



**HAL**  
open science

# Heterogeneity and Non-Constant Effect in Two-Stage Quantile Regression

Christophe Muller

► **To cite this version:**

Christophe Muller. Heterogeneity and Non-Constant Effect in Two-Stage Quantile Regression. 2017.  
halshs-01157552v3

**HAL Id: halshs-01157552**

**<https://shs.hal.science/halshs-01157552v3>**

Preprint submitted on 10 Apr 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Heterogeneity and Non-Constant Effect in Two-Stage Quantile Regression

Christophe Muller

WP 2015 - Nr 22

# Heterogeneity and Non-Constant Effect in Two-Stage Quantile Regression

Christophe Muller<sup>b</sup>

<sup>b</sup> Aix-Marseille Univ, CNRS, EHESS, Ecole Centrale Marseille, Greqam, Marseille, France.  
14, Boulevard Jules Ferry, Aix-en-Provence, France.  
christophe.muller@univ-amu.fr

February 2017

**Abstract:** Heterogeneity in how some independent variables affect a dependent variable has become a major topic of study in econometrics and statistics. In this respect, this paper addresses the question of constant versus non-constant effect through quantile regression modeling. For linear quantile regression under endogeneity, it is often believed that the fitted-value setting (i.e., replacing endogenous regressors with their exogenous fitted-values) implies constant effect (that is: the coefficients of the covariates do not depend on the considered quantile, except for the intercept). Here, it is shown that, under a weakened instrumental variable restriction, the fitted-value setting can allow for non-constant effect, even though only the constant-effect coefficients of the model can be identified. An application to food demand estimation in 2012 Egypt shows the practical potential of this approach.

**JEL Codes:** C13, C21, C31.

**Key Words:** Two-Stage Estimation, Quantile Regression, Fitted-Value Setting, Non Constant Effect, Partial Identification.

This work was carried out thanks to the support of the A\*MIDEX project (No. ANR-11-IDEX-0001-02) funded by the "Investissements d'Avenir" French Government program, managed by the French National Research Agency (ANR). I am grateful to P. Lavergne for a useful discussion and to participants in a seminar at the Toulouse School of Economics and at the 2015 ERCIM conference in London for their comments. Remaining errors are mine.

# 1 Introduction

## 1.1 The issue

Heterogeneity in how some independent variables affect a dependent variable has become a major topic of study in econometrics and statistics. Allowing for heterogeneity is potentially important because it may correct for misspecifications of regression models in studies where the scientific theory does not imply that different statistical individuals should be characterized by the same level of coefficient (i.e., effect) of a given independent variable. Allowing for heterogeneity also yields a much broader and more flexible set of models. Moreover, it generally contributes to explain a larger proportion of the variance of the dependent variable, notably through non-constant effect of regressors. Finally, it opens possibilities of complex and subtle causality features, involving different subpopulations with distinct levels of effects of variables (e.g., the ‘treated’, the ‘untreated’, the ‘compliers’). We discuss in Sub-Section 1.2 some practical economic examples in which heterogeneity of effect can be spelt out according to the constancy and non-constancy of some effects.

In this respect, this paper addresses the question of constant versus non-constant effect through quantile regression modeling. It exhibits a particular form of non-constant effect models through two-stage quantile regressions based on the fitted-value setting under endogeneity. The proposed approach can be regarded as intermediate between the constant effect quantile model and the fully non-constant effect quantile model. For this, the paper considers the identification of a structural linear equation using quantile regression.

Since the seminal work by Koenker and Bassett (1978), there are two trends in the literature on quantile regressions when dealing with endogeneity. The first one,

denoted the ‘structural setting,’ corresponds to models directly specified in terms of the conditional quantile of the structural equation of interest. In that case, semiparametric restrictions needed for identification are directly imposed on the structural errors terms, or on the structural model if there are no separable errors. The second trend, denoted the ‘fitted-value setting,’ is based on the conditional quantile of the reduced-form equation. Accordingly, the restrictions are imposed on the reduced-form errors when they can be separated. In this later setting, the analysts substitute the endogenous regressors with the fitted-values obtained from ancillary regressions based on some exogenous variables. The fitted-value setting corresponds to quantile restrictions on the reduced form. Beyond being a calculation device, examining the reduced form makes sense for several reasons. First, as noted in Blundell and Powell (2006): “The reduced form for  $y_t$  may be of interest if the values of instrumental variables are control variables for the policy maker.” Second, the fitted-value setting also enjoys some algebraic and computational advantages. Thus, the reduced form can also be viewed as a convenient intermediary stage for calculations. Amemiya (1974) pointed out that, while substitution of fitted values in nonlinear structural functions generally yields inconsistent estimates of the structural parameters, consistent estimation methods that substitute fitted values into the structural function can rely on linearity of the regression, when the model is based on the reduced-form error, with similar stochastic properties to the structural error.

However, even in the context of linear models for quantile regressions, it is believed that the fitted-value setting corresponds to constant effect models, a little attractive characteristic. That is: it is believed that all coefficients, except the intercept, must be the same for all quantiles, as discussed in Lee (2007, p1138).<sup>1</sup> The latter corresponds

---

<sup>1</sup>We quote: “ ‘the fitted value’ approach, which is developed by Amemiya (1982) and Powell (1983), replaces  $X$  with the fitted value of  $\mu + Z'\pi$  in the system:  $Y = X\beta(\tau) + Z_1'\gamma(\tau) + U$  and

to the constant effect case for the second stage of the estimated quantile regression model. This criticism has not been addressed in the recent literature, other than by falling back to the structural setting, or by assuming constant effect for the true quantile process. We deal with this gap. In this paper, we exhibit a particular case of non-constant effect (i.e., quantile-dependent coefficients) for linear quantile regression with the fitted-value setting. However, we still require that the causal effect of an endogenous regression is constant.

The literature on the structural setting for linear quantile regressions is abundant,<sup>2</sup> while it meets computation costs for correcting endogeneity issues.<sup>3</sup> In contrast, the fitted-value setting corresponds to a simple two-step quantile regression procedure, analogous to the 2SLS method, and has been readily employed by empirical researchers who are not always expert econometricians or programmers, thereby conveniently avoiding computation burden.<sup>4</sup> Namely, with the fitted-value setting, no control function nonparametric estimation, no simulation, no computation iteration or

---

$X = \mu(\alpha) + Z'\pi(\alpha) + V$ . To see how the fitted value approach works, consider the reduced-form equation for  $Y$ :  $Y = \beta[\mu + Z'\pi] + Z_1'\gamma + \eta$ , where  $\eta = U + \beta V$ . In order to estimate  $\beta$  and  $\gamma$  consistently, the fitted value approach requires that  $Q_{\eta|Z}(\tau|z)$  be independent of  $z$ .” on page 1138, and further “under the independence assumption,  $\beta$  and  $\gamma$  are constant over  $\tau$ ” on page 1139. The correspondence with the notations we introduce later reads: “replaces  $Y_t'$  with the fitted value of  $x_t'\Pi$  in the system:  $y_t = Y_t'\gamma_\theta + x_{1t}'\beta_\theta + u_{t\theta}$  and  $Y_t' = x_t'\Pi + V_t'$ . To see how the fitted value approach works, consider the reduced-form equation for  $y_t$ :  $y_t = [x_t'\Pi]\gamma_\theta + x_{1t}'\beta_\theta + v_{t\theta}$ , where  $v_{t\theta} = u_{t\theta} + V_t'\gamma_\theta$ . In order to estimate  $\beta$  and  $\gamma$  consistently, the fitted value approach requires that  $Q_{v_{t\theta}|x_t}(\theta|x_t)$  be independent of  $x_t$ .” and “under the independence assumption,  $\gamma_\theta$  and  $\beta_\theta$  are constant over  $\theta$ ”

<sup>2</sup>See for example: Abadie et al. (2002), Chen et al. (2003), Chesher (2003), Hong and Tamer (2003), Chernozhukov and Hansen (2005, 2006, 2008), Imbens and Newey (2006), Ma and Koenker (2006), Chernozhukov, Imbens and Newey (2007), Lee (2007), Jun (2009), Chernozhukov, Fernandez-Val and Kowalski (2015).

