LATENT STRUCTURES AND QUANTILES OF THE TREATMENT EFFECT DISTRIBUTION*

ERICH BATTISTIN, CARLOS LAMARCHE and ENRICO RETTORE†

Preliminary and Incomplete
Please do not cite without permission

Keywords: Factor Models; Quantile Regression; Treatment Effects.

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 one or more outcomes of interest (see, for example, Heckman *et al.*, 1999, for a review). The evidence from almost all empirical studies points to heterogeneous effects.

The existence of such heterogeneity notwithstanding, drawing causal inference on the effects of a policy intervention has been traditionally concerned with the measurement of the *mean* effect of the intervention. The role played by heterogeneity is typically investigated by comparing the mean effect for different subgroups of the population identified by observable characteristics. However, this strategy does not allow to draw definitive conclusions about the distribution of the effects of the policy, that can only be inferred by considering alternative parameters.

The common practise of looking at mean effects is mostly the result of a pragmatic approach to the evaluation problem. On the one hand, well established statistical techniques that have been developed in the literature over the years mainly focus on averages. Most importantly, identification of characteristics of the effect distribution other than the mean is precluded on a logical ground, since the assumptions required to achieve point identification of the mean effect are not enough to retrieve the effect distribution. In the last decade, the empirical relevance of investigating characteristics of the treatment effect distribution was discussed by very many authors, starting from the seminal work by Heckman *et al.* (1997).

This paper derives new conditions for identification of features of the treatment effect distribution other than the mean, thus allowing to draw policy conclusions that are more

*First draft: December 12, 2010. This Version: March 30, 2012.
†Erich Battistin and Enrico Rettore: Department of Statistics, University of Padova, Via Cesare Battisti 241, 35123 Padova - Italy. Email: erich.battistin@unipd.it and enrico.rettore@unipd.it. Carlos Lamarche: Department of Economics, University of Oklahoma. 202 Hester Hall, 729 Elm Avenue, Norman, OK 73019. Tel.: +1 405 325 5857. Email: lamarche@ou.edu.

general than standard average treatment effect analysis. We assume that potential outcomes are related through latent individual ability or skills, and frame the identification problem considering a factor model that generalizes the standard specification widely used in microeconometrics, psychology and other social sciences. This represents the first contribution of the paper. The conditions needed are very general in nature, or at least as general as those already presented in other studies estimating economic models in this setup (see, for example, Carneiro *et al.*, 2003, and Heckman *et al.*, 2006). Under the factor model considered, the second contribution of this paper is to provide identification results for quantiles of the effect distribution for participants with particular outcomes in the non-participation state.

We derive an estimation strategy that makes use of the most recent results in the literature on quantiles. The methodology developed is then used to evaluate the distributional effects of an Italian policy that combines income support to eligible dismissed employees with benefits to employers who hire them.

2. Notation and parameter of interest

The notation employed in the potential outcome approach to causal inference is used throughout. Assume that the 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, $\mathbf{W}$ is a *vector* of K random variables, D is the *binary* treatment or policy status defining the corresponding potential outcomes (Y_1, Y_0), and $\mathbf{X}$ are control variables exogenous to the model. To fix ideas, in what follows the vector $\mathbf{W}$ will include outcome measurements which are pre-determined with respect to the policy roll out, although the approach that we take can trivially be extended to allow for post-programme outcomes that are different from Y. It is assumed throughout that the treatment status is randomly allocated to units of the relevant population.[1]

Assumption 1. (*Policy assignment mechanism*). The random variables involved are independent of the policy status:

$$F_{Y_1 Y_0 \mathbf{W}\mathbf{X}|D}[y_1, y_0, \mathbf{w}, \mathbf{x}|d] = F_{Y_1 Y_0 \mathbf{W}\mathbf{X}}[y_1, y_0, \mathbf{w}, \mathbf{x}].$$

This assumption is made for convenience, as it rules out by construction any source of selection that might result from alternative programme participation processes. Conditional versions of this assumption may also be considered, for example reflecting selection on the

[1] The notation $F_{A|B}[A|b]$ indicates the conditional distribution of the random variable A given $B = b$. A similar notation will be employed for the conditional τ-quantile function $Q_{A|B}[\tau|b]$.

observables $\mathbf{X}$ or randomness of the treatment status locally with respect to a selection threshold.

