DISCUSSION PAPER SERIES

IZA DP No. 6862

Education and Freedom of Choice: Evidence from Arranged Marriages in Vietnam

Shahe Emran
Fenohasina Maret-Rakotondrazaka
Stephen C. Smith

September 2012

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Education and Freedom of Choice: Evidence from Arranged Marriages in Vietnam

Shahe Emran
George Washington University

Fenohasina Maret-Rakotondrazaka
George Washington University

Stephen C. Smith
George Washington University
and IZA

Discussion Paper No. 6862
September 2012

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 6862
September 2012

ABSTRACT

Education and Freedom of Choice: Evidence from Arranged Marriages in Vietnam*

Using household data from Vietnam, we provide evidence on the effects of education on freedom of spouse choice. We use war disruptions and spatial indicators of schooling supply as instruments. The point estimates indicate that a year of additional schooling reduces the probability of an arranged marriage by about 14 percentage points for an individual with eight years of schooling. We also estimate bounds on the effect of education on arranged marriage when exclusion restrictions are violated locally (the lower bound is six to seven percentage points). The impact of education is strong for women, but significantly weaker for men.

NON-TECHNICAL SUMMARY

We find strong evidence that enhanced education reduces the incidence of arranged marriage, at least for the case of Vietnam. The impact is especially great for completion of eighth grade. We argue that this finding has broader implications for the notion that education enhances human freedom, in addition to enhancing economic development in a more narrow sense.

JEL Classification: I2, O12, D1, J12

Keywords: arranged marriage, education, schooling, freedom of choice, development, Vietnam, Red River delta, labour markets, social interactions

Corresponding author:

Stephen C. Smith
Department of Economics
George Washington University
Monroe Hall 306
2115 G St. NW
Washington, DC 20052
USA
E-mail: ssmith@gwu.edu

* We would like thank Robert Pollak, James Foster, Bryan Boulier, Justin May, Leonardo Magnusson, Chris Taber, Todd Elder, Forhad Shilpi, Shawn McHale, Tony Castleman, Zhaoyang Hou, and participants at the AEA conference in San Francisco and Southern Economic Association Annual Conference in Washington DC for helpful comments on an earlier version of the paper. The standard disclaimer applies.

(I) Introduction

There is a broad consensus among development practitioners, policy makers, and academic researchers that education is one of the most effective interventions to reduce poverty and promote development. A large empirical literature on economics of education, in the context of both developed and developing countries, focuses on the labour market returns to education (Angrist and Krueger, 1990, 1992; Card,1995,2001; Alderman et al.,1996; Glewwe 1996; Duflo, 2001; Heckman and Li, 2003; Heckman, 2005; among others). There is a relatively small literature that focuses on the direct productivity effects of education, especially in the context of self-employment in agriculture and non-farm activities (Jamison and Lau, 1982; Yang and An, 2002; Kurosaki and Khan, 2006). A substantial literature in growth theory points to human capital as a critical ingredient for long-run growth (see, for example, Romer, 1989, 1994; Aghion et al., 1998; Acemoglu, 2009), although the empirical literature has been more mixed in identifying the effects of human capital in cross-country growth regressions (Durlauf et al. 2005). A related literature addresses whether net social returns to education diverge from net private returns, due to public subsidies and/or to positive externalities (Acemoglu and Angrist, 2000; Basu et al., 2001; Psacharopoulos and Patrinos 2004).

However, education confers many benefits to an individual (and the society) beyond the returns in the labour market or its direct productivity and growth effects. As Amartya Sen noted '[Education] can add to the value of production in the economy and also to the income of the person who has been educated. But even with the same level of income, a person may benefit from education – in reading, communicating, arguing, in being able to choose in a more informed way, in being taken more seriously by others and so on' (Sen, 1999, p. 294). The returns to education in non-market activities, especially as an input in the household production function have been the focus of an additional strand of literature (Haveman and Wolfe, 1984, 2002; Behrman and Wolfe, 1987; Rosenzweig and Schultz, 1989; Kenkel, 1991). The focus of this paper is on non-market returns to education beyond the health and productivity gains addressed in the current literature. Our emphasis is on returns to education in social interactions; specifically, we address the question of whether education improves an individual's freedom of choice by strengthening his or her bargaining position in social interactions, within the household and in other social settings. We use the freedom to choose ones own spouse, one of the most

important decisions in human life, as an indicator of freedom of choice.

The existing literature on intra-household bargaining focuses on conflict over preferences between husband and wife, and analyses its implications for control over resources (Haddad and Kanbur, 1990; Lundberg and Pollak, 1996; Browning, 1998). In contrast, our focus is on the bargaining between the parents and their children under the assumption that the preferences of parents with respect to the spouse choice are not fully aligned with the preferences of children. Since our interest is in understanding the implications of education for freedom of choice, we define arranged marriage as the cases where parents wield the primary decision making power; they choose the spouse of children with or without children's input. Conversely, freedom of choice refers to the cases where children are the primary decision makers; they choose their spouse with or without parental input.

Parents in developing countries make decisions about the education of their children when they are young, taking into account factors including old-age support, social status, and pure altruism. There is, however, a trade-off from the parents' perspective: a more educated child may be more able to take care of them in old age and bring in social prestige, but he/she is also likely to exert more bargaining power in decision making (notably because of better outside options derived from improved economic opportunities and social networks). The bargaining game between children and parents thus changes significantly after education is acquired. This implies that, ceteris paribus, better-educated children would be more likely to choose their own spouse. We use household survey data from the Vietnam Longitudinal Survey (VLS) to examine whether better education among children reduces the incidence of arranged marriage.

To the best of our knowledge, this is the first empirical analysis of the causal effects of education on arranged marriage in the economics literature. There is an anthropological and sociological literature that uses case study methods and multivariate statistical models to investigate the relationship between arranged marriage and education of children (for example, Jejeebhoy, 1995). However in general those studies do not disentangle the causality issue.

Identification of the causal effects of education on choice of spouse can be challenging. Addressing omitted variables bias due to unobserved individual ability as well as unobserved individual and parental preferences is important in estimating the effects of education on spouse choice. Any observed negative partial correlation between the probability of arranged marriage and the level of education might be driven by unobserved heterogeneity that affects both

schooling attainment and spouse choice decisions. For example, more 'progressive' parents are more likely to invest in the education of children and also find it acceptable to allow the children to choose their own spouse. A 'stubborn' child might be more successful in withstanding the pressure to drop out of school for early marriage, and thus may have higher education. He/she is also likely to be more emphatic in defending his/her own choice of spouse. However it is also possible that a stubborn child drops out of school as a rebellion to parental preference for education, and also does not allow the parents to choose the spouse for him/her. Such unobserved heterogeneity would lead to a biased estimate of the effects of children's education on spouse choice in an OLS regression.[1] Thus, it is not possible to pin down the direction of any such bias from a priori reasoning.

The household survey from Vietnam used in this paper is especially suitable for isolating the effects of education. There is good information on individual and parental characteristics and labour market opportunities. Also, the data come from 10 communes located in the Red River Delta region in Vietnam, and exhibit fewer variations in cultural practices than many other data sets.[2] This implies that cultural heterogeneity is not likely to drive the results we report in this paper. The results provide robust evidence in favour of a statistically significant and numerically important negative effect of education on the probability of arranged marriage.[3]

First, we follow a large and mature literature on the direct or financial returns to education and use indicators of supply of schooling and exogenous shocks to schooling attainment as instruments (see Card, 2001; Blundell et al., 2005). We use birthplace to represent variations in schooling supply across geographic locations and use a cohort dummy for 1955 and earlier births to capture positive shocks to educational attainment due to lower intensity of conflicts in North Vietnam during their relevant school going years (see below). Regarding birthplace in our data set, 98.7 per cent of individuals grew up in their place of birth.[4] Such lack of geographic mobility is a traditional feature of this region and, from the 1950s, one likely to be strongly reinforced by the Ho Khau internal registration system that restricted geographic mobility in Vietnam.[5]

To ensure that birthplace qualifies as a reasonable instrument in our context, we also control for differences in labour market opportunities. Such opportunities are likely to vary across geographic locations, and are important determinants of both the incidence of arranged marriage (Goode, 1963), and investment in human capital.[6] Better labour market opportunities improve children's outside option if they face having to leave their parental house and are

deprived from parental wealth as a consequence of choosing their own spouse. We include an indicator of non-agricultural occupational opportunities at the time of marriage, a rural vs. urban dummy, and household's wage income to control for local labour market opportunities.[7] The rural vs. urban dummy also captures differences in living costs, especially housing costs. As mentioned before, in addition we control for a number of individual and parental characteristics that also help isolate the causal effects of education on spouse choice.