<sup>3</sup>As documented in Kim and Muller (2013).

<sup>4</sup>Arias et al. (2001), Garcia et al. (2001), Chortareas et al. (2012) and Chepatrakul et al. (2012).

grid are necessary. Partial theoretical results had been obtained by Amemiya (1982) and Powell (1983), who analyze the two-stage least-absolute-deviations estimators in simple settings, and by redefining the dependent variable. Chen and Portnoy (1996) and Kim and Muller (2004, 2015) investigate such two-stage quantile regressions with diverse first-step estimators - least squares (LS), least absolute deviation (LAD), and trimmed least squares estimators - and in general settings.

However, according to Lee (2007)'s claim, all these authors, under the fitted-value setting, deal with constant effect specifications. In contrast, this paper focuses on the occurrence of non-constant effect with the fitted-value setting, although heterogeneity will still not be allowed for some model coefficients.

For this, it is first shown that any separable model can be made to satisfy a quantile restriction for any quantile  $\theta$ , provided it allows for an inconsistency term, which is characterized. Second, it is shown how the influence of the inconsistency term can be weakened in terms of its link with covariate effects. This is done by assuming some weakened instrumental variable (IV) restrictions, which may even allow for endogenous regressors in the reduced form. Under these new IV restrictions, it is shown that non-constant effect can arise in linear quantile regression even under endogeneity dealt with the fitted-value setting. However, in that case, only the constant effect coefficients can be identified. Finally, it is shown that the constant effect (respectively, non-constant effect) coefficients of the reduced form can be simply transmitted to constant effect (respectively, non-constant effect) coefficients in the structural equation.

## 1.2 Practical examples

Let us now illustrate our reflections with a few economic examples in which constant effect may occur at least for some variables in the regressions. That is: in some cases, one can neglect the heterogeneity of causal effects for some variables in the model. A first illustration is the assessment of a policy rule that, first, is constant for an identifiable population, and second, affects an outcome variable that is shifted by a quantity proportional to the policy rule. For example, one may wish to assess the impact of a cash transfer program on family earnings. Let  $y_t$  be the total earnings of family  $t$  and  $Y_t$  be the policy treatment, which is here a dummy variable describing whether or not the family is covered by the program. The justification for assuming that there is no heterogeneity in some effects is here based on legal obligations that impose equal treatments of ‘equals’ in that case. Assume that the treated families can be identified through some observable characteristics. For example, in France, family allowances are implemented through cash transfers calculated as a precise function of the number of children by age class in the family (ADECRI, 2008), not allowing for any other heterogeneity in the transfer rule. The information on family composition can generally be observed in household survey data, which can be used in a regression of family earnings. In that case, the transfer amount, which is the constant parameter  $\gamma$  to estimate, is generally not, or is ill observed in the survey data. However, the treatment dummy variable  $Y_t$  can be observed, and used as a covariate in this regression, while it may be endogenous. For example, families with unobserved low motivation for work may have more children and lower income. This may be the case if having children is used by some families as a mean to access social aid and to compensate for insufficient incomes.

Another illustration is the assessment of the impact on household income of taxes



calculated from observable categories of household. Assume that these unobserved tax amounts are fixed within each household category. Then, when running a quantile regression with the household net income as a dependent variable, the effect of taxes can be described by a vector of fixed coefficients  $\gamma$  for the dummy variables specifying the household categories. The legal tax schedule justifies assuming no heterogeneity in this effect. Another example is that of racketing extraction tariffs imposed by mafia organizations, which may have the same property of being fixed for some given categories of businesses, while their actual amounts are generally unobservable by researchers. In that case, the reason for no heterogeneity of the causal effect could stem from some ‘social customs’, rather than a legal obligation.

Another example is a structural wage equation for a labor market study for a sample of workers, in which the dependent variable ( $y_t$ ) is the logarithm of individual wage rate, while the two independent variables in this equation are the industrial sector dummy ( $x_{1t}$ ) and the worker’s education level ( $Y_t$ ), and the regressors in the reduced form are the industrial sector dummy and the worker’s birth quarter ( $x_{2t}$ ). The birth quarter is used as an instrument for the education level that is typically suspected to be the sole endogenous independent variable in the structural model. For example, Angrist and Krueger (1991) and Angrist and Pischke (2009) find constant effect for this variable. Then, in that case the justification for no heterogeneity of this effect is empirical. However, one may expect the log wage rate to be positively correlated with the capital of the firm, which is omitted from the model, and therefore should be included in the error  $v_t$ , and furthermore is correlated with the industrial sector dummy. Then,  $x_{1t}$  and  $v_t$  may be correlated, while  $x_{2t}$  and  $v_t$  should be independent according to the usual justification for using quarter of birth as an instrument in wage equations.<sup>5</sup> The proposed modeling approach in this paper will

---

<sup>5</sup>Conditioning on the industrial sector might make the hypothesis of independence of the birth

somewhat mimic such independence setting.

A final illustration is the assessment of the contribution to total family expenditure of some unobserved expenses for a discrete good, or a discrete service, when its price is fixed (e.g., when provided by public institutions), which implies no heterogeneity in these public tariffs. For example, access to a given public service with fixed characteristics might be observable in some survey data, while not its officially fixed price. Estimating the corresponding constant coefficient  $\gamma$  for this spending in an equation for total expenditure would allow some inference about the unobserved price.

In all of these illustrations, we have constant effect for a treatment variable of interest. Other exogenous variables  $x_{1t}$  that determine the studied outcomes may also be included.<sup>6</sup> In the considered examples, the coefficients  $\beta$  of these exogenous variables generally correspond to non-constant effect. Indeed, the economic theory does not provide any reason for imposing constant effect for these variables. Such non-constancy is likely to generate a specification bias in quantile regressions that would wrongly assume constant coefficients for these variables. However, if the interest of the researcher is exclusively in the coefficients  $\gamma$  of the structural endogenous regressors, the fact that the coefficients  $\beta$  of the structural exogenous regressors have non-constant effect that cannot be identified in the fitted-value setting is not an issue as long as constant effect coefficients can be identified, which is what is shown in this paper.

Finally, some observed heterogeneity in the treatment effect can easily be incorporated in these cases by interacting the treatment variable with some observed characteristic and the error more plausible if the sector was a common determinant of the latter two variables, although it is unclear why this should be the case.

---

<sup>6</sup>We shall be able to relax the exogeneity assumption later on.

acteristics, or by considering subpopulations defined in terms of these characteristics. This is an easy way to relax too stringent specifications of constant effect.

