IV Estimation with weak instruments: an application to the determinants of school attainment in Peru

Luis García Núñez, Ph.D.*
Departamento de Economía
PUCP

RESUMEN

El método de variables instrumentales (VI) se ha convertido en uno de los más utilizados en econometría aplicada. Sin embargo, la literatura reciente ha mostrado que si los instrumentos no cumplen con las condiciones de relevancia y exogeneidad, al menos en una forma «fuerte», los resultados de la estimación podrían estar muy lejos de los verdaderos parámetros poblacionales, y que la interpretación de los resultados podría no tener ningún sentido. Este documento presenta este problema y muestra algunos *tests* para la verificación de la validez de los instrumentos usados. Adicionalmente, se realiza una aplicación de los mismos en la estimación del efecto de las horas de trabajo infantil y el trabajo en quehaceres del hogar sobre el logro educativo, usando un panel de 913 niños extraídos de la Encuesta Nacional de Niveles de Vida (ENNIV) de los años 1997 y 2000. En esta estimación, los instrumentos seleccionados pasaron las pruebas de relevancia y exogeneidad, por lo que podemos confiar en que los resultados son consistentes. Estos muestran un impacto positivo del trabajo en quehaceres del hogar sobre el logro educativo observado tres años después. Por otro lado, se advierte un impacto negativo del trabajo infantil pero no es significativo.

Palabras clave: instrumentos débiles, variables instrumentales, trabajo infantil, trabajo doméstico, educación.

ABSTRACT

The instrumental variables (IV) method has recently become widely popular in applied econometrics. However, recent work in the literature has shown that if instruments do not hold the relevance and exogeneity conditions at least in a «strong» way, the estimation results could be quite different from the population parameters, and an interpretation of the results would be meaningless. This paper presents this problem and shows some recent tests to verify the validity of the instruments employed. Additionally, I apply those tests to the estimation of the effects of child labor and household work on school attainment, by using a sample of 913 children from the Peruvian Living Standard Measurement Survey (LSMS) of 1997 and 2000. In this estimation, the instruments selected passed all the tests of exogeneity and relevance, so we can trust the consistency of these results. They show that household work performed by children has a positive impact on the children's observed educational performance three years later. In addition, the impact of child labor is negative, but it is not significantly different from zero.

Keywords: weak instruments, instrumental variables, child labor, housework, education.

* Correo electrónico: lgarcia@pucp.edu.pe

INTRODUCTION

The instrumental variables (IV) method has recently become widely popular in applied econometrics. As is well-known, such an econometric method consists of the use of a set of variables (called instruments) that —under some conditions— produce consistent estimators of the parameters in the linear regression model. However, if those conditions do not hold at least in a «strong» way, the estimation results could be quite different from the population parameters, and an interpretation of the results would be meaningless.

This paper summarizes that problem and shows some recent tests to verify how good the instruments could be. Additionally, I apply those tests of IV estimation to the determinants of school attainment using a panel of children from the Peruvian Living Standard Measurement Survey (LSMS) of 1997 and 2000. I regress the grade-age distortion in 2000 on the hours of child labor and the total hours spent on housework in 1997, plus additional socioeconomic and demographic characteristics. Clearly, those two variables on the right-hand side are endogenous since they are correlated to unobserved characteristics of the household, such as parent's preferences, so the IV method seems appropriate in this estimation.

The paper is organized in the following way. Section 1 summarizes the IV method for the case of two-stage least squares. Section 2 discusses the relevance condition and presents some tests to detect if the instruments selected meet this condition. Section 3 focuses on the exogeneity problem and presents some tests to verify the accomplishment of this condition. Section 4 presents an application of this technique in the estimation of the impact of child labor and household work on school attainment. Finally, the last section concludes.

1. THE PROBLEM OF ENDOGENOUS REGRESSORS AND THE IV METHOD

One of the assumptions of the classical linear regression model is that the conditional expectation of the error term vector (u) is equal to zero, conditioned to the observation of regressors data matrix X. This assumption implies that the covariance of X and u is zero.

A problem arises when the researcher suspects that this assumption is not true and that some of the variables in the data matrix are correlated to the error term. In this setting, the OLS (ordinary least squares) estimator is biased and inconsistent.

To explain the IV (instrumental variables) solution to this problem, let's write an econometric model using Stock and Yogo's (2002) notation, where n endogenous regressors Y have been included on the right-hand side plus K_1 exogenous regressors X:

$$y = Y\beta + X\gamma + u \tag{1}$$

where y is a $Tx1$ matrix, Y is a Txn data matrix of endogenous regressors, X is a TxK_1 data matrix of included instruments and is the $Tx1$ vector of errors.

The IV method in its form of the two-stage least squares (TSLS) proposes to use a set of Z variables not included in the regression used as instruments to predict the values of Y, so the auxiliary equation is:

$$Y = Z\Pi + X\Phi + V \tag{2}$$

where Z is a TxK_2 matrix of excluded instruments. Identification requires that $K_2 > n$.

The TSLS method says that each endogenous regressor should be regressed against the set of included and excluded regressors as shown in equation (2), and then their predicted values, $\hat{Y}$, should be used in equation (1) in place of Y. Finally, the estimation by OLS of equation (1) will produce consistent estimates of the parameters if Z holds two important conditions:[1]

(a) Relevance: Z must be correlated to Y. Formally, $\frac{1}{T}Z'Y \xrightarrow{p} \boldsymbol{\Sigma}_{ZY}$, a finite matrix with rank n.

(b) Exogeneity: Z must not be correlated to u. Formally, $\frac{1}{\sqrt{T}}Z'u \xrightarrow{d} N(0,\sigma^2\boldsymbol{\Sigma}_{ZZ})$, where $\boldsymbol{\Sigma}_{ZZ}$ is the probability limit of $\frac{1}{T}Z'Z$ and is also a finite matrix with rank K_2.

