I Z A

Returns to Education in Russia: Where There Is Risky Sexual Behaviour There Is Also an Instrument

G. Reza Arabsheibani
Anita Staneva

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Returns to Education in Russia:
Where There Is Risky Sexual Behaviour
There Is Also an Instrument

G. Reza Arabsheibani
Swansea University
and IZA

Anita Staneva
Swansea University

ABSTRACT

Returns to Education in Russia: Where There Is Risky Sexual Behaviour There Is Also an Instrument

Finding an instrument that is orthogonal to the disturbance term in the wage equation has been a topic of great deal of debate. Recently, Chesson *et al.* (2006) proposed that higher discount rates are significantly associated with a range of sexual behaviours, including having sex before age of 16 years. Following their paper, we use unprotected sexual behaviour and earlier age of sexual intercourse as alternative instrumental variables to account for endogeneity in schooling. The Bound et al. (1995) F-test indicates that our instruments are strongly correlated with schooling. We fail to reject the test of over-identifying restrictions, which demonstrates that the proposed instruments are valid and not correlated with the current earnings. In line with previous studies, our results suggest that the IV estimates of the returns to schooling are higher than the OLS estimates. In addition to the conditional mean, the proposed instruments are applied over the wage distribution using Chernozhukov and Hansen (2008) instrumental variable quantile approach.

Corresponding author:

G. Reza Arabsheibani
School of Business, Economics and Law
Richard Price Building
Swansea University
Singleton Park
Swansea, SA2 8PP
United Kingdom
E-mail: G.Arabsheibani@swan.ac.uk

I. INTRODUCTION

The human capital literature focuses on ability bias as one of the central problem in estimating the rate of returns to education. If people of higher ability have the capacity to earn more (at a given education level) and if they also tend, significantly, to acquire more education than others, the failure to take ability differences explicitly into account has two consequences: it makes standard Ordinary Least Squares (OLS) estimates to overstate the gross contribution of education to earnings and to understate the opportunity cost of foregone earnings to high-ability persons who attain high levels of education (Hause, 1972). One solution to the ability bias problem is the method of instrumental variables (IV). Finding an instrument that is orthogonal to the disturbance term in the wage equation has been a topic of a great deal of debate as it is difficult to identify demand-side variation in education uncorrelated with individual earnings (Arcand et al., 2005). A suitable instrumental variable must meet two conditions: relevance and exogeneity. The relevance condition requires that the instrument be correlated with the number of years of education that an individual receives. The exogeneity condition requires that the instrument affects earnings only through the channel of education, and therefore the instrument is uncorrelated with the error term in the earnings equation (Leigh and Ryan, 2008).

In this paper we estimate the rate of returns to education in Russia using two alternative instruments for education: risky sexual behaviour, defined as unprotected intercourse at age of 16 and individual behaviour related with earlier age of sexual intercourse. The intuition in choosing risky sexual behaviour as an instrument is based on the same strategy as Chevalier and Walker's (2001) suggestion that smoking habits are a good predictor for discount rates[1]. Our study was motivated by

Chesson et al. (2006) findings that young individuals are associated with higher discount rate, and higher discount rates are significantly associated with a range of risky sexual behaviours. A higher discount rate would then mean a higher weight for current income and thus lower propensity to invest in education (Arai and Kjellstom, 1999). For instance, young girls who become mothers as a teenager confront lowered educational and occupational outcomes for themselves and their children (Furstenberg et al., 1989). We propose that risky sexual behaviour could be used as an instrument for education in estimating the link between educational outcomes and earnings in the sense that unprotected sexual intercourse and earlier age of sexual intercourse reflect discount rates indicating time preferences.

This paper attempts to contribute to the literature in several ways. First, we propose novel instruments in the IV literature and assess their empirical validity within the Russian context. Thus, when we estimate the return to education using risky sexual behaviour as an instrument, what we are recovering is the average marginal return to education for the group of individuals who are having higher discount rates because of their time preference and because they favour more the present time. Second, the two instruments are tested in a quantile regression context, by using a method recently formulated by Chernozhukov and Hansen (2008) and usually referred to as instrumental variable quantile regression model (IVQR). The IVQR approach proposed by Chernozhukov and Hansen (2008) allows us to uncover the heterogeneous effect across the earnings distribution and to take into account the endogeneity problem. Third, by arguing that risky sexual behaviour is likely to be correlated with educational attainment, we address some important issues in terms of policy implications. Nowadays, abortion and birth control have become widespread (Avdeev et al., 1995). The sexual behaviour of adolescents has changed and the

tendency is towards a reduction in age of the first intercourse. Those who have early sex are more likely to experience early pregnancy and low academic achievements. Finally, we believe that Russia is an interesting case with its transition away from planned economy. The Soviet period was officially conservative, traditional and restrictive in areas related to sexual behaviour. However, with Gorbachov and the Perestroika, all restrictions regarding sexuality were removed and western patterns of behaviour became role models (Chervyakov and Kon, 2000). At the same time the attitude of women has been changing particularly as they have become more economically independent.

The rest of the paper is organized as follows. Section 2 briefly introduces the Russian education system; Section 3 describes the data; Section 4 presents the empirical methodology employed; Section 5 discusses the results and provides an evaluation of the empirical validity of the instrumental variables and finally Section 6 concludes.

II. THE RUSSIAN EDUCATION SYSTEM

The Russian Federation is unusual in many ways, but one vital fact is that compared to developing economies, the country has a high human capital relative to GDP per capita (Spagat, 2006). Education was one of the top priorities of the communist system, serving both ideological and economic role: to promote the goals of the Soviet Union and to meet the skills needs of the centrally planned economy (Banzragch, 2010). People assigned high value to education not usually because of its future wage rewards (which were low), but primary because of prestige and self-esteem associated with education itself and with a qualified white-collar job (Cheidvasser and Benitez-Silva, 2007).