The cohort dummy represents differences in schooling attainment of the cohort born after 1955 compared to the older cohort (above 40 years of age in 1995). In our sample 43 per cent of individuals are born before 1955. A plausible interpretation of this dummy variable is that it captures the effects of varying intensity of conflicts and war experienced by different age cohorts when they were school aged. The history of war and conflict in Vietnam is long and complex; the intensity of conflict varied over time and space significantly. The cohorts born before 1955 in the Red River Delta region experienced a relatively stable period from 1954 to 1965, which is expected to improve their educational attainment compared to the later cohorts, other things equal (such as family income and access to information).[8] For the Red River Delta region, the disruptions caused by US bombing in the late sixties to mid seventies were severe and the later cohorts plausibly would have suffered negative shocks to their educational attainment.[9] The ten communes surveyed for our data set were bombed particularly heavily because of their proximity to important railroad links and industrial installations (Merli, 2000, p. 4). The empirical results reported later show that the earlier cohorts did experience systematically higher educational attainment *after* controlling for a vector of individual and household level variables and the rural-urban dummy.[10] One might worry that intensity of conflict can have direct effects on the marriage market because it affects the sex ratio. The changing sex ratio may also affect the incentives for human capital investment by parents if such investments respond to potential returns in the marriage market (LaFortune, 2009). We thus include age cohort specific sex ratios as additional controls in the regressions (calculated using the 1979 census in Vietnam).[11]

The evidence from the formal test of exogeneity discussed below (see section IV.1 and Table 3) show that conditional on the set of control variables, the birthplace and age cohort dummy satisfy the exclusion restrictions unambiguously (P-value of Hansen's J statistic is 0.81).[12] Thus, we control for the major factors that are discussed in the literature on arranged marriage that can also potentially affect parents' decisions about children's education, such as

labour market opportunities and cost of living. Note that even if we are unable to control for some of the determinants of arranged marriage, it does not affect our identification as long as they are not important determinants parental decision regarding children's education.[13]

The empirical results indicate that education significantly reduces the likelihood of having an arranged marriage. Averaging the point estimates from linear probability (2SLS, GMM, CUE-GMM) and Probit (Control Function and IV Probit) models imply that one year of additional schooling reduces the probability of arranged marriage by about 14 percentage points for an individual with eight years of education. The probability that an individual will have an arranged marriage is close to zero when he/she has about ten years of schooling.

As pointed out recently by Conley et al. (2012) it may be more credible to assume that the instruments satisfy the exclusion restrictions approximately, that is, the instruments may have non-zero but very small effects in the structural equation. To address any remaining concern that there might still be some (arguably very weak) effects of birthplace and cohort dummy on arranged marriage through some indirect and complex channels, we report results from the Conley et al. (2012) bounds approach, under the assumption that the instruments may violate the exclusion restrictions locally. None of the bounds contain zero, providing strong evidence of a negative causal effect of education on the probability of arranged marriage. The lower bound estimates indicate a smaller effect of education; one year of additional schooling reduces the probability of arranged marriage by about six to seven percentage points for an individual with 8 years of education. But the central conclusion in this paper does not depend on the exclusion restrictions imposed in the IV approach.

The results also suggest that the effect of education has important gender dimensions. The impact of education is very strong for women; the marginal effect is about 17 percentage points for a woman with the average level of education (7.77 years of schooling).[14] The evidence in favour of a significant causal effect of education on arranged marriage in case of men is, however, much weaker.

The remainder of the paper is organized as follows. Section (II) discusses the conceptual framework, which helps us identify the relevant control variables. Section (III) provides a brief discussion of the data source and variables descriptions. Section (IV), arranged in subsections, is devoted to the empirical strategy used in this paper to identify and estimate the causal effects of education on freedom of spouse choice. Section (V) presents estimates of causal effects of

education on arranged marriage using an instrumental variables approach and the Conley et al. (2012) bounds approach. In the conclusion, we provide a summary and context of the results and contributions of the paper.

(II) Conceptual Framework

Arranged marriage is common in many developing countries, especially in Asia and Africa. There are reasons to believe that in many cases it is not a cooperative outcome where parents and children jointly choose a spouse to maximize family welfare.[15] There are no a priori reasons to expect that the preferences of parents and children would be perfectly aligned.[16] Parents might be interested in arranged marriage to further their objectives of strengthening their own social and business network, to improve their standing in the community, and to uphold cultural traditions. Another important motive is old-age support; if the parents choose the spouse it is less likely that the spouse of their children will skew the distribution of resources against them, especially in old age. It has been argued that the transition from arranged marriage to own choice redistributes resources away from the parents to the children, and might have implications for savings and growth (Edlund and Lagerlof, 2006).

The practice of arranged marriage seems to decline generally with economic development; the expansion of education and labour market opportunities for children are negatively correlated with the incidence of arranged marriage (for a discussion, see Goode, 1963). There is a substantial literature in sociology and anthropology that provides suggestive evidence of a negative correlation between education and the probability of having arranged marriage in a society (for a summary, see Jejeebhoy, 1995). Education can influence spouse choice through a variety of channels. It may mould children's preference, enrich the pool of potential partners, and alter the threat point in the bargaining game. The changes in preference can be due to interactions with people with different attitudes to arranged marriage, exposure to other cultural norms, and in general through a broadening of outlook. Social interactions at school and college may increase the pool of potential partners with similar values and thus of better compatibility. Education alters the threat point in the bargaining game between parents and children, because educated children in general have better outside options (labour market prospects and higher permanent income). In the event of a conflict regarding spouse choice, the parents can use a prospective bequest as a threat, that is, they might redistribute away from a son or daughter who

chooses his/her own spouse. They can also threaten to drive the newly married couple from the parental home, an effective deterrent when the housing and other related costs of starting a separate household are sufficiently high.[17] Educated children are also more equipped to handle pressure from parents, which sometimes may take the form of legal and other forms of harassment, when they choose their own spouse. Finally, extended schooling tends to delay the age at marriage, thereby placing the spouse choice decision at a time when the children are more mature and capable of making better choices.

The fact that education may weaken parent's influence on children in general, and on spouse choice in particular, also implies that parents will take into account this additional cost when making the education decisions for their children. This has interesting implications for allocation of resources across different children for education. Parents might trade off potentially higher earnings for more compliant behaviour of children, assuming that the correlation between ability and assertiveness is positive. This is likely to result in a negative selection effect in our context because parents would invest more in educating the more pliant children even though they may be of lower ability, and the more pliant a child is, the easier it is for parents to overrule his/her spouse choice.[18]

(III) Data

We use data from the Vietnam Longitudinal Survey (VLS, 1995-1998) conducted by researchers from the University of Washington and Institute of Sociology in Hanoi.[19] The choice of this data set is motivated by the fact that that it has information on choice of spouse in addition to information on individual and parental characteristics, and on labour market opportunities. The data set is comprised of 1185 households and covers 4464 individuals.[20] Because of some missing observations for variables such as years of schooling, age at marriage, and number of parental siblings, our analysis is based on a sample of 3219 observations.