Let $\underline{Z} = [X\ Z]$ be the TxK matrix of all instruments, $K = K_1 + K_2$, and $\underline{X} = [Y\ X]$ the $Tx(n+K_1)$ matrix of all the right-hand side variables in the main equation. Then the IV estimator of the parameters in equation (1) is:

$$\hat{\delta}_{IV} = \begin{bmatrix} \hat{\beta} \\ \hat{\gamma} \end{bmatrix}_{IV} = \left(\underline{X}'P_{\underline{Z}}\underline{X}\right)^{-1}\underline{X}'P_{\underline{Z}}y \tag{3}$$

where $P_{\underline{Z}} = \underline{Z}\left(\underline{Z}'\underline{Z}\right)^{-1}\underline{Z}'$. In particular, we are interested in the vector of parameters $\hat{\beta}_{IV}$. Following Stock and Yogo's (2002) notation, let the superscript «$\perp$» denote residuals from the projection on X, so $Y^{\perp} = M_X Y$ and $Z^{\perp} = M_X Z$ are the residuals from OLS regression, where $M_X = I - X\left(X'X\right)^{-1}X'$. Then,

$$\hat{\beta}_{IV} = \left(Y^{\perp}{}'P_{Z^{\perp}}Y^{\perp}\right)^{-1}Y^{\perp}{}'P_{Z^{\perp}}y. \tag{4}$$

When the two conditions mentioned above hold, it is not difficult to show that $P\lim(\hat{\beta}_{IV}) = \beta$. This is not the case for the OLS estimator $\hat{\beta}_{OLS} = \left(Y^{\perp}{}'Y^{\perp}\right)^{-1}Y^{\perp}{}'y$, which is biased and inconsistent.

Obviously, if the problem of endogenous regressors does not exist, there is no need to use the IV method since the OLS method would give us consistent estimators and

[1] Taken from Hall *et al.* (1996).

more efficient estimators than the IV method. One way to check if the endogeneity problem exists is to conduct a test on exogeneity. The Wu-Hausman test is based on the discrepancy between the IV and the OLS estimator. The null hypothesis is that all variables on the right-hand side of the main equation are not correlated to the error term. The first step of the test is to save the OLS residuals from the first-stage regression. In the second step, those residuals are included in the main equation. If there are multiple endogenous regressors, a joint test of inclusion of those residuals is conducted. If those parameters are statistically significant, that would lead us to reject the null hypothesis of exogeneity; consequently, an IV regression is needed. Otherwise, the OLS estimation is preferred.

2. WEAK INSTRUMENTS

The relevance condition states that the excluded instruments must be correlated with the endogenous variables Y. If this condition does not hold (i.e., the instruments are weak), the linear IV estimates will be inconsistent, and the limiting distribution of the parameters may not be normal.

Going over equation (2), we observe that the instruments will be totally irrelevant if $\Pi = 0$. Consequently, a way to measure the weakness of the instruments is through the «concentration parameter»,[2] defined in the case of $n = 1$ as $\mu^2 = \Pi' Z' Z \Pi / \sigma_r^2$. This value can be understood as a population counterpart of a Wald test applied in the first stage that tests the null hypothesis that $\Pi = 0$. It is useful to express this value in its F-statistic form, which is μ^2 / K_2. Hence, small values of μ^2 / K_2 reveal that the instruments are weak.In the case of $n > 1$, instead of a single value, the analog of the F-statistic is a K_2 x K_2 matrix $G = \Sigma_{VV}^{-1/2}{}' \Pi' Z' Z \Pi \Sigma_{VV}^{-1/2} / K_2$, where Σ_{vv} is the covariance matrix of V. Small values of the minimum eigenvalue of G will mean that the instruments are weak.

It has been shown in the literature that $\hat{\beta}_{IV}$ is biased toward $P\lim(\hat{\beta}_{OLS})$. Indeed, if the instruments are irrelevant for predicting Y (which is equivalent to saying that $\Pi = 0$ in equation (2)), then $E\left(\hat{\beta}_{IV} - \beta\right) = P\lim \hat{\beta}_{OLS} - \beta$.

2.1. Detection of weak instruments

2.1.1. The «rule of thumb»

Some tests have been proposed to detect if the instruments are weak or not. The natural test when $n = 1$ is the F-test in the first stage, which tests the null hypothesis $\Pi = 0$. Staiger and Stock (1997: 557-586) proposed a «rule of thumb» when we have only one endogenous variable on the right-hand side of the equation. If the F-statistic is greater than ten, the instruments are strong; otherwise, they are weak. However, when we have

---

2 For further details, see Stock *et al.* (2002: 518-529).

more than one endogenous variable on the right-hand side, the F test and the «rule of thumb» are not good tests of relevance.

2.1.2. The Stock-Yogo test

Cragg and Donald (1993: 222-240) proposed a test on the identification of the model that Stock and Yogo (2002) used as a generalization of the F-test. The Cragg-Donald statistic is the minimum eigenvalue of

$$G_T = \hat{\Sigma}_{VV}^{-1/2}\,'\hat{\Pi}\,'Z^{\perp}\,'Z^{\perp}\hat{\Pi}\hat{\Sigma}_{VV}^{-1/2} / K_2 = \hat{\Sigma}_{VV}^{-1/2}\,'Y^{\perp}\,'P_{Z^{\perp}}Y^{\perp}\hat{\Sigma}_{VV}^{-1/2} / K_2$$

where $\hat{\Sigma}_{VV} = Y^{\perp}\,'M_{Z^{\perp}}Y^{\perp} / (T - K_2)$. If this statistic is greater than the critical values computed by Stock and Yogo, the null hypothesis of weak instruments is rejected.

To tabulate the critical values, Stock and Yogo (2002) defined the relative bias *b* of the IV estimator as a fraction or percentage of the OLS bias. For example, if *b=0.1*, it means that the IV bias (due to the weakness of the instruments) is 10% or less of the OLS bias (due to the endogeneity of the regressors). The critical values for the Stock-Yogo test vary across the values of *b.* As Stock and Yogo (2002: 15-16) noted, «[this definition] allows the researcher to return to the natural units of the application using expert judgment about the possible magnitude of the OLS bias». Lower values of *b* require greater values of the minimum eigenvalue to reject the null hypothesis.

