Child Labor and the Wealth Paradox: the role of altruistic parents.

Luiz Renato Lima*, Shirley Mesquita†, Marianne Wanamaker‡

September 18, 2014

1 Appendix to Child Labor and the Wealth Paradox: The Role of Altruistic Parents.

In this appendix, we provide additional information on the econometric model used in the paper, the full list of covariates with their descriptive statistics, and additional empirical analysis regarding other covariates.

As stated in main text of the paper, we extended the Bhalotra and Heady (2003) approach towards quantile regression by using the Censored Quantile Instrumental Variable (CQIV) developed by Chernozhukov, Fernández-Val and Kowalski (2011). This estimator deals with the problems of censoring and endogeneity without assuming the classical linear normality and homoscedasticity assumptions. The equation of working hours will be estimated using the MCMC-simulated censored quantile regression (Chernozhukov and Hong, 2003), which is used to optimize the Powell (1986) objective function. Indeed, the MCMC is faster and always find the global minimum of that objective function. We use the same database for Pakistan and the regression specification as in Bhalotra and Heady (2003).

2 Econometric Methodology

Setting aside the censoring problem for now, the equation for hours of labor (H) can be written as:

$$\begin{aligned} H_i &= Z_i\beta + X_i\gamma + e_i \quad (1) \\ i &= 1, ..., N, \end{aligned}$$

*Corresponding author. Department of Economics, University of Tennessee at Knoxville, 527A Stokely Management Center (e-mail: llima@utk.edu).

†Department of Economics, Federal University of Paraiba-Brazil.

‡Department of Economics, University of Tennessee at Knoxville, 527A Stokely Management Center.

where (Z) is a vector of exogenous variables, (X) is an endogenous variable and N is the sample size.

Censoring and endogeneity are common problems in economic analysis, but in these circumstances the traditional OLS regression gives inconsistent estimates of the parameters of interest. To deal with both problems we will estimate the main equation using the Censored Quantile Instrumental Variable (CQIV) estimator, introduced by Chernozhukov et. al. (2011). This estimator combines censored quantile regression (CQR), developed by Powell (1986), to deal with censoring semiparametrically, with a control variable approach to allow for endogenous regressors.

According to Chernozhukov et. al. (2011), the basic idea is to add a variable to the main regression (1) such that, once we condition the model on this variable, regressors and error terms become independent. In brief, the CQIV estimator is obtained in two stages that are nonadditive in the unobservables. The additional variable is called the control variable and it is usually unobservable, thereby the first stage estimates a nonadditive model for this variable.

In the first stage, we define an auxiliary equation that describes (X) in terms of exogenous variables (Z_1):

$$X_i = Z_{1,i}\pi + u_i$$

And let:

$$e_i = u_i\alpha + \epsilon_i$$

As discussed in Bhalotra and Heady (2003), Z_1 is a vector of instrumental variables to consumption expenditure per capita. This vector includes the community unemployment rate together with indicators of the level of infrastructural development of the community, namely, the presence of railway, market and piped water and their interactions with the education of the head of household, in order not to lose the effect of variation in income within communities. We estimate the first stage using a standard quantile regression.

Substituting e_i in (1) we have the following conditional equation:

$$H_i = Z_i\beta + X_i\gamma + u_i\alpha + \epsilon_i$$

This conditional equation will be estimated in the second stage using a censored quantile regression model and including the estimated control variable to deal with the endogeneity problem. We also consider the exogeneity test developed by Smith and Blundel (1986), where the null hypothesis is that (X) is exogenous, i.e., $\alpha = 0$.

Estimation of censored quantile regression is done by minimizing the following criterion function first formulated by Powell (1986):

$$\min_{\beta,\gamma,\alpha} \sum_{i=1}^{N} \rho_\tau \left(H_i - \max\left(0, Z_i\beta + X_i\gamma + u_i\alpha\right)\right) \tag{2}$$

where $\rho_\tau(u) = u\left(\tau - I\left(u < 0\right)\right)$.