The survey covers ten communes around the Red River Delta, partitioned into four groups: urban communes (three), within three km of highway or inter-provincial highway (two), within three to ten km (two), and more distant than ten km (three). About 80 per cent of the 4,464 individuals interviewed were married. The characteristics of the sample used for estimation (3219 observations) are reported in Table 1. For our variable of interest, 'arranged marriage,' the parents chose the spouse with or without consultation with their children for 25 per cent of

married individuals. Interestingly, there is no evidence of gender differences in the *incidence* of arranged marriages. The average age in the survey year (1995) is 39 years, with 63 per cent of the sample born before 1960. The average education is about 8.07 years of schooling, and 82 per cent of individuals had more than five years of schooling. The educational attainment for men is higher; the average education in the sample of men is 8.58 years of schooling, while the average for women is 7.77 years. The education system in Vietnam is standard in that students graduate from high school after 12 years of schooling. We recoded the years of schooling variable by assigning 13.5 years of education if the person has attended college but did not finish, and 15 if the person has attended and finished college. The average age at marriage is 22.54 years, which indicates that early marriage may be uncommon (6.5 per cent below 18 years). Only four per cent of individuals met their spouse at school. This implies that schools do *not* play an important role as a meeting place for potential marriage partners; and the effects we estimate and discuss later in the paper reflect primarily strengthening of bargaining power rather than an information and matching-at-the-school effect.

(IV) Empirical Strategy

To test the hypothesis that education has a causal effect on the incidence of arranged marriage, we estimate the following Probit model:[21]

$$P(Y_i = 1 | X_i) = \varphi(\beta_0 + \beta_i E_i + X_i \Pi + \varepsilon_i) \qquad (1)$$

Y_i is a binary variable that takes on the value of one if the marriage was decided by the parents for individual *i* and zero if he/she chose his/her own spouse with or without parent approval and $\Phi(.)$ is the standard normal CDF. The level of educational attainment by individual *i* is represented by E_i, and X_i is a vector of relevant control variables. The error term in equation (1) captures unobserved heterogeneity of both parents and individual *i*. Educational attainment is measured by 'years of schooling'. The vector of control variables includes a rich set of household and individual level characteristics. As mentioned before, an advantage of the data set used in this paper is that it has good information on characteristics of an individual, his/her parents (both mother and father) and also on labour market opportunities.[22] The parental

characteristics include occupation, birth order (if first born), and number of siblings. Occupation is a choice variable of the parents and thus would reflect their preference and ability. The number of siblings and birth order of a child might plausibly affect the nature of the bargaining game between parents and children. A smaller family size may be an indicator of more 'progressive' views in general.[23] Individual level variables for children include age, gender, age at marriage, a dummy for no religious affiliation, and number of siblings. It is important to control for the gender of the child as parents may have son preference. The age at marriage may influence the nature of bargaining between parents and children. Religious affiliation may influence views about acceptability of parental choice.[24]

We also control for a set of household level variables that might affect investment in human capital and may also influence preference regarding arranged marriage (that is, both education and freedom of choice may be normal goods). Higher income households are, ceteris paribus, more likely to invest in children's education because of relaxed credit constraints. We use household's agricultural and wage income as controls for income. In addition, a dummy for brick-wall house and dummies for ownership of radio and TV are also included. These are indicators for household wealth; and ownership of TV and radio also controls for access to information which might affect views regarding acceptability of arranged marriage. As we discuss in the following, children's own age, wage income and the rural-urban dummy play a role in our identification strategy.

As discussed above, education may affect the probability of arranged marriage through changes in the bargaining game, but also through better search and information regarding the pool of partners in school. To isolate the role played by bargaining, we control for the fact that at least part of the effect of education would capture the information and matching channel by including a dummy for the case when a person met his/her spouse at school.

(IV.1) Approaches to Identification

Identification of the causal effects of education on probability of arranged marriage is challenging due to multi-level unobserved heterogeneity in ability and preference. We use an instrumental variables strategy and the recent Conley et al. (2012) bounds approach under the assumption that instruments are 'plausibly exogenous' to provide evidence on the causal effects of education on arranged marriage.

(IV.1.1) Instrumental Variables Approach

Following a large literature on returns to education in the labour market, we use indicators of schooling supply and shocks due to war as instruments for children's education. More specifically, the instruments used in this paper are birthplace and a dummy for birth cohort above 40 years of age (born in or before 1955). The birthplace represents variations in the supply of schooling across geographic locations, while the cohort dummy represents variations over time due to varying intensity of conflict. Birthplace is a good indicator of schooling supply only if the respondent grew up in the place of birth, as did 98.7 per cent of the respondents in our data set. This lack of geographic mobility may seem surprising, but it is not uncommon in low-income countries; and it is also a natural outcome of restrictions on geographic mobility through the Ho Khau (internal registration) system. Labour market opportunities and housing costs have been identified in the sociological literature as important determinants of arranged marriage (Goode, 1963). We use an indicator of non-farm opportunities at the time of marriage (share of non-farm employment), wage income, and a rural-urban dummy to capture labour market opportunities and housing market conditions. Controlling for labour market opportunities is important for consistent estimation of the causal effects of education, because both education and outside options of a child in the bargaining game depend on it. Better market opportunities might induce the parents to invest more in children and also allow the children to assert their preferences ex post by improving their outside option. The rural-urban dummy also controls for differences in exposure to cultural factors (for example, modernization).

One might argue that the birthplace of children is the location chosen by parents (especially by the father in a patriarchal and patrilocal society like Vietnam); and some fathers might choose location close to the school if they value education more. If the fathers who value education more also differ systematically from other parents in their attitude to arranged marriage (for example, they may be more ‘progressive’), then this will weaken the case for the exclusion restriction. However, there are good reasons to believe that, in the specific context of our data set, this is not a concern. First, in our sample, the household location seems static and historically determined for most of the parental generation; about 70 per cent of fathers were born in the same place as the children. For the other 30 per cent of cases, the birthplace of children reflects parental choice. However, unlike some developed countries, the location choice

in Vietnam is primarily determined by parental occupation and labour market opportunities, and children's school choice usually plays a minor role. In the regressions, we thus control for parental occupation (both father's and mother's occupation).[25] The evidence presented below shows that conditional on the set of observed covariates, the instruments comfortably satisfy the formal test of exogeneity (P-value for Hansen's J statistics is 0.81).

There are 27 birthplace dummies we can use as instruments along with the cohort dummy. This, however, creates weak IV problems due to too many instruments, as we have only one endogenous regressor. We thus need to reduce the dimension of the instruments. Although the respondents in the survey were born in 27 different birthplaces, most of the sample is concentrated in a few places. We define two dummies, one for the case when the respondent was born in the same commune he/she is currently located, and the second one for the cases where respondent was not born in the current commune, and the birthplace is the location with the largest number of respondents among all the 'other' birthplaces. Together, these two dummies represent 86 per cent of the sample. The other 25 birthplaces together represent the excluded category. We also use an alternative approach where the first two principal components (based on Eigen values) of the 27 birthplace dummies represent the schooling supply variations across space. As noted by Amemia (1966) and emphasized more recently by Bai and Ng (2008), using principal components to reduce the dimension of the set of instruments works well in such cases. The results using this alternative set of instruments are consistent with the conclusions reported in this paper. For the sake of brevity, we do not report these alternative results.[26]

The cohort dummy (that equals one if born in or before 1955) reflects the difference in schooling attainment between the cohorts born in 1955 and earlier and those born after 1955. Although Vietnam suffered a series of wars and conflicts, there was a relative calm in North Vietnam from 1954 to 1965 (Harrison, 1989; Young, 1991; Merli, 2000). The age cohorts born in 1955 and earlier, and in school during this period experienced a positive shock to their schooling attainment. Later cohorts (born after 1955) experienced a negative shock to education attainment, attributable, among other things, to heavy U.S. bombings[27] in the Red River Delta region (including communes in the VLS survey area) during the Vietnam War, and to related disruptions such as moving school locations and the enlistment of fathers of school age children. Using the same VLS data set as in our paper, Merli (2000) shows that war mortality was much higher during the U.S. bombings in the Red River Delta region compared to earlier periods.

An alternative is to use a cohort dummy for birth year between 1945-1955 to capture the positive shock to education due to the stability from 1954-1965 in the Red River Delta region. The results from this alternative formulation of the cohort dummy are very similar to the ones based on the 1955 cohort dummy, and thus are omitted to save space.

