Pi}(\cdot, \tau_d) \equiv E(\psi_{\tau_d}(\cdot)(\boldsymbol{X}', \boldsymbol{W}'_{(2)}))$ where $\psi_{\tau_d} = \tau_d - I(u \leq 0)$, and let $\boldsymbol{\pi}_0 = (\alpha, \boldsymbol{\beta}', \boldsymbol{\eta}')'$ and $\boldsymbol{\pi}_1 = (\alpha, \boldsymbol{\beta}')'$. The Jacobian matrices $\nabla\mathbf{\Pi}(\boldsymbol{\pi}_0, \tau_d)$ and $\nabla\mathbf{\Pi}(\boldsymbol{\pi}_1, \tau_d)$ have full rank and are continuous.

Assumption 9. There exist limiting positive definite matrices $\boldsymbol{S}(\tau_d)$ and $\boldsymbol{J}(\tau_d)$:

$$\boldsymbol{J}(\tau_d) = \lim_{n_d \to \infty} [\boldsymbol{W}_2, \boldsymbol{X}]' \boldsymbol{\Psi} [\boldsymbol{W}_1, \boldsymbol{X}]; \quad \boldsymbol{S}(\tau_d) = \lim_{n_d \to \infty} [\boldsymbol{W}_2, \boldsymbol{X}]' [\boldsymbol{W}_2, \boldsymbol{X}],$$

where $\boldsymbol{\Psi}$ is a diagonal matrix with typical element $f_{id}(\xi_{id}(\tau_d))$ and $\xi_{id}(\tau_d)$ is the τ_d-th conditional quantile function.

Assumption 10. $\max \|\boldsymbol{z}\|/\sqrt{n_d} \to 0$, for $\boldsymbol{z} = \{\boldsymbol{W}_1, \boldsymbol{X}, \boldsymbol{W}_2\}$.

We employ conditions that are standard in the literature of instrumental variable quantile regression. The first part of Condition 5 is standard and the second part of the assumption ensures a well-defined asymptotic behavior of the quantile regression estimator. The behavior of the conditional density in a neighborhood of the conditional quantile function is crucial for the asymptotic behavior of the quantile regression estimator. Assumption 6 is important for the existence, convergence in probability and law of the two-stage estimators previously defined, and consequently, it is needed for the behavior of the quantile regression estimator. Assumptions 7 and 8 are similar to Conditions R2 and R3 in Chernozhukov and Hansen (2006). Condition 9 leads to asymptotic covariance matrices that have a representation similar to the matrices found in Chernozhukov and Hansen (2006). Condition 10 is also standard and ensures the finite dimensional convergence of the objective function.

Theorem 3. *Under Assumptions 1-10, the estimators defined in (5.3) and (5.6),* $(\hat{\boldsymbol{\lambda}}'_d, \hat{\boldsymbol{\gamma}}_d)$*, are consistent and asymptotically normally distributed. Moreover, under these conditions and provided that* $\gamma_{dl} \neq 0$ *for at least one* $l = 1, \ldots, r$*, the quantile regression estimator defined in* (5.9)*,* $(\hat{Q}_{U_d}(\tau_d), \hat{\boldsymbol{\beta}}(\tau_d)')$*, is consistent and asymptotically normal with mean zero and covariance matrix* $\boldsymbol{J}(\tau_d)^{-1}\boldsymbol{S}(\tau_d)\boldsymbol{J}(\tau_d)^{-1}$*.*

Theorem 3 implies that the components of the asymptotic covariance matrix, $\boldsymbol{J}(\tau_d)$ and $\boldsymbol{S}(\tau_d)$, can be estimated using standard sample counterparts. For instance,

$$\hat{\boldsymbol{S}}(\tau_d) = \tau_d(1 - \tau_d)\frac{1}{n_d}\sum_{i=1}^{n_d} \boldsymbol{Z}_i\boldsymbol{Z}'_i, \tag{5.11}$$

$$\hat{\boldsymbol{J}}(\tau_d) = \frac{1}{2n_d h_{n_d}}\sum_{i=1}^{n_d} I(|\hat{u}_i(\tau_d)| \leq h_{n_d})\boldsymbol{D}_i\boldsymbol{Z}'_i, \tag{5.12}$$

where $\boldsymbol{Z}_i = [\boldsymbol{W}_{2i}, \boldsymbol{X}_i]$, $\boldsymbol{D}_i = [\boldsymbol{W}_{1i}, \boldsymbol{X}_i]$, $\hat{u}_i(\tau_d) = y_{id} - \hat{\boldsymbol{\lambda}}_d' \boldsymbol{W}_{1i} - \boldsymbol{X}_i' \hat{\boldsymbol{\beta}}(\tau_d)$, $\hat{\boldsymbol{\lambda}}_d(\tau_d) = \hat{\boldsymbol{\lambda}}_d + \hat{\gamma}_d \hat{Q}_{U_d}(\tau_d)$, and h_{n_d} is a properly chosen bandwidth (see Koenker 2005 for additional details). Note that the classical delta method and equations (5.11) and (5.12) lead to inference for the parameters of interest associated with the quantile specific loadings.

Inference for the QDGs and its partial derivatives requires a different strategy. We follow closely Jin, Ying and Wei (2001) and Ma and Kosorok (2005) proposing to use a re-sampling approach designed to perturb the objective function. This method maintains the endogenous structure of the model and can be easily implemented by drawing independent variables from an exponential distribution with mean 1 and variance 1. We first obtain a draw from the specified distribution. Then, we perturb the objective functions associated with (5.3) and (5.6), and also (5.9) to obtain $(\boldsymbol{\lambda}_{db}^*, \boldsymbol{\gamma}_{db}^*, Q_{U_d}^*(\tau_d))$. Using these estimates, we obtain $\boldsymbol{\lambda}_{1b}^*(\tau_1) - \boldsymbol{\lambda}_{0b}^*(\tau_0)$ and repeat the procedure B times. The sample variance of (3.2) can be obtained from $\{\boldsymbol{\lambda}_{1b}^*(\tau_1) - \boldsymbol{\lambda}_{0b}^*(\tau_0)\}_{b=1}^B$. Moreover, conditional on $\boldsymbol{\theta}$, the sample variance and $100(1-2q)$ confidence interval of the QDG as in (3.1) can be obtained by constructing the q-th quantile and $(1-q)$-th quantile of $\{QDG_b^*(\tau_1; \boldsymbol{\theta}, Q_{U_d}^*(\tau_d))\}_{b=1}^B$.

6. Monte Carlo evidence

This section reports results from several simulation experiments designed to evaluate the performance of the method in finite samples. We generate the dependent variable considering the following basic equations:

$$W_{ik} = \sum_{l=1}^{2} \delta_{kl}\Theta_{il} + (1 + \sum_{l=1}^{2} \xi_{kl}\Theta_{il})V_{ik}, \quad k = 1, \ldots, 4; \tag{6.1}$$

$$Y_{id} = \beta_0 + \beta_1 X_i + \sum_{l=1}^{2} \lambda_{dl}\Theta_{il} + (1 + \sum_{l=1}^{2} \gamma_{dl}\Theta_{il})U_{id}, \quad d = 0, 1; \tag{6.2}$$

$$Y_i = D_i Y_{1i} + (1 - D_i) Y_{0i}, \tag{6.3}$$

where $X_i \sim \chi_3^2$, $D_i \sim \mathcal{B}(1, 0.5)$, and $\boldsymbol{V} \sim \mathcal{N}(0, 0.5\boldsymbol{I}_4)$. Motivated by the empirical application considered in the next section, the parameter values for the factor loadings are assumed to be equal to $\boldsymbol{\lambda}_0 = (0.7, 0.5)$ and $\boldsymbol{\lambda}_1 = (0.9, 0.8)$, and the parameter values for the scale parameter are equal to $\boldsymbol{\gamma}_0 = (0.2, 0.1)$ and $\boldsymbol{\gamma}_1 = (0.25, 0.15)$. We assume that $\beta_0 = \beta_1 = 0$. Moreover, we adopt the normalization $\boldsymbol{L}_{(1)} = \boldsymbol{I}_2$ and

$\boldsymbol{G}_{(1)} = \mathbf{0}_2$, and then we set

$$\boldsymbol{L}' = \begin{bmatrix} 1 & 0 & 1 & 0.5 \\ 0 & 1 & 0.5 & 1 \end{bmatrix}', \ \boldsymbol{G}' = \begin{bmatrix} 0 & 0 & 0.1 & 0.1 \\ 0 & 0 & 0.1 & 0.1 \end{bmatrix}'.$$

The evidence presented in this section is based on 1000 randomly generated samples, three sample sizes equal to $n_d = \{2500, 5000, 10000\}$ and the following three designs:

Design 1: We assume that $\boldsymbol{\Theta} = (\Theta_1, \Theta_2)'$ and $\boldsymbol{U} = (U_0, U_1)'$ are i.i.d. $\mathcal{N}(\mathbf{0}, \boldsymbol{I})$.