Powell's estimator provides valid inference in Tobin-Amemiya models without distributional assumptions and with heteroscedasticity of unknown form. However, as noticed by Buchinsky (1994) and Fitzenberger (1997), Powell's estimator suffers from computational instability when the number of parameters to be estimated is large which is exactly the case we have in hand because of the large amount of explanatory variables (see Appendix Table 1). We address this problem by estimating (2) with the MCMC-simulated censored quantile regression developed by Chernozhukov and Hong (2003).

This algorithm is attractive both theoretically and computationally. As explained by Chernozhukov and Hong (2003), the MCMC approach uses an adaptive Markov chain Monte Carlo sampling from the conditional parameter distribution. The method involves a class of simulation techniques from Bayesian statistics. In brief, the computational process simulates a series of parameter draws such that the marginal distribution of the series is (approximately) the quasi-posterior distribution of the parameters. The estimator is therefore a function of this series, and may be given explicitly as the mean or a quantile of the series, or implicitly as the minimizer of a smooth globally convex function.

This technique is more suitable than standard Tobit procedure, because it allows for the characterization of the whole conditional distribution of the dependent variable as a function of covariates. Thereby, it is possible to get a more complete mapping of the impact of a change in the size of the land on child labor as well as other explanatory variables. The classical linear assumptions of normality and homoscedasticity are no longer assumed, corresponding to an additional advantage of the proposed econometric technique. Finally, we reinforce that this proposed approach extends the method used by Bhalotra and Heady (2003) from the traditional estimation of the mean function towards estimation of the quantile functions.

3 Control Variables and Additional Empirical Results

The database for this study is built upon information from the Pakistan Integrated Household Survey (PIHS) 1991. This survey was done by the government of Pakistan in cooperation with the World Bank, as a part of the series of the Living Standards Measurement Research Study (LSMS) to developing countries with the aim of providing subsidy for policy makers and researchers. The discussion here on the data is the same found in Bhalotra and Heady (2003). The main regressors are the size of agricultural land measured in acres and its quadratic term. Following the literature, we included dummies to represent renting and sharecropping arrangements, which are the most common in Pakistan. The sample selected in this study is of 1,139 children between 10 and 14 years old living in a rural area and whose their households operate in a family farm, either as landowners or non-owner (rent or sharecrop). These families represent 49.8% of the total and of this total, 31% are land owners. Appendix Table 2 shows that in the group of selected families, the participation rate and the working hours are higher for both sexes when compared with the total sample of children in rural Pakistan.

Household food expenditure per capita is a proxy for family income, as is the educational level of each parent. Both are included to capture the effects of household resources. Since the incentive to put a child to work on the farm could be related to the available pool of family labor, household size and composition also appear as regressors. According to Bhalotra and Heady (2003), dummy variables for religion of the household are included to capture attitudinal differences in the valuation of work, in which the Muslim religion is assumed to be the reference category. We also include the relation of the child to the household head where the reference is child of the household head. In order to capture some local effects of differences in wages and prices, the estimating equation includes a set of dummies that identify the provinces of residence of the family and the average wage rate of adults in the agriculture sector. We also include dummies to identify whether primary schools are present in the community where the child lives as a proxy for school costs, and an indicator for public transportation in the community, as it may affect access to school. A summary of the variables used in the empirical model and their descriptive statistics are shown in Appendix Table 1.

As for our empirical results (see Appendix Table 3), we can add that regarding other covariates, the variable per capita food expenditure showed a negative effect on child labor. Regarding the characteristics of the children, we note that age is positively correlated with the probability of working. There is evidence linking the increase in child labor with age. This fact is generally associated with better employment opportunities and pay, and therefore the high opportunity cost of exclusive dedication of the child to education. The effect of the level of relationship to the head of the family changes with the sex of the child. When the child is the son of the head, this reduces the probability of working among boys and increases in girls. The land tenure type (mode of operation) has significant effects on child labor for a given acreage, mainly among boys. Female headship significantly increases child labor, while household size reduces child labor in every case. Father's secondary education significantly reduces both boys and girls' work. The mother's middle/secondary education tends to increase child hours of work, but this result may be biased, since only 0.3% of mothers have exactly this level of education. Overall, higher levels of parental education, result in lower probabilities of working. The presence of primary school and public transport both reduce the likelihood of girls to work and raise the likelihood for boys.