The Russian educational system provides free and compulsory education from the age of 6 or 7 to 15 years. Education is split into a compulsory basic education, and on-going higher education. The stages of compulsory schooling are: primary education for ages 6-7 to 9-10 inclusive; senior school for ages 10-11 to 12-13 inclusive, and senior school for ages 13-14 to 14-15 inclusive. If a secondary school pupil wishes to go on to higher education, he or she must stay at school for another two years. Primary and secondary schooling together account for 11 years of study, split into elementary (grades 1-4), middle (grades 5-9) and senior (grades 10-11) classes. Higher education is provided by state and non-state higher education institutions. There are three levels of higher education – incomplete higher education (2 years at least), 4 year programmes leading to bachelor degree and postgraduate study with duration of 1-2 years leading to master degree. There are two types of research degrees: Kandidat nauk, which takes up to 4 years and Doktor nauk, which takes up to 10 years.

III. THE DATA

The data in this study comes from the Russian Longitudinal Monitoring Survey (RLMS), which is a nationally representative survey providing a range of demographic and location characteristics together with rich information on labour market activity. The RLMS was designed to measure the effect of economic and political reforms on the economic welfare of the Russian population. We use Round 10 (2001) survey data, which includes additional sexual history information. Everyone aged 14-49 should have received the sexual history questionnaire, the 'Sexual Behaviour Module'. The purpose of the 'Sexual Behaviour Module' survey is to provide information about the knowledge level and attitudes towards Human

Immunodeficiency Virus (HIV) and sexual behaviour that might increase the risk of HIV infection. We merge individuals' employment and education module data with the sexual behaviour data. Table 1 reports the sample statistics associated with the relevant variables. After eliminating those who reported zero earnings and excluding individuals who were not of working age and still in education, the sample consists of 3,047 individuals. For our analysis, the key advantage of the RLMS sexual history data is that it allows us to construct our two instruments – unprotected sex and earlier age at sexual intercourse. More specifically, we define earlier age at sexual intercourse as equal to 1 if individuals report that they had their first sexual intercourse at age of less than 16. We measure a worker's unprotected sexual behaviour by a dichotomous variable which determines whether individuals report unsafe sex at their first intercourse. The dependent variable is the logarithm of hourly earnings resulting from the primary occupation of the individuals and excludes earnings from secondary jobs, or from agricultural production, and non-monetary benefits[2]. Education is defined as a total number of years achieved by an individual. The vector of exogenous control variables used in the estimations also includes potential experience[3] (linear and quadratic terms), a dummy variables for entrepreneurship and a set of regional dummy variables. We have also included a government sector dummy so as to control for the remaining effect of the old public sector wage structure. The specifications are also supplemented with an additional variable that captures wages arrears effect[4].

[Table 1 here]

The average log wage rate is higher for men. Both genders exhibit average levels of schooling above 12 years but the average length of time spent in education is slightly higher for women than for men. Not surprisingly, women's employment is more

concentrated in the public sector compared to men's. Finally, 14.7% of employed men and 6.7% of employed women report they have their first sexual intercourse at age less than 16.

IV. METHODOLOGY

Various instruments for schooling have been widely proposed in the literature. Recently, much attention has focused on supply-side source of variation in education, attributable to such instruments as the minimum school leaving age, or tuition costs. As Card (1999) observes, there is also a long tradition of utilising family background data, such as father's and mother's educational attainment, to control for unobserved ability. There is high correlation between children's educational attainments and those of their parents. Recent papers have used measures of the cost of education such as distance to school (Card, 1994) or constraints such as the interaction between compulsory schooling and the quarter of birth (Angrist and Krueger, 1991). An alternative is to group observations according to childhood smoking behaviour on the grounds that those who choose to smoke at an early age have a higher discount rate in terms of time preference that they place considerable weight on satisfying current wants at the expense of the future (Evans and Montgomery, 1994; Chevalier and Walker, 1999). If an individual has a higher discount rate because of his/her rate of time preference, he/she cares less about the future and more about the present and will therefore quit formal education at a younger age. Therefore, the mechanism through which the discount rate affects the process of education and earnings suggests a negative bias in the OLS estimates (Dickson, 2009).

The role of individual discount rates in predicting human capital investment has been widely discussed in the labour economics literature (Becker, 1975; Mincer, 1974). If individuals differ in their preferences and their financial resources, this will

result in different discount rates and leads to variation in the point at which they stop acquiring education. Therefore, higher discounting is associated with lower educational outcomes[5] (Lang, 1993) and individuals with lower discount rates are more likely to invest in both education and health (Fuchs, 1982). This implies that the choice of quantity of schooling may differ amongst individuals of the same ability because of differences in individual discount rates (Lang, 1993).

We follow Chesson et al.'s (2006) argument that unprotected sexual behaviour and earlier age of sexual intercourse demonstrates that individuals have a high discount rate. Whether one started sex before age of 16, an age where schooling decisions are actually taken strongly correlates with years of education. Much of the variation in years of education, especially among younger workers, is generated by the educational decisions made during the teen years (Evans and Montgomery, 1994). Such association may not involve causal effect: early sexual intercourse and decline in academic aspirations and performance during adolescence may lie on the same development trajectory, with shared antecedent risk factors (Parkes et al., 2010). Other theoretical models suggest direct effects of early sexual intercourse. Specific effects include possibility of disruption of education caused by pregnancy or sexually transmitted infections, and less time and concentration on educational objectives if teenagers are pre-occupied with sexual activity (Safron et al., 2001). However, whilst Rector and Johnson (2005) indicate that early sexual debut does not necessarily causes poor academic achievement, delaying sex seems to be associated with higher educational aspirations and achieving better grades. The authors also consider that sexually active teens who used contraception have somewhat better academic outcomes than sexually active teens who did not use contraception.