Design 2: To examine the robustness to departures from Gaussian conditions, we maintain Design 1 but we change the distribution of the factors and uniqueness. We assume that $\boldsymbol{\Theta} = (\Theta_1, \Theta_2)'$ and $\boldsymbol{U} = (U_0, U_1)'$ are i.i.d. t-student distribution with 10 degrees of freedom, t_{10}.

Design 3: To examine the robustness to the assumption on the independence of the factors Θ_l, we maintain Design 1 but we allow Θ_1 to be correlated with Θ_2. The correlation coefficient is equal to 0.2.

Table 6.1 presents the bias and root mean square error (RMSE) for the factor loadings and scale parameters. The first columns correspond to the factor loadings and scale parameters of the equation for Y_1 and the last columns correspond to the factor loadings and scale parameters of the equation for Y_0. The top panel presents the bias and RMSE in the Gaussian case (Design 1), the middle panel shows results in the t_{10} case (Design 2), and finally the bottom panel presents results in the case of correlated factors (Design 3).

While it is not surprising to see that the estimator for the factor loading is unbiased, it is interesting to find out that the method offers, in general, an excellent performance when the scale parameters are estimated with a moderate sample size of 2500 observations. We also find that, as expected, the estimation of scale parameters improves dramatically as the sample size increases. The method proposed in this paper continues to perform well when we depart from Gaussian conditions (Design 2) and when the factors are correlated (Design 3).

Table 6.2 present the bias (in absolute value) and RMSE for the quantile of the uniqueness. We selected three quantiles, $\tau = \{0.10, 0.50, 0.75\}$, and we present results when $\boldsymbol{\Theta}$ and $\boldsymbol{U}$ are i.i.d Gaussian variables as in Design 1. Because the results in Table 6.1 do not seem to vary across designs, we restrict attention to only three quantiles and one design.

	Factor loadings and scale parameters:								
	n_d	λ_{11}	λ_{12}	γ_{11}	γ_{12}	λ_{01}	λ_{02}	γ_{01}	γ_{02}
					Design 1				
Bias	2500	0.004	-0.002	-0.037	-0.013	0.000	0.000	-0.003	-0.009
RMSE	2500	0.042	0.041	0.645	0.310	0.038	0.036	2.289	0.327
Bias	5000	-0.001	0.000	-0.016	-0.004	-0.001	0.000	-0.007	0.004
RMSE	5000	0.029	0.029	0.376	0.217	0.026	0.026	0.430	0.200
Bias	10000	0.000	0.000	0.000	0.000	0.000	-0.001	-0.025	-0.007
RMSE	10000	0.020	0.020	0.262	0.154	0.018	0.018	0.295	0.145
					Design 2				
Bias	2500	0.000	0.002	-0.005	0.011	0.000	-0.001	0.167	-0.002
RMSE	2500	0.038	0.037	0.496	0.269	0.034	0.033	6.929	0.266
Bias	5000	0.000	0.000	-0.015	-0.001	0.000	0.000	-0.036	-0.007
RMSE	5000	0.026	0.028	0.325	0.185	0.023	0.025	0.424	0.190
Bias	10000	0.000	-0.001	-0.002	0.001	0.000	0.000	0.008	0.008
RMSE	10000	0.019	0.018	0.211	0.127	0.017	0.017	0.274	0.135
					Design 3				
Bias	2500	0.004	-0.002	-0.052	-0.014	0.000	0.000	-0.184	-0.009
RMSE	2500	0.045	0.043	0.839	0.371	0.040	0.038	2.922	0.390
Bias	5000	-0.001	0.000	-0.016	-0.002	-0.001	0.000	-0.004	0.006
RMSE	5000	0.031	0.031	0.454	0.256	0.028	0.027	0.522	0.238
Bias	10000	0.000	0.000	0.003	0.001	0.000	-0.001	-0.030	-0.008
RMSE	10000	0.022	0.021	0.312	0.182	0.019	0.019	0.352	0.171

TABLE 6.1. *Small sample performance of Instrumental Variable estimators.*

We find that the quantile estimator performs well at the lower tail of the conditional distribution of the counterfactual response variable, with small biases that range from 5.6 percent to 8.8 percent (Table 6.2). As expected, the estimator continues to perform well at the 0.10 and 0.75 quantiles, with significant improvements in terms of bias and RMSE.

Lastly, Table 6.3 shows the finite sample performance of the proposed method in comparison with other candidate methods. We consider the following quantile regression estimators: (1) The quantile regression (QR) estimator for cross-sectional

		Quantiles of Uniqueness:					
		$\tau_d = 0.10$		$\tau_d = 0.50$		$\tau_d = 0.75$	
	n_d	Q_{U_1}	Q_{U_0}	Q_{U_1}	Q_{U_0}	Q_{U_1}	Q_{U_0}
Bias	2500	-0.073	0.004	-0.032	-0.016	-0.025	-0.057
RMSE	2500	0.108	0.072	0.067	0.051	0.079	0.090
Bias	5000	-0.106	-0.036	-0.040	-0.023	-0.001	-0.029
RMSE	5000	0.122	0.070	0.057	0.046	0.054	0.056
Bias	10000	-0.109	-0.046	-0.055	-0.037	0.015	-0.014
RMSE	10000	0.120	0.066	0.061	0.045	0.044	0.035

TABLE 6.2. *Small sample performance of Instrumental Variable estimators.*

data; (2) The instrumental variable quantile estimator proposed by Chernozhukov and Hansen (2006) which is denoted by IVQR; (3) The proposed approach in this paper which is labeled QDG. The upper block of the table presents results for the first quantile-specific factor loading and the bottom part of the table shows results for the second quantile-specific factor loading. By simple inspection of the performance of the method to estimate the factor loadings, Table 6.3 is informative of the bias in the estimation of the QDG and its derivative with respect to Θ_1 and Θ_2. This is naturally important since it might illustrate potential biases in empirical applications.

It is not surprising perhaps to find out that the QR method is biased and has poor MSE performance. On the other hand, the application of the IVQR method offers a considerable improvement in terms of both bias and RMSE, since $(W_{11}, W_{12}$ is a vector of endogenous variables. Note however that IVQR ameliorates the biases, but it does not completely eliminate them even in large samples. The proposed approach has biases that are closed to zero and low RMSE. Our estimator offers the best finite sample performance and continues to perform very well at the quantile 0.1, only with a minor increase in MSE.

7. Empirical application(s)

This section provides an empirical application using administrative data collected for the evaluation of an income support programme in Italy. The eligible population consists of employees affected by collective redundancy, provided that they have completed one year of continuous, regular employment at firm. The programme has