Supported by the relevance of such examples, this paper brings to the fore a new way of allowing for partially restricted heterogeneity of effect through quantile regression modeling under endogeneity. The paper is organized as follows. Section 2 presents the model and the assumptions. Section 3 exhibits and analyzes a case of non-constant effect for the fitted-value setting. Section 4 reports an empirical illustration. Finally, Section 5 concludes.

## 2 The Model

Assume that our interest lies in the parameter vector  $\alpha_\theta \equiv (\beta'_\theta, \gamma'_\theta)'$  in the following linear equation for  $T$  observations and an arbitrary quantile index  $\theta \in (0, 1)$  that will denote quantile restrictions introduced later on.

$$y_t = x'_{1t}\beta_\theta + Y'_t\gamma_\theta + u_{t\theta} = z'_t\alpha_\theta + u_{t\theta}, \quad (1)$$

where  $[y_t, Y'_t]$  is a  $(G + 1)$ -row vector of endogenous variables,  $x'_{1t}$  is a  $K_1$ -row vector of exogenous variables,  $z_t = [x'_{1t}, Y'_t]'$  and  $u_{t\theta}$  is an error term.

Since we wish to study non-constant effect models, we emphasize that the coefficient vector and the errors may vary with the considered quantile index  $\theta$ . We denote by  $x'_{2t}$  the row vector of the  $K_2$  exogenous variables excluded from (1), and we assume  $K_2 \geq G$ . We further assume that  $Y_t$  can be linearly predicted from the exogenous variables through the following equation, which we assume to be correctly specified.

$$Y'_t = x'_t\Pi + V'_t, \quad (2)$$

where  $x'_t = [x'_{1t}, x'_{2t}]$  is an unbounded  $K$ -rows vector with  $K = K_1 + K_2$ ,  $\Pi$  is a  $K \times G$  matrix of unknown parameters, and  $V'_t$  is a  $G$ -rows vector of unknown error terms.

Again, some stochastic restrictions on the errors  $V_t$  must be assumed so as to complete (2) for defining a correctly specified model. For example, one may assume that the conditional expectation of  $V_t$  is zero, as for ordinary least squares (OLS) estimation, thus ensuring the consistency of the fitted-value for  $Y_t$ ; or alternatively some quantile restriction at a quantile  $\theta$  as in Kim and Muller (2004). To avoid absurdities, we assume that the columns in  $x_t$  are linearly independent. Using (1) and (2),  $y_t$  can also be expressed as:

$$y_t = x_t' \pi_\theta + v_{t\theta}, \quad (3)$$

where

$$\pi_\theta = H(\Pi) \alpha_\theta \text{ with } H(\Pi) = \left[ \left( \begin{array}{c} I_{K_1} \\ 0 \end{array} \right), \Pi \right] \quad (4)$$

and

$$v_{t\theta} = u_{t\theta} + V_t' \gamma_\theta.$$

Again here, we allow for the vector of coefficients  $\pi_\theta$  to vary with the quantile  $\theta$ . We first consider the following quantile restriction on the reduced-form errors for a given quantile  $\theta_0$ , which has been used in the literature in Amemiya (1982), Powell (1983), Chen and Portnoy (1996), and Kim and Muller (2004, 2015).

**Condition 1:**  $E(\psi_{\theta_0}(v_{t\theta_0})|x_t) = 0$ , for an arbitrary given  $\theta_0 \in (0, 1)$ , where  $\psi_\theta(z) \equiv \theta - 1_{[z \leq 0]}$ , and  $1_{[\cdot]}$  is the indicator function.

Condition 1 imposes that zero is the given  $\theta_0^{\text{th}}$ -quantile of the conditional distribution of  $v_{t\theta_0}$ . It is associated with the fitted-value setting in which, first, the conditional quantile restriction is placed on the reduced-form error  $v_{t\theta_0}$ , and second, the information set used for the conditional restriction exclusively consists of the exogenous variables  $x_t$ .

One traditional advantage of the fitted-value setting is that it allows simpler algebraic analyses of the asymptotic representations of estimators, with decomposition of formulae into simple components corresponding to each estimation step. It has additional virtues when the reduced form has some interesting economic interpretation or use, for example as in Blundell and Powell (2006) who suggest that it can be used for policy simulations. This setting can also be seen as a convenient way to synthesize the information to deal with sufficient conditions for consistent estimation. As a matter of fact, weaker sufficient conditions can be obtained by restricting the error  $v_{t\theta_0}$  (e.g.,  $v_{t\theta_0}$  independent of  $x_t$ ) in the reduced-form equation, than by trying to impose similar restrictions simultaneously on  $u_{t\theta_0}$  and  $V_t$  (for example,  $u_{t\theta_0}$  and  $V_t$  independent of  $x_t$ ). Indeed, the latter would involve  $G + 1$  conditions, to be compared with the more parsimonious unique condition imposed on the reduced-form error.

One issue is understanding how quantile restrictions for different quantiles come together so as to define a unique quantile process for a single model; that is, so that all these restrictions are compatible. The analysis will clarify this point, which is central for clarifying the possible non-constancy of effects. Condition 1 is not what is used to establish any results of the paper, but it is something we shall try to understand better, by weakening it in Assumption 2 below.

The link of structural and reduced-form parameters is described by (4). Identification of the structural parameters is obtained, as often, from assuming that  $H(\Pi)$  is of full column rank. A first-stage estimator of  $\Pi$  in (2),  $\hat{\Pi}$ , allows the construction of the fitted value  $\hat{Y}'_t = x'_t \hat{\Pi}$ , which is substituted for  $Y'_t$  in the final estimation stage. For any quantile  $\theta$ , the two-stage quantile regression estimator  $\hat{\alpha}_\theta$  of  $\alpha_\theta$  is defined as a solution to:

$$\min_{\alpha} S_T(\alpha, \hat{\Pi}, \theta) = \sum_{t=1}^T \rho_{\theta}(y_t - x_t' H(\hat{\Pi}) \alpha), \text{ where } \rho_{\theta}(z) = z \psi_{\theta}(z). \quad (5)$$

In the next section, we exhibit some non-constant effect with the fitted-value setting.

### 3 Non-Constant Effect in the Fitted-Value Setting

#### 3.1 Regularities and quantile restrictions

Let us start again with Equations (1) and (2), with possible non-constant effect in structural and reduced-form equations, *but without a priori imposing Condition 1*. In order to deal with unique quantile values so as to simplify the discussion, we make the following continuity and monotonicity assumption, for a starting value  $\theta_0$  of the quantile index.

**Assumption 1:** *For a given quantile index  $\theta_0$ , the cdf of  $v_{t\theta_0}$  conditional on  $x_t$ , denoted  $F_{v_{t\theta_0}|x_t}$ , the cdf of  $v_{t\theta_0}$  conditional on  $x_{1t}$ , denoted  $F_{v_{t\theta_0}|x_{1t}}$ , and the marginal cdf of  $x_{2t}$ , denoted  $F_{x_{2t}}$ , are continuous and strictly increasing.*

Under Assumption 1, an inverse cdf term can always be isolated in the reduced-form equation, for  $\theta_0$ , by denoting:  $y_t = x_t' \pi_{\theta_0} + v_{t\theta_0} = F_{v_{t\theta_0}|x_t}^{-1}(\theta) + x_t' \pi_{1\theta_0} + v_{t\theta}^*$  and  $v_{t\theta}^* \equiv v_{t\theta_0} - F_{v_{t\theta_0}|x_t}^{-1}(\theta)$ . Then, by construction,  $v_{t\theta}^*$  is characterized by the conditional quantile restriction:  $E(\psi_{\theta}(v_{t\theta}^*)|x_t) = \theta - \Pr[v_{t\theta}^* \leq 0|x_t] = \theta - \Pr[v_{t\theta_0} \leq F_{v_{t\theta_0}|x_t}^{-1}(\theta)|x_t] = \theta - \theta = 0$ . As a consequence, one can always obtain a quantile regression restriction at  $\theta$ , even distinct from  $\theta_0$ , for the reduced-form, provided one accepts a possible nuisance inconsistency term  $F_{v_{t\theta_0}|x_t}^{-1}(\theta)$  that can affect all the coefficients of the model. Note that  $v_{t\theta}^*$  depends both on  $\theta$  and on  $\theta_0$ . In the next subsection, we weaken the quantile

restriction at quantile index  $\theta_0$ , so as to allow enough flexibility for generating non-constant effect.