Additionally, in deciding whether to resist an opportunity for a sexual encounter, young individuals desire an immediate sexual gratification rather than the long-term relationship (Reimers et al., 2009)[6]. A decision about sexual behaviour might be spontaneously made in a moment, at which time little thought may be given to the possible future risky consequences. According to this perspective, an individual's choice to abstain from or restrict sexual activity is a logical consequence of conventional goals or values (e.g., educational and career pursuits) (Ohannessian and Crockett, 1993). Chesson et al. (2006) find that higher discount rates are associated with sexual behaviour and sexual health outcomes, such as ever having sex, earlier age of sexual debut and pregnancy. Also, persons with higher discount rates would be more likely to have had sex, to have more sexual partners, to have unprotected sex, and to acquire a sexually transmitted disease, than those with lower discount rate.

Our instruments should satisfy the independence condition, which implies that the selected instruments are uncorrelated with the ability factors and the reduced form schooling residual. One can argue that both instruments might be correlated with economic disadvantage and given the intergenerational transmission of income, therefore likely to be correlated with factors that cause earnings directly. However, recent study by Blum et al. (2000) suggests that family income explains only a very small amount of the variation in whether or not a teen has ever had sex[7]. Their results show that the impact of family income on 'ever had sex' diminishes with age. More specifically, an analogue of R^2 for logistic regression indicated that race/ethnicity, income, and family structure explains 9.7% of the variation in whether a younger teen had ever had sexual intercourse and 2.9% of the variability in sexual intercourse among older teens (Connellan et al., 2001). Finally, as in Arcand et al. (2005), we

believe that in the case of Russia these instruments should meet the orthogonality conditions because under communism the intergeneration mobility of social background were only obtained through educational attainment.

To address the problem of endogeneity of education the following Two Stage Least Squares (2SLS) method describing log of earnings $(\ln w_i)$ and years of schooling (S_i) at the mean is applied:

$$\ln w_i = \beta S_i + Z_{1i}\delta + u_i \tag{1}$$

$$S_i = Z_i\alpha + v_i \tag{2}$$

where Z_{1i} is a vector of exogenous variables that determine individual earnings (Y_i), Z_i is a vector of exogenous variables that influence the educational decision, $E(Z_i, v_i) = 0$, Z_{1i} is a subset of Z_i. The returns to education is measured by β and since u_i is defined as the residual from a regression of v_i on Z_i, it is uncorrelated with Z_{1i} by construction, and $Z_i\alpha = Z_{1i}\alpha_1 + Z_{2i}\alpha_2$, $\alpha_2 \neq 0$ is the rank condition needed for identifying equation (1), or as discussed by Card (1994), identification in the IV is provided by including variables in the vector Z_i that are not contained in Z_{1i}. Equation (1) is then estimated by substituting the fitted values from the first-stage regression of S_i on Z_i as defined in equation (2).

In practise, to model the relationship between education and earnings, we need to compute the predicted values S^* for education S_i and then replace the education variable in the earnings function with this predicted value. Since S^* is not correlated with the error term in the earnings function, the estimated return based on predicted education is unbiased. The addition of S^* is sufficient to eliminate endogeneity bias arising from the possible correlation between education and unobservable influences on earnings.

The assumption that sexual risk-taking (i.e. those who did not use condoms and those who start sex at early age) could be adopted as alternative instrumental variables is valid if risky sexual behaviour (the discount rate) does not correlate with individual ability. As for the exogeneity test, to ensure that the selected instruments have no relationship with the unobserved error term, the overidentifying restrictions test is used as the orthogonality condition for all the instruments. The relevant instruments are highly correlated with the endogenous regressor even after controlling for the exogenous regressors. The relevance of the instruments in our specification is tested in the first-stage regression. As suggested by Bound et al. (1995), we report F-tests on the excluded variables and the partial R^2 from the first stage regression[8]. Intuitively, the stronger is the association, the stronger will be the identification of the model. As shown by Staiger and Stock (1997), the first-stage F-statistics on the excluded instruments serve as an indicator for weak identification. Additionally, for a joint statistics Stock et al. (2002) propose using the minimum eigenvalue of a matrix analogue of the first-stage test F-statistic. Stock and Yogo (2005) present relevant critical values for this eigenvalues as the desired degree of bias, the number of endogenous variables, and the number of overidentifying restrictions vary. These tables include the single endogenous regressor as a special case and presume at least two overidentifying restrictions, so they do not apply to just-identified models. We also test the set of instruments for redundancy as shown by Breusch et al. (1999). In an overidentified context, using several instruments can cause the estimator to have poor finite-sample performance and dropping redundant instruments may lead to more reliable estimation (Baum et al., 2003).

The study further extends the analysis by estimating endogeneity adjusted returns to education across the entire earnings distribution using the recently

developed instrumental variable quantile regression approach (IVQR) proposed by Chernozhukov and Hansen (2008). Their model differs from that in Abadie et al. (2002) who consider a quantile treatment effects (QTE) model for the subpopulation of 'compliers', which applies only to binary treatment variables. The Chernozhukov and Hansen (2008) model is designed for studying heterogeneous QTE over the entire population and applies to binary, discrete and continuous treatment variables. The identification of their approach relies on the existence of instrumental variables Z that are statistically related to S.

Chernozhukov and Hansen (2008) consider the following model:

$$Y = D'\alpha(U) + X'\beta(U), \qquad U \mid X, Z \sim Uniform\,(0, 1), \tag{3}$$

$$D = \delta(X, Z, V), \tag{4}$$

$$\theta \to D'\alpha(\theta) + X'\beta(\theta) \text{ is strictly increasing in } \theta \tag{5}$$

In these equations, Y is scalar outcome of interest or in our case log of earnings; U is a scalar unobserved random variable; D is a vector of endogenous variable determined by (4) and correlated with U, X is a vector of exogenous control variables, Z is a vector of instrumental variables that are statistically related and V is a vector of unobserved random variables possibly correlated with U.