Comparisons of potential outcomes for well defined populations of individuals define 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. However, even when individuals are randomized into/out of the 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 *marginal distributions* of potential outcomes. The unrestricted set of joint distributions consistent with the marginals can be exploited to partially-identify the distribution of β via classical probability inequalities. However, the resulting identification set is generally way too wide (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.

In what follows, the parameter of interest is represented by features of the treatment effect distribution other than the mean. In particular we focus on the quantile function of the distribution of Δ conditional and $Y_0 = y_0$ within cells defined by $\mathbf{X}$, which is defined as $Q_{\Delta|Y_0\mathbf{X}}[\tau|y_0, \mathbf{x}]$.

3. General formulation of the problem

3.1. **The model.** The following latent factor structure is assumed, which we state as conditional on cells defined by $\mathbf{X}$:

$$W_1 = \Theta + V_1, \tag{3.1}$$

$$W_k = \delta_k\Theta + (1 + \xi_k\Theta)V_k, \quad k = 2, \ldots, K \tag{3.2}$$

$$Y_1 = \lambda_1\Theta + (1 + \gamma_1\Theta)U_1, \tag{3.3}$$

$$Y_0 = \lambda_0\Theta + (1 + \gamma_0\Theta)U_0. \tag{3.4}$$

The uniqueness $(U_1, U_0, \mathbf{V})$, where $\mathbf{V} \equiv [V_1, V_2, \ldots, V_K]$, are mutually independent and independent of the latent indicator Θ. It is assumed that the loadings λ_1 and λ_0 have both the same sign, and we conventionally assume that it is positive. It is also assumed that all the random variables involved in the factor structure are continuous, with cdf's that are strictly increasing on the corresponding supports. The conditions embedded in the model can be summarized as follows.

Assumption 2. (*Factor model*). The random variables defining the factor model are mutually independent:

$$F_{U_1 U_0 \Theta \mathbf{V}|D\mathbf{X}}[u_1, u_0, \theta, \mathbf{v}|d, \mathbf{x}] = F_{U_1|\mathbf{X}}[u_1|\mathbf{x}] F_{U_0|\mathbf{X}}[u_0|\mathbf{x}] F_{\Theta|\mathbf{X}}[\theta|\mathbf{x}] \prod_{k=1}^{K} F_{V_k|\mathbf{X}}[v_k|\mathbf{x}].$$

The advantages of using latent structures to establish identifying correspondences between data and the joint distribution of potential outcomes have already been discussed in previous work (see, amongst others, Carneiro *et al.*, 2003). The research question is under which conditions realizations of $(Y, D, \mathbf{W}, \mathbf{X})$ revealed by the data and the assumptions made on the factor structure are enough to point identify the joint distribution of the potential outcomes, and thus any feature of the treatment effect distribution.

Previous results discussed in the literature provide an answer by considering the standard factor model, which follows by imposing $\gamma_0 = 0$, $\gamma_1 = 0$ and $\xi_k = 0$, for $k = 2, \ldots, K$, in the expressions above (see Carneiro *et al.*, 2003, for a discussion). The informational contents of the model are easily understood by considering $K = 2$, which we maintain as the working example throughout the paper. First note that the model can be written as:

$$\begin{aligned}
\tilde{W}_1 &= \Theta + \tilde{V}_1, \\
\tilde{W}_2 &= \Theta + \tilde{V}_2, \\
\tilde{Y}_1 &= \Theta + \tilde{U}_1, \\
\tilde{Y}_0 &= \Theta + \tilde{U}_0,
\end{aligned}$$