2.1.3. The Anderson canonical correlations test[3]

There is another test of instrument relevancy, the Anderson canonical correlations test. It was developed based on the idea that the relevance of the instruments and identification of the model are closely related.

To exemplify this idea, suppose that the model presented in equations (1) and (2) does not have included instruments. Then the alternative equations are:

$$y = Y\beta + u \qquad (1')$$

$$Y = Z\Pi + V \qquad (2')$$

In this model, the IV estimator of *b* is just $\hat{\beta}_{IV} = \left(Y\,'P_Z Y\right)^{-1} Y\,'P_Z y$, with asymptotic covariance matrix $AVar(\hat{\beta}_{IV}) = \sigma^2 (Y\,'P_Z Y)^{-1}$.[4] Since $Y\,'P_Z Y$ is a *n×n* matrix, if the rank of Σ_{ZY} is less than *n*, then we would have an identification problem in the estimation of the covariance matrix. Therefore, to examine the relevance condition, it is necessary to study the diagonalization of *Z'Y*.

[3] The argument presented here was taken from Hall *et al.* (1996: 287-288). In most of the explanation, I follow their notation.

[4] Notice that $AVar(\hat{\beta}_{IV}) = \sigma^2 (Y\,'Z(Z\,'Z)^{-1}Z\,'Y)^{-1}$.

To do this task, we define the squared canonical correlations r_i^2, $i=1,...,n$ of Z and Y, as the eigenvalues of $(Y'Y)^{-1}(Y'Z)(Z'Z)^{-1}(Z'Y)$ with associated eigenvectors α_i, and let A be the matrix whose columns are those eigenvectors. Those squared canonical correlations are also the nonzero eigenvalues of $(Z'Z)^{-1}(Z'Y)(Y'Y)^{-1}(Y'Z)$ whose associated eigenvectors are ξ_i, and the matrix of those eigenvector is G. Then,

$$A'Y'ZG = \Lambda_n$$

where $\Lambda_n = diag(r_1, r_2, \ldots, r_n)$, where the correlations have been ordered from higher to lower. As a result, the analysis of poor instruments is reduced to the observation of the canonical correlations. If at least one of them is close to zero, we will say that the instruments are poor and the model is nearly under-identified.

At this point, it is easy to imagine that the test of the relevance of the instruments is a test of the significance of the smallest canonical correlation. Hall *et al.* (1996) proposed the Likelihood Ratio test (LR):

$$LR = -T\log\left(1 - r_n^2\right)$$

where r_n is the smallest sample canonical correlation. Under the null hypothesis of under-identification/low relevance (the smallest canonical population equals zero), LR is asymptotically distributed as $\chi^2(K_2 - n + 1)$. Rejection of the null hypothesis is a signal that the instruments are relevant.

However, Hall *et al.* (1996) did not suggest the LR statistic as a criterion to select instruments. Hall *et al.* showed that choosing instruments from a large set of candidates in order to maximize the LR statistic could exacerbate the effects caused by the endogenous regressors.

2.1.4. The redundancy test[5]

Based on canonical correlations, Hall and Peixe (2000) constructed a test to check if some of the excluded instruments are redundant. An instrument is redundant if its inclusion in the excluded instruments set has no impact on the asymptotic variance-covariance matrix of the estimators.

To present the test, let us partition matrix Z in equation (2') into

$$Y = Z_1\Pi_1 + Z_2\Pi_2 + V \qquad (2'')$$

where Z_1 is a Txq_1 matrix, Z_2 is a Txq_2 matrix, and $q_1 + q_2 = K_2$.

Given the Z_1 data matrix of the relevant instruments, we would like to test if the variables contained in Z_2 are irrelevant. Assuming that the error terms are independently

[5] All the ideas in this section can be found in Hall and Peixe (2000): 10-11. I have adapted their notation to that used in this paper.

normally distributed and under the null hypothesis that $\Pi_2 = 0$ (the variables contained in Z_2 are redundant given the variables in Z_1), the statistic is

$$LR_T = -T\sum_{i=1}^{n} \ln(1 - r_i^2) + T\sum_{i=1}^{n} \ln(1 - \tilde{r}_i^2)$$

where r_i are the sample canonical correlations between Y and Z, and $\tilde{r}_i$ are the sample canonical correlations between Y and Z_1. Under the null hypothesis, $LR_T \xrightarrow{d} \chi^2_{nq2}$. Rejection of the null hypothesis would mean that the evaluated instruments are not redundant.

2.2. Inference

When instruments are weak, inference should be carried out with caution. The typical t-statistic may be meaningless if the weakness is severe. Two approaches have arisen in recent literature to deal with this problem. One requires correcting the bias in estimates and standard errors to improve the normal approximation. The second approach just presents the estimates and uses confidence intervals and tests that are fully robust to weak instruments (they have the correct size regardless of the value of the concentration parameter) (Stock *et al.* 2002: 523).

The Anderson-Rubin test is an example of this type of test.[6] The statistic that tests the null hypothesis $H_0 : \beta = \beta_0$ is

$$AR(\beta_0) = \frac{(y - Y\beta_0)' P_{Z^\perp} (y - Y\beta_0) / K_2}{(y - Y\beta_0)' M_{Z^\perp} (y - Y\beta_0) / (T - K_2)}$$

Under the null hypothesis and weak instruments, $AR(\beta_0) \xrightarrow{d} \chi^2_{K_2} / K_2$. If $\beta_0 = 0$, it tests if the parameters of the endogenous variables on the right-hand side of the main equation are jointly significant. This test is good when we have only one endogenous regressor, but the test's power declines when two or more endogenous regressors are present (Andrews and Stock 2005: 10).

Other tests have been proposed in this field, such as the Lagrange multiplier (or score) Kleibergen's statistic, the Wald statistic and Moreira's likelihood ratio test. All show similar results to the Anderson-Rubin test.[7]

---

6 See also Cruz and Moreira (2005).

7 See Cruz and Moreira (2005: 401- 402). Unfortunately, they will not be conducted in this paper because the current STATA v.9 do-files do not estimate them with more than one endogenous regressor. See also Stock *et al.* (2002: 523).