One might be concerned that the cohort dummy may also reflect changes in the informal social norms or formal legal codes regarding acceptability of arranged marriage; changes in such norms or codes over the relevant time period would result in heterogeneity in the incidence of arranged marriage across different cohorts of children. To account for slowly changing social norms, we include age and age squared of children. Age controls for possible generational changes in views regarding arranged marriage. There was also change in the marital law in Vietnam; the practice of arranged marriage was declared illegal in 1960. Not surprisingly, however, the formal law only took hold on the ground very gradually (Van Bich, 1999).[28]

At the same time, the 1960 Marriage and Family Law may itself also be viewed in part as a more direct factor behind the trend toward an increase in educational attainment as well as a decline in arranged marriage.[29] In addition to banning arranged marriage, the 1960 law set a minimum age at marriage for women at 18 and men at 20. Concurrently, the government also expanded basic educational opportunities for both girls and boys as well as for illiterate adults (i.e., the parental generation who may make decision about mate selection). Thus, despite the slow adjustments, we include a dummy for marriage before 1960 to capture the possible impact of this change in the formal legal code on the incidence of arranged marriage.

Another potential objection to the exclusion restriction on the cohort dummy mentioned before could be that the conflict might affect the marriage market directly through its effects on the sex ratio. We thus include the cohort specific sex ratio as an additional control.

(IV.1.2) Set Identification: Bounds Under Approximate Exclusion Restrictions

The instrumental variables approach outlined above is attractive because it provides us with point identification.[30] However, the IV approach relies on exact exclusion restrictions on the instruments, that is, the excluded instruments are assumed to have coefficients of exactly zero in the structural equation (in our case, the equation for arranged marriage). While this is plausible as we argue above, still some readers may find this too restrictive. We thus use the recent approach developed by Conley et al. (2012) where the exclusion restrictions on the instruments

are relaxed, but point identification is not possible. In the context of our model, the exact exclusion restrictions imposed in the IV approach implies that $H_o : \theta_1 = \theta_2 = 0$ in the following specification of the spouse choice function:

$$P(Y_i = 1 | X_i) = \varphi(\beta_0 + \beta_i E_i + X_i \Pi + \theta_1 V_i + \theta_2 D_c + \varepsilon_i) \quad (2)$$

where V_i is birthplace, and D_i is a dummy for age cohort (40 years cut-off). It may be more plausible to argue that these exclusion restrictions hold only approximately, that is, $H_o : \theta_1 = \theta_2 \cong 0$, in an application like ours where there are potentially multiple sources of unobserved heterogeneity, and the instruments might have very small direct effects on arranged marriage if they are proxy variables for omitted heterogeneity. Under the approximate exclusion restrictions, the instruments are 'plausibly exogenous' in the terminology of Conley et al. (2012) who develop a set of approaches under this weaker exogeneity condition, and show that one can estimate bounds on the causal effect of the endogenous variable.

We implement a straightforward and intuitive approach to modelling 'plausible exogeneity' of the instruments as developed in Conley et al. (2012); it specifies a support of possible values for $\theta_j \in [-\delta, +\delta]$ where $\delta > 0$ can be arbitrarily small. Under this assumption, we estimate the lower and upper bounds for the estimate of the parameter of interest, β_1. We report results for a number of alternative values of δ.

(V) Empirical Results

(V.1) Baseline Results

Figure 1 shows the predicted probability of arranged marriage (dummy equals one if marriage was arranged by parents with or without children's inputs) against different years of schooling from a simple Probit regression of arranged marriage on education, without any controls. There is a clear negative relationship between education and the probability of arranged marriage. As discussed before, this bivariate correlation – while interesting in its own right – cannot inform us about the possible causal effect of education, due to omitted variables bias. Table 2 presents results from estimating equation (1) above using Probit analysis for different sets of control variables. The corresponding results from OLS estimation are very similar, and are reported in the Table A.1 in the appendix. The first column reports estimates with only individual level controls; then, we progressively include parental characteristics, household level

variables, labour market opportunities at the time of marriage, and a rural-urban dummy. The estimates show a consistently negative and statistically significant effect of education on the probability of arranged marriage.[31] Although suggestive, these results may suffer from omitted variables bias because of unobserved heterogeneity in ability and preference of both parents and children. Another source of possible bias is measurement error in schooling, which would result in downward-biased estimates. It is not possible to pin down the direction of bias from a priori reasoning.

(V.2) Estimates from the IV Approach

The estimated effects of education on the probability of arranged marriage from the IV approach are reported in Table 3. For estimation, we use IV Probit and a control function approach for the Probit model (Blundell and Smith, 1986; Rivers and Vuong, 1988); while for the linear probability model 2SLS and GMM are used. With a binary dependent variable, control function with Probit may provide more efficient estimates compared to 2SLS and GMM that ignore the binary nature of the dependent variable. The linear probability model, on the other hand, has the advantage that it does not require any distributional assumption.

The IV diagnostics show that the instruments pass the test of exogeneity convincingly; the P-value of Hansen's J statistic is 0.81. The results thus provide strong evidence that the exclusion restrictions imposed are satisfied comfortably. The estimates from first stage show that the instruments have good power; the F statistic for exclusion of the instruments is 12.49, which is higher than the Bound et al. (1995) rule of thumb of ten for one endogenous regressor. In the first stage regression, the cohort dummy is significant at the five per cent level, and the birthplace dummies are significant at the one per cent level. Consistent with the discussion above, the estimated coefficient on the cohort dummy shows that the cohort born after 1955 has significantly lower educational attainment.

The first column in Table 3 (Panel A) shows the estimated effect of education using the control function approach and the second column presents the results from IV Probit (estimates are marginal effects evaluated at the sample mean). The control function results show that education is clearly endogenous in the arranged marriage equation, as the residual from the first stage is significant at the one per cent level. The estimated marginal effect of education on arranged marriage is significant both statistically and numerically. According to the control

function estimate, one year of additional schooling reduces the probability of arranged marriage by 17 percentage points for an individual with 8.07 years of education (the average level of education in the sample). The estimated marginal effect from IV Probit is somewhat smaller; one year of additional schooling reduces the probability of arranged marriage by 13 percentage points. The third (2SLS) and fourth (GMM) columns in Table 3 report the results from the linear probability model. The estimated marginal effects from these alternative estimators are identical to that from the IV Probit. The estimates are statistically significant at the one per cent level. The IV estimates from Probit and linear probability models thus provide robust evidence in favour of a strong negative effect of education on the probability of arranged marriage.

Thus, we have strong IVs; but to allay any remaining concerns regarding weak instruments issues, we report results from two alternative checks. First, we use CUE-GMM to estimate the linear probability model and compare the estimates with those from 2SLS. As pointed out by Hansen et al, (1996) and emphasized recently by Stock and Watson (2008), if the instruments are weak, then the CUE-GMM estimates perform better than 2SLS, and would differ significantly from the 2SLS estimates. The results from CUE-GMM are reported in the last column of Table 3. The estimated marginal effect of education is 13 percentage points, the same as the common estimate from IV Probit, 2SLS, and GMM. Second, we use the recently developed tests for weak IV robust inference in binary choice models with heteroskedasticity developed by Magnusson (2008) and Finlay and Magnusson (2009). Panel B in Table 3 reports the results from the Finlay and Magnusson (2009) approach to weak IV robust inference for the linear probability and the Probit models. Note that the Probit results are for the estimated coefficients, not for the marginal effects. According to the conditional likelihood ratio (CLR) and Anderson-Rubin (A-R) tests, the null of no effect of education on arranged marriage is rejected at the one per cent level (P-value=0). The 95 per cent confidence intervals for CLR and Anderson-Rubin tests do not include zero. The lower bound estimate from the 95 per cent confidence interval gives an estimated marginal effect of about seven percentage points reduction in the probability of arranged marriage (estimates from linear probability model).

(V.3) Relaxing the Exclusion Restrictions: Estimated Bounds

In this section, we report results from the recent bounds approach developed by Conley et al. (2012). As discussed earlier, the exclusion restrictions are relaxed and we model ‘plausible

exogeneity' of the instruments by assuming that the coefficients on the instruments in the arranged marriage equation belong to an interval, that is, $\theta_k \in [-\delta,+\delta] \forall k$ with $\delta > 0$. The estimated bounds are reported for 95 per cent confidence intervals in Table 4 with $\delta = 0.0001$, 0.0005, 0.001, 0.005, and 0.01. The reported results are from the 2SLS estimator.