It then follows under (3) and (5) that

$$P[Y \le D'\alpha(\theta) + X'\beta(\theta) \mid Z, X] = \theta \tag{6}$$

thus providing the moment conditions

$$E[(\theta - 1(Y \le D'\alpha(\theta) + X'\beta(\theta)))(Z', X')'] = 0 \tag{7}$$

The instrumental variable quantile regression (IVQR) estimator proposed by Chernozhukov and Hansen (2008) can be viewed as computationally attractive method of approximately solving the moment condition (7). For a given value of the structural parameter, say α, we could estimate $\beta(\theta)$ consistently by regressing

$Y - D'\alpha(\theta)$ on X and Z with traditional quantile regression. If the model is identified, the true value of $\alpha(\theta)$ is the only one for which the coefficients on Z are zero.

V. ESTIMATES OF RETURNS TO EDUCATION IN RUSSIA

Our results are reported in Table 2, where the first column shows the OLS estimates, followed by the first-stage estimates and IV results in column three. The bottom panel of the table presents results from different tests assessing the relevance of our instruments: the first-stage F-statistic for the significance of the instruments, the Kleibergen-Paap rank Wald statistics, and the test of redundancy of specified instrument. Table 3 displays the overall QR and IVQR estimates of returns to education for the 10th, 50th and 90th quantiles. Separate IVQR estimates for males and females are presented in Table 4.

Turning to the estimates in Table 2, we obtain significantly higher IV estimates than the corresponding OLS. Overall, the OLS results give a rate of return of 0.056, which indicates that earnings rise by 5.6% on average with each extra year of education, whereas using IV we get estimates twice as large. The overidentified IV estimates of returns to education are 13.4% for males and 12.4% for females.

Turning to the first stage estimation results, we can see that earlier age of sexual intercourse and unprotected sexual behaviour at age of 16 have a very strong and negative effect on individuals' educational outcomes. Overall, individuals with unprotected sexual behaviour have, on average, 1.36 fewer years of schooling than their counterparts. Those who report that they start their first sexual intercourse before age of 16 have, on average, 1.28 fewer years of education which suggests that risky sexual activity in early age is associated with lower educational outcomes.

Although encouraging, statistical significance alone is not sufficient to rule out a weak instrument problem. A test on excluding our potential instruments from the reduced form equation, yields an F-statistics of 34.40 and partial R^2 of 0.022. The F-statistics of 15.03 for men and 19.10 for women reject the null hypothesis that the instruments are weak. Both of these compare favourably with those reported in Bound et al. (1995). In addition, we perform Stock and Yogo (SY) test comparing the minimum eigenvalue statistic (MES) with the critical values (we are willing to accept at most a rejection rate of 10% of a nominal 5% Wald test). Since the overall MES statistics of 34.40 exceed the critical value of 19.93, our results pass the SY test. The Kleibergen-Paap rank Wald Statistics reject the null hypotheses, which also support the significance of the instruments. This is true for both males and females. The redundancy test indicates that both instruments provide useful information to identify the equations. On the basis of these tests we confirm validity of the instruments and reject the null hypothesis that these two instruments are weak.

[Table 2 here]

However, our main concern lies in the direct effect of the instruments on earnings. Bound et al. (1995) also recommend conducting an F-test of the effect of the instrument on the earning residuals to ensure that the instruments are not directly correlated with the wage once the other control variables are included. Therefore, we regress the earnings residuals from the IV estimates against the instruments in order to ensure that the instruments are not directly correlated with earnings (the t-values associated with the coefficients on the instruments and the corresponding p-values are reported in Table 2 as Bound et al. (2)). Our results here are encouraging – the instruments do not explain any significant variation in the residuals. This test was passed for both males and females. This conclusion is strengthening by Sargan's test

of overidentifying restriction, which result in failure to reject the null hypothesis[9] (p-value=0.6445 in the overall equation) and therefore the two instruments (unprotected sex and earlier age at sexual intercourse) do not independently affect one's current earnings.

The overall QR estimates, reported in Table 3, range from 5.2% to 7.2% between the 10th and 90th quantiles and returns to education is monotonically increasing across the conditional earnings distribution. The IVQR point estimates suggest considerable heterogeneity in the returns to schooling. The effect for adjusting for endogeneity by utilizing the two instruments is most marked at the top end of the earnings distribution, where the estimated coefficient rises to nearly 15%. This suggests that gains are more pronounced for individuals at the upper tail of the earnings distribution that for those in the lower tail. Following the interpretations common in the prior literature, we interpret the quantiles as the level of individual ability (e.g. Arias et al., 2001). The earnings increment associated with education is higher for those individuals whose unobservable characteristics place them at the top of the conditional earnings distribution. In line with previous studies, the results suggest that the endogeneity adjusted IVQR estimates of returns to schooling are higher than the QR estimates.

VI.   CONCLUSION

Teenage behaviour such as early sexual involvement or unprotected sexual intercourse is an important social problem, given the fact that such youth, are at higher risk for probability of early pregnancy, sexually transmitted diseases including HIV. As the Russian society continues to undergo significant transformation, a shift towards increased use of modern contraceptive methods may be of significant benefit to the health and well-being of individuals.

In this paper we estimate returns to education in Russia through the effect of risky sexual behaviour among youths. Following Chesson et al. (2006) findings that higher discount rates are significantly associated with the sexual behaviour, we argue that unprotected sex at age 16 and the earlier age of first sexual debut could be adopted as alternative instruments for education. The study indicates that the two instruments do indeed have very strong and negative effect on schooling outcomes. Those who report early sexual intercourse and unprotected sexual behaviour have on average 1.3 fewer years of education.