Statistic	n_d	τ	QR	IVQR	QDG	QR	IVQR	QDG
				Quantile Regression Methods				
				$\lambda_{11}(\tau_1)$			$\lambda_{01}(\tau_0)$	
Bias	2500	0.10	-0.148	0.060	-0.018	-0.129	0.020	0.001
RMSE	2500	0.10	0.151	0.088	0.027	0.132	0.065	0.014
Bias	5000	0.10	-0.150	0.058	-0.027	-0.128	0.025	-0.007
RMSE	5000	0.10	0.151	0.073	0.030	0.129	0.045	0.014
Bias	10000	0.10	-0.149	0.058	-0.027	-0.128	0.027	-0.009
RMSE	10000	0.10	0.149	0.066	0.030	0.129	0.039	0.013
Bias	2500	0.50	-0.296	0.005	-0.008	-0.231	0.002	-0.003
RMSE	2500	0.50	0.297	0.049	0.017	0.232	0.045	0.010
Bias	5000	0.50	-0.298	0.003	-0.010	-0.229	0.004	-0.005
RMSE	5000	0.50	0.298	0.035	0.014	0.230	0.029	0.009
Bias	10000	0.50	-0.298	0.003	-0.014	-0.231	0.006	-0.007
RMSE	10000	0.50	0.298	0.026	0.015	0.231	0.023	0.009
Bias	2500	0.75	-0.378	-0.047	-0.006	-0.290	-0.025	-0.011
RMSE	2500	0.75	0.379	0.072	0.020	0.291	0.055	0.018
Bias	5000	0.75	-0.380	-0.049	0.000	-0.288	-0.024	-0.006
RMSE	5000	0.75	0.380	0.062	0.013	0.288	0.043	0.011
Bias	10000	0.75	-0.380	-0.047	0.004	-0.288	-0.021	-0.003
RMSE	10000	0.75	0.380	0.055	0.011	0.289	0.032	0.007
				$\lambda_{12}(\tau_1)$			$\lambda_{02}(\tau_0)$	
Bias	2500	0.10	-0.177	0.037	-0.011	-0.115	0.013	0.000
RMSE	2500	0.10	0.179	0.072	0.016	0.118	0.058	0.007
Bias	5000	0.10	-0.177	0.040	-0.016	-0.114	0.012	-0.004
RMSE	5000	0.10	0.178	0.061	0.018	0.115	0.040	0.007
Bias	10000	0.10	-0.176	0.044	-0.016	-0.114	0.011	-0.005
RMSE	10000	0.10	0.176	0.054	0.018	0.115	0.031	0.007
Bias	2500	0.50	-0.266	-0.008	-0.005	-0.165	-0.007	-0.002
RMSE	2500	0.50	0.267	0.052	0.010	0.167	0.044	0.005
Bias	5000	0.50	-0.265	-0.004	-0.006	-0.164	-0.004	-0.002
RMSE	5000	0.50	0.265	0.036	0.008	0.165	0.030	0.005
Bias	10000	0.50	-0.266	-0.006	-0.008	-0.165	-0.005	-0.004
RMSE	10000	0.50	0.266	0.027	0.009	0.165	0.022	0.005
Bias	2500	0.75	-0.314	-0.046	-0.004	-0.194	-0.023	-0.006
RMSE	2500	0.75	0.315	0.075	0.012	0.196	0.052	0.009
Bias	5000	0.75	-0.314	-0.039	0.000	-0.194	-0.021	-0.003
RMSE	5000	0.75	0.314	0.055	0.008	0.195	0.042	0.006
Bias	10000	0.75	-0.314	-0.043	0.002	-0.194	-0.021	-0.001
RMSE	10000	0.75	0.315	0.052	0.007	0.194	0.033	0.004

TABLE 6.3. *Small sample performance of quantile estimators.*

both passive and active components. A redundancy pay is offered only to individuals dismissed by firms with more than 15 employees. In addition, a substantial reduction in labour costs is offered to firms hiring individuals of the eligible population, whether or not they are on benefits. Crucial to our analysis, the duration of the two components varies with worker age at dismissal. With our application in mind, we consider only two age groups. Workers aged 40 or below are eligible for at most one year, the redundancy pay being conditional on unemployment; the eligibility period is extended to two years for workers older than 40. The institutional details of the programme are discussed at large in Paggiaro *et al.* (2009).

We use social-security records for all eligible individuals in the years 1995-1998. Due to data quality problems, our analysis is limited to selected areas of Northern Italy. This admittedly limits the external validity of our conclusions, as the region considered is characterized by per-capita GDP some 15 percent higher than the national average and male unemployment rate of 6.1 percent in 1998 (compared to 11.4 in the country). Our sample consists of $2,223$ males aged between 35 and 45. Summary statistics are reported in Table 7.1.

The following equations for earnings are considered (see for example Meghir and Pistaferri, 2004):

$$w_{it} = f_i + p_{it} + v_{it},$$

where f_i is an individual specific effect, p_{it} is a permanent component of income and v_{it} is the transitory shock. The time index t denotes years from entrance in the programme, with negative values indexing pre-programme years. We follow Heckman and Cunha (2007) and use the following parametrization:

$$p_{it} = \lambda_t \theta_i,$$

where θ_i is individual specific endowment and λ_t is a time varying coefficient constant across individuals. This specification implies that innovations to the permanent income process can be written as:

$$p_{it} - p_{it-1} = (\lambda_t - \lambda_{t-1})\theta_i.$$

Thus the variance of permanent shocks depends on $(\lambda_t - \lambda_{t-1})$, the case of constant coefficients ruling out innovations to the process. Our application considers two alternative specifications depending on how individual specific effects are modeled.

In the first specification, we impose that f_i is totally spanned by a set of exogenous regressors $\mathbf{X}$, so that there is $f_i = \beta x_i$. This set includes a cubic polynomial in age, a

	All Workers	With Benefits	Without Benefits
	Individuals aged $35 - 40$		
Age	36.929	37.021	36.830
	(1.388)	(1.369)	(1.401)
Made redundant in 1995	0.180	0.198	0.160
	(0.384)	(0.399)	(0.367)
Made redundant in 1996	0.265	0.262	0.268
	(0.441)	(0.440)	(0.443)
Made redundant in 1997	0.317	0.320	0.314
	(0.465)	(0.466)	(0.464)
Made redundant in 1998	0.239	0.221	0.258
	(0.426)	(0.415)	(0.438)
Earnings 3 years after	801.823	838.319	762.774
	(319.471)	(341.562)	(289.280)
Earnings 1 year before	768.627	813.778	720.316
	(256.119)	(252.462)	(251.385)
Earnings 2 years before	766.613	807.305	723.073
	(243.422)	(243.973)	(235.396)
Number of Observations	1,036	535	501
	Individuals aged $41 - 45$		
Age	42.492	42.595	42.377
	(1.709)	(1.715)	(1.695)
Made redundant in 1995	0.203	0.229	0.174
	(0.402)	(0.420)	(0.379)
Made redundant in 1996	0.243	0.270	0.212
	(0.429)	(0.444)	(0.409)
Made redundant in 1997	0.300	0.266	0.338
	(0.458)	(0.442)	(0.473)
Made redundant in 1998	0.254	0.235	0.276
	(0.436)	(0.424)	(0.447)
Earnings 3 years after	812.420	849.486	771.200
	(318.724)	(323.488)	(308.431)
Earnings 1 year before	799.862	842.473	752.475
	(265.767)	(268.269)	(254.951)
Earnings 2 years before	801.644	841.408	757.422
	(259.824)	(256.535)	(256.509)
Number of Observations	1,187	625	562

TABLE 7.1. *Summary statistics for the Income Support Program in Italy.*

dummy for redundancy pay recipients (i.e. firms over of below 15 at time of dismissal), and a full set of dummies for year of redundancy. Our identification strategy exploits the sharp discontinuity in duration of benefits around the age 40 cutoff. Bearing in mind the sample restrictions on the age range, the maintained assumption is that individuals aged 40 or below proxy the counterfactual scenario for individuals aged 41 or above. This allows us to investigate the effects of lengthening eligibility by one year for individuals around the discontinuity cutoff. In this setting, D represents a dummy for individuals aged 41 or above at the time of dismissal. We allow the variance of transitory shocks to vary across individuals through its dependence on the component θ_i:

$$v_{it} = (1 + \gamma_t \theta_i) u_{it}.$$

The outcome is (logged) annual earnings three years after entrance in the programme ($t = 3$), for which the corresponding equation is:

$$w_{i3} = \beta x_i + \lambda_3 \theta_i + (1 + \gamma_3 \theta_i) u_{i3}.$$

We let β and γ vary depending on age (above and below age 40). Note also that our procedure allows for groups specific variances of the transitory shock u_{i3}. Potential outcomes defined from w_{i3} for the $D = 0$ and the $D = 1$ groups represent Y_0 ad Y_1, respectively, in the notation employed above. The availability of a long pre-programme panel allows us to employ lagged values of the outcome as additional measurements. In particular, we set W_1 and W_2 to (logged) annual earnings XX and XX years before entrance in the programme, respectively. The time lag between the two measurements adds credibility to the assumption that temporary shocks in the earning equations are orthogonal.

The second specification controls for f_i by first differencing the model, and allows for the following form of heteroskedasticity:

$$\Delta v_{it} = (1 + \gamma_t \theta_i) \Delta u_{it}.$$

In this setting, the post-programme equation becomes:

$$w_{i3} - w_{i2} = (\lambda_3 - \lambda_2) \theta_i + (1 + \gamma_3 \theta_i)(u_{i3} - u_{i2}).$$

The additional measurements employed for W_1 and W_2 are as in the first specification.