The results show that the estimated bounds do not vary significantly with the value of δ. More importantly, none of the 95 per cent confidence intervals contain zero. This is strong evidence in favour of a robust negative effect of education on the probability of arranged marriage. The negative causal effect identified earlier in Table 3 is thus robust to relaxation of the exclusion restrictions imposed on the instruments for the IV estimates. It is reassuring that even if we allow for non-zero but low-level direct influence of the instruments on the probability of arranged marriage, the central conclusion that education reduces the probability of having an arranged marriage remains intact. Even if one uses the most conservative estimate from the lower bound, one year of additional schooling still reduces the probability of arranged marriage substantially, by six to seven percentage points. Note that $\delta = 0.01$ implies that each of the instruments can affect the probability of arranged marriage by up to one percentage point. Since we use three instruments, together they can affect the probability of arranged marriage by three percentage points. This allows for substantial direct effect of the instruments, and thus the estimated lower bound should probably be interpreted as a conservative estimate of the causal effects of education on arranged marriage.

(V.4) Gender, Education and Arranged Marriage

The results discussed so far do not consider possible gender differences in the effects of education on probability of arranged marriage. The results reported in Table 2 from the Probit model show that the gender of the respondent matters for arranged marriage; however, these results cannot answer the question whether the effects of education on arranged marriage are significantly different for a women compared with men. In this section we explore the issue of gender differences in the effects of education. Table 5 reports the results from estimating equation (1) separately for male and female sub-samples using control function, IV Probit, 2SLS, GMM and CUE-GMM estimators. The instruments used are the same as in Table 3. The IV diagnostics show that, for both the sub-samples, the instruments satisfy the exogeneity condition comfortably (the lowest P-value for Hansen's J statistics is 0.57). The strength of the instrument

set is reasonably good for the female sub-sample (Cragg-Donald F= 8.76), but the instruments lack power in the case of the male sub-sample (Cragg-Donald F=4.32). Thus weak IV bias is a concern for the male sub-sample. To address this concern, we report estimates using CUE-GMM, which is robust to weak instruments, and also report results from Finaly-Magnusson (2008) weak IV robust inference procedure.

The results differ significantly between male and female sub-samples. The marginal effect of education on the probability of arranged marriage is much higher for a woman compared with a man. A natural question at this point is how can the effect across genders be different? There are at least two reasons behind possible gender differences in the effects of education. First, in the context of a traditional society, arranged marriage does not necessarily imply a symmetric arrangement for both the bride and the groom. It is common for the parents of a prospective groom to do the preliminary screening, and then present a menu of choices to the groom who chooses from among them. This implies that, for the groom, it is not an arranged marriage, but for the bride it is. The second reason is that even if the marginal effect of education differs across gender, there are other factors that can potentially offset the differential effects; so one would not observe any imbalance in the arranged marriages.

For a woman with an average level of education (7.77 years), one year of additional schooling reduces the probability of arranged marriage by about 17 percentage points (averaging over estimates from alternative estimators in Table 5). The corresponding estimate for a male with average education (8.58 years of schooling) is about seven percentage points. The marginal effect of education in the male sub-sample is, however, not always precisely estimated; it is not significant at the 10 per cent level using to 2SLS and Control function estimates, but is significant according to GMM and CUE-GMM and IV Probit estimates. The results from weak IV robust inference are also not unambiguous. According to the Conditional Likelihood Ratio (CLR) test for linear probability model, the null of no effect of education on arranged marriage for men can be rejected at the 10 per cent level (P-value equals 0.07), but the null cannot be rejected at the 10 per cent level according to the Anderson-Rubin test. The weak IV robust results for IV Probit also indicate that education does not have a significant effect on arranged marriage for men at the 10 per cent level. Taken together, the evidence of a causal effect of education on arranged marriage for men is thus much weaker than the evidence for women.

(VI) Conclusions

This paper provides evidence of broader returns to education beyond the labour market and direct productivity effects, with a focus on benefits of education in social interactions. As noted by Amartya Sen, among others, an educated person is treated with more respect ('taken more seriously') in social interactions. Thus education may enable an individual to assert his/her choices within the household and in broader social interactions. We focus on the bargaining between parents and children, and provide estimates of the causal effects of education on freedom to choose ones own spouse. Education improves the outside option of children in bargaining with parents due to better labour market opportunities, among other things. The choice of spouse is among the most important decisions in human life, and freedom in spouse choice is used here as an indicator of freedom of choice in general in social interactions.

To the best of our knowledge, this is the first paper in the literature to provide evidence on the causal effects of education on freedom of spouse choice. Using data from ten communes in the Red River Delta region in Vietnam, we estimate the causal effect of education on the probability of arranged marriage where parents choose the spouse with or without inputs from children. For identification, we rely on instruments representing variations in school supply across geographic space and exogenous shocks to schooling attainment for different birth cohorts because of varying intensity of war and conflicts and associated disruptions. To capture geographic variations in schooling supply we use birthplace. Birthplace is a good indicator of relevant schooling supply in our data set as more than 98 per cent of respondents grew up in their place of birth, at least in part reflecting lack of geographic mobility due to the Ho Khau internal registration system. We use a dummy for cohorts born in 1955 and earlier to capture the variations in schooling attainment over time. The cohort dummy represents the positive shock to the educational attainment of the older cohorts (born in 1955 or earlier) because of relative stability during 1954-1965 in the Red River Delta region. This also captures the negative shock to educational attainment of later cohorts caused by the US bombing (and other disruptions) in the Red River delta region during the Vietnam War.

The empirical results provide strong evidence in favour of a both numerically and statistically significant negative effect of education on the probability of arranged marriage. According to the estimates from the instrumental variables approach, one year of additional

schooling reduces the probability of arranged marriage by approximately 14 percentage points for an individual with eight years of education. The estimates from Probit model imply that ten years of schooling for an individual would reduce the probability of arranged marriage to close to zero in Vietnam. There are significant gender differences in the causal effects of education; the impact of education is numerically higher and statistically stronger in the case of women. The evidence of a causal effect of education on arranged marriage in the case of men is much weaker, both in terms of numerical magnitude and statistical significance.

The conclusion that education has a negative causal effect on the probability of arranged marriage does not depend on the exclusion restrictions imposed in the IV approach. The results from the Conley et al. (2012) bounds approach under weaker exclusion restrictions also support this conclusion. Although the lower bound estimate from the Conley et al.(2012) approach is smaller, the effect of education is still substantial; one year of additional schooling leads to an approximately six to seven percentage point reduction in the probability of arranged marriage for an individual with average level of education (eight years of schooling). The empirical evidence presented in this paper thus points to important benefits of education to individuals in social interactions. The focus on labour market returns to education in the economics literature thus might be understating the full private and social net returns to education.

In addition to its inherent country-specific and historical interest, this paper also provides insights, and motivation for further research, well beyond the current study. In particular, in other countries, incidence of arranged marriage remains high; and average educational attainment by children remains well below standards achieved in Vietnam. Moreover, the study points to methodological strategies for addressing broader questions of the causal effects of education on various forms of freedom, providing a framework for further development of this research. Finally, the study also provides a general framework for the analysis of the broader notion of nonmarket returns to education.

Notes

[1] As pointed out by an anonymous referee, the presence of government policies both encouraging education and discouraging arranged marriage in the study period suggest another plausible reason for the correlation; this also motivates the use of an IV approach.

[2] The Red River Delta is considered the center of ethnic Vietnamese culture. Vietnamese families are usually characterized as part of a Confucian patrilineal, patrilocal, and patriarchal cultural heritage (Keyes, 1996). The main variation pointed up by anthropologists is the view of Condominas (1957), and Keith (2011), that 'Catholics in Vietnam are unique.' In fact our data show that Catholics have a somewhat higher incidence of arranged marriage than non-Catholics. We control for Catholic adherence in our regression analysis (we also control for Buddhist and other religious adherence).

[3] In our sample, the incidence of arranged marriage is approximately 25 per cent. The survey was carried out by a group of Vietnamese and American sociologists in 1995, and a large number of respondents (63 per cent) belong to birth cohorts before 1960. The incidence of arranged marriage has declined significantly in recent years in Vietnam because of expansion of education and improved labour market opportunities, plus a 1959 law that had statistically significant impact, among other factors. In fact, the results presented in this paper imply that the probability of arranged marriage should become very small once an individual in Vietnam attains ten years of schooling, controlling for other factors.