The *F*-test on excluding our instruments from the reduced form equation indicates that education is well correlated with the instruments and we do not have a weak instruments problem. The *p*-values for the exogeneity test are greater than 0.05 and therefore our instruments (unprotected sex and earlier age at sexual intercourse) do not significantly correlate with wages. As expected, and in line with previous literature, our estimates of returns to education in Russia, controlling for discount rate bias, indicate higher rate of return than the OLS estimates. The IV estimate of returns to education in Russia is about 12%, which is consistent with other IV studies. We also use recently developed instrumental variable quantile regression approach to assess the educational effect across the conditional earnings distribution. Our results show, that there exists heterogeneity in returns to education across the earnings distribution. In particular, gains are more pronounced for individuals in the upper tail of the earnings distribution than for those in the lower tail.

ACKNOWLEDGMENTS

1. Several studies suggest a link between higher discount rates and risky health behaviours, such as smoking (Fuchs, 1982; Munasinghe and Sicherman, 2001; Viscusi and Hersch, 2001).
2. Hourly earnings are defined as reported monthly net earnings divided by 4.34 and then divided by reported weekly hours of work.
3. Potential experience is imputed as follows: age – schooling – 6.
4. Wage arrears or unprecedented delays in the payment of wages have become an endemic feature of the Russian labour market. There are several forms wage arrears can take in Russia: 1) not paid wages 2) delayed but paid in full wages, 3) paid in time but not in full or 4) paid in part and not in time wages.
5. Haley (1973) states that individual with the higher discount rate will stop specializing in the production of human capital sooner than the individual with lower discount rate.
6. Reimers et al. (2009) show that given the choice, people generally prefer to receive pleasant things, sooner rather than later.
7. For instance, whether earlier age at sexual intercourse among black youths is the result of persistent poverty and learned behaviour, as suggested by Geronimus and Korenman (1992), remains to be clarified. Perhaps, as Dash (1989) argued, the social context of young and poor blacks may determine age at sexual intercourse more than any other set of factors. However, Donenberg et al. (2003) conclude that parental behaviour and peer influence are the most important variables associated with adolescents' initial sexual activity.
8. To examine whether the instrument is sufficiently correlated with the amount of education, Bound et al. (1995) have proposed the researchers to report the partial-R^2 and the F-statistics of the identifying instruments in the first-stage estimation. As a rule of thumb, the F-statistic of a joint test whether all additional instruments are significant should be bigger than 10.
9. The null hypothesis of over-identifying restriction test is that all the including instrumental variables are jointly exogenous.

REFERENCES

Abadie, Alberto, Angrist, Joshua, and Imbens, Guido, 2002. Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica* 70(1): 91-117.

Angrist, Joshua, D., and Krueger, Alan, B. 1991. Does compulsory school attendance affects schooling and earnings? *The Quarterly Journal of Economics* 106(4): 979-1014.

Arai, Mahmood, and Kjellsstrom, Christian, 1999. Returns to human capital in Sweden. In *Returns to Human Capital in Europe-A Literature Review*, ed. Asplund, R. and Pereira, P. ETLA, Taloustieto Oy: 299-324.

Arcand, Jean-Louis, D'hombres, Béatrice, and Gyselinck, Paul, 2005. Instrument choice and the returns to education: New evidence from Vietnam. Working Paper N 200422, CERD.

Arias, Omar, Sosa-Escudero, Walter, and Hallock, Kevin F., 2001. Individual heterogeneity in the returns to schooling: Instrumental variables quantile regression using twins' data. *Empirical Economics* 26(1): 7-40.

Avdeev, Alexandre, Blum, Alain., and Troitskaya, Irina, 1995. The history of abortion statistics in Russia and the USSR from 1900 to 1991. *Population: An English Selection* 7: 39-66.

Banzragch, Otgontugs, 2010. Education and the labour market in Central Asia: The cases of Mongolia and Tajikistan. PhD Thesis, Columbia University, 3420851.

Baum, Christopher F, Schaffer, Mark E., and Stillman, Steven, 2003. Instrumental variables and GMM: Estimation and testing. *Stata Journal* 3: 1-31.

Becker, Gary S., 1975. Human capital: A theoretical and empirical analysis, with special reference to education. NBER Books. The University of Chicago Press.

Blum, R., Beuhring, T., Shew, M., Bearinger, L., Sieving, R., and Resnick, M., 2000. The effects of race/ethnicity, income, and family structure on adolescent risk behaviours. *American Journal of Public Health* 90(12): 1879-1884.

Breusch Trevor, Qian Hailong, Schmidt Peter, and Wyhowski Donald. 1999. Redundancy of moment conditions. *Journal of Econometrics* 91 (1): 89-111.

Bound, John, Jaeger, David, and Baker, Regina, 1995. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association* 90: 443-450.

Card, David, 1994. Earning, schooling and ability revisited. NBER Working Paper N 4832.

Card, David, 1999. The causal effect of education on earnings, In Handbook of Labor Economics, ed. Ashenfelter, O., and Card, D., 3b, North Holland, Amsterdam.

Cheidvasser, Sofia, and Benitez-Silva, Hugo, 2007. The educated Russian's curse: Returns to education in the Russian Federation during the 1990s. *Labour* 21(1): 1-41.

Chernozhukov, Victor, Hansen, Christian and Michael Jansson, 2008. Inference approaches for instrumental variable quantile regression. *Economics Letters* 95: 272-277

Chervyakov, Valeriy, and Kon, Igor, 2000. Sexual revolution in Russia and the tasks of sex education. In *AIDS in Europe: New challenges for the social sciences*, ed. Sandford Theo. London: Routledge, 119 –134.

Chesson, Harrell W., Leichliter, Jami S., Zimet, Gregory D., Rosenthal, Susan L., Bernstein ,David I., and Fife, Kenneth H. 2006. Discount rates and risky sexual behaviours among teenagers and young adults. *Journal of Risk and Uncertainty* 32(3): 217-230.