where $\tilde{U}_j \equiv U_j/\lambda_j$, $\tilde{Y}_j \equiv Y_j/\lambda_j$ for $j = 0, 1$, and $\tilde{W}_k \equiv W_k/\gamma_k$ and $\tilde{V}_k \equiv V_k/\gamma_k$, for $k = 1, 2$. A straightforward application of the Kotlarski's (1967) theorem implies that the distributions of the uniqueness and of Θ are non-parametrically identified from knowledge of $(\tilde{Y}, D, \tilde{W}_1, \tilde{W}_2)$, where $\tilde{Y} = \tilde{Y}_0 + D(\tilde{Y}_1 - \tilde{Y}_0)$.[2] Since factor loadings are identified from the correlation structure of $(Y, D, \mathbf{W})$ (conditional on $\mathbf{X}$), this in turn implies that the distribution of (Y_0, Y_1) is identified from observed data. The availability of more $\mathbf{W}$'s would add a certain degree of over-identification to the model, on top of that already defined from the availability of the $D = 1$ and $D = 0$ groups.

3.2. **Contribution to the existing literature.** The classical formulation of the factor model represents a useful tool to learn about policy parameters that are different from the average treatment effect. Yet it is worth noting that it also has some implications that, depending on the application, may be violated in the data. To see this, note that the

[2]The results follows from Lemma 1, Remark 3 and Remark 4 in Kotlarski (1967). The distribution of θ is actually over-identified.

system of equations that we consider implies:

$$\Delta = \left(\frac{\lambda_1 + \gamma_1 U_1}{\lambda_0 + \gamma_0 U_0} - 1 \right) Y_0 - \left(\frac{\lambda_1 + \gamma_1 U_1}{\lambda_0 + \gamma_0 U_0} \right) U_0 + U_1,$$

which in turn yield:

$$\begin{aligned} Q_{\Delta|Y_0,U_0}[\tau_1|y_0, Q_{U_0}[\tau_0]] &= \left(\frac{\lambda_1[\tau_1]}{\lambda_0[\tau_0]} - 1 \right) y_0 - \frac{\lambda_1[\tau_1]}{\lambda_0[\tau_0]} Q_{U_0}[\tau_0] + Q_{U_1}[\tau_1], \\ \lambda_1[\tau_1] &\equiv \lambda_1 + \gamma_1 Q_{U_1}[\tau_1], \\ \lambda_0[\tau_0] &\equiv \lambda_0 + \gamma_0 Q_{U_0}[\tau_0], \end{aligned} \tag{3.5}$$

which is the quantile function of the distribution of returns conditional on having $Y_0 = y_0$ and $U_0 = Q_{U_0}[\tau_0]$ in the policy off scenario. Similarly there is:

$$\nabla_{Y_0} Q_{\Delta|Y_0,U_0}[\tau_1|y_0, Q_{U_0}[\tau_0]] = \left(\frac{\lambda_1[\tau_1]}{\lambda_0[\tau_0]} - 1 \right). \tag{3.6}$$

It is clear that, when the γ's are set to zero, the value y_0 of the base state distribution enters as a location shift which is constant across quantiles of the treatment effect distribution. Our model thus generalizes the factor structure considered in previous papers allowing for heteroskedasticity that depends on latent ability.

4. Identification

Identification of (3.5) is established by the following correspondence:

$$Q_{Y_1 - Y_0|Y_0 U_0}[\tau_1|y_0, Q_{U_0}[\tau_0]] = \mathcal{H}_{\lambda_0[\tau_0],\lambda_1[\tau_1]}\{Q_{U_0}[\tau_0], Q_{U_1}[\tau_1]\}, \tag{4.1}$$

where $\mathcal{H}$ is a (known) point identifying functional. For known values of the loadings $\lambda_1[\tau_1]$ and $\lambda_0[\tau_0]$, this establishes a correspondence between the unobserved term on the left hand side of the equation and the two quantile functions on the right hand side. Similarly, knowledge of the loadings is sufficient to retrieve (3.6). The relationship in (4.1) sets the stage for analogue estimation of the quantity of interest.