Note that in this instance imposing the restriction of conditional indepedendence between the two potential outcomes amounts to say that the transitory shock hitting a specific individual in case she were treated is independent of the transitory shock

Group	Model 1		Model 2		Model 3	
	λ	γ	λ	γ	λ	γ
Policy off	0.723	0.162	0.749	0.156	-0.277	0.162
	(0.039)	(0.019)	(0.036)	(0.004)	(0.040)	(0.019)
Policy on	0.775	0.150	0.779	0.149	-0.225	0.150
	(0.034)	(0.002)	(0.033)	(0.001)	(0.033)	(0.002)

TABLE 7.2. *Instrumental variable results for the effect of the Income Support Program in Italy. Standard errors are in parentheses.*

Group	Quantile				
	0.1	0.25	0.50	0.75	0.90
	Model 1: Dependent variable Y				
Policy off	0.497	0.605	0.758	0.817	0.864
	(0.067)	(0.064)	(0.055)	(0.050)	(0.063)
Policy on	0.639	0.669	0.788	0.869	0.924
	(0.047)	(0.041)	(0.041)	(0.037)	(0.048)
	Model 3: Dependent variable $Y - W$				
Policy off	-0.502	-0.401	-0.236	-0.183	-0.136
	(0.066)	(0.065)	(0.054)	(0.050)	(0.062)
Policy on	-0.367	-0.328	-0.216	-0.131	-0.069
	(0.056)	(0.039)	(0.041)	(0.037)	(0.056)

TABLE 7.3. *Instrumental variable quantile regression results for the effect of the Income Support Program in Italy. Standard errors are in parentheses.*

hitting the same individual in case she were not treated. Something that might be hard to find a theoretical motivation for. As an alternative, one might consider the restriction (**??**) according to which the treatment adds a random component to the transitory shock hitting the individual no matter for her treatment status.

Table 7.2 shows results from three different models. In model 1, the values of λ_1 and β_1 are estimated from a 2SLS regression of Y on W_1 for the group aged 41 or above ($D = 1$), using W_2 as instrument. Individuals in the younger group are used to estimate λ_0 and β_0. In model 2, we use additional instruments W_3, W_4, W_5 and

W_6. Model 3 is the model for earnings in difference. In all the models, the coefficient a is positive as implied by the theory.

Table 7.3 and Figure 7.1 show results obtained from using Chernozhukov and Hansen (2006, 2008).

The 2SLS procedure yields estimates of γ and the possibility of implementing the approach suggested in Section 5.2. We therefore obtain estimates of λ, γ and $Q_U(\tau)$. To compare the results of this approach with the results shown in Figure 7.1, we offer a comparison of the estimates produced using Chernozhukov and Hansen (2006, 2008) and our approach. Results are shown in Figure 7.2.

The main results are apparent. First, the standard factor model is rejected since both γ_0 and γ_1 are by far statistically different from zero. Second, all the parameters in the model appear to be independent of the treatment status D, i.e. $\gamma_0 = \gamma_1$ and $\lambda_0 = \lambda_1$, $Q_{U_1}[\tau] = Q_{U_0}[\tau]$. That is one cannot reject the hypothesis that the distribution of Y_1 is equal to the distribution of Y_0. Unless one is willing to take into consideration the rather implausible hypothesis that the program had an impact just by permuting the outcomes across individuals, this is enough to conclude that the sharp null hypothesis of non causal effect to anyone is not rejected by the data. This is consistent with previous findings on average effects (Paggiaro *et al.*, 2009). Note that for this conclusion to hold it is irrelevant whether it is U_1 or $U_1 - U_0$ to be independent of U_0.

8. Conclusion

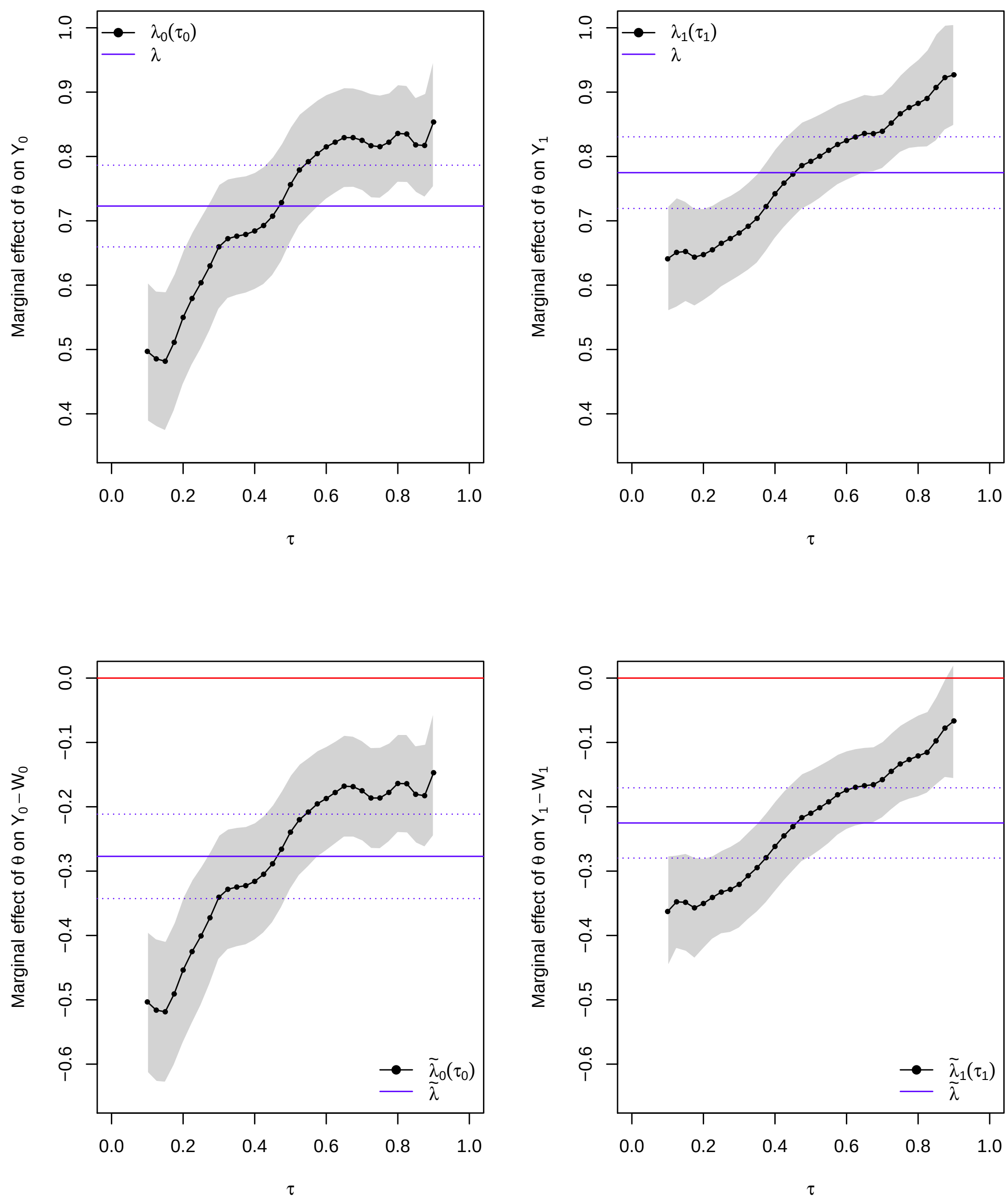

FIGURE 7.1. *Instrumental variable quantile regression results for the effect of the policy.*