[4] There are 27 birthplace dummies we can use as instruments. With only one endogenous regressor, this is likely to create a weak IV problem. As a result, we reduce the dimension of the instruments sets, as described in section four.

[5] This is similar to the Hukou system in China. The fact that Ho Khau was implemented together with land reform suggests a further reason for its effectiveness. The Ho Khau system did not become more circumventable until the 1990s, thus providing an additional explanation of why mobility remained so very low throughout the relevant sample period. For details on 1990s channels of circumvention see Hardy (2001). While the presence of the Ho Khau system helps explain the very low mobility, it is not the institution but rather the fact that almost all individuals did grow up in their place of birth that is relevant for our present econometric purposes.

[6] For an analysis of the role of labour market opportunities in the transition from arranged marriage to free spouse choice in medieval Europe, see de Moor and Van Zanden (2005).

[7] Since most of agriculture consists of self-employment, a non-farm occupation is a reasonable indicator of labour market opportunities. If alternatively we use finer occupational categories at the time of marriage, the main results of the paper remain robust.

[8] For discussions on wars in Vietnam, see Harrison (1989) and Young (1991).

[9] The US aerial bombing of North Vietnam started on 2 March 1965 with Operation Rolling Thunder. Another major air campaign was the December 1972 in Operation Linebacker II. Potential sources of disruption include the moving of schools to different locations and enlistment of fathers of school age children into the military. The impact on education is found despite the longer term rising trend in educational attainment. It is not that years of schooling significantly decreased during this period, but rather that the increase became slower than the longer-term trend. This meant that otherwise similarly situated children – taking account of the trend and other variables for which we control – likely attained fewer years of schooling in this period other things equal. It is that source of variation that we are using to demonstrate the causal effect of education on the incidence of arranged marriage through this instrumental variable.

[10] One could argue that the oldest age cohort, born long before 1954, would not experience the fruits of relative calm during the 1954-1965 period, as they would have been past their school age. Thus, as one of our several robustness checks, we focus on an interval of age cohort that also excludes the oldest age cohorts; and define a dummy for year of birth between 1945-1955. The main conclusions of this paper are not sensitive to this alternative definition of the cohort dummy and we omit the results for sake of brevity.

[11] Observe that the changing sex ratio primarily affects the bargaining between the bride and bridegroom. Since war is likely to negatively affect the availability of eligible men in the marriage market, it would tend to improve the bargaining position of men (and their parents) vis a vis the women (and their parents). However, there are no compelling reasons to believe that it would affect the bargaining between the parents and children, which is the focus of this paper. An excellent discussion on the sex ratio in Vietnam, especially during the periods after the war, is provided by Goodkind (1997), who notes 'The marriage squeeze in Vietnam against women was most severe in the 1970s and 1980s.' Another potential objection to the exclusion restriction imposed on the cohort dummy is that it might proxy for changing formal rules or informal norms regarding marriage. We control for age of the respondent to account for changing views regarding arranged marriage across different generations. Age also controls for any other age-specific changes that can potentially affect both marriage choices and parental choices in education. In addition, a dummy for marriages before 1960 is included to account for the reform of marital law in Vietnam (officially enacted 27 December, 1959) - although it seems to have had limited effect on the ground for many years, as arranged marriages continued to take place albeit at a declining rate (see author appendix Figure A.1).

[12] For more complete discussion of identification issues and the roles played by different control variables, see Section IV below.

[13] In other words, such omitted variables do not cause any omitted variables bias.

[14] Such asymmetrical response to education may reflect, among other things, the fact that in many cases the groom chooses from a 'menu' of brides screened initially by the parents. This is not pure arranged marriage for the groom, but still is arranged for the bride who does not have any say.

[15] To the best of our knowledge there is no evidence on the suitability of a unitary model to explain intra-household decisions

between parents and children. There is, however, overwhelming evidence against the unitary model in the context of intra-household allocations between husband and wife (see, for example, Haddad and Kanbur, 1990; Lundberg and Pollak, 1996).
[16] We would also note that even in the cases of apparent general agreement concerning the principle of arranged marriage, there is plenty of room for specific disagreement.
[17] For a theoretical analysis of arranged marriage where fixed costs of new household formation plays an important role, see Dasgupta et al., (2008).
[18] A related issue which might create a spurious negative correlation between education and arranged marriage is the possibility that the parents arrange marriage while the child is attending school, and his/her education is truncated as a result of the marriage. This, however, seems not to be a problem in our data set, as there are only a few observations (21 observations out of 4464) where the timing of marriage corresponds to that of withdrawal from school.
[19] http://csde.washington.edu/research/projects/hirschman/vietnam/vls.html
[20]Although in general it would have been econometrically advantageous to exploit the four-year panel data set features, for our variables of interest, in particular regarding marital status, there are no significant changes across various years. For example, only 42 among the 4464 changed marital status during those four years. Our empirical analysis is based on the 1995 data.
[21] In the empirical implementation, we also report results from the corresponding linear probability model, which do not depend on distributional assumptions.
[22] The data set also has information on the characteristics of the spouse. The characteristics of spouse can potentially be proxy variables for unobserved preference of parents and children. In the context of arranged marriages, the characteristics of the spouse might capture parental unobserved heterogeneity, as they reflect parental preference. In the case of own choice of spouse, they might reveal the preferences of children. However, the characteristics of the spouse are not exogenous as they depend on the nature of spouse choice. We thus chose not to use them as controls in the presentation. However, we note that the estimates of the causal effect of education on arranged marriage presented remain virtually identical if we include spousal characteristics.
[23] We thank James Foster for suggesting this interpretation.
[24] The sample consists of 17 per cent Christian, 20 per cent Buddhist, and 63 per cent with no identified religion. Some of those professing no religion engaged in ancestral worship practices; we thank an anonymous referee for pointing this out.
[25] As mentioned earlier, the lack of spatial mobility in Vietnam was reinforced, especially in the rural areas, by an identity card system known as Ho Khau (residence registration book), initiated in 1954 in conjunction with land reform (Hardy, 2001).
[26] The results are available from the authors.
27 The Red River Delta was a focus of bombing due to the presence of important rail links or industrial installations.
[28] According to one observer, '[I]n fact, the implementation of these laws has not been simple, and the social reality is often far cry from the ideals posited by the law-makers. A massive effort is needed to modify people's thinking and to establish in the minds and behavior of the masses the new ideas about marriage and the family. This requires education, and the elimination of not only long-standing traditions, but also misconceptions, and even strong opposition. The question is how to educate people to renounce the traditions of polygamy, child marriages, arranged marriages…' (Van Bich, 1999, p.59). This is consistent with a well-established literature in development economics regarding differences between changes in formal and informal institutions; as Douglass North (1994) stresses, even if formal rules 'may be changed overnight, the informal rules usually change only ever so gradually.' The 1960 law also banned bride price; although this topic is outside the scope of our research, Teerawichitchainan and Knodel (2011) argue 'The socialist attempts to eradicate bride price had moderate impacts in the North' and note 'the persistence of traditional values and cultural resilience of marriage payments in the Vietnamese society.' Their data show that substantial levels of bride price transfers are present despite their illegality, which is consistent with our empirical finding that the 1960 law can explain only a rather limited fraction of the decline in arranged marriage. Their evidence suggests a strong cultural resilience of 'brideprice, dowry, and bidirectional transfers'; in future research it would be interesting to examine whether their incidence is impacted by education, other things equal. Further details on the law are found in Scornet (2009).
[29] We would like to thank an anonymous referee for suggesting that we emphasize this point and for pointing out that the North Vietnamese government at the time sought to prohibit also 'other Confucian-based family practices and viewed this as an essential step towards the elimination of private ownership and inequality.' In this period (late 1950s and early 1960s), the government 'also expanded basic educational opportunities for both girls and boys as well as for illiterate adults,' which would extend to parents who may make decisions about arranging marriage. Our data show that a pre-1960 marriage was statistically significantly more likely to be arranged. We also find that arranged marriage was declining prior to, and continued its ongoing decline well after, the promulgation of the law. In particular, the data indicate that a substantial percentage of marriages were arranged in the period after 1960; this can be seen in Figure A1, in an Appendix available from the authors upon request.
[30] The randomized assignment of one year of additional schooling is not possible, a point emphasized by Angrist and Krueger (2001). Thus one has to rely on other sources of potential exogenous variations in years of schooling for identification.
[31] The Probit (Table 2) and linear probability model (appendix Table A.1) estimates show that parents' number of siblings (both father and mother) exerts a robust positive effect on the probability of arranged marriage, while own number of siblings is not significant. Also, wage income is a robust negative determinant of arranged marriage. The 1960 legal reform dummy is significant indicating the legal change had some impact on the incidence of arranged marriage. Access to information (TV and radio) seems to reduce the probability of arranged marriage.