### 3.2 Generating non-constant effect

Let us now return to our maintained model, but *instead of Condition 1*, we now consider the following weaker restriction.

**Assumption 2:** *For a given quantile  $\theta_0$  and under Assumption 1:*

$$v_{t\theta_0} \text{ is independent of } x_{2t}, \text{ conditionally on } x_{1t}. \quad (6)$$

It is easy to translate this assumption in terms of conditional quantiles. Under Assumption 1, Assumption 2 is equivalent to:  $F_{v_{t\theta_0}|x_t}^{-1}(\theta) = F_{v_{t\theta_0}|x_{1t}}^{-1}(\theta)$ , for all  $\theta \in (0, 1)$ . Note also that Assumption 2 is implied by the general assumption of independence of  $v_{t\theta_0}$  with all the exogenous variables,  $x_{1t}$  and  $x_{2t}$ .

In this setting, Assumption 2 is akin to conditions in the control function literature, with here the control function known to depend on  $x_{1t}$  only. Complete exogeneity is not required, as discussed later. Assumption 2 is also related to the notion of ‘conditional exogeneity’ in White and Chalak (2010). Note that the restriction in Assumption 2 implies that the OLS in the reduced form may be inconsistent in the allowed case where the  $x_{1t}$  are endogenous. One may also have  $E(\psi_\theta(v_{t\theta})|x_{1t}) \neq 0$  under this hypothesis. This means that  $x_{1t}$  may be endogenous in the sense of the quantile regressions of quantile  $\theta$  for equation (3). In such situation, equation (3) no longer characterizes a typical ‘reduced form’ based only on exogenous regressors, although to simplify we still denote it the ‘*reduced-form equation*.’

In order to present more clearly our point, consider the following simplified model with  $x_{1t} = 1$  and  $\gamma_\theta = \gamma = -1$  for all  $\theta$ , and only one endogenous regressor  $Y_t$ . Assume

that  $(u_{t\theta}, V_t)$  are independent of  $x_{2t}$  and jointly normal centered with variances 1 and covariance  $\rho$ . In this case,  $F_{v_{t\theta}|x_{2t}}(v_{t\theta}|x_{2t}) = \Pr(v_{t\theta} \leq 0|x_{2t}) = \Pr(u_{t\theta} \leq V_t|x_{2t}) = \Pr(u_{t\theta} \leq V_t) = E(\Phi(V_t)) = 0.5$  that does not depend on  $\theta$  when conditioning only on  $x_{2t}$ . The reduced-form equation is then:

$y_t = \beta_\theta - (\Pi_1 + x'_{2t}\Pi_2 + V_t) + u_{t\theta} = (\beta_\theta - \Pi_1 + c_\theta) - x'_{2t}\Pi_2 + (u_{t\theta} - V_t - c_\theta)$ , where  $\Pr(u_{t\theta} - V_t - c_\theta \leq 0|x_{2t}) = \theta$ , which defines the constant  $c_\theta$ . The variation in  $\theta$  is clearly associated to terms where  $x_{2t}$  does not appear, here for the intercept.

Then, in this case it is possible to have a coefficient of  $x_{1t}$  (the intercept) that varies with  $\theta$  and coefficients of  $x_{2t}$  that are constant, under the independence assumption that has been stated. A similar argument can be made by relaxing the restrictions on values of parameters, on variables  $x_{1t}$ , and on independence conditions. As a matter of fact, a sufficient independence condition for consistent estimation can be parsimoniously expressed in terms of the reduced-form error  $v_{t\theta}$ , as we shall show. Of course, similar points can be made about the standard single step quantile regression.

In the remainder of this section, we show how non-constant effect can be obtained for conditional quantiles of the reduced-form, and then conveyed to the conditional quantiles of the structural form.

Since  $v_{t\theta} = u_{t\theta} + V'_{t\theta}\gamma_\theta$ , Assumption 2 for all  $\theta$  could be secured by assuming the stronger restriction that both  $u_{t\theta}$  and  $V_t$  are independent of  $x_{2t}$ , conditionally on  $x_{1t}$ . Thus, Assumption 2 can also be seen as the consequence of a natural, while strong, instrumental variable characterization of  $x_{2t}$  for the structural model.

We now show that Assumptions 1 and 2 imply that there is constant effect in the quantile regressions of the reduced-form equation for  $x_{2t}$ , but not necessarily for  $x_{1t}$ .

**Proposition 1:** *Under Assumptions 1 and 2, for a quantile regression process of the reduced form (3):*



- (a) *There is constant effect for the variables in  $x_{2t}$ .*
- (b) *Non-constant effect is possible for the variables in  $x_{1t}$ .*
- (c) *For all  $\theta$ ,  $F_{v_{t\theta_0}|x_{1t}}^{-1}(\theta)$  is linear in  $x_{1t}$ .*

The proof is in the Appendix. Result (c) is implied in particular by the popular ‘linear location-scale hypothesis’ in the quantile regression literature on non-constant effect (e.g., Card and Lemieux, 1996, and Koenker, 2005, pp 104-105). In cases where Result (c) would be judged a too unrealistic consequence, this can be easily relaxed by incorporating polynomial terms in  $x_{1t}$ , or splines, in the model, as is usual for approximating nonlinear functions. Alternatively, one could specify a reduced form (3) as being partially linear in  $x_{2t}$ , and possibly nonlinear in  $x_{1t}$ , with an unknown nonlinear functional form. In this case, the above approach delivering constant effect for  $x_{2t}$  and unrestricted (nonlinear) effect for  $x_{1t}$  would remain valid, and Result (c) would not be necessary anymore. However, this would drive us toward nonparametric estimation methods, which is not what we discuss in this paper. Finally, instead of imposing Assumption 2, one may first test for which coefficients the hypothesis of constant effect is not rejected in typical quantile regression estimation. This would guide the specification of Assumption 2 in the considered data. In the next subsection, we discuss the identification of the reduced-form parameters under Assumptions 1 and 2.

### 3.3 Identification

**Proposition 2:** *Under Assumptions 1 and 2, the components of  $\pi_\theta$  in the reduced form (3) for any quantile index  $\theta$  can be identified, for the coefficients of  $x_{2t}$ ,  $\pi_{2\theta} = \pi_2$ , while not necessarily for the coefficients of  $x_{1t}$ ,  $\pi_{1\theta}$ .*

The proof is in the Appendix. The reason for the results in Proposition 2 is that under  $v_t$  independent on  $x_{2t}$  conditionally on  $x_{1t}$ , only the coefficient of  $x_{2t}$  can be identified through solving these  $K_2$  orthogonality conditions, even conditionally on  $x_{1t}$ .