4.1. Identification of quantile specific factor loadings.

We first consider identification of the factor loadings in the standard setting. This helps clarify the approach that we will take to allow for quantile specific loadings. Consider set of equations for individuals in the 'policy on' regime, that is equations (3.1), (3.2) and (3.3). By substituting (3.1) into (3.2) and (3.3), and rearranging terms there is:

$$Y_1 = (\lambda_1 + \gamma_1 U_1) W_1 - (\lambda_1 + \gamma_1 U_1) V_1 + U_1, \tag{4.2}$$

$$W_2 = (\delta_2 + \xi_2 V_2) W_1 - (\delta_2 + \xi_2 V_2) V_1 + V_2, \tag{4.3}$$

which define the (feasible) regressions of Y_1 and W_2 on W_1. It is easy to see that the following moment conditions are defined:

$$COV\{(\lambda_1+\gamma_1 U_1)V_1 - U_1, W_2\} = 0, \tag{4.4}$$

$$COV\{(\delta_2+\xi_2 V_2)V_1 - V_2, Y_1\} = 0, \tag{4.5}$$

from which there is:

$$(\lambda_1+\gamma_1 E\{U_1\}) = \frac{COV\{Y_1,W_2\}}{COV\{W_1,W_2\}}, \tag{4.6}$$

$$(\delta_2+\xi_2 E\{V_2\}) = \frac{COV\{W_2,Y_1\}}{COV\{W_1,Y_1\}}. \tag{4.7}$$

The empirical counterpart of the ratios on the right hand side of the equations define the IV estimand from the regressions of Y_1 on W_1 and of W_2 on W_1, using W_2 and Y_1 as instruments, respectively.

Remark 1. It is possible to obtain the empirical counterpart on the right hand side of equation (4.6) using standard arguments. We write (4.2) as $\dot{Y}_1 = Y_1 - \pi_1 W_1 = \pi_2 W_2 + U_1'$, where $\pi_1 = (\lambda_1 + \gamma_1 E(U_1))$. We have then that $\hat{\pi}_2 = (W_2' W_2)^{-1} W_2' \dot{Y}_1$, and

$$\hat{\pi}_1 = \underset{\pi_2\in\Pi_2}{\operatorname{argmin}}\{\hat{\pi}_2'(W_2'W_2)\hat{\pi}_2\}. \tag{4.8}$$

By analogy, the first part of the paper is devoted to extending the identification result to the case of quantile specific loadings, employing instrumental quantile regressions. The result is contained in Theorem 1 and Theorem 2, for which conditions are discussed in the next section. The results in the theorems imply that the quantity in (3.6) can be retrieved from the data. Because of this, the standard factor model can be tested against the data, as in this case the derivative should be constant across quantiles.

Theorem 1 (Latent structures and conditional quantile functions)**.** *If the conditions of the model hold,* $P(Y_1 \le q(\theta,\tau_1)) = P(Y_1 < q(W_1,\tau_1)|W_2) = \tau_1$ *for a linear quantile function* $q(\cdot)$*. Similarly,* $P(Y_0 \le q(\theta,\tau_0)) = P(Y_0 < q(W_1,\tau_0)|W_2) = \tau_0$.

Proof. We show the result for the conditional quantiles of Y_1. Consider:

$$\begin{aligned} P(Y_1 \le q(\theta,\tau_1)) &= P(q(\theta,U_1) \le q(\theta,\tau_1)) = P(q(W_1,V_1,U_1) \le q(W_1,V_1,\tau_1)) \\ &= P(q(W_1,U_1') \le q(W_1,\tau_1)|W_2) \\ &= P(q(W_1,U_1) \le q(W_1,\tau_1)|W_2) \\ &= P(U_1 \le \tau_1|W_2) = \tau. \end{aligned}$$

By Theorem 1 in Chernozhukov and Hansen (2005), $P(U_1 < \tau_1|W_2) = \tau$. The proof for Y_0 follows immediately from previous derivations. $\square$

Theorem 2 (Quantile specific loadings)**.** *Under the previous conditions, there exists a correspondence between* $Q_{Y_1-Y_0|Y_0U_0}[\tau_1|y_0, Q_{U_0}[\tau_0]]$ *and* $\mathcal{H}_{\lambda_0[\tau_0],\lambda_1[\tau_1]}$.