Bibliography

Acemoglu, D. and Angrist, J.D. (2000) How Large Are Human-Capital Externalities? Evidence from Compulsory Schooling Laws. *NBER Macroeconomics Annual*, pp. 9-59.

Aghion, P., Howitt, P., Brant-Collett, M. and García Peñalosa, C. (1998) *Endogenous Growth Theory*. (Cambridge, MA: MIT Press).

Alderman, H., Behrman, J.R., Ross, D.R. and Sabot, R. (1996). The Returns to Endogenous Human Capital in Pakistan's Rural Wage Labour Market. *Oxford Bulletin of Economics and Statistics*, 58(1), pp. 29-55.

Amemia, T. (1966) On the Use of Principal Components of Independent Variables in Two-Stage Least Square Estimation. *International Economic Review,* 7 (3).

Angrist, J.D. and Krueger, A.B. (1991) Does Compulsory School Attendance Affect Schooling and Earnings? *Quarterly Journal of Economics*, 199, pp. 979-1014.

Angrist, J.D. and Krueger, A.B. (1992) The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples. *Journal of the American Statistical Association*, 418, pp. 328-36.

Angrist, J. D. and Krueger, A. B. (2001) Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. *Journal of Economic Perspectives*, 15, pp. 69-85.

Bai, J and Ng, S. (2008) Selecting Instrumental variables in a Data Rich Environment. *Journal of Time Series Econometrics*,1 (1).

Barbieri, M. and Belanger, D. (2009) *Reconfiguring Families in Contemporary Vietnam.* (Stanford: Stanford University Press).

Basu, K., Narayan, A. and Ravallion, M. (2001) Is Literacy Shared within House- Holds? Theory and Evidence for Bangladesh. *Labour Economics*, 8(6), pp. 649-65.

Behrman, J.R. and Wolfe, B.L. (1987) How Does Mothers Schooling Affect Family Health, Nutrition, Medical Care Usage and Household Sanitation? *Journal of Econometrics,* 36(1/2), pp. 185-204.

Blundell, R., Dearden, L. and Sianesi, B. (2005) Evaluating the Effect of Education on Earnings: Models, Methods and Results from the National Child Development Survey. *Journal of the Royal Statistical Society Series A*, 168(3), pp. 473-512.

Blundell, R. and Smith, R. J. (1986) An Exogeneity Test for a Simultaneous Tobit Model. *Econometrica,* 54, pp. 679-85.

Bound, J., Jaeger, D., and Baker, R. (1995) Problems with Instrumental Variables Estimation when the Correlation between the Instruments and the Endogenous Variables is Weak. *Journal of the American Statistical Association* 90, 443-50

Browning, M. and Chiappori, P.A. (1998) Efficient Intra-Household Allocations: A General Characterization and Empirical Tests. *Econometrica*, 66(6), pp. 1241-78.

Card, D. (1995) Earnings, Ability and Schooling Revisited. *Research in Labour Economics*, 14, pp. 23-48.

Card, D. (2001). Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems. *Econometrica*, 69, pp. 1127-60.

Condominas, G. (1997) *We have eaten the forest: the story of a Montagnard village in the central highlands of Vietnam* **(**New York: Hill and Wang)

Conley, T., Hansen, C.B., and Rossi, P. (2012) Plausibly Exogenous. *Review of Economics and Statistics*, 94(1), pp. 260-272.

Dasgupta, I., P. Maitra, and D. Mukherjee. (2008) Arranged Marriage, Co-Residence and Female Schooling: A Model with Evidence from India. *CREDIT Working paper, Univ. of Nottingham.*

De Moor, T. and van Zanden, J.L. (2005) Girl Power: The European Marriage Pattern and Labour Markets in the North Sea Region in the Late Medieval and Early Modern Period. *IISG Amsterdam.*

Duflo, E. (2001) Schooling and Labour Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment. *American Economic Review*, 91, pp. 795-813.

Durlauf, S.N., Johnson, P.A. and Temple, J. (2005) Growth Econometrics. *Handbook of economic growth*, 1, pp. 555-677.

Edlund, L. and Lagerlof, N.P. (2006) Individual Versus Parental Consent in Marriage: Implications for Intra-Household Resource Allocation and Growth. *American Economic Review*, pp. 304-07.

Finlay, K. and Magnusson, L. (2009) Implementing Weak Instrument Robust Tests for a General Class of Instrumental Variables Models. *Stata Journal*, 9, pp. 398-421.

Glewwe, P. (1996) The Relevance of Standard Estimates of Rates of Return to Schooling for Education Policy: A Critical Assessment. *Journal of Development Economics*, 1(2), pp. 267-90.

Goode, W.J. (1963*), World Revolution and Family Patterns*. (Oxford: Free Press Glencoe).

Goodkind, D. (1997) The Vietnamese Double Marriage Squeeze. *International Migration Review*, 31(1), pp.108-127

Haddad, L. and Kanbur, R. (1990) How Serious Is the Neglect of Intra-Household Inequality? *The Economic Journal*, 100(402), pp. 866-81.

Hansen, L.P., Heaton, J. and Yaron, A. (1996) Finite-Sample Properties of Some Alternative GMM Estimators. *Journal of Business & Economic Statistics*, pp. 262-80.

Hardy, A. (2001) Rules and Resources: Negotiating the Household Registration System in Vietnam under Reform. *SOJOURN: Journal of Social Issues in Southeast Asia* 16, 2, 187-212.

Harrison, J. P. (1989) *The Endless War: Vietnam's Struggle for Independence.* (New York: Columbia University Press)

Haveman, R.H. and Wolfe, B.L. (1984) Schooling and Economic Well-Being: The Role of Non Market Effects. *Journal of Human Resources*, pp. 377-407.

Haveman, R.H. and Wolfe, B.L. (2002) Social and Nonmarket Benefits from Education in an Advanced Economy. *Conference Series-Federal Reserve Bank of Boston. 97-131.*

Heckman, J.J. (2005) China's Human Capital Investment. *China Economic Review*, 16(1), pp. 50-70.

Heckman, J.J. and Li, X. (2003) Selection Bias, Comparative Advantage and Heterogeneous Returns to Education: Evidence from China in 2000. *NBER Working Paper Series.*

Jamison, D.T. and Lau, L.J. (1982) *Farmer Education and Farm Efficiency*. (Washington: World Bank).

Jejeebhoy, S.J. (1995) *Women's Education, Autonomy, and Reproductive Behaviour: Experience from Developing Countries*. (Oxford: Clarendon Press).

Kenkel, D.S. (1991) Health Behaviour, Health Knowledge, and Schooling. *Journal of Political Economy*, pp. 287-305.

Keyes, C.F. (1996) Being Protestant Christians in Southeast Asian Worlds. *Journal of Southeast Asian Studies*, pp. 280-92.

Kurosaki, T. and Khan, H. (2006) Human Capital, Productivity, and Stratification in Rural Pakistan. *Review of Development Economics*,10(1), pp. 116-34.

Lafortune, J. (2009) Making Yourself Attractive: Pre-marital Investments and returns to Education in the marriage market. *Working paper, University of Maryland*.

Lundberg, S. and Pollak, R. A. (1996) Bargaining and Distribution in Marriage. *Journal Of Economic Perspectives*, 10(4), pp. 139-58.

Magnusson, L. (2008) Inference in Limited Dependent Variable Models Robust to Weak Identification. *Tulane Economics Working Paper Series, Working Paper 0801*.