Chevalier, Arnaud and Walker, I. 2001., Further results on the returns to education in the UK. In *Education and earnings in Europe: a cross country analysis of the returns to education*, ed. Harmon, Colm, Walker, Ian and Westergaard-Nielsen, Niels, E. Elgar Pub, Cheltenham, UK.

Connellan, J., Baron-Cohen, S., Wheelwright, S., Ba'tki, A., Ahluwalia, J., 2001. Sex differences in human neonatal social perception. *Infant Behaviour and Development* 23: 113-118.

Dash, Leon, 1989. When children want children. New York, NY: William Morrow and Co.

Dickson, Matt, 2009. The causal effect of education on wages revisited, IZA Working Paper N 4419.

Donenberg, G., Bryant, F., Emerson, E., Wilson, H., Pasch, K., 2003. Tracing the roots of early sexual debut among adolescents in psychiatric care. *Journal of American Academic Child Adolescent Psychiatry* 42: 594-608.

Evans, William, and Montgomery, Edward, 1994. Education and health: Where there's smoke there's an instrument. NBER Working Paper N 4949.

Fuchs, Victor, 1982. Time preference and health: An exploratory study. In Economic aspects of health, ed. Fuchs,Victor. University of Chicago Press.

Furstenberg, Frank F., Brooks-Gunn, J., and Morgan, Philip S., 1989. Adolescent mothers in later life. Cambridge University Press.

Geronimus, Arline T., and Korenman, Sanders, 1992. The socioeconomic consequences of teen childbearing reconsidered. *The Quarterly Journal of Economics* 107(4): 1187-1214.

Haley, William J., 1973. Human capital: The choice between investment and income. *The American Economic Review* 63(5): 929-944.

Hause, John C., 1972. Earnings profile: Ability and schooling. *Journal of Political Economy* 80: 108-138.

Lang, Kevin, 1993. Ability bias, discount rate bias and the return to education. MPRA Working Paper N 24651.

Leigh, Andrew, and Ryan, Chris, 2008. Estimating returns to education using different natural experiment techniques. *Economics of Education Review* 27(2): 149-160.

Mincer, Jacob A., 1974. *Schooling, experience, and earnings*. Columbia University Press, New York.

Munasinghe, Lalith., and Sicherman, Nachum, 2001. Why do dancers smoke? Time preference, occupational choice, and wage growth. NBER Working Paper N 7542.

Ohannessian, Christine McCauley and Crockett, Lisa J., 1993. A longitudinal investigation of the relationship between educational investment and adolescent sexual activity. *Journal of Adolescent Research* 8(2): 167-182.

Parkes, A., Wight, D., Henderson, M., and West, P., 2010. Does early sexual debut reduce teenagers' participation in tertiary education? Evidence from the SHARE longitudinal study. *Journal of Adolescence* 33(5): 741-754.

Rector, Robert, and Johnson, Kirk, 2005. *Teenage sexual abstinence and academic achievement*. Washington, D.C. The Heritage Foundation.

Reimers, Stian, Elizabeth Maylor A., Stewart, Neil, and Chater, Nick, 2009. Associations between a

one-shot delay discounting measure and age, income, education and real-world impulsive behaviour. *Personality and Individual Differences* 47(8): 973-978.

Safron, D.J, Schulenberg, J.E. and Bachman, J.G., 2001. Part-time work and hurried adolescence: The link among work intensity, social activities, health behaviours, and substance use. *Journal of Health and Social Behaviour* 42(4): 425-449.

Spagat, Michael, 2006. Human capital and the future of transition economies. *Journal of Comparative Economics* 34(1): 44-56.

Staiger, Douglas, and Stock, James, 1997. Instrumental variables regression with weak instruments. *Econometrica* 65(3): 557-586.

Stock, James H., Wright, Jonathan H. and Yogo, Motohiro, 2002. A survey of weak instruments, weak identification in GMM. *Journal of Business and Economic Statistics* 20: 518-529.

Stock, James H., and Yogo, Motohiro, 2005. Testing for weak instruments in linear IV regression. In *Identification and inference for econometric models: Essays in Honour of Thomas Rothenberg*, ed. Andrews D. and Stock J., Cambridge: Cambridge University Press: 80-108.

Viscusi Kip, W. and Hersch Joni, 2001. Cigarette smokers as job risk takers. *Review of Economics and Statistics* 83(2): 269-280.

TABLE 1

Descriptive statistics of variables used in the analysis

Variable	Description	ALL		MALES		FEMALES	
		Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
lhearning	Log of hourly wage	2.490	0.967	2.658	0.999	2.348	0.915
education	Years in education	12.606	2.799	12.338	2.839	12.833	2.744
exp	Potential experience	16.844	11.605	16.526	11.943	17.114	11.306
exp^2	Potential experience squared	418.352	453.257	415.647	480.867	420.648	428.560
single	=1 if single	0.096	0.294	0.098	0.297	0.093	0.291
female	=1 if female	0.541	0.498	-	-	-	-
enterpren	=1 if entrepreneurship	0.020	0.141	0.030	0.171	0.012	0.110
govern	=1 if in public sector	0.709	0.454	0.663	0.473	0.748	0.435
foreign	=1 if company is foreign owned	0.049	0.216	0.063	0.243	0.038	0.190
arrears	=1 if wage arrears effect	0.175	0.380	0.188	0.391	0.163	0.370
moscow	=1 if Moscow	0.210	0.407	0.197	0.398	0.220	0.415
northern	=1 if Northern region	0.071	0.257	0.073	0.260	0.070	0.255
east	=1 if East	0.065	0.247	0.067	0.250	0.064	0.244
volga	=1 if Volga	0.100	0.301	0.103	0.304	0.098	0.298
agesex_16	=1 if sexual intercourse before age 16	0.104	0.305	0.147	0.354	0.067	0.250
unpr_sex	=1 if unprotected sex	0.011	0.104	0.011	0.106	0.010	0.101
N		3047		1399		1648	

Source: RLMS Round 10, 2001, sexual history data.