In these conditions, as discussed before, non-constant effect is possible for the conditional quantiles of the reduced form for  $\pi_{1\theta}$  for any  $\theta$ , even though Assumption 2 is made only for the value  $\theta_0$ . For example, a special case of non-constant effect quantile regression can be parameterized such as:  $\pi_{1\theta} = \pi_1 + \theta$ . However, even though the exhibited non-constant effect in the reduced form allows for generalization of the model as compared to constant effect cases, the non-constant effect coefficients cannot be identified with usual quantile regression settings. Though, the other coefficients of the model - in  $\pi_2$  - can be consistently estimated, while they do not vary with  $\theta$ .

One may also want to consider fully specified quantile regressions, while only at some quantiles. In that case, one may require Condition 1 for these quantiles only, keeping Assumption 2 for other quantiles. Similarly, Jun (2008) studies variations across quantiles of identification procedures through instrument variables. This suggests an interest in applying different semi-parametric IV restrictions at different quantiles. This would allow for many distinct models with (partial) non-constant effect.

Typically, Condition 1 or Assumption 2 can be imposed a priori to define the model, along with the linear functional form. Instead, one may wonder how to test them. We have seen that it is always possible to introduce the bias term  $F_{v_t\theta_0|x_t}^{-1}(\theta)$  so as to satisfy Condition 1. As a matter of fact, testing Condition 1 is tantamount to test for endogeneity related to the omission of this term in the original model. This can be done using Hausmann-type tests, for example based on the comparison of the

quantile regression estimator with the IVQR estimator of Chernozhukov and Hansen (2006, 2008). The test can also be seen as a test of linearity since what is tested is jointly the functional form of the model. In turn, Assumption 2 can be obtained by incorporating a restricted bias term in the model, this time  $F_{v_{t\theta_0}|x_{1t}}^{-1}(\theta)$ , while keeping a linear component in  $x_{1t}$  in the model. Testing Assumption 2 boils down to testing  $E(\psi_\theta(v_{t\theta})|x_{2t}) = 0$ , which can be done with a similar test as for Condition 1, adjusting these restrictions with IVs, and leaving aside the components in  $x_{1t}$  in the test null hypothesis, while not in the estimation.

Note that a well-known feature of the fitted-value approach is that it assumes that the reduced form is well specified. Misspecification of (3) jointly with Assumption 2 may yield to inconsistent estimation if some omitted variables imply that  $v_{t\theta_0}$  is not independent of  $x_{2t}$ , even conditionally on  $x_{1t}$ .

### 3.4 Transmitting the non-constant effect to the structural form

We now assess the consequences for the structural model of the partial occurrence of non-constant effect in the reduced form. As pointed out before, Assumption 2 implies that the parameter  $\pi_\theta = (\pi'_{1\theta}, \pi'_{2\theta})'$  in the reduced-form model (3) is such that  $\pi_{2\theta}$  is identified and corresponds to constant effect, while  $\pi_{1\theta}$  is not identified and may allow for non-constant effect.

**Proposition 3:** *Under Assumptions 1 and 2, for the structural quantile model (1) with the fitted-value setting based on the predictive equation (2):*

(a) *The coefficient vector  $\gamma_\theta$  of the endogenous independent variables  $Y_t$  in the structural model does not vary with the quantile index  $\theta$ :  $\gamma_\theta = \gamma$ , for all  $\theta \in (0, 1)$ .*

The endogenous independent variables have constant effect and the coefficient vector  $\gamma$  is identified.

(b) The coefficient vector  $\beta_\theta$  of the exogenous independent variables  $x_{1t}$  in the structural model can vary with the quantile index  $\theta$ . The exogenous independent variables may have non-constant effect. Moreover, the coordinates of  $\beta_\theta$  include exactly the same unknown inconsistency term as their corresponding coefficients in  $\pi_{1\theta}$  in the reduced-form model. The coefficient vector  $\beta_\theta$  is not identified in general.

The proof is in the Appendix. When a first-stage estimation is performed based on (2), the independent variables  $x_t$  consist of vectors  $x_{1t}$  and  $x_{2t}$ . If non-zero asymptotic inconsistency terms are present only in the coefficients of  $x_{1t}$  in the reduced-form estimator  $\hat{\pi}_\theta$ , which is the case on which we focus, then the non-zero asymptotic inconsistency in the second-stage estimator  $\hat{\alpha}_\theta = [\hat{\beta}'_\theta, \hat{\gamma}'_\theta]'$  is exclusively confined to the coefficients of  $x_{1t}$ ; that is, only  $\beta_\theta$  is not identified. Therefore, the coefficient  $\gamma_\theta$  of the endogenous variables  $Y_t$  in the structural equation can be identified. In this setting, because  $\pi_2$  is characterized by constant effect, and because  $\gamma_\theta$  is not connected to  $\pi_{1\theta}$ , we have also constant effect for  $\gamma_\theta$ . In contrast, a non-constant effect may occur for  $\beta_\theta$ .

Thus, allowing for the weakened IV condition in Assumption 2 has enabled us to introduce non-constant effect on the vector  $\beta_\theta$ , even though this parameter is not identified. This alone is a generalization of the *stricto sensu* constant effect structural quantile regression, which may be useful if the researcher's interest is concentrated on vector  $\gamma$  that is identified and can be estimated consistently. Indeed, this setting avoids misspecification of the quantile regression, if the true data generating process involves non-constant effect for  $\beta_\theta$  and constant effect for  $\gamma$ .

What if the interest of the researcher is also in the effect of exogenous regressors in

the structural model? Assuming instead that  $v_{t\theta_0}$  is independent of  $x_{1t}$ , conditionally on  $x_{2t}$ , and using the same proof technique, would deliver an identified constant effect for the exogenous variables, associated with a non-identified non-constant effect for the endogenous variables, which may respond to specific interests in the effect of exogenous variables. However, a strong and simultaneous interest in all the coefficients of the model would bring us back to the use of constant effect for all variables. The next section reports a real data example with numerical results.

## 4 An Empirical Example

Let us now consider an application to the study of food household expenditure in Egypt. The data is taken from the 2012 Egypt HIECS and cover 7527 households (United Arab Republic, 2014). The structural model (7) is a linearization of the model in Blundell, Chen, and Kristensen (2007).

$$y_t = \beta_{0\theta} + \beta_{1\theta}x_{1t} + \gamma_{\theta}Y_t + u_{t\theta}, \quad (7)$$

where  $\beta_{0\theta}$ ,  $\beta_{1\theta}$  and  $\gamma_{\theta}$  are parameters to estimate. Eq. (7) represents an Engel curve in which the household food budget share  $y_t$ , for each household  $t$ , is determined by household log total expenditure  $Y_t$  and by the number of children of age 14 years old or less within the household,  $x_{1t}$ . In these data, the food budget share varies from 1.4 percent to 81.5 percent, with a mean of 40.6 percent and a standard deviation of 11.0 percent. The mean annual total expenditure is 25,815 Egyptian Pound, not very far from the mean annual male income is of 29,965 Egyptian Pound. At this period, most incomes in Egypt still come from male workers. According to the Central Agency for Public Mobilization and Statistics, 70.4 percent of family income comes from work, while 15.9 percent is from cash and commodity transfers,

10.6 percent is the estimated rental value of the household dwelling. On average, 1.3 children are present in an Egyptian household, with a coefficient of variation close to 1. The log total expenditure variable  $Y_t$  is suspected to be endogenous, for example because of measurement errors in the expenditure data. Following Blundell, Chen, and Kristensen (2007) and Chen and Pouzo (2009), the (mostly male) log-earnings of household  $t$ , here denoted  $x_{2t}$ , can be used as an instrument for  $Y_t$ .