3. INSTRUMENT EXOGENEITY

The second condition for good instruments says that they must not be correlated with the error term in the main equation, or in other words, they have been correctly excluded from the main equation. If they have any effect on the endogenous variable on the left-hand side, that effect should occur through the effect on the endogenous variables on the right-hand side. This condition is also related to the identification of the model.

3.1. The Sargan test

The Sargan statistic is used to test if instruments are correlated or not with the error term where the null hypothesis is that instruments are exogenous. From the parameter estimated in equation (3), the Sargan statistic is:

$$\text{Sargan statistic} = T \cdot \frac{(y - X\hat{\delta}_{IV})' P_{Z^{\perp}} (y - X\hat{\delta}_{IV})}{(y - X\hat{\delta}_{IV})'(y - X\hat{\delta}_{IV})}$$

which equals TxR^2 from a regression of the IV residuals on the set of instruments. Under the null hypothesis that all the instruments are valid (not correlated to the error term and that the excluded instruments were correctly excluded from the main equation), Sargan's statistic converges in distribution to $c^2(K_2 - n)$. If the value of this statistic is large, it would lead us to reject the null hypothesis, which would imply that some instruments *could be* correlated to the error term.[8] It is important to point out that this test is constructed under the assumption of conditional homoskedasticity of the error term.

3.2. The «difference-in-Sargan» test

On the basis of the Sargan statistic, we could also conduct the «difference-in-Sargan» statistic or C-statistic.[9] We use this statistic to perform a test on a subset of the orthogonality conditions. Suppose that we suspect a set of instruments may not be valid, i.e., may not be exogenous. As before, let $\underline{Z}$ be the TxK matrix of the full set of instruments and $\underline{Zs}$ be a $TxKs$ matrix containing a subset of valid instruments, where $K>Ks$. Our intention is to test if the remaining K-Ks instruments are valid or not. Intuitively, when instruments are exogenous, Sargan statistic is a small number, so the inclusion of non-valid instruments should increase the magnitude of this statistic. Let $\hat{\delta} = (X' P_Z X)^{-1} X' P_Z y$ be the IV estimator using $P_{\underline{Z}} = \underline{Z}(\underline{Z}'\underline{Z})^{-1}\underline{Z}'$ and $\bar{\delta} = (\underline{X}' P_{\underline{Zs}} \underline{X})^{-1} \underline{X}' P_{\underline{Zs}} y$ the IV estimator using $P_{\underline{Zs}} = \underline{Zs}(\underline{Zs}'\underline{Zs})^{-1}\underline{Zs}'$. Then the C-statistic is:

---

[8] Failure of other assumptions in the model could also lead to a rejection of the test, even if the instruments are exogenous.

[9] See Hayashi (2000: 232-233).

$$C = \frac{\hat{u}\,'P_{\underline{Z}}\hat{u} - \overline{u}\,'P_{\underline{Zs}}\overline{u}}{\hat{\sigma}^2}$$

where $\hat{u} = y - \underline{X}\hat{\delta}$, $\hat{u} = y - \underline{X}\hat{\hat{\delta}}$ and $\hat{\sigma}^2 = \hat{u}\,'\hat{u}\,/\,T$. Under the null hypothesis that the *K* and the *Ks* sets of instruments are valid, *C* is asymptotically c^2 (*K*- *Ks*).

Notice that the variables tested may be either excluded or included instruments. If excluded instruments are being tested, rejection of the null hypothesis means that those instruments are not a valid set of instruments because they do not meet the exogeneity condition. On the other hand, if included instruments are being tested, rejection of the null hypothesis would lead us to conclude that they are endogenous variables in the main equation because they would be correlated to the error term.[10]

4. APPLICATION: THE EFFECT OF CHILD LABOR AND HOUSEHOLD WORK ON SCHOOL ATTAINMENT

In this section, I apply the instrumental variables method to the estimation of the determinants of school achievement measured by the school-age distortion. More specifically, I am interested in the effect of the hours of child labor and household work on this variable.

It is widely accepted that child labor is dangerous to children younger than fourteen years old, who are exposed to exploitation and abuse. In addition, the time and effort spent on these activities negatively affect school performance and eventually may cause children to repeat grades and drop out of school. In addition, data show that many children in developing countries also spend a considerable amount of time working or doing household chores. For example, in Peru, one out of every four children is engaged in some kind of economic activity, and three of every four children do household work, which is understood as those chores performed at home, such as cleaning, cooking, etc.

There are some positive effects of work at an early age such as a better management of time and resources, the acquisition of experience, maturation, and independence; however, it is accepted that the negative effects might surpass the positive ones.

It is important to highlight the effect on human capital accumulation. The current educational indicators depend on the flow of past time allocations. For example, it is likely that a child who was a child worker will be behind compared to other students of the same age group who were always fulltime students. As a consequence, we could observe the effect of child labor and household work on a child's future performance at school.

---

[10] See STATA help for *ivreg2* command.

4.1. Literature review

Most of the studies in this area focus on short-term effects when they regress, for example, school attendance on hours of child labor using cross-section data. Only few papers in the child labor and youth labor literature have studied directly the consequences of past child labor on current education attainment. In a recent paper, Beegle *et al.* (2005: 19-20) use data from Vietnam to find the impact of child labor five years later on school participation, educational attainment (highest grade completed), occupation, earnings and health. This empirical model includes an explanatory dummy variable whether the child did or did not work five years ago. According to the authors, that dummy may cause some potential biases in estimates, so they use an instrumental variable approach to deal with this problem. Neither household work nor contemporary values of child labor are considered in the regressions. Their results show a negative and significant effect of child labor on school attendance and the highest grade completed five years later.