TABLE 2

Estimated returns to education in Russia, OLS and IV estimates

Dependent Variable	OLS Earnings (1)	First stage Education (2)	IV$_1$ Earnings (3)	First stage Education (4)	IV$_1$ Earnings (5)	First stage Education (6)	IV$_1$ Earnings (7)
	Overall sample			**Males**		**Females**	
education	0.0562***		0.1247***		0.1337**		0.1243**
	(0.0059)		(0.0401)		(0.0633)		(0.0514)
exp	0.0141***	-0.0426***	0.0154***	-0.0453**	0.0174**	-0.0394**	0.0132**
	(0.0043)	(0.0141)	(0.0043)	(0.0215)	(0.0068)	(0.0187)	(0.0055)
exp^2	-0.0004***	-0.0005	-0.0004***	-0.0003	-0.0004**	-0.0008	-0.0003**
	(0.0001)	(0.0004)	(0.0001)	(0.0005)	(0.0002)	(0.0005)	(0.0002)
single	-0.1254**	-0.2907*	-0.1126*	-0.4754*	-0.2070**	-0.1275	-0.0323
	(0.0550)	(0.1757)	(0.0595)	(0.2569)	(0.0968)	(0.2413)	(0.0750)
female	-0.3366***	0.3750***	-0.3686***	-	-	-	-
	(0.0324)	(0.0986)	(0.0377)				
enterpren	0.1583	0.7845**	0.1034	0.9863**	0.2042	0.2789	-0.0750
	(0.1636)	(0.3397)	(0.1210)	(0.4410)	(0.1649)	(0.4807)	(0.1917)
govern	-0.1968***	0.3333***	-0.2203***	0.1497	-0.1237**	0.5212***	-0.3095***
	(0.0361)	(0.1047)	(0.0396)	(0.1570)	(0.0572)	(0.1382)	(0.0561)
foreign	0.3811***	-0.2180	0.3908***	0.0997	0.3408***	-0.6439**	0.4523***
	(0.0754)	(0.2249)	(0.0761)	(0.3123)	(0.1067)	(0.3101)	(0.1142)
arrears	-0.4045***	-0.2726**	-0.3849***	-0.3663*	-0.4920***	-0.1714	-0.2893***
	(0.0474)	(0.1257)	(0.0446)	(0.1930)	(0.0702)	(0.1644)	(0.0574)
moscow	0.5453***	0.6205***	0.4999***	0.8279***	0.3752***	0.4632**	0.5958***
	(0.0362)	(0.1385)	(0.0494)	(0.2049)	(0.0881)	(0.1875)	(0.0577)
northern	0.5224***	0.0318	0.5222***	-0.1609	0.5264***	0.1725	0.5294***
	(0.0706)	(0.1792)	(0.0644)	(0.2507)	(0.1011)	(0.2540)	(0.0833)
east	0.0481	-0.4845**	0.0790	-0.5720*	-0.0021	-0.3879	0.1604*
	(0.0800)	(0.1999)	(0.0694)	(0.3028)	(0.1089)	(0.2659)	(0.0890)
volga	-0.0921*	0.2483	-0.1065*	0.0547	-0.1047	0.3709*	-0.1124
	(0.0549)	(0.1637)	(0.0560)	(0.2759)	(0.0862)	(0.1908)	(0.0740)
agesex_16		-1.2789***		-1.0788***		-1.5381***	
		(0.1460)		(0.1970)		(0.2180)	
unpr_sex		-1.3608***		-1.4438**		-1.3719***	
		(0.3810)		(0.5792)		(0.4921)	
constant	1.9511***	13.1923***	1.0805**	13.2507***	0.9374	13.4892***	0.7609
	(0.0915)	(0.1656)	(0.5120)	(0.2412)	(0.8141)	(0.2001)	(0.6755)
N	3047	3047	3047	1399	1399	1648	1648
F	55.24	18.69	48.94	9.55	18.28	12.25	26.33
Bound et al (1)		34.40***		15.03***		19.10***	
p-value		0.0000		0.0000		0.000	
Partial-R^2		0.0222		0.0212		0.0228	
Kleibergen-Paap rk Wald stat		97.40		39.56		39.845	
p-value		0.0000		0.000		0.000	
IV redundancy test (agesex_16)		72.112		26.860		37.150	
p-value		0.000		0.000		0.000	
IV redundancy test (unpr_sex)		9.303		4.529		5.469	
p-value		0.0023		0.0333		0.0194	
Bound et al (2) p-value	-1.41	0.158		-0.24 0.807		0.45 0.655	
Bound et al (2) p-value	-1.08	0.282		0.20 0.841		-0.49 0.625	
Sargan overidentifying test (χ^2)	0.2129			0.081		0.115	
p-value	0.6445			0.7763		0.7347	

Notes: Standard errors in parentheses; ***, ** and * denote significance at the 1%, 5% and 10% level respectively.
Bound *et al* (1): F-test for the exclusion of the instruments in the first stage (education) equation. Bound *et al* (2): t-value and corresponding p-values associated with two instruments when earnings residuals from the IV estimation are regressed against the instrument.