We are interested in the conditional quantile of  $y_t$  in (7). Hence, the error term  $u_{t\theta}$  should satisfy a conditional quantile restriction associated with a given quantile index  $\theta$ . However, as mentioned before, this restriction is rather made through the reduced-form model, accordingly with the fitted-value setting.

With this empirical setting, Assumption A1 reads:  $v_{t\theta}$  is independent of male earnings, conditionally on the number of children. That is: the correlation of any exogenous shock on total expenditure with male income exclusively goes through a change in the number of children. This allows for example, that an exogenous increase in the number of children changes the household preferences in the direction of relatively more food spending (e.g., on milk) and simultaneously stimulates male economic activity (e.g., because a larger family implies more needs to satisfy). In contrast, the usual approach of merely assuming the exogeneity of male income (i.e.,  $v_{t\theta}$  is independent of male earnings and of the number of children) does not allow such simultaneous causal channels.

Table 1 reports the estimates for the first-stage OLS; and the second-stage 2SLS and two-stage quantile regressions (based on the fitted-value setting), for the deciles. The standard errors, and the variance-covariance matrix of the quantile estimators across deciles, are computed using a bootstrap procedure, based on 300 replications. The bootstrap procedure encompasses the two estimation stages. That is: it also

accounts for the imprecision in estimating the fitted-value of the log total expenditure.

The estimated coefficients of total expenditure are significantly negative, which is consistent with food being an inferior good, as expected. Also corresponding to common notions, the number of children is related to a higher budget share devoted to food. This should be partly a consequence of lower living standards in larger families with the same total expenditure. This result may also stem from different household preferences, when children are present, and a stronger taste for food, e.g. for milk. The results of the usual Hausman test of exogeneity based on the comparison of OLS and 2SLS confirms the suspected presence of endogeneity in mean (Khi-Square statistic of 37.93 with two degrees of freedom, yielding a P-value almost equal to zero).

The results of Fisher tests of equality of coefficients across quantiles (based on the above bootstrap estimation of the variance-covariance matrix of the whole quantile process) show that the hypothesis of constant effect seems to be clearly rejected for the number of children (P-value of 0.000, Khi-Square statistic of 39.51 with 8 degrees of freedom), while not for the log total expenditure (P-value of 0.828).

A few words of comment are of order here. Since the full constant effect model appears to be rejected, the asymptotic formulae for the usual variance-covariance matrix estimator for the two-stage quantile regression based on fitted-value should not apply, since this approach is only valid under constant effect. Using the bootstrap procedure encompassing both estimation stages partly helps us avoiding this issue by not committing to any a priori functional form for the asymptotic variance-covariance matrix estimator. However, for the same reason, even the consistency of the estimators is in doubt under the typical quantile restriction in Condition 1. Yet, with assuming Assumption 1 instead, the consistency of the estimator of the constant effect of the log

total expenditure can be achieved. As a result, the bootstrap test of the hypothesis of constant effect for the log total expenditure is valid. In contrast, the akin bootstrap test applied to the hypothesis of constant effect of the number of children may be inconsistent since the estimator of this coefficient can only be consistent under the constant effect case. However, since the full constant effect model is rejected<sup>7</sup>, this can only come from a variable different from log total expenditure (which has constant effect). The only remaining regressor being the number of children, this implies that we can deduce that this variable has non-constant effect. Then, the initial impression is vindicated, even though the non-constant effect coefficient cannot be estimated consistently.

## 5 Conclusion

Individual heterogeneity in how some independent variables affect a dependent variable is nowadays central in econometrics and statistics. We have proposed a new way to incorporate heterogeneity in the effect of regressors through quantile regression. Namely, we have shown how some particular cases of non-constant effect can be obtained with two-stage quantile regressions based on the fitted-value setting under endogeneity. However, we have also established that only the coefficients of constant-effect variables can be identified, even though non-constant effect can be present for the other variables. Finally, we have discussed a few practical examples where our approach would be useful, including an empirical illustration.

Our results are based on relatively little demanding instrumental variable conditions, for example summarized so that the reduced-form errors are independent of

---

<sup>7</sup>Otherwise the two bootstrap test statistics would be consistent and would have indicated constant effects for all coefficients except the intercept.



*SOME excluded variables*, conditionally on the other independent variables. Such weakening of the usual IV conditions is potentially useful since convincing instruments are typically difficult to find and, as a consequence, reducing requirements for IV methods is valuable.

A few limitations still remain in the conducted analysis, and may deserve further investigation. First, some practical applications based on actual policy data characterized by constant effect of treatment variables would better elicit the concrete potential of the proposed method as compared to traditional approaches. Second, it would be desirable to relax the current partition between constant effect and non-constant effect, for endogenous and exogenous regressors, so that it would allow for more flexible specifications of the model. Third, as for traditional methods, finding instruments, even for weakened conditions, may remain a practical challenge in many fields of study.

Other challenges would arise from possible extensions of this work. First, more general specifications, especially nonlinear or involving some nonparametric component would be of interest, as would be some more varied weakened instrument hypotheses. A second interrogation is about whether some conclusions of this study would carry over other estimation methods than quantile regressions. Finally, designing test strategies for exploring restrictions in the heterogeneity of effects emerges as a fruitful research avenue.

## References

- [1] Abadie, A., J. Angrist and G. Imbens (2002), “Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings,” *Econo-*

- metrica*, 70, 91-117.
- [2] ADECRI (2008), “*The French Social System*,” ADECRI, Paris.
- [3] Amemiya, T. (1974), “The Nonlinear Two Stage Least Square Estimator,” *Journal of Econometrics*, 2, 105-110.
- [4] Amemiya, T. (1982), “Two Stage Least Absolute Deviations Estimators,” *Econometrica*, 50, 689-711.
- [5] Angrist, J. D., V. Chernozhukov and I. Fernandez-Val, “Quantile Regression under Misspecification, with an Application to the US Wage Structure,” *Econometrica*, Vol. 77, N. 2, 539-569.
- [6] Angrist, J. D. and A. B. Krueger (1991), “Does compulsory school attendance affect schooling and earnings?” *The Quarterly Journal of Economics*, November 1991, 106 (4), 979-1014.
- [7] Angrist, J. D. and J. Pischke (2009), “*Mostly Harmless Econometrics: An Empiricist’s Companion*,” Princeton University Press, Princeton.
- [8] Arias, O., K.F. Hallock and W. Sosa-Escudero (2001), “Individual Heterogeneity in the Returns to Schooling: Instrumental Variables Quantile Regression Using Twins Data,” *Empirical Economics*, 26, 7-40.
- [9] Amemiya, T. (1974), “The Nonlinear Two-Stage Least-Square Estimator,” *Journal of Econometrics*, 2, 105-110.
- [10] Blundell, R., X. Chen and D. Kristensen (2007), “Semi-nonparametric IV estimation of shape invariant Engel curves,” *Econometrica*, 75, 1613-1669.