4 The Importance of Controlling for Endogeneity

We find that the estimated coefficient on the control variable, (α), is statistically significant, indicating that the exogeneity was rejected for food expenditure. In addition, the endogenous regressor and the error term became independent after the inclusion of α in the estimating equation, indicating that the models are robust to endogeneity.

In terms of inference, Appendix Table 4 shows that the results reported in the paper are significantly changed in terms of the value of the parameters and in its statistical significance when endogeneity is ignored, indicating the importance of the use of the method of instrumental variable to assess the impact of family resources on child labor.

References

[1] Bhalotra, S. and Heady, C. (2003). Child Farm Labor: The Wealth Paradox. *World Bank Economic Review*, 17(2): 197-227.

[2] Buchinsky, Moshe, "Changes in the U.S. wage structure 1963-1987: Application of quantile regression," *Econometrica* 62 (1994), 405-458.

[3] Chernozhukov, V., Fernández-Val, I., and Kowalski, A. (2011). Quantile Regression with Censoring and Endogeneity. NBER Working Paper, No. 16997.

[4] Chernozhukov, V. and Hong, H. (2003). An MCMC approach to Classical Estimation. *Journal of Econometrics*, 115: 293-346.

[5] Fitzenberger, Berned, "A Guide to Censored Quantile Regressions," In: Maddala, Gangadharrao and Calyampudi Rao (eds.), *Handbook of Statistics, Robust Inference*, North-Holland, Amsterdam (1997)

Tables and Figures

Table 1: Description of variables used in regressions

	Boys				Girls			
Variable	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
Dependent Variable								
Hours of Farm Work	7.7	15.6	0	77	5.2	11.0	0	77
Child Characteristics								
Age[1]	11.8	1.4	10	14	11.9	1.4	10	14
Child of Head[2]	0.8	0.4	0	1	0.8	0.4	0	1
Household resources								
Ln per capita food expenditure	4.8	0.6	2.1	6.5	4.8	0.6	3.4	6.5
Land[3]	8.1	18.3	0	157	6.4	14.1	0	150
Land Square[3]	398.5	2307	0	24649	240.2	1577.6	0	22500
Farm organization								
Rent[2]	0.1	0.3	0	1	0.1	0.4	0	1
Sharecrop[2]	0.4	0.5	0	1	0.3	0.5	0	1
Household structure								
Household size	10.9	5.9	3	40	10.7	5.1	3	40
Female head[2]	0.0	0.1	0	1	0.0	0.2	0	1
Male<9[2]	0.8	0.4	0	1	0.8	0.4	0	1
Male 15-19[2]	0.5	0.5	0	1	0.5	0.5	0	1
Male 20-59[2]	1.0	0.2	0	1	1.0	0.2	0	1
Male >60[2]	0.3	0.5	0	1	0.3	0.5	0	1
Female<9[2]	0.8	0.4	0	1	0.8	0.4	0	1
Female 15-19[2]	0.4	0.5	0	1	0.4	0.5	0	1
Female 20-59[2]	1.0	0.1	0	1	1.0	0.1	0	1
Female >60[2]	0.2	0.4	0	1	0.2	0.4	0	1
Parents' education								
Mother middle/secondary[2]	0.0	0.1	0	1	0.0	0.1	0	1
Father secondary[2]	0.0	0.2	0	1	0.1	0.2	0	1
Community variables								
Primary school. girls[2]	0.8	0.4	0	1	0.8	0.4	0	1
Primary school. boys[2]	0.9	0.3	0	1	0.9	0.3	0	1
Public transportation[2]	0.7	0.5	0	1	0.6	0.5	0	1
Ln male wage	3.6	0.3	3	4.3	3.6	0.3	3.0	4.3
Religion[2]	1.0	0.2	0	1	1.0	0.2	0	1
Punjab Province[2]	0.3	0.4	0	1	0.3	0.4	0	1
NWFP Province[2]	0.2	0.4	0	1	0.2	0.4	0	1
Balochistan Province[2]	0.0	0.2	0	1	0.1	0.2	0	1
Instrumental Var.								
Unem. Rate	4.4	0.4	4	5.2	4.4	0.5	4.0	5.2
Railway[2]	0.1	0.2	0	1	0.0	0.2	0	1
Market[2]	0.1	0.3	0	1	0.1	0.3	0	1
Piped water[2]	0.2	0.4	0	1	0.2	0.4	0	1
Unem. Rate * Head Education[1]	8.4	13.7	0	51.5	8.3	13.7	0	64.1
Sample Size		**595**				**544**		