TABLE 3
Estimated returns to education in Russia, QR and IVQR estimates

	QR			IVQR		
	10th	50th	90th	10th	50th	90th
education	0.0524***	0.0522***	0.0716***	0.0495	0.1549***	0.1511**
	(0.0096)	(0.0065)	(0.0102)	(0.0690)	(0.0514)	(0.0710)
exp	0.0110*	0.0194***	0.0048	0.0097	0.0258***	-0.0006
	(0.0065)	(0.0046)	(0.0072)	(0.0074)	(0.0055)	(0.0076)
expsq	-0.0003	-0.0006***	-0.0002	-0.0002	-0.0005***	-0.0000
	(0.0002)	(0.0001)	(0.0002)	(0.0002)	(0.0001)	(0.0002)
single	-0.0644	-0.0491	-0.1745*	-0.0826	-0.0150	-0.2330**
	(0.0888)	(0.0642)	(0.0988)	(0.1025)	(0.0764)	(0.1055)
female	-0.2199***	-0.3587***	-0.3676***	-0.1977***	-0.3908***	-0.4681***
	(0.0493)	(0.0356)	(0.0560)	(0.0649)	(0.0484)	(0.0668)
enterpren	-0.2182	0.0918	0.7237***	-0.2317	0.0194	0.8224***
	(0.1743)	(0.1261)	(0.1991)	(0.2084)	(0.1554)	(0.2147)
govern	-0.1155**	-0.2049***	-0.2256***	-0.1292*	-0.2745***	-0.1943***
	(0.0551)	(0.0403)	(0.0618)	(0.0682)	(0.0508)	(0.0702)
foreign	0.4983***	0.3767***	0.2511**	0.4908***	0.4758***	0.3012**
	(0.1138)	(0.0820)	(0.1277)	(0.1310)	(0.0977)	(0.1349)
arrears	-0.6549***	-0.3795***	-0.2818***	-0.6620***	-0.3323***	-0.2171***
	(0.0640)	(0.0469)	(0.0732)	(0.0768)	(0.0573)	(0.0791)
moscow	0.6196***	0.5698***	0.4359***	0.6234***	0.4760***	0.4781***
	(0.0627)	(0.0455)	(0.0699)	(0.0851)	(0.0635)	(0.0877)
northern	0.3286***	0.4614***	0.8199***	0.3385***	0.5247***	0.7237***
	(0.0968)	(0.0699)	(0.1084)	(0.1109)	(0.0827)	(0.1142)
east	-0.6000***	0.1671**	0.2352**	-0.4505***	0.1358	0.2788**
	(0.0993)	(0.0725)	(0.1122)	(0.1195)	(0.0891)	(0.1231)
volga	-0.2427***	-0.0402	0.0063	-0.2581***	-0.0856	-0.0750
	(0.0834)	(0.0600)	(0.0926)	(0.0964)	(0.0719)	(0.0993)
constant	0.8805***	1.9738***	2.8342***	0.9209	0.6296	1.9283**
	(0.1438)	(0.0995)	(0.1524)	(0.8817)	(0.6576)	(0.9081)
N	3047	3047	3047	3047	3047	3047

Notes: Standard errors in parentheses; ***, ** and * denote significance at the 1%, 5% and 10% level respectively.

TABLE 4
IVQR estimates of returns to education in Russia, males and females

	Males			Females		
	10th	50th	90th	10th	50th	90th
education	0.1575*	0.1618**	0.2959***	0.0634	0.1693**	0.2114**
	(0.0886)	(0.0813)	(0.1048)	(0.0692)	(0.0669)	(0.0974)
exp	0.0181*	0.0177**	0.0161	0.0242***	0.0213***	-0.0055
	(0.0095)	(0.0087)	(0.0112)	(0.0074)	(0.0071)	(0.0104)
expsq	-0.0004*	-0.0003*	-0.0001	-0.0006***	-0.0004**	0.0002
	(0.0002)	(0.0002)	(0.0003)	(0.0002)	(0.0002)	(0.0003)
single	-0.2075	-0.1924	-0.1718	0.0946	0.1199	-0.2227
	(0.1353)	(0.1243)	(0.1600)	(0.1011)	(0.0976)	(0.1421)
enterpren	0.1994	0.0759	0.2070	-0.3163	-0.0805	0.3243
	(0.2306)	(0.2118)	(0.2728)	(0.2583)	(0.2494)	(0.3632)
govern	0.0063	-0.1511**	-0.2533***	-0.2834***	-0.3457***	-0.2362**
	(0.0800)	(0.0735)	(0.0946)	(0.0755)	(0.0730)	(0.1062)
foreign	0.1794	0.3849***	0.3054*	0.3479**	0.5797***	0.4409**
	(0.1493)	(0.1371)	(0.1765)	(0.1538)	(0.1486)	(0.2164)
arrears	-0.5592***	-0.4241***	-0.2875**	-0.2861***	-0.2586***	-0.1393
	(0.0982)	(0.0902)	(0.1162)	(0.0773)	(0.0747)	(0.1087)
moscow	0.3483***	0.3584***	0.2736*	0.7163***	0.5686***	0.5393***
	(0.1232)	(0.1131)	(0.1457)	(0.0778)	(0.0751)	(0.1094)
northern	0.3350**	0.5161***	0.7238***	0.4706***	0.4953***	0.5563***
	(0.1413)	(0.1298)	(0.1671)	(0.1123)	(0.1085)	(0.1579)
east	0.0898	-0.0174	0.2924	0.0091	0.2205*	0.6519***
	(0.1522)	(0.1398)	(0.1801)	(0.1199)	(0.1158)	(0.1687)
volga	-0.1149	-0.0977	0.0272	-0.0780	-0.1189	-0.2000
	(0.1205)	(0.1107)	(0.1425)	(0.0997)	(0.0963)	(0.1402)
constant	0.0398	0.6108	-0.3950	0.9537	0.0736	0.7215
	(1.1383)	(1.0455)	(1.3464)	(0.9101)	(0.8790)	(1.2800)
N	1399	1399	1399	1648	1648	1648

Notes: Standard errors in parentheses; ***, ** and * denote significance at the 1%, 5% and 10% level respectively.