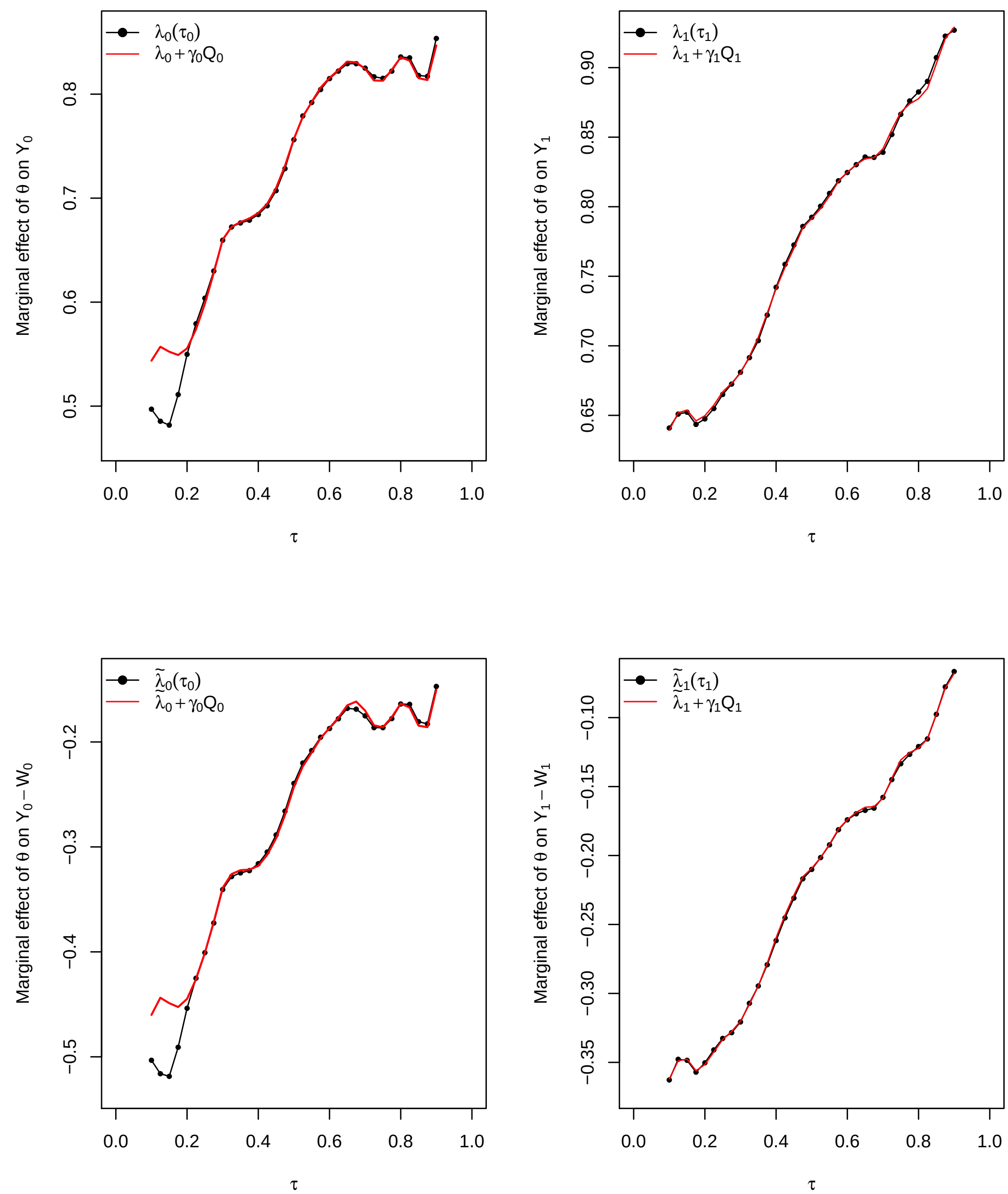

FIGURE 7.2. *A comparison of quantile regression results obtained by IQR and IVQR.*

References

Aakvik, A., J. J. Heckman, and E. J. Vytlacil (2005): "Estimating treatment effects for discrete outcomes when responses to treatment vary: an application to Norwegian vocational rehabilitation programs," *Journal of Econometrics*, 125, 15–51.

Abadie, A., J. Angrist, and G. Imbens (2002): "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings," *Econometrica*, 70(1), 91–117.

Abbring, J. H., and J. J. Heckman (2007): "Econometric Evaluation of Social Programs, Part III: Distributional Treatment Effects, Dynamic Treatment Effects, Dynamic Discrete Choice, and General Equilibrium Policy Evaluation," in *Handbook of Econometrics, Vol6, Part B*, chap. 72, pp. 5145–5303. North-Holland, Amsterdam.

Angrist, J. D., and M. Rokkanen (2013): "Wanna Get Away? RD Identification Away from the Cutoff," National Bureau of Economic Research Working Paper Series 18662.

Arcidiacono, P., E. Aucejo, H. Fang, and K. Spenner (2011): "Does Affirmative Action Lead to Mismatch? A New Test and Evidence," *Quantitative Economics*, 2(3), 303–333.

Athey, S., and G. W. Imbens (2006): "Identification and Inference in Nonlinear Difference-in-Differences Models," *Econometrica*, 74, 431–497.

Bitler, M. P., J. B. Gelbach, and H. W. Hoynes (2006): "What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments," *American Economic Review*, 96, 988–1012.

Bonhomme, S., and J.-M. Robin (2010): "Generalized Non-Parametric Deconvolution with an Application to Earnings Dynamics," *The Review of Economic Studies*, 77, 491–533.

Carneiro, P. M., K. Karsten, and J. J. Heckman (2003): "Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College," National Bureau of Economic Research Working Paper Series 9546.

Chernozhukov, V., and C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245–261.

——— (2006): "Instrumental Quantile Regression Inference for Structural and Treatment Effects Models," *Journal of Econometrics*, 132, 491–525.

——— (2008): "Instrumental Variable Quantile Regression: A Robust Inference Approach," *Journal of Econometrics*, 142, 379–398.

Cooley, J., S. Navarro, and Y. Takahashi (2011): "How the Timing of Grade Retention Affects Outcomes: Identification and Estimation of Time-Varying Treatment Effects," .

Cunha, F., and J. J. Heckman (2007): "Identifying and Estimating the Distributions of Ex Post and Ex Ante Returns to Schooling," *Labour Economics*, 14(6), 870–893.

DiNardo, J., and D. S. Lee (2011): "Program Evaluation and Research Designs," in *Handbook of Labor Economics, Chapter 5, Volume 4, Part A*, pp. 463–536.

Doksum, K. (1974): "Empirical Probability Plots and Statistical Inference for Nonlinear Models in the Two-Sample Case," *Annals of Statistics*, 2, 267–277.

Firpo, S. P. (2007): "Efficient Semiparametric Estimation of Quantile Treatment Effects," *Econometrica*, 75(1), 259–276.

FIRPO, S. P., S. PINHEIRO, AND G. RIDDER (2010): "Bounds on functionals of the distribution treatment effects," Textos para discussao 201, Escola de Economia de Sao Paulo, Getulio Vargas Foundation, Brazil.

HECKMAN, J., R. LALONDE, AND J. SMITH (1999): "The Economics and Econometrics of Active Labor Market Programs," in *Handbook of Labor Economics, Vol 3*, ed. by A. Ashenfelter, and D. Card. Amsterdam: Elsevier Science.

HECKMAN, J., J. STIXRUD, AND S. URZUA (2006): "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior," *Journal of Labor Economics*, 24(3), 411–482.

HECKMAN, J. J., H. F. LOPES, AND R. PIATEK (2014): "Treatment Effects: A Bayesian Perspective," *Econometric Reviews*, 33(1-4), 36–67.

HECKMAN, J. J., J. SMITH, AND N. CLEMENTS (1997): "Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts," *Review of Economic Studies*, 64, 487–535.

HECKMAN, J. J., AND J. VYTLACIL (2007): "Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation," in *Handbook of Econometrics, Vol6, Part B*, chap. 70, pp. 4779–4874. North-Holland, Amsterdam.

JACKSON, C. K. (2013): "Non-Cognitive Ability, Test Scores, and Teacher Quality: Evidence from 9th Grade Teachers in North Carolina," .

KOENKER, R. (2005): *Quantile Regression*. Cambridge University Press, New York, New York.

MANSKI, C. F. (1990): "Nonparametric Bounds on Treatment Effects," *American Economic Review, P&P*, 80(2), 319–323.

MEGHIR, C., AND L. PISTAFERRI (2004): "Income variance dynamics and heterogeneity," *Econometrica*, 72(1), 1–32.

PAGGIARO, A., E. RETTORE, AND U. TRIVELLATO (2009): "The Effect of a Longer Eligibility to a Labour Market Programme for Dismissed Workers," *Labour*, 23(1), 37–66.

ROKKANEN, M. (2013): "xXx," .

ROSENBAUM, P. R., AND D. B. RUBIN (1983): "The central role of the propensity score in observational studies for causal effects," *Biometrika*, 70(1), 41–55.

SCHENNACH, S. M. (2008): "Quantile Regression with Mismeasured Covariates," *Econometric Theory*, 24, 1010–1043.

WEI, Y., AND R. J. CARROLL (2009): "Quantile Regression With Measurement Error," *Journal of the American Statistical Association*, 104.

Appendix A. Proofs

In this Appendix, we use notation that is standard: $\rightsquigarrow$ denotes weak convergence, $\to$ denotes convergence in probability, o and O denote the usual order of magnitudes, CLT denotes ordinary Central Limit Theorem, $\tilde{y} = Y - \boldsymbol{\lambda}'\boldsymbol{W}_1$ and $\Phi = \boldsymbol{\gamma}'\boldsymbol{W}_1$.