Notes: in years[1]; dummy[2]; in acres[3]

Table 2: Child Farm Work in Pakistan - 10 to 14 years

	Participation		Weekly work hours	
	Total	Sample Selected	Total	Sample Selected
Boys	22%	31.7%	22.6 (18.5), N=266	24.8 (19.0), N=197
Girls	28%	39.1%	12.05 (12.9), N=307	13.53 (14.1), N=221

Notes: The weekly work hours values are conditional on participation in the activity during the reference week. The values in parentheses are standard deviations around the means. N is the number of working children..

Table 3: Child Work on the Household Farm - Model Coefficients

Variable	**Boys**			**Girls**		
	Tau 0.25	*Tau 0.5*	*Tau 0.75*	*Tau 0.25*	*Tau 0.5*	*Tau 0.75*
Child Characteristics						
Age	0.8453**	1.5481***	2.3221***	-0.1346	4.2424	0.6412***
	(0.4182)	(0.4150)	(0.0351)	(1.0209)	(3.0225)	(0.0233)
Child of Head	-2.5793*	1.3800	-1.2085***	0.1139	2.1690**	1.5213***
	(1.4528)	(0.8731)	(0.0494)	(0.8082)	(0.9611)	(0.0318)
Household resources						
Ln per capita food expenditure	-2.8514	-5.7157***	-6.5991***	1.0578	-7.9860	-2.5444***
	(1.7539)	(1.5667)	(0.0629)	(0.8117)	(6.6911)	(0.0312)
Land	-1.8873*	-2.2620***	0.1967**	-0.3018	-1.2971***	0.1429***
	(0.9773)	(0.7663)	(0.0946)	(0.5313)	(0.5014)	(0.0298)
Land Square	-3.2850***	-2.9710***	-0.0093**	-3.3605***	-4.2354**	-0.0013***
	(1.1534)	(0.9947)	(0.0042)	(0.9499)	(2.0465)	(0.0003)
Farm organization						
Rent	-6.5982**	-3.6586**	-0.9079***	4.2888***	-10.7977	1.5187***
	(2.7762)	(1.4973)	(0.0460)	(1.2467)	(7.2830)	(0.0188)
Sharecrop	6.6899***	6.1189***	5.8856***	-0.7633	30.9674	1.4317***
	(1.3337)	(1.1185)	(0.0699)	(1.4684)	(22.7282)	(0.0292)
Household structure						
Household size	-0.1132	-0.9522***	-0.2257***	-2.2022**	-1.2469	0.1045***
	(0.2899)	(0.2675)	(0.0497)	(1.0419)	(1.3432)	(0.0074)
Female head	13.0274***	13.8708***	11.4349***	0.5817	3.9405***	0.2251***
	(1.0807)	(0.8697)	(0.0490)	(0.4464)	(1.3294)	(0.0134)
Male<9	-0.8926	1.4133	-0.3532***	-2.2466***	7.0639	-1.3089***
	(0.7450)	(1.1466)	(0.0560)	(0.3944)	(7.2706)	(0.0122)
Male 15-19	-6.7718*	-3.1543***	-1.1573***	-3.3683***	-3.0708	-1.7854***
	(3.9910)	(1.2113)	(0.0846)	(0.6025)	(1.9178)	(0.0127)
Male 20-59	-0.2040	-1.0559	1.7629***	-5.2946***	-26.9911	-4.1191***
	(0.8203)	(1.3601)	(0.0285)	(0.6788)	(16.4332)	(0.0157)
Male>60	-3.1428***	-3.0340***	-0.9429***	2.8114***	8.0741	1.8788***
	(0.7159)	(0.7104)	(0.0495)	(1.0145)	(5.8524)	(0.0251)
Female<9	3.4235	4.6964***	0.6474***	4.3492***	6.1137***	3.1170***
	(2.2170)	(1.0844)	(0.0947)	(0.8717)	(1.6373)	(0.0169)