In a study using data from the United States, Carr *et al.* (1996: 72-73) study the effect a decade later of high school work on educational attainment, wages and participation in the labor market. They find a negative effect of high school work on the level of education attained but a positive effect on wages, labor force participation and employment status. In contrast, using a shorter horizon, other studies have found different results such as D'Amico (1984: 160-161) who found that, for the female group, working at high school reduces the probability of dropping out school (except for white females who worked more than 20 hours a week). In the case of males, working less than 20 hours a week reduces the probability of dropping out school, but for individuals who work more than 20 hours a week it is more likely to drop out high school (except for white males).

There are several papers that focus their attention on the short-run effect of child labor on school achievement. All of them use cross-section data and, as a consequence, include contemporary values of child labor and school achievement only. For example, Psacharopoulos (1997: 377-386) studies the effect of child labor on years of schooling using data from two Latin American countries. He finds that child labor reduces schooling by two years on average. In other work, Jensen and Skyt (1997: 407-424) analyze the determinants of school attendance, under the assumption that there is an exact trade-off between child labor and hours of study (provided there are no more activities during the day). Ravallion and Wodon (2000: c158-c175) do not make this assumption and ask if child labor displaces schooling. They find that a school-price subsidy increases schooling but has limited effect on child labor. Patrinos and Psacharopoulos (1997: 387-405), using data from Peru, do not find evidence that child labor influences the age-grade distortion. Finally, Levison and Moe (1998: 339-356) state that child labor is not the only obstacle to schooling. Rather, household work is also an important deterrent to schooling, especially in the case of girls.

Another group of papers measures the impact of child labor on learning achievement. Gunnarson *et al.* (2003: 17-22) analyze the impact of child labor on language and math scores in eleven Latin American countries, finding a negative impact. In a similar work and using data from Ghana, Heady (2003: 385-398) includes in his regressions two kinds of child labor (at home and outside home) and child housework. He finds that child labor has a negative and significant effect on math and reading scores. Household work also has a negative effect, but only has a significant effect on the easy math test. In an earlier work, Akabayashi and Psacharopoulos (1999: 132-138) also analyze the effect on reading and math scores in the Tanzanian case. They find a negative effect of child labor on these scores, but their results are not robust when other variables are included in the regression such as school attendance and hours of study.

4.2. The age-grade distortion index

The age-grade distortion, or SAGE, is measured as

$$SAGE = \frac{S}{A-E} * 100$$

where S is years of schooling, A is age and E is entry age to school. Usually, SAGE is a real number between 0 and 100, where SAGE=100 means that the individual has a good performance and he has not repeated any year or dropped out of school. However, in a few cases, it could be the case that SAGE>100 because some children might start their education at an earlier age than the entry age. If SAGE is low (close to zero), it is a sign that this child has stopped studying for some years or has had a very low performance.[11] It is desirable that SAGE be close to 100.

4.3. The econometric model

Let y_{it} be the value of the age-grade distortion index for individual i at time t, l_{it-1} be the total hours of child labor (market work) of individual i at time $t-1$, h_{it-1} be the hours of household work of individual i at time $t-1$, and X_i be a vector of other variables. The main equation to be estimated is:

$$y_{it} = \beta_1 l_{it-1} + \beta_2 h_{it-1} + \gamma' X_{it} + u_{it}$$

In this model, the effect of child labor and household work on school attainment is assessed by the significance of the parameters β_1 and β_2.

[11] In the Peruvian education system, a student who has a very low performance and his/her grades are low or below a minimum standard during the year must enroll in the same grade the next academic year.

Both l_{it-1} and h_{it-1} are endogenous because they depend on some unobserved characteristics related to household preferences. For example, these variables depend on the head of household attitude to send his/her child to study or work, which could be related to the expected educational level he or she desires for his/her child.

This endogeneity problem is treated in this model using the IV method. To do so, the equivalent expressions for equation (2) in this model are

$$l_{it} = \Pi_l \text{'} Z_{it} + \Phi_l \text{'} X_{it} + v_{it}^l$$

$$h_{it} = \Pi_h \text{'} Z_{it} + \Phi_h \text{'} X_{it} + v_{it}^h$$

where v_{it}^l and v_{it}^h are error terms that could be correlated between them, and could also be correlated to u_{it}.

4.4. Data

The data were taken from a panel of individuals who participated in the Peruvian Living Standards Measurement Survey (LSMS) 1997 and 2000. These surveys include detailed information of hours of child work, hours of child household work, household characteristics, years of schooling, wages, as well as other socioeconomic characteristics.

From that sample, a sub-sample of individuals in age ranging from eleven to sixteen in 2000 was selected; hence, the unit of analysis is the child. Due to insufficient information or errors in the codification of the sample, many observations were lost. The total number of individuals selected for the sample is 926, where two or more could belong to the same family. Due to missing data in some of the variables, the regressions were estimated with 913 individuals. Finally, 209 clusters were found.

Table 1 shows a brief summary of the descriptive statistics of the variables used in the estimation.

4.5. Results

The model was estimated using OLS and IV under a couple of different specifications. In addition to the estimation of the parameters of the model, special attention has been paid to the tests mentioned in previous sections. Since these are weighted data from a clustered survey, weights were used plus the command that takes into account the clusters. In addition, all regressions were run with robust standard errors.