Proof of Theorem 1. Use (4.1) to write:

$$E_Y(Y|d, \boldsymbol{w}_{(2)}) = E_{Y_d}(Y_d|d, \boldsymbol{w}_{(2)}) = \boldsymbol{\lambda}_d' E_{\boldsymbol{W}_{(1)}}(\boldsymbol{W}_{(1)}|d, \boldsymbol{w}_{(2)}).$$

Consider:

$$\begin{aligned}(Y_d - \boldsymbol{\lambda}_d'\boldsymbol{W}_{(1)})^2 &= (-\boldsymbol{\lambda}_d'\boldsymbol{V}_{(1)} + (1 + \boldsymbol{\gamma}_d'\boldsymbol{W}_{(1)} - \boldsymbol{\gamma}_d'\boldsymbol{V}_{(1)})U_d)^2,\\ &= (\boldsymbol{\lambda}_d'\boldsymbol{V}_{(1)})^2 + (1 + \boldsymbol{\gamma}_d'\boldsymbol{W}_{(1)} - \boldsymbol{\gamma}_d'\boldsymbol{V}_{(1)})^2 U_d^2\\ &- 2\boldsymbol{\lambda}_d'\boldsymbol{V}_{(1)}(1 + \boldsymbol{\gamma}_d'\boldsymbol{W}_{(1)} - \boldsymbol{\gamma}_d'\boldsymbol{V}_{(1)})U_d,\end{aligned}$$

where there is:

$$\begin{aligned}(1 + \boldsymbol{\gamma}_d'\boldsymbol{W}_{(1)} - \boldsymbol{\gamma}_d'\boldsymbol{V}_{(1)})^2 &= 1 + \boldsymbol{\gamma}_d'\boldsymbol{W}_{(1)}\boldsymbol{W}_{(1)}'\boldsymbol{\gamma}_d + \boldsymbol{\gamma}_d'\boldsymbol{V}_{(1)}\boldsymbol{V}_{(1)}'\boldsymbol{\gamma}_d\\ &+ 2\boldsymbol{\gamma}_d'\boldsymbol{W}_{(1)} - 2\boldsymbol{\gamma}_d'\boldsymbol{V}_{(1)} - 2\boldsymbol{\gamma}_d'\boldsymbol{W}_{(1)}\boldsymbol{V}_{(1)}'\boldsymbol{\gamma}_d.\end{aligned}$$

The conditioning on $\boldsymbol{W}_{(2)}$ yields the following terms:

$$\begin{aligned}E_{\dot{Y}}(\dot{Y}|d, \boldsymbol{w}_{(2)}) &= \boldsymbol{\lambda}_d'\Omega_{\boldsymbol{V}_{(1)}}\boldsymbol{\lambda}_d + \sigma_{U_d}^2 E[(1 + \boldsymbol{\gamma}_d'\boldsymbol{W}_{(1)} - \boldsymbol{\gamma}_d'\boldsymbol{V}_{(1)})^2|d, \boldsymbol{w}_{(2)}],\\ &= \boldsymbol{\lambda}_d'\Omega_{\boldsymbol{V}_{(1)}}\boldsymbol{\lambda}_d + \sigma_{U_d}^2(1 + \boldsymbol{\gamma}_d'\Omega_{\boldsymbol{V}_{(1)}}\boldsymbol{\gamma}_d)\\ &+ \sigma_{U_d}^2\boldsymbol{\gamma}_d' E(\boldsymbol{W}_{(1)}\boldsymbol{W}_{(1)}'|d, \boldsymbol{w}_{(2)})\boldsymbol{\gamma}_d\\ &+ 2\sigma_{U_d}^2\boldsymbol{\gamma}_d' E(\boldsymbol{W}_{(1)}|d, \boldsymbol{w}_{(2)}) - 2\sigma_{U_d}^2\boldsymbol{\gamma}_d' E(\boldsymbol{W}_{(1)}\boldsymbol{V}_{(1)}'|d, \boldsymbol{w}_{(2)})\boldsymbol{\gamma}_d,\end{aligned}$$

where $\Omega_{\boldsymbol{A}}$ is the covariance matrix of $\boldsymbol{A}$. Use (2.8) to write:

$$E(\boldsymbol{W}_{(1)}\boldsymbol{V}_{(1)}'|d, \boldsymbol{w}_{(2)}) = \Omega_{\boldsymbol{V}_{(1)}},$$

so that there is:

$$\begin{aligned}E_{\dot{Y}}(\dot{Y}|d, \boldsymbol{w}_{(2)}) &= \boldsymbol{\lambda}_d'\Omega_{\boldsymbol{V}_{(1)}}\boldsymbol{\lambda}_d + \sigma_{U_d}^2(1 - \boldsymbol{\gamma}_d'\Omega_{\boldsymbol{V}_{(1)}}\boldsymbol{\gamma}_d)\\ &+ \sigma_{U_d}^2\boldsymbol{\gamma}_d' E(\boldsymbol{W}_{(1)}\boldsymbol{W}_{(1)}'|d, \boldsymbol{w}_{(2)})\boldsymbol{\gamma}_d\\ &+ 2\sigma_{U_d}^2\boldsymbol{\gamma}_d' E(\boldsymbol{W}_{(1)}|d, \boldsymbol{w}_{(2)}).\end{aligned}$$

By re-arranging terms, it follows that the coefficients on W_l^2 identify $a_l = \sigma_{U_d}^2\gamma_{dl}^2$, the coefficients on W_l identify $b_l = 2\sigma_{U_d}^2\gamma_{dl}$, and the coefficients on W_lW_s identify $c_{ls} = 2\sigma_{U_d}^2\gamma_{dl}\gamma_{ds}$ for $l = 1, \ldots, r$ and $s = 1, \ldots, r$. □

Proof of Theorem 2. We show the result for the conditional quantile of $Y = DY_1 + (1-D)Y_0$ for $d = \{0,1\}$. Consider:

$$\begin{aligned} P(Y \leq q_d(\boldsymbol{\theta}, \tau_d)|\boldsymbol{\theta}) &\stackrel{1}{=} P(q_d(\boldsymbol{\theta}, U_d) \leq q_d(\boldsymbol{\theta}, \tau_d)|\boldsymbol{\theta}) \stackrel{2}{=} P(U_d \leq Q_{U_d|\boldsymbol{\theta}}(\tau_d)|\boldsymbol{\theta}) \\ &\stackrel{3}{=} P(U_d \leq Q_{U_d|\boldsymbol{W}_{(1)},\boldsymbol{V}_{(1)}}(\tau_d)|\boldsymbol{W}_{(2)}, \boldsymbol{V}_{(2)}) \\ &\stackrel{4}{=} P(U_d \leq Q_{U_d|\boldsymbol{W}_{(1)}}(\tau_d)|\boldsymbol{W}_{(2)}, \boldsymbol{V}_{(2)}) \\ &\stackrel{5}{=} \int \ldots \int P(U_d \leq Q_{U_d|\boldsymbol{W}_{(1)}}(\tau_d)|\boldsymbol{W}_{(2)}, \boldsymbol{V}_{(2)}) dP(\boldsymbol{V}_{(2)}|\boldsymbol{W}_{(2)}) \\ &\stackrel{6}{=} P(U_d \leq Q_{U_d|\boldsymbol{W}_{(1)}}(\tau_d)|\boldsymbol{W}_{(2)}) \\ &\stackrel{7}{=} F_{U_d|\boldsymbol{W}_{(2)}}\left(Q_{U_d|\boldsymbol{W}_{(1)}}(\tau_d)\right) = \tau_d. \end{aligned}$$

The first two equalities follow by definition and Assumption 4. The third equality follows by definition and by the fact that $\boldsymbol{\theta}$ is a function of $(\boldsymbol{W}_{(1)}, \boldsymbol{V}_{(1)})$ as well as $(\boldsymbol{W}_{(2)}, \boldsymbol{V}_{(2)})$. The fourth equality follows from independence between U_d and $\boldsymbol{V}_{(1)}$ stated in Assumption 1.ii. Moreover this independence assumption between U_d and $\boldsymbol{V}_{(2)}$ is used to obtain $P(U_d \leq Q_{U_d|\boldsymbol{W}_{(1)}}(\tau_d)|\boldsymbol{W}_{(2)})$. The last equality holds by definition of conditional quantiles and $U_d|\boldsymbol{W}_{(2)}$ and $U_d|\boldsymbol{W}_{(1)}$ equals in distribution. By definition, we also have $P(U_d \leq Q_{U_d|\boldsymbol{W}_{(1)}}(\tau_d)|\boldsymbol{W}_{(2)}) = P(Y \leq q_d(\boldsymbol{W}_{(1)}, \tau_d)|\boldsymbol{W}_{(2)})$ and the result follows. □