Female 15-19	-0.9219	-2.8917***	-1.8423***	0.9520	3.1537	-0.5333***
	(1.0827)	(0.5266)	(0.0311)	(1.1005)	(3.1973)	(0.0272)
Female 20-59	9.3740***	4.2630***	5.8716***	-3.2604**	-13.2107**	-5.7646***
	(2.4591)	(0.7372)	(0.1125)	(1.3328)	(5.4439)	(0.0099)
Female >60	-0.2909	-2.9420	-2.4776***	-1.7102**	2.4488	-2.8129***
	(2.3610)	(1.8736)	(0.0599)	(0.7523)	(3.4230)	(0.0452)
Parents' education						
Mother middle/secondary	53.2078***	50.5229***	48.8426***	0.5833	9.3286	1.4625***
	(2.4700)	(0.9363)	(0.0668)	(0.7601)	(5.9148)	(0.0179)
Father secondary	-2.3938**	-2.2113***	-0.8143***	-3.3569***	-13.3394**	-3.1594***
	(1.1457)	(0.5934)	(0.0486)	(0.9329)	(6.4406)	(0.0053)
Community variables						
Primary school, girls	-7.9248***	-2.6854	-5.3821***	-10.6320***	-6.2082**	-9.7471***
	(1.3251)	(2.4779)	(0.0710)	(0.4517)	(2.7637)	(0.0224)
Primary school, boys	5.7851***	5.3869***	4.1328***	2.3246	-6.3929	5.4656***
	(1.0294)	(0.5251)	(0.0429)	(1.5606)	(9.1241)	(0.0154)
Public transportation	4.1138**	3.9311**	0.4217***	-3.5169**	-4.9455*	-0.5497***
	(1.7523)	(1.5867)	(0.0890)	(1.7486)	(2.6319)	(0.0433)
Ln male wage	1.9137	2.0745**	0.1198	1.2064	0.0806	-0.4036***
	(1.7627)	(1.0057)	(0.0904)	(0.8089)	(1.2258)	(0.0096)
Religion	-3.0451*	-1.8281***	-1.9476***	-12.9737***	-14.3926***	-10.6367***
	(1.5955)	(0.5286)	(0.0567)	(0.9202)	(2.4358)	(0.0097)
Punjab Province	-1.5949	-1.7174*	-0.3313***	6.3181***	18.6707	4.1931***
	(1.4265)	(0.9027)	(0.0474)	(1.2819)	(12.5576)	(0.0192)
NWFP Province	1.0859	-4.9217*	1.4195***	0.0239	-7.4593	0.6154***
	(2.3705)	(2.7181)	(0.0684)	(0.8411)	(6.0268)	(0.0125)
Balochistan Province	-6.2395***	-5.1399**	-1.8187***	-7.9412***	-9.9927***	-7.2626***
	(2.3858)	(2.2184)	(0.0547)	(0.8325)	(3.8297)	(0.0259)
Constant	-13.6144***	-0.2624	11.2857***	21.4588***	24.8422***	31.8793***
	(3.7035)	(0.8847)	(0.0472)	(1.5246)	(2.9503)	(0.0242)
Log consumption (α)	1.7710**	3.3685***	4.9548***	1.4045**	4.2902*	2.8574***
	(0.8232)	(0.5037)	(0.0956)	(0.6126)	(2.5195)	(0.0217)
Sample Size	595	595	595	544	544	544

Notes: *p<0.10, **p<0.05, ***p<0.01; standard errors in parentheses.

Table 4: Model Coefficients - Estimation without control variable

	Boys			Girls		
	Tau 0.25	Tau 0.50	Tau 0.75	Tau 0.25	Tau 0.50	Tau 0.75
Land	-6.2664	-2.2669	0.2696**	-4.1167	-2.6849***	0.1407***
	(4.7636)	(1.5677)	(0.1347)	(3.1409)	(0.5391)	(0.0333)
Land Squared	-5.3015	-6.8770	-0.0103**	-5.4089*	-0.2447	-0.0010***
	(3.7871)	(5.4669)	(0.0048)	(2.8319)	(0.6283)	(0.0003)

Notes: *p<0.10, **p<0.05, ***p<0.01; standard errors in parentheses.