Proof. By Lemma 1, the τ_1-specific loading $\lambda_1(\tau_1)$ in $Q_{Y_1}(\tau_1|\theta) = \lambda_1(\tau_1)\theta$ can be identified from the following system of equations:

$$\begin{aligned} (4.9) \quad Y_1 &= \lambda_1 W_1 - (\lambda_1 + \gamma_1 U_1)V_1 + (1 + \gamma_1 W_1)U_1 = q(W_1, V_1, U_1) = q(W_1, U_1'), \\ (4.10) \quad W_1 &= \delta(W_2, V), \end{aligned}$$

where $\delta(\cdot)$ is an unknown function and $V = (V_1, V_2)$. To see this, we write the moment condition $P(Y_1 \le q(W_1, \tau_1)|W_2) = P(Y_1 - q(W_1, \tau_1) \le 0|W_2) = P(\dot{Y}_1 \le 0|W_2) = \tau_1$ as $0 = Q_{\dot{Y}_1}(\tau_1|W_2)$ for all $\tau_1 \in \mathcal{T}_1$. Letting $\mathcal{S}(\tau_1, \beta) = E\rho_{\tau_1}(\dot{Y} - W_2'\beta)$, we have that $\beta(\tau_1, \lambda_1) = \text{argmin}_{\beta\in\mathcal{B}}\{\mathcal{S}(\tau_1, \beta)\}$ and therefore:

$$\lambda_1(\tau_1) = \underset{\lambda\in\mathcal{L}}{\text{argmin}}\{\beta(\tau_1, \lambda_1)' A \beta(\tau_1, \lambda_1)\}. \tag{4.11}$$

The identification of the quantile-specific factor loading corresponding to the conditional distribution of Y_0 can be shown similarly following the steps of the proof sketched before. □

Corollary 1. *Under the previous assumptions and provided that the correspondence* $\mathcal{H}$ *has continuos derivarives* h *with respect to a point in the distribution of* Y_0*,*

$$\nabla_{Y_0} Q_{\Delta|Y_0,U_0}[\tau_1|y_0, Q_{U_0}[\tau_0]] = \nabla_{Y_0}\mathcal{H}_{\lambda_0[\tau_0],\lambda_1[\tau_1]}\{Q_{U_0}[\tau_0], Q_{U_1}[\tau_1] = h(\lambda_0[\tau_0], \lambda_1[\tau_1]).$$

4.2. **Identification of quantiles of uniqueness.** Levels of (3.5) are identified once the quantile functions of U_1 and U_0 can be retrieved from the data. The aim of this section is to provide sufficient conditions for this to hold.

Remark 2. Under conditions of the location-shift model for predetermined outcome variables $(W_1, W_2, ..., W_K)$, if $0 \in \Theta$, then,

$$\begin{aligned} Q_{U_1}(\tau_1) &= \nabla_{\tau_1} Q_{Y_1}(\tau_1|\theta)|_{\theta=0} \\ Q_{U_0}(\tau_0) &= \nabla_{\tau_0} Q_{Y_0}(\tau_0|\theta)|_{\theta=0}. \end{aligned}$$

In the model presented before, note that $Q_{U_1}(\tau_1) = Q_{Y_1}(\tau_1|\theta)$ at $Q_{W_1}(\tau_v|\theta) = Q_{V_{W_1}}(\tau_v))$. We can use the other measures of W in a model with an intercept and obtain $Q_{Y_1}(\tau_1|\theta)$ for which $Q_{W_1}(\tau_v|\theta) = Q_V(\tau_v)$.

5. Exact calculations

To be written.

6. Montecarlo evidence

To be written.

7. Estimation

To be written.

8. Data

To be written.

References

Carneiro, P.M., Karsten, K.T. and Heckman, J.J. (2003) Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College, *NBER Working Paper* No. W9546.

Heckman, J.J. Smith, J. and Clements, N. (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. Lalonde, R. and Smith, J. (1999) *The Economics and Econometrics of Active Labor Market Programs*, *Handbook of Labor Economics*, Volume 3, Ashenfelter, A. and Card, D. (eds.), Amsterdam: Elsevier Science.

Heckman, J.J., Stixrud, J. and Urzua, S. (2006) The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior, *Journal of Labor Economics* 24 (3), 411-482.