Proof of Corollary 1. Let $\tilde{Y} := Y - q_d(\boldsymbol{W}_{(1)}, \boldsymbol{\lambda}_d, \boldsymbol{\gamma}_d, \tau_d)$, where $q_d(\boldsymbol{W}_{(1)}, \boldsymbol{\lambda}_d, \boldsymbol{\gamma}_d, \tau_d) = \boldsymbol{\lambda}_d'\boldsymbol{W}_{(1)} + (1 + \boldsymbol{\gamma}_d'\boldsymbol{W}_{(1)})Q_{U_d}(\tau_d)$. By Theorem 2, $P(Y \leq q_d(\boldsymbol{W}_{(1)}, \boldsymbol{\lambda}_d, \boldsymbol{\gamma}_d, \tau_d)|\boldsymbol{W}_{(2)}) = \tau_d$ which implies that $P(\tilde{Y}_d \leq 0|\boldsymbol{W}_{(2)}) = \tau_d$ as $0 = Q_{\tilde{Y}_d}(\tau_d|\boldsymbol{W}_{(2)})$ for all $\tau_d \in (0,1)$. Letting $\mathcal{S}(\boldsymbol{\lambda}_d, \boldsymbol{\gamma}_d, \tau_d, \boldsymbol{\eta}) = \mathbb{E}\rho_{\tau_d}(\tilde{Y} - \boldsymbol{W}_{(2)}'\boldsymbol{\eta})$, we have that $\boldsymbol{\eta}(\boldsymbol{\lambda}_d, \boldsymbol{\gamma}_d, \tau_d) = \text{argmin}_\eta\{\mathcal{S}(\boldsymbol{\lambda}_d, \boldsymbol{\gamma}_d, \tau_d, \boldsymbol{\eta})\}$ and therefore $Q_{U_d}(\tau_d)$ is the solution of:

$$\min\{\boldsymbol{\eta}(\boldsymbol{\lambda}_d, \boldsymbol{\gamma}_d, \tau_d)'\boldsymbol{A}\boldsymbol{\eta}(\boldsymbol{\lambda}_d, \boldsymbol{\gamma}_d, \tau_d)\}$$

for a given positive definite matrix $\boldsymbol{A}$. The continuity of derivatives implied by the factor model and monotonicity conditions of the quantile model give the second result.

The second result follows immediately from Theorem 1 and Theorem 2. By Theorem 1, $\boldsymbol{\lambda}_d$ and $\boldsymbol{\gamma}_d$ are identified for $d = \{0,1\}$. By Theorem 2 and the first part of the proof, the quantiles of the uniqueness are identified. Then $\boldsymbol{\lambda}_d(\tau_d) = \boldsymbol{\lambda}_d + \boldsymbol{\gamma}_d Q_{U_d}(\tau_d)$ gives identification of the quantile specific loadings. □

Proof of Corollary 2. Since $U_1 = U_0 + (U_1 - U_0)$, orthogonality between U_0 and U_1 is replaced by orthogonality between U_0 and $U_1 - U_0$ Then the following representation for the quantile function of the treatment effect follows:

$$QDG[\tau_1; \boldsymbol{\theta}, Q_{U_0}(\tau_0)] = [\boldsymbol{\lambda}_1 - \boldsymbol{\lambda}_0]'\boldsymbol{\theta} + [\boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_0]'\boldsymbol{\theta} Q_{U_0}(\tau_0) + (1 + \boldsymbol{\gamma}_1'\boldsymbol{\theta}) Q_{U_1 - U_0}(\tau_1)$$

$Q_{U_1-U_0}(\tau)$ can be written as a functional of $Q_{U_0}(\tau)$ and $Q_{U_1}(\tau)$ through deconvolution so that it is identified at the condition given in the next section for the identification of $Q_{U_0}(\tau)$ and $Q_{U_1}(\tau)$. As in (3.1) if $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_0$ are zero $\boldsymbol{\Theta}$ enters the quantile function only as a location shifter, otherwise it affects the whole shape of the distribution of the causal effect. □

Proof of Theorem 3. The proof is completed in two steps. We first show the consistency of the quantile regression estimator and then asymptotic normality. The proof of consistency and asymptotic normality for the 2SLS estimators are standard under Assumptions 1, 3, and 6 and it is not discussed here (see, e.g., White 2000).

Under regularity conditions, consistency results follow immediately from Corollary 3.1 in Newey (1991), Corollary 3.2.3 in van der Vaart and Wellner (1996) and Proposition 2 in Chernozhuknov and Hansen (2008). Using standard arguments under conditions 1, 3, and 6, it is straightforward to show that $\hat{\boldsymbol{\lambda}}_d \to \boldsymbol{\lambda}_d$ and $\hat{\boldsymbol{\gamma}}_d \to \boldsymbol{\gamma}_d$. Moreover, consider $q_n(\boldsymbol{\lambda}_d) = n_d^{-1} \sum_{i=1}^{n_d} \boldsymbol{W}'_{1i)} \boldsymbol{\lambda}_d$ and $\bar{q}_n(\boldsymbol{\lambda}_d) = n_d^{-1} \sum_{i=1}^{n_d} E(\boldsymbol{W}'_{1i} \boldsymbol{\lambda}_d)$. We note that Assumption 1 and 2 of Newey (1991) are trivially satisfied since $\boldsymbol{\Lambda}$ is a compact set and there is point-wise convergence $q_n(\boldsymbol{\lambda}_d) - \bar{q}_n(\boldsymbol{\lambda}_d)$ for each $\lambda \in \boldsymbol{\Lambda}$. Moreover, Assumption 3A is also satisfied since for any $\tilde{\boldsymbol{\lambda}}_d$ and $\boldsymbol{\lambda}_d$ in $\boldsymbol{\Lambda}$,

$$|\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{W}'_{ij(1)} (\tilde{\boldsymbol{\lambda}}_j - \boldsymbol{\lambda}_j)| \leq |\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{W}'_{ij(1)}||\tilde{\boldsymbol{\lambda}}_j - \boldsymbol{\lambda}_j|, \tag{A.1}$$

and, consequently, $\bar{q}_n(\boldsymbol{\lambda}_j)$ is equicontinuous and $\sup_{\lambda \in \boldsymbol{\Lambda}} |q_n(\boldsymbol{\lambda}_j) - \bar{q}_n(\boldsymbol{\lambda}_j)| = o_p(1)$. By proposition 2 in Chernozhukov and Hansen (2008), we have that $\sup_{\alpha \in \mathcal{A}} \|\hat{\boldsymbol{\vartheta}}(\alpha, \cdot) - \boldsymbol{\vartheta}(\alpha, \cdot)\| \to 0$ for $\boldsymbol{\vartheta} = (\boldsymbol{\beta}', \boldsymbol{\eta}')'$. This implies that $\sup_{\alpha \in \mathcal{A}} \|\hat{\boldsymbol{\eta}}(\alpha, \cdot) - \boldsymbol{\eta}(\alpha, \cdot)\| \to 0$, and that $\|\hat{\boldsymbol{\alpha}}(\cdot) - \boldsymbol{\alpha}(\cdot)\| \to 0$. Consider a small ball $\boldsymbol{\alpha}_n$ of radius r_n centered at $\alpha(\tau)$. Then for any $\boldsymbol{\alpha}_n \to \boldsymbol{\alpha}(\tau)$, we have that $\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}_n, \cdot) \to \boldsymbol{\beta}(\boldsymbol{\alpha}(\tau), \cdot) = \boldsymbol{\beta}(\tau)$, and $\hat{\boldsymbol{\eta}}(\boldsymbol{\alpha}_n, \cdot) \to \boldsymbol{\eta}(\boldsymbol{\alpha}(\tau), \cdot) = \boldsymbol{\eta}(\tau) = 0$. Hence $\hat{\boldsymbol{\vartheta}}(\boldsymbol{\alpha}_n, \cdot) \to \boldsymbol{\vartheta}(\boldsymbol{\alpha}(\tau), \cdot)$ for any $\boldsymbol{\alpha}_n \to \boldsymbol{\alpha}(\tau)$.

Asymptotic normality follows by Andrews (1994, pp. 2263-65) and Theorem 3 in Chernozhuknov and Hansen (2008). Note that the asymptotic distribution of our quantile estimator depends on $q_n(\boldsymbol{\lambda}_d)$ evaluated at $\boldsymbol{\lambda}_d = \hat{\boldsymbol{\lambda}}_d$ and $q_n(\boldsymbol{\gamma}_d)$ evaluated at $\boldsymbol{\gamma}_d = \hat{\boldsymbol{\gamma}}_d$. Because the conditions of Corollary 3.1 in Newey (1991) are met, the asymptotic behavior of the estimator depends on $\boldsymbol{\lambda}_d$ and $\boldsymbol{\gamma}_d$ and then standard CLT results can be applied. The argument is similar to the one offered in Chernozhuknov and Hansen (2008) who obtain the asymptotic distribution of a quantile estimator using weights and transformation of instruments obtained in a first stage.