Table 1
Definition of variables and descriptive statistics

Variables	Definition	Mean	Std.	N° obs.
SAGE	School-Age ratio in year 2000	73.202	1.066	913
Hours of housework	Average weekly hours of housework in 1997 (including zeroes)	9.118	0.319	913
Hours worked	Average weekly hours of work in an economic activity in 1997 (including zeroes)	3.854	0.504	913
Hours of housework (b)	Average weekly hours of housework in 1997 (excluding zeroes)	10.565	0.317	806
Hours worked (b)	Average weekly hours of work in an economic activity in 1997 (excluding zeroes)	13.787	0.842	230
Sex	Dummy: 1 if child is male, 0 if child is female	0.502	0.016	913
Age	Child age in year 2000	13.412	0.060	913
Head's education	Head of household years of schooling in year 2000	7.628	0.272	913
Household size	Household size in year 2000	6.485	0.111	913
Child's chronic disease	Dummy: 1 if child suffers a chronic disease, 0 otherwise	0.071	0.011	913
Head's acute disease	Dummy: 1 if household head suffered of acute illness in the past 4 weeks, 0 if not	0.062	0.010	913
Head's chronic disease	Dummy: 1 if household head suffers of chronic illness, 0 otherwise	0.174	0.019	913
Number of adults	Number of adults in household, in year 2000 (parents included)	3.003	0.085	913
No spouse	Dummy: 1 if there is no spouse in household, 0 if there is spouse in 1997	0.164	0.019	913
Farming	Dummy: 1 if the child works in farming, cattle raising, hunting and fishing; 0 otherwise	0.191	0.030	913
Sales	Dummy: 1 if the child works in wholesales, retail sales; 0 otherwise	0.050	0.008	913
Firewood	Dummy: 1 if household used firewood as source of energy for cooking in 1997, 0 otherwise	0.392	0.040	913
Public school	Dummy: 1 if the child studied primary in a public school, 0 otherwise	0.927	0.013	913
Rural	Dummy: 1 if the child lives in a rural area, 0 if the child lives in an urban area	0.435	0.045	913
Log(non-labor income)	Log of weekly non labor income per head in 1997	2.156	0.043	913
Log (vpagr)	Log of monthly value of production assigned for agricultural economic activity in year 2000	1.857	0.225	913
Log (vpnagr)	Log of monthly value of production assigned for non agricultural economic activities in year 2000	3.519	0.185	913
Excessive housework	Dummy: 1 if the child does housework for more than fourteen hours a week, 0 otherwise	0.123	0.013	913
Excessive child labor	Dummy: 1 if the child works more than fourteen hours a week, 0 otherwise	0.091	0.017	

Source: LSMS (Instituto Cuánto 1997, 2000).
Elaboration: owner.

All these results are shown in table 2. The first column shows the regression by OLS, and the other three columns are IV regressions, called models 1, 2, and 3. The objective is to show how the results change as we switch from an OLS estimation (biased and inconsistent) to an IV estimation (consistent). The model was estimated under three alternative models. The first one includes the same variables as those in column one. Model 2 adds a couple of variables that capture the idea that excessive housework and child labor may negatively affect the SAGE formula. Model 3 adds to model 1 the dummy of the child's gender. In model 1, that dummy is an excluded instrument, but in model 3, the dummy is an included instrument, since it may be argued that there could be gender differences in school attainment in Peru, i.e., gender, not only child labor and housework, affects directly school attainment.

I tried several possible instruments that meet the relevance and exogeneity conditions and that were available in the data. After several attempts, four instruments were used in models 1 and 2: the child's gender, if the child works on a farm or something related to agriculture, if the child works in wholesale or retail sales, and if the child studied primary school education in a public school. It is well-known in the literature of child labor that there are some important gender differences in child work and household work; nonetheless, it is not totally clear that child gender may directly affect school attainment. Consequently, child gender was used as an excluded instrument for models 1 and 2, and was introduced (just to show that it does not work well) as an included instrument in model 3.

In the case of the economic sector to which the child belongs, I chose those instruments because in Peru it is common to observe non-paid family workers in those activities. Therefore, they could be used as good predictors of child labor and child housework, but there is nothing that tells me that those two variables directly affect school attainment. Finally, concerning the last instrument (if the child received primary school education in a public school), one could reasonably think that it may directly affect the level of school attainment since it is known in Peru that this kind of education has a lower quality level compared to private schools. However, statistic tests (shown later) show that this instrument worked better as an excluded instrument rather than as an included one. Perhaps, children who study in public schools (a cheap and low-quality alternative for the poor) are also exposed to a higher incidence of child labor.

The results in table 2 show that *hours of housework* have a positive and significant effect (at the 5% level) on school attainment in models 1 and 2. This result is consistent with the idea that housework at early ages promotes a sense of responsibility and maturity, which would have a positive effect on schooling results. On the other hand, the impact of child labor is not significant; however, the sign of the estimated parameters is negative, as expected. A plausible explanation for this weak effect of child labor on school attainment is the small number of hours that children work at ages eight to thirteen.[12]

[12] Hours of work were measured three years before the schooling result. Since our group of study is children ages 11 to 16 in 2000, they were in the range 8 to 13 years in 1997. To give some figures, the average weekly

Table 2
Results of estimation by OLS and IV

Dep. Variable: SAGE 2000	OLS	IV		
		Model 1	Model 2	Model 3
Hours of housework	0.095 (0.096)	0.758* (0.359)	1.357* (0.582)	0.784 (0.689)
Hours worked	0.081 (0.113)	-0.055 (0.189)	-0.144 (0.317)	-0.056 (0.186)
Age	2.326** (0.521)	1.656** (0.612)	1.680** (0.575)	1.633* (0.823)
Head's education	1.159** (0.190)	1.229** (0.193)	1.172** (0.206)	1.232** (0.207)
Household size	-2.177** (0.600)	-2.247** (0.587)	-2.128** (0.610)	-2.248** (0.590)
Child's chronic disease	-8.359** (2.698)	-7.623* (3.041)	-7.025* (2.996)	-7.612* (3.037)
Head's acute disease	-6.777* (3.142)	-6.901* (3.158)	-5.256+ (3.015)	-6.940* (3.201)
Head's chronic disease	1.327 (2.001)	1.942 (2.049)	0.669 (1.996)	1.969 (2.083)
Number of adults in household	2.641** (0.855)	3.098** (0.879)	3.222** (0.888)	3.120** (1.075)
No spouse	-2.038 (2.340)	-3.016 (2.385)	-3.396 (2.596)	-3.044 (2.589)
Rural	-0.296 (2.453)	-0.718 (2.710)	-1.697 (2.841)	-0.744 (2.797)
Firewood	-6.081* (2.371)	-4.885+ (2.547)	-4.323+ (2.453)	-4.835+ (2.653)
Log(non-labor income)	1.983* (1.003)	1.994* (1.007)	1.945+ (1.005)	5.05+ (1.034)
Log (vpagr)	-0.468+ (0.274)	-0.578* (0.291)	-0.582* (0.285)	-0.582+ (0.304)
Log (vpnagr)	-0.131 (0.243)	-0.097 (0.255)	-0.241 (0.272)	-0.097 (0.255)
Excessive housework	- -	- -	-24.257* (9.971)	- -
Excessive child labor	- -	- -	3.658 (7.292)	- -
Child sex	- -	- -	- -	0.119 (2.530)
Constant	38.565** (7.871)	40.666** (7.968)	38.128** (7.993)	40.583** (8.136)
R^2	0.2457	0.1968	0.1920	0.1929
F	14.70	14.33	13.71	13.37
P-value	0.00	0.00	0.00	0.00