Kotlarski, I. (1967) On Characterizing the Gamma and the Normal Distribution, *Pacific Journal of Mathematics* 20 (1), pp. 69-76.

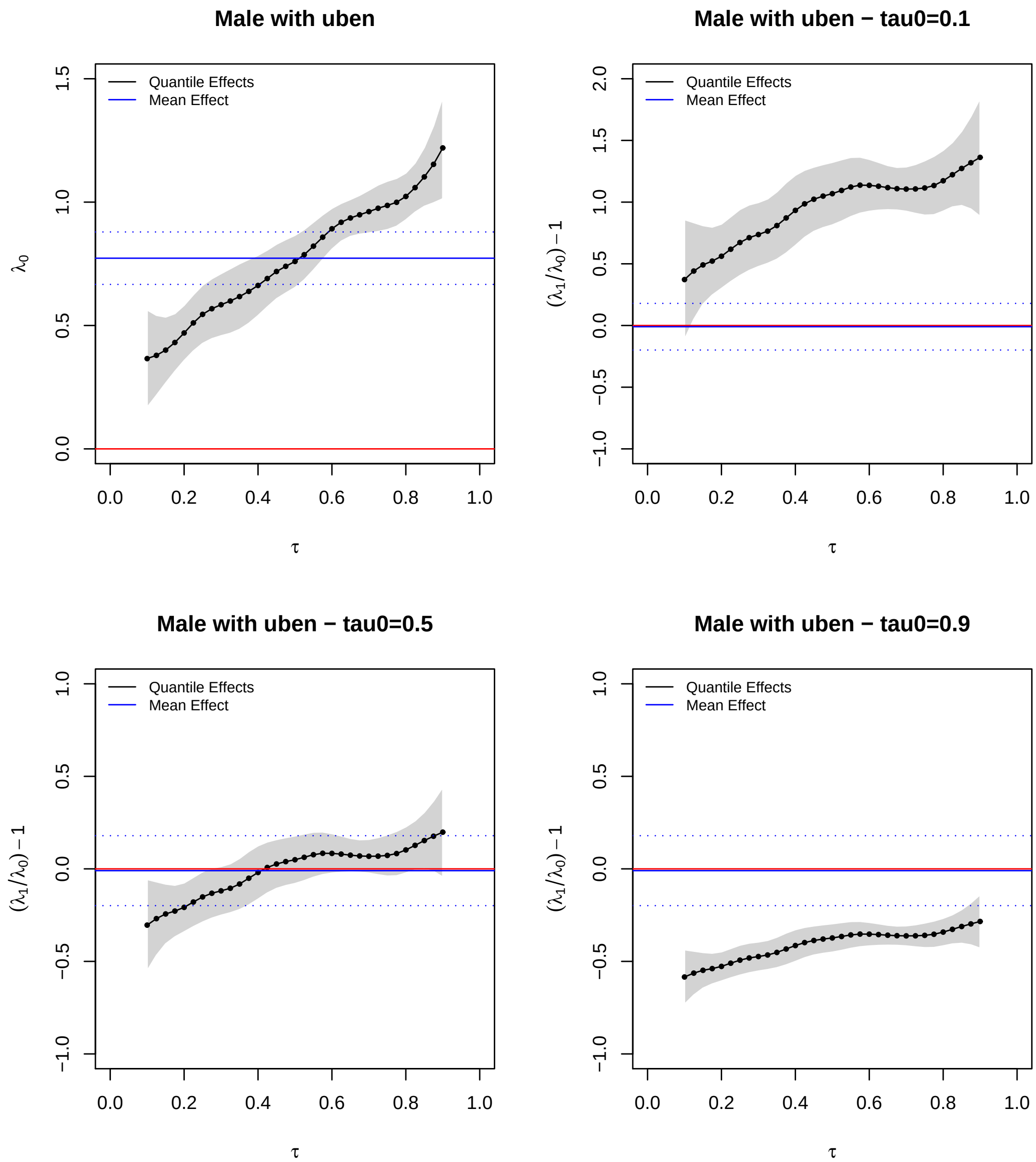

FIGURE 8.1. Preliminary results using Italian data.