- [11] Blundell, R. and J. Powell (2006), “Endogeneity in Nonparametric and Semiparametric Regression Models” in M. Dewatripont, L.P. Hansen and S.J. Turnovsky, eds., “*Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress, Vol. II,*” Cambridge University Press, Cambridge.
- [12] Card, D. and T. Lemieux, “Wage dispersion, returns to skill, and black-white wage differentials,” *Journal of Econometrics*, Elsevier, vol. 74(2), pages 319-361, October
- [13] Chen, L.-A. and S. Portnoy (1996), “Two-Stage Regression Quantiles and Two-Stage Trimmed Least Squares Estimators for Structural Equation Models,” *Commun. Statist.-Theory Meth.*, 25, 1005-1032.
- [14] Chen, X., O. Linton and I. Van Keilegem (2003), “Estimation of Semi-Parametric Models When the Criterion Function is not Smooth,” *Econometrica*, 71, 1591-1608.
- [15] Chernozhukov, V., Fernández-Val, I. & Kowalski, A. E., 2015. “Quantile regression with censoring and endogeneity,” *Journal of Econometrics*, Elsevier, vol. 186(1), pages 201-221
- [16] Chernozhukov, V. and C. Hansen (2005), “An IV Model of Quantile Treatment Effects,” *Econometrica*, 73, 245-261.
- [17] Chernozhukov, V. and C. Hansen (2006), “Instrumental Quantile Regression Inference for Structural and Treatment Effect Models,” *Journal of Econometrics*, 132, 491-525.
- [18] Chernozhukov, V. and C. Hansen (2008), “Instrumental Variable Quantile Regression: A Robust Inference Approach,” *Journal of Econometrics*, 142, 379-398.

- [19] Chernozhukov, V., G.W. Imbens and W.K. Newey (2007), “Instrumental Variable Estimation of Nonseparable Models,” *Journal of Econometrics*, 139, 4-14.
- [20] Chevapatrakul, T., T.-H. Kim and P. Mizen (2009), “The Taylor principle and monetary policy approaching a zero bound on nominal rates: quantile regression results for the United States and Japan,” *Journal of Money, Credit and Banking*, 41, 1705-1723.
- [21] Chesher, A. (2003), “Identification in Nonseparable Models,” *Econometrica*, 71, 1405-1441.
- [22] Chortareas, G., G. Magonis and T. Panagiotidis (2012), “The asymmetry of the New Keynesian Phillips Curve in the euro area,” *Economics Letters*, 114, 161-163.
- [23] Garcia, J., P.J. Hernandez and A. Lopez (2001), “How Wide is the Gap? An Investigation of Gender Wage Differences Using Quantile Regression,” *Empirical Economics*, 26, 149-167.
- [24] Hong, H. and E. Tamer (2003), “Inference in Censored Models with Endogenous Regressors,” *Econometrica*, 71, 905-932.
- [25] Imbens, G.W. and W.K. Newey (2006), “Identification and Estimation of Triangular Simultaneous Equations Models without Additivity,” Discussion paper, MIT Department of Economics.
- [26] Jun, S.J. (2008), “Weak identification robust tests in an instrumental quantile model,” *Journal of Econometrics*, 144, 118-138.
- [27] Jun, S.J. (2009), “Local structural quantile effects in a model with a nonseparable control variable,” *Journal of Econometrics*, 151, 82-97.

- [28] Kim, T. and C. Muller (2004), “Two-Stage Quantile Regressions when the First Stage is Based on Quantile Regressions,” *The Econometrics Journal*, 18-46.
- [29] Kim, T. and C. Muller (2013), “A Test for Endogeneity in Quantiles,” Working Paper Aix-Marseille School of Economics.
- [30] Kim, T. and C. Muller (2015), “Inconsistency Transmission and Variance Reduction in Two-Stage Quantile REgression,” Discussion Paper Aix-Marseille School of Economics.
- [31] Koenker, R. (2005), “*Quantile Regression*,” Cambridge University Press, New York.
- [32] Koenker, R. and G. Bassett (1978), “Regression Quantiles,” *Econometrica*, 46, 33-50.
- [33] Lee, S. (2007), “Endogeneity in Quantile Regression Models: A Control Function Approach,” *Journal of Econometrics*, 141, 1131-1158.
- [34] Ma, L. and R. Koenker (2006), “Quantile Regression Methods for Recursive Structural Equation Models,” *Journal of Econometrics*, 134, 471-506.
- [35] Powell, J. (1983), “The Asymptotic Normality of Two-Stage Least Absolute Deviations Estimators,” *Econometrica*, 51, 1569-1575.
- [36] United Arab Republic (2014), “Income and Expenditure Household Consumption Survey 2012/13. Vol. 111: Patterns of Household Expenditure According to Socio-Economic Characteristics,” January, Central Agency for Public Mobilization of Statistics, Cairo.

- [37] White, H. and K. Chalak (2010), “Testing a Conditional Form of Exogeneity,” *Economics Letters*, 109, 88-90.

## Proofs:

**Proof of Proposition 1:** Let us consider *another* quantile index  $\theta$  different from the previously chosen  $\theta_0$ , and let us impose Condition 1 for this  $\theta$ , that is: the typical quantile restriction for a quantile  $\theta$ . In this way, we can investigate the features of a quantile regression model centered on a quantile index  $\theta$ . Then, we examine how this conditional quantile restriction can coexist with Assumption 2 assumed with the chosen  $\theta_0$ .

The conditional quantile restriction at  $\theta$  is:  $P[v_{t\theta} \leq 0 | x_t] = \theta$ , i.e.,  $P[y_t \leq x_t' \pi_\theta | x_t] = \theta$ , which implies that  $P[x_{1t}' \pi_{1\theta_0} + x_{2t}' \pi_{2\theta_0} + v_{t\theta_0} \leq x_{1t}' \pi_{1\theta} + x_{2t}' \pi_{2\theta} | x_t] = \theta$ ,

where  $\pi_{1\theta_0}, \pi_{2\theta_0}, \pi_{1\theta}, \pi_{2\theta}$  are the respective components of  $\pi_{\theta_0}$  and  $\pi_\theta$  according to the partition of  $x_t$  into  $x_{1t}$  and  $x_{2t}$ . By regrouping, we obtain

$$P[v_{t\theta_0} \leq x_{1t}' (\pi_{1\theta} - \pi_{1\theta_0}) + x_{2t}' (\pi_{2\theta} - \pi_{2\theta_0}) | x_t] = \theta,$$

which in turn implies that  $F_{v_{t\theta_0}|x_t} [x_{1t}' (\pi_{1\theta} - \pi_{1\theta_0}) + x_{2t}' (\pi_{2\theta} - \pi_{2\theta_0})] = \theta$ . Finally, under Assumption 1, we have

$$x_{1t}' (\pi_{1\theta} - \pi_{1\theta_0}) + x_{2t}' (\pi_{2\theta} - \pi_{2\theta_0}) = F_{v_{t\theta_0}|x_t}^{-1}(\theta) = F_{v_{t\theta_0}|x_{1t}}^{-1}(\theta), \quad (8)$$

where the latter equality is obtained using Assumption 2. (a) Since  $F_{v_{t\theta_0}|x_{1t}}^{-1}(\theta)$  does not depend on  $x_{2t}$ , which has no zero component, equation (8) implies  $\pi_{2\theta} = \pi_{2\theta_0}$ . As a consequence, we can drop the dependence of  $\pi_{2\theta}$  on  $\theta$ , which becomes  $\pi_2$ . (b) Since there is no restriction on the effect of the variables in  $x_{1t}$ , their coefficients may vary with  $\theta$  in that case. To be consistent with (8), we must also have (c), which is therefore a consequence of the hypotheses. QED.