Source: LSMS (Instituto Cuánto 1997, 2000).
Elaboration: owner.
Note: standard errors in parentheses. ** = significant at the 1% level, * = significant at the 5% level, + = significant at the 10% level.

hours of work for children in the range 8 to 13 years were 13.03 in 1997 and 13.66 in 2000. The average weekly hours for children between 14 to16 years old was 15.47 in 1997 and 23.27 in 2000.

At a glance, the main determinants of school attainment observed in regressions are child's age, head of household education, household size, if the child suffers from a chronic disease, if the head of household suffers from an acute disease, and the number of adults in the household. In the case of head of household education, a better educated head has a positive effect on the age-degree distortion (less distortion). This makes sense because a higher educated head of household would give more weight to education, pay more attention to a child's education, and increase the probability of success (reducing the probability of failure).

It is not surprising that the household size increases the age-grade distortion because a larger household usually implies fewer resources for its members. We also observe an important impact of child health. If the child suffers from a chronic disease, it reduces his/her capabilities to study, which eventually affects his/her SAGE index. Curiously, head of household's chronic disease has no significant impact on this index, but the variable head of household's acute disease affects the school attainment. This result is not intuitive because we usually associate chronic diseases with long-term effects, and acute diseases with short-term effects. The age-grade distortion should be mainly affected by long-run variables. The number of adults in a household affects positively the SAGE index, perhaps because the intra-household dependency rate may be reduced (more workers, fewer non-workers) when the number of adults rises.

There are other variables that also significantly affect the SAGE index. For example, if the household used firewood as domestic fuel, we observe lower levels of school attainment. I also expected a negative impact of the dummy variable that assigns 1 to rural areas and 0 to urban areas, because rural areas in Peru are the poorer and less developed areas of the country. The estimated sign is correct, but it is not significant. Perhaps, since firewood is used as source of fuel in many rural households, this variable captures the effect of lower educational performance in rural areas.

The logarithm of non-labor income is also important in the regression. The results indicate that households with higher unearned income show a higher educational performance (less age-grade distortion). Finally, the logarithm of the value of agricultural production has a negative sign, which means that households with larger values of this production exhibit lower levels of the SAGE index. A possible explanation for this result is that higher values of agricultural production are related to higher amounts of work of family members (including children), which may affect school performance and may cause a higher age-grade distortion.

To assess the validity of instruments, all tests described in sections 2 and 3 were applied to the models. Table 3 shows the results of those tests. Concerning the relevance tests, the F-test tells us that the instruments are relevant; however, the Stock-Yogo test—a more precise test—says that instruments are relevant only for model 1 with 5% significance and tolerating a relative bias of 5% of the OLS bias. On the other hand,

Table 3
Tests of relevance and exogeneity

	Null hypothesis H_0:	Model 1	Model 2	Model 3
Relevance tests				
F-test in first stage regression				
Housework equation	All coefficients of excluded instruments in the reduced form equation equal zero	11.45 (0.0000)	8.53 (0.0000)	4.47 (0.0046)
Work equation		37.48 (0.0000)	19.14 (0.0000)	48.64 (0.0000)
Stock-Yogo test (Cragg-Donald stat.)	Excluded instruments are «weak»	15.39 cv=11.04 b=0.05 sig.=0.05	10.92 cv=11.04 b=0.05 sig.=0.05	7.24 cv=N.A. b=0.05 sig.=0.05
Anderson canonical correlation stat.	Underidentification/low relevance of instruments	60.75 (0.0000)	43.59 (0.0000)	21.89 (0.0000)
Hall-Peixe redundancy test on excluded instruments				
Child's sex	Instrument tested is «redundant»	35.975 (0.0000)	37.582 (0.0000)	- -
Farming		227.902 (0.0000)	152.783 (0.0000)	227.902 (0.0000)
Sales		97.284 (0.0000)	47.104 (0.0000)	97.284 (0.0000)
Public school		9.317 (0.0000)	4.902 (0.0862)	9.317 (0.0095)
Exogeneity tests				
Sargan statistic	Full set of instruments are not correlated with the error term and excluded instruments were correctly excluded from the main equation	0.209 (0.9009)	0.550 (0.7595)	0.208 (0.6485)
Difference-in-Sargan (C-statistic)				
Child's sex	Instrument tested is exogenous	0.001 (0.9701)	0.550 (0.4584)	- -
Farming		0.072 (0.7884)	0.219 (0.6397)	N.A.
Sales		0.108 (0.7420)	0.235 (0.6276)	N.A.
Public School		0.157 (0.6923)	0.300 (0.5836)	N.A.
Significance of endogenous regressors				
Anderson-Rubin stat. (Chi-sq.)	Coefficients of endogenous regressors in main equation equal zero	8.82 (0.0657)	9.02 (0.0605)	5.12 (0.1634)

Source: LSMS (Instituto Cuánto 1997, 2000).

Elaboration: owner.

Note: P-values in parentheses. The C-statistic could not be calculated for Model 3 because excluding one instrument makes the model exactly identified. In the Stock-Yogo test: cv = critical value, b= relative bias, sig.=significance level.

the Anderson canonical correlation statistics are high enough to reject the hypothesis of under-identification, and that the group of excluded instruments selected is relevant. The Hall-Peixe test of the individual relevance of the instruments shows that all are relevant, except the variable *public school,* which is not relevant at the 5% level of significance for model 2.