For any $\boldsymbol{\alpha}_n$, $C_{id} = \rho_{\tau_d}(\tilde{y}_{id} - \xi_{id}(\tau_d) - \Phi'_{id}\hat{\delta}_\alpha/\sqrt{n} - \boldsymbol{X}'_{id}\hat{\boldsymbol{\delta}}_\beta/\sqrt{n} - \boldsymbol{W}'_{2i)}\hat{\boldsymbol{\delta}}_\eta/\sqrt{n})$, where $\xi_{id}(\tau_d) = \alpha_d(\tau_d)\Phi_{id}+\boldsymbol{X}'_i\boldsymbol{\beta}(\tau_d)+\boldsymbol{w}'_{2i}\boldsymbol{\eta}(\tau_d)$, $\hat{\delta}_\alpha(\alpha_n,\tau_d) = \sqrt{n}(\hat{\alpha}(\alpha_n,\tau_d)-\alpha(\tau_d))$, $\hat{\boldsymbol{\delta}}_\beta(\alpha_n,\tau_d) = \sqrt{n}(\hat{\boldsymbol{\beta}}(\alpha_n,\tau_d) - \boldsymbol{\beta}(\tau_d))$, $\hat{\boldsymbol{\delta}}_\eta(\alpha_n,\tau_d) = \sqrt{n}\hat{\boldsymbol{\eta}}(\alpha_n,\tau_d)$. It follows that,

$$\text{(A.2)} \qquad \sup ||v(\delta_\alpha, \boldsymbol{\delta}_\vartheta) - v(0,\mathbf{0}) - \mathbb{E}(v(\delta_\alpha, \boldsymbol{\delta}_\vartheta) - v(0,\mathbf{0})|| = o_p(1)$$

where $||\cdot||$ denotes the standard Euclidean norm of a vector, $\psi_{\tau_d}(u) = \tau_d - I(u<0)$ and,

$$v(\delta_\alpha, \boldsymbol{\delta}_\vartheta) = \frac{-1}{\sqrt{n}}\sum_{i=1}^n \boldsymbol{H}_i\psi_{\tau_j}\left(\tilde{y}_{ij} - \xi_{ij}(\tau_j) - \delta_\alpha/\sqrt{n}\Phi_{ij} - \boldsymbol{H}'_{ij}\boldsymbol{\delta}_\vartheta/\sqrt{n}\right)$$

with $\boldsymbol{H}_i = (\boldsymbol{X}'_i, \boldsymbol{W}'_{2i})'$. Expanding we obtain,

$$\begin{aligned}
&\mathbb{E}(v(\delta_\alpha, \boldsymbol{\delta}_\vartheta) - v(0,\mathbf{0}) = \\
&= \ -\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbb{E}\boldsymbol{H}_i\left(\psi_{\tau_d}\left(\tilde{y}_{id} - \xi_{id}(\tau_d) - \delta_\alpha/\sqrt{n}\Phi_i - \boldsymbol{H}'_i\boldsymbol{\delta}_\eta/\sqrt{n}\right) - \psi_{\tau_d}(\tilde{y}_{id} - \xi_{id}(\tau_d))\right) \\
&= \ -\frac{1}{\sqrt{n}}\sum_{i=1}^n \boldsymbol{H}_i\left(F_{id}\left(\xi_{id}(\tau_d) + \delta_\alpha/\sqrt{n}\Phi_{id} + \boldsymbol{H}'_{id}\boldsymbol{\delta}_\vartheta/\sqrt{n}\right) - \tau\right) \\
&= \ -\frac{1}{\sqrt{n}}\sum_{i=1}^n \boldsymbol{H}_i f_{id}(\xi_{id}(\tau_d))\left(\Phi_{id}\delta_\alpha/\sqrt{n} + \boldsymbol{H}'_i\boldsymbol{\delta}_\vartheta/\sqrt{n}\right) + o_p(1)
\end{aligned}$$

Note that $v(\hat{\delta}_\alpha, \hat{\boldsymbol{\delta}}_\vartheta) \to 0$, and thus $\mathbb{E}(v(\boldsymbol{\delta}_\alpha, \boldsymbol{\delta}_\eta) - v(0,\mathbf{0})) = v(0,\mathbf{0}) + o_p(1)$. Letting $\boldsymbol{\delta}_\vartheta = (\boldsymbol{\delta}'_\beta, \boldsymbol{\delta}'_\gamma)'$, we write the last expression as,

$$-\frac{1}{\sqrt{n}}\sum_{i=1}^n \boldsymbol{H}_i f_{id}(\xi_{id}(\tau_d))\left(\Phi_i\delta_\alpha/\sqrt{n} + \boldsymbol{H}'_i\boldsymbol{\delta}_\vartheta/\sqrt{n}\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \boldsymbol{H}_i\psi_{\tau_d}\left(\tilde{y}_{id} - \xi_{id}(\tau_d)\right) + o_p(1)$$

Alternatively, using more convenient notation, we write the last expression as,

$$\boldsymbol{J}_\alpha\boldsymbol{\delta}_\alpha + \boldsymbol{J}_\vartheta\boldsymbol{\delta}_\vartheta = \mathbb{J}_\psi + o_p(1)$$

where $\boldsymbol{J}_\alpha = \lim_{n\to\infty}\boldsymbol{H'\Psi\Phi}$, $\boldsymbol{J}_\vartheta = \lim_{n\to\infty}\boldsymbol{H'\Psi H}$, $\boldsymbol{\Psi} = diag(f_{id}(\cdot))$ and $\mathbb{J}_\psi$ is a mean zero random variable with covariance $\tau_d(1-\tau_d)\boldsymbol{H'H}$. Letting $[\bar{\boldsymbol{J}}'_\beta, \bar{\boldsymbol{J}}'_\eta]'$ be a conformable partition of $\boldsymbol{J}_\vartheta^{-1}$, we have that $\hat{\boldsymbol{\delta}}_\eta = \bar{\boldsymbol{J}}'_\eta(\mathbb{J}_\psi - \boldsymbol{J}_\alpha\delta_\alpha)$ and $\hat{\boldsymbol{\delta}}_\beta = \bar{\boldsymbol{J}}'_\beta(\mathbb{J}_\psi - \boldsymbol{J}_\alpha\delta_\alpha)$. Letting $\boldsymbol{Z} = \bar{\boldsymbol{J}}'_\eta\boldsymbol{A}\bar{\boldsymbol{J}}_\eta$ as in Chernozhukov and Hansen (2006), we have that

$$\hat{\delta}_\alpha = (\boldsymbol{J}'_\alpha\boldsymbol{Z}\boldsymbol{J}_\alpha)^{-1}\boldsymbol{J}'_\alpha\boldsymbol{Z}\mathbb{J}_\psi.$$

Replacing it in the previous expression,

$$\hat{\boldsymbol{\delta}}_\eta = \bar{\boldsymbol{J}}_\eta'(\mathbb{J}_\psi - \boldsymbol{J}_\alpha\delta_\alpha) = \bar{\boldsymbol{J}}_\eta'(\boldsymbol{I} - \boldsymbol{J}_\alpha(\boldsymbol{J}_\alpha'\boldsymbol{Z}\boldsymbol{J}_\alpha)^{-1}(\boldsymbol{J}_\alpha'\boldsymbol{Z}))\mathbb{J}_\psi = \bar{\boldsymbol{J}}_\eta'(\boldsymbol{I} - \boldsymbol{L})\mathbb{J}_\psi = \bar{\boldsymbol{J}}_\eta'\boldsymbol{M}\mathbb{J}_\psi$$

where $\boldsymbol{L} = \boldsymbol{J}_\alpha[\boldsymbol{J}_\alpha'\boldsymbol{Z}\boldsymbol{J}_\alpha]^{-1}\boldsymbol{J}_\alpha'\boldsymbol{Z}$ and $\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{L}$. Due to invertibility of $\bar{\boldsymbol{J}}_\eta$, $\hat{\boldsymbol{\delta}}_\eta = \mathbf{0} \times O_p(1) + o_p(1)$. Similarly, substituting back δ_α, we obtain that $\hat{\boldsymbol{\delta}}_\beta = \bar{\boldsymbol{J}}_\beta'(\boldsymbol{I} - \boldsymbol{L})\boldsymbol{J}_\psi$. By the regularity conditions, we have that,

$$\begin{pmatrix} \delta_\alpha(\alpha_n, \tau_d) \\ \hat{\boldsymbol{\delta}}_\beta(\alpha_n, \tau_d) \end{pmatrix} = \begin{pmatrix} \sqrt{n}(\hat{\alpha}(\alpha_n, \cdot) - \alpha(\tau_d)) \\ \sqrt{n}(\hat{\boldsymbol{\beta}}(\alpha_n, \cdot) - \boldsymbol{\beta}(\tau_d)) \end{pmatrix} \rightsquigarrow \mathcal{N}\left(\mathbf{0}, \boldsymbol{J}(\tau_d)^{-1}\boldsymbol{S}(\tau_d)\boldsymbol{J}(\tau_d)^{-1}\right).$$

□