**Proof of Proposition 2:**

Recall that Assumptions 1 and 2 imply the linearity property ( $F_{v_{t\theta_0}|x_{1t}}^{-1}(\theta) = x'_{1t}\pi_{1F\theta}$ ). We only need to check what happens when applying Assumption 2 to the identification of the quantile regression estimator, instead of applying the complete quantile regression restrictions in Condition 1. Indeed, under Assumption 2, we have the following theoretical restrictions:

$$E(x_{2t}\psi_\theta(y_t - x'_{2t}\pi_{2\theta_0} - x'_{1t}(\pi_{1\theta_0} + \pi_{1F\theta}))) = 0, \quad (9)$$

while

$$E(x_{1t}\psi_\theta(y_t - x'_{2t}\pi_{2\theta_0} - x'_{1t}(\pi_{1\theta_0} + \pi_{1F\theta}))), \quad (10)$$

which is typically assumed to be zero for quantile regressions under Condition 1, is now left undetermined.

As in usual settings of quantile regressions, these orthogonality restrictions with respect to  $x_{2t}$ , (9), are satisfied by the quantile regression estimator under Assumption 2. In contrast, the restrictions of orthogonality with respect to  $x_{1t}$ , (10), may or may not be satisfied. Therefore, there is no full identification of the usual quantile regression estimator under Assumption 2. A formal way to see this is that under

Assumptions 1 and 2, we have:  $0 = E(\psi_\theta(v_{t\theta}^*)|x_t) = E(\psi_\theta(v_{t\theta_0} - F_{v_{t\theta_0}|x_t}^{-1}(\theta))|x_t)$

$= E(\psi_\theta(y_t - x'_t\pi_{\theta_0} - F_{v_{t\theta_0}|x_{1t}}^{-1}(\theta))|x_t) = E(\psi_\theta(y_t - x'_t\pi_{\theta_0} - x'_{1t}\pi_{1F\theta})|x_t) = 0$ . This implies:

$$E(x_t\psi_\theta(y_t - x'_{2t}\pi_{2\theta_0} - x'_{1t}(\pi_{1\theta_0} + \pi_{1F\theta}))) = 0. \quad (11)$$

This is the ‘linearity property’ of  $F_{v_{t\theta_0}|x_{1t}}^{-1}(\theta)$  that allows the definition of a non-constant effect ( $\pi_{1\theta_0} + \pi_{1F\theta}$ ). In contrast, the restriction ( $E(x_t\psi_\theta(y_t - x'_t\pi_\theta)) = 0$ ) defines what is estimated. Eq. (11) shows that  $\pi_{2\theta_0}$  and  $\pi_{1\theta_0} + \pi_{1F\theta}$  are identified for

quantile regression estimation at quantile  $\theta$ . Therefore,  $\pi_{1\theta}$  may be unidentified. The issue is that, at that stage, one does not want to identify  $\pi_{1\theta_0}$  but rather  $\pi_{1\theta}$ .

The reduced form for quantile index  $\theta$  is:  $y_t = x'_{1t}\pi_{1\theta} + x'_{2t}\pi_{2\theta} + v_{t\theta}$ . There is no reason why  $\pi_{1\theta}$  and  $\pi_{1\theta_0} + \pi_{1F\theta}$  should coincide in that case.  $\pi_{1\theta_0} + \pi_{1F\theta}$  is not necessarily the parameter one wants to estimate. Therefore, there may be an estimation bias in general for the coefficients of  $x_{1t}$ . Similarly, there is no reason why we should equal  $v_{t\theta}$  and  $v_{t\theta}^*$ . QED.

**Proof of Proposition 3:**

We decompose the link of the reduced-form and the structural-form parameters, by splitting system (4) into two blocks of equations, partitioning  $\Pi = \begin{bmatrix} \Pi_1 \\ \Pi_2 \end{bmatrix} = [\Pi'_1, \Pi'_2]'$ , according to the partition of  $\pi_\theta$ . We obtain:

$$\pi_{1\theta} = \beta_\theta + \Pi_1\gamma_\theta \tag{12}$$

and

$$\pi_{2\theta} = \Pi_2\gamma_\theta, \tag{13}$$

where  $\pi_{2\theta} = \pi_2$ , which does not include any inconsistency term, is identified and does not depend on  $\theta$ , as seen before. If the system is exactly-identified (i.e.,  $K_2 = G$ ), then  $\gamma_\theta$  can be directly expressed in terms of  $\Pi_2$  and  $\pi_2$ , which implies that  $\gamma_\theta$  is identified and does not depend on  $\theta$ .

For the remaining over-identified case (i.e.,  $K_2 > G$ ), there are more equations than are necessary to identify  $\gamma_\theta$ . However, similarly to the well-known analyzes of the two-stage least squares and the general method of moments,  $\gamma_\theta$  is still identified, providing the hypotheses made are valid, and the considered models are well specified. In that case, for example, only  $G$  arbitrary equations in (13) may be kept to define  $\gamma_\theta$ . Also, when estimation is considered, one can enhance the asymptotic performance of



the resulting estimators by choosing appropriate weighting matrices to combine the equations in (13). In either the exactly-identified case or the over-identified case,  $\gamma_\theta$  depends only on  $\Pi_2$  and  $\pi_2$ . As a consequence, the properties of identification and of constant effect of  $\pi_{2\theta} = \pi_2$ , which is without inconsistency term, are conveyed to  $\gamma_\theta \equiv \gamma$ .

Since  $\gamma_\theta = \gamma$  results from (13), and since  $\Pi_1$  is given and does not depend on  $\theta$ , (12) implies that  $\beta_\theta = \pi_{1\theta} - \Pi_1\gamma$  includes the non-constant effect from  $\pi_{1\theta}$ , with exactly the same non-identification problem and inconsistency term as for  $\pi_{1\theta}$ . Indeed, since  $\pi_{1\theta}$  includes an unknown inconsistency term,  $\beta_\theta$  includes also this term and is not identified, even though  $\Pi_1$  and  $\gamma$  are identified. QED.

**Table 1. Estimation Results for OLS, 2SLS and Quantile regressions**

	Intercept	Number of Children	Log Male Income
<i>First Stage</i>			
OLS	1.454 (.0465)	.0269 (.00181)	.8403 (.00459)
<i>Second Stage</i>			
OLS	1.312 (.0230)	.0159 (.000853)	-.0925 (.00230)
2SLS	1.249 (.0254)	.0156 (.000855)	-.0861 (.00255)
Quantile Regression (index $\theta$ )			
10	1.081 (.0351)	.0207 (.00127)	-.0827 (.00351)
20	1.155 (.0295)	.0185 (.00123)	-.0858 (.00296)
30	1.201 (.0373)	.0184 (.00114)	-.0873 (.00378)
40	1.242 (.0296)	.0172 (.000948)	-.0886 (.00294)
50	1.290 (.0387)	.0157 (.00105)	-.0907 (.00381)
60	1.321 (.0351)	.0149 (.00110)	-.0910 (.00348)
70	1.351 (.0387)	.0137 (.00106)	-.0910 (.00386)
80	1.384 (.0504)	.0129 (.00152)	-.0906 (.00501)
90	1.398 (.0381)	.01096 (.00121)	-.0869 (.00383)

7527 observations. Standard errors are in parentheses. All coefficients are significant at the 1 percent level.