The other tests are the exogeneity tests. The results of the Sargan test tell us that we cannot reject the null hypothesis of exogeneity in all the models. Something similar occurred with the «difference-in-Sargan» statistic, whose values are low enough to not reject the null hypothesis that individual instruments are exogenous.

Finally, the Anderson-Rubin test tells us that the two parameters of the endogenous regressors are jointly significant at the 10% level only in models 1 and 2. They are not significant in model 3. This result is similar to the t-statistics shown in table 2, where the parameter of child labor was not significant.

5. SUMMARY AND CONCLUSIONS

This paper presents some new advances in testing the validity of instruments when we estimate models using the instrumental variables (IV) method. These new techniques are especially useful when we have more than one endogenous regressor on the right-hand side of the equation to be estimated. This requires more refined techniques beyond the well-known «rule of thumb» (F-test).

In addition, I applied those techniques to estimate a model of the effect of child labor and household work on school attainment, as measured by the age-grade distortion. Using a panel of 913 Peruvian children ages eleven to sixteen years old in 2000, the main results show that household work performed by children has a positive impact on the observed educational performance three years later. However, when the hours of housework exceed the threshold of fourteen hours per week, a negative impact is observed on the age-grade distortion. On the other hand, the impact of child labor is negative, as we expected, but surprisingly, it is not significantly different from zero.

Concerning the validity of the instruments, we can trust the results obtained in this paper because all the instruments passed the tests of exogeneity and relevance, which makes the instrumental variable estimation a consistent method for estimating parameters in linear regression models.

REFERENCES

AKABAYASHI, Hideo and George PSACHAROPOULOS

1999 «The Trade-off Between Child Labour and Human Capital Formation: A Tanzanian Case Study». *Journal of Development Studies*, Vol. 35, N° 5, pp. 120-140, London.

ANDREWS, Donald and James STOCK

2005 *Inference with Weak Instruments*. Cowles Foundation Discussion Paper N° 1530. New Haven: Cowles Foundation, Yale University.

BEEGLE, Kathleen, Rajeev DEHEJIA and Roberta GATTI

2005 *Why Should We Care About Child Labor? The Education, Labor Market and Health Consequences of Child Labor*. Working Paper N° 3479. Washington, DC.: World Bank.

CARR, Rhoda, James WRIGHT and Charles BRODY

1996 «Effects of High School Work Experience a Decade Later: Evidence from the National Longitudinal Survey». *Sociology of Education*, Vol. 69, N° 1, pp. 66-81, Albany.

CRAGG, John and Stephen DONALD

1993 «Testing Identifiability and Specification in Instrumental Variable Models». *Econometric Theory*, Vol. 9, N° 2, pp. 222-240, New York.

CRUZ, Luiz and Marcelo MOREIRA

2005 «On the Validity of Econometric Techniques with Weak Iinstruments. Inference on Rreturns to Education using Compulsory School Attendance Laws». *The Journal of Human Resources*, Vol. 40, N° 2, pp. 393 – 410, Madison, WI.

D'AMICO, Ronald

1984 «Does Employment During High School Impair Academic Progress?». *Sociology of Education*, Vol. 57, N° 3, pp. 152-164, Albany.

GUNNARSON, Victoria, Peter OZAREM and Mario A. SANCHEZ

2003 *Child Labor and School Achievement in Latin America*. Working Paper N° 03023. Ames: Iowa State University, Department of Economics.

HALL, Alastair and Fernanda PEIXE

2000 *A Consistent Method for the Selection of Relevant Instruments*. Econometric Society World Congress 2000 Contributed papers 0790. Evanston: Econometric Society.

HALL, Alastair, Glenn RUDEBUSCH and David WILCOX

1996 «Judging Instrument Relevance in Instrumental Variables Estimation». *International Economic Review*, Vol. 37, N° 2, pp. 283-298, Philadelphia.

HAYASHI, Fumio

2000 *Econometrics*. New Jersey: Princeton University Press.

HEADY, Christopher

2003 «The Effect of Child Labor on Learning Achievement». *World Development*, Vol. 31, N° 2, pp. 385-398, Oxford.

INSTITUTO CUÁNTO

1997 *Encuesta Nacional de Niveles de Vida* (ENNIV)

2000 *Encuesta Nacional de Niveles de Vida* (ENNIV)

Jensen, Peter and Helena Skyt Nielsen
1997 «Child Labor or School Attendance? Evidence from Zambia». *Journal of Population Economics*, Vol. 10, N° 4, pp. 407-424, Heidelberg.

Levison, Deborah and Karine Moe
1998 «Household Work as a Deterrent to Schooling: An Analysis of Adolescent Girls in Peru». *Journal of Developing Areas*, Vol. 32, N° 3, pp. 339-356, Nashville.

Patrinos, Harry Anthony and George Psacharopoulos
1997 «Family Size, Schooling and Child Labor in Peru - An Empirical Analysis». *Journal of Population Economics*, Vol. 10, N° 4, pp. 387-406, Heidelberg.

Psacharopoulos, George
1997 «Child Labor versus Educational Attainment. Some Evidence from Latin America». *Journal of Population Economics*, Vol. 10, N° 4, pp. 377-386, Heidelberg.

Ravallion, Martin and Quentin Wodon
2000 «Does Child Labor Displace Schooling? Evidence on Behavioral Responses to an Enrollment Subsidy». *The Economic Journal*, Vol. 110, N° 462, pp. C158-C175, London.

Staiger, Douglas and James Stock
1997 «Instrumental Variables Regression with Weak Instruments». *Econometrica*, Vol. 65, N° 3, pp. 557-586, Evanston.

Stock, James and Motohiro Yogo
2002 *Testing for Weak Instruments in Linear IV Regressions*. NBER Technical Working Paper N° 284. Cambridge: NBER.

Stock, James, Jonathan Wright and Motohiro Yogo
2002 «A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments». *Journal of Business & Economic Statistic*, Vol. 20, N° 4, pp. 518-529, Alexandria.