Merli, M. G. (2000) Socioeconomic Background and War Mortality During Vietnam's Wars. *Demography*, 37(1), pp. 1-15

North, D. (1994) Economic Performance through Time. *American Economic Review*, 84, 359-36.

Psacharopoulos, G. and Patrinos, H. (2004) Returns to investment in education: a further update. *Education economics*, 12(2):111-134.

Rivers, D. and Vuong, Q. H. (1988) Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models. *Journal of Econometrics*, 39(3), pp. 347-66.

Romer, P.M. (1989) Human Capital and Growth: Theory and Evidence. *NBER Working paper.*

Romer, P.M. (1994) The Origins of Endogenous Growth. *Journal of Economic Perspectives*, 8(1), pp. 3-22.

Rosenzweig, M.R. and Schultz, T.P. (1989) Schooling, Information and Nonmarket Productivity: Contraceptive Use and Its Effectiveness. *International Economic Review*, 30(2), pp. 457-477.

Scornet, C., (2009) State and Family: Reproductive Policies and Practices in: Barbieri, M., Belanger, D. (eds) *Reconfiguring Families in Contemporary Vietnam* (Stanford, CA: Stanford University Press), pp. 41-74

Sen, A. K. (1999) Development as Freedom. (Oxford: Oxford University Press)

Stock, J. and Watson, M. (2008) What's New in Econometrics. *NBER Summer Institute.*

Teerawichitchainan, B. (2009) Trends in Military Service in Northern Vietnam, 1950–1995: A Sociodemographic Approach. *Journal of Vietnamese Studies*, Vol. 4, No. 3, pp. 61-97

Teerawichitchainan, B. and Knodel, J. (2011) Change and Persistence in Marriage Payments in Vietnam, 1963-2000. *PSC Research Report* No. 11-733. University of Michigan, March 2011.

Teerawichitchainan, B. (2009) Household Headship in the Red River Delta, Vietnam: The Political Construction of the Family, in: Barbieri, M., Belanger, D. (eds) *Reconfiguring Families in Contemporary Vietnam* (Stanford: Stanford University Press), pp. 329-364

Van Bich, P. (1999) *The Vietnamese Family in Change: The Case of Red River Delta*. (Richmond, UK: Curzon Press).

Yang, D. T. and An, M. Y. (2002) Human Capital, Entrepreneurship, and Farm Household Earnings. *Journal of Development Economics*, 68(1), pp. 65-68.

Young, M. B. (1991) *The Vietnam Wars, 1945-1990*. (New York: Harper Collins).

Table 1 Summary Statistics

	Mean	Std.	Min	Max
Arranged Marriage	0.25	0.43	0	1
Arranged Marriage (male)	0.24	0.43	0	1
Arranged Marriage (female)	0.25	0.43	0	1
Years of schooling	8.07	2.91	0	15
Years of schooling (male)	8.59	2.78	0	15
Years of schooling (female)	7.77	2.86	0	15
Schooling of (above 8 years)	0.71	0.45	0	1
Schooling of (above 5 years)	0.82	0.38	0	1
Age	39.18	10.77	17	65
Age at marriage	22.54	4.12	11	53
Proportion of married in or before 1960	0.10	0.30	0	1
Number of siblings	5.02	2.16	0	11
Has no religion	0.63	0.48	0	1
Met spouse in school	0.04	0.19	0	1
Male	0.47	0.5	0	1
Father is Farmer	0.67	0.47	0	1
Father is born	0.52	0.50	0	1
Father's number of siblings	3.75	2.17	0	14
Mother is Farmer	0.85	0.36	0	1
Mother is first born	0.46	0.50	0	1
Mother's number of siblings	3.62	2.13	0	14
HH's number of children	3.24	1.80	0	14
HH's has brickwall	0.97	0.16	0	1
HH's has access to information (tv, radio)	0.61	0.49	0	1
HH's agricultural income	1.89	1.62	0	12
HH's Wage	2.99	2.72	0	16
Cohort Dummy (equals 1 if born in or before 1955)	0.43	0.5	0	1
Proportion of born in or before 1960	0.63	0.48	0	1
Birthplace Dummy 1	0.70	0.46	0	1
Birthplace Dummy 2	0.20	0.40	0	1
Sex Ratio (Males to Females)	93.64	7.57	84.70	106.60
Labor Market Opportunities$_1$	0.40	0.11	0	0.64

$_1$Indicates the proportion in the Non-Farm sector at each year of marriage
Number of Observations: 3219
Vietnam Longitudinal Survey 1995

Figure 1: Schooling and Probability of Arranged Marriage (Probit without any controls)

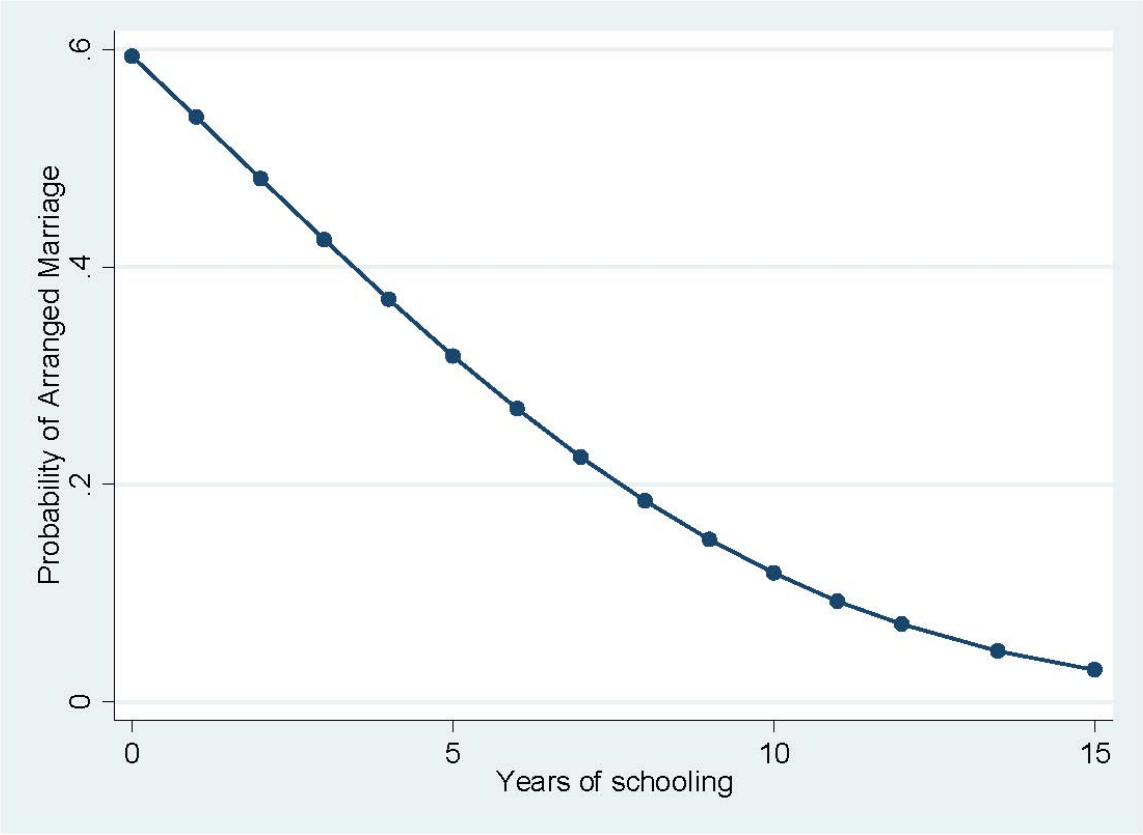

Table 2: Estimates from Probit (Coefficients and Marginal Effects)

Dependent Variable: Dummy for arranged marriage	(1)		(2)		(3)		(4)	
	Coefficient	M.E.$_1$	Coefficient	M.E.	Coefficient	M.E.	Coefficient	M.E.
Years of schooling	**-0.07*** **	**-0.02*** **	**-0.07*** **	**-0.02*** **	**-0.07*** **	**-0.02*** **	**-0.05*** **	**-0.01*** **
	(-6.76)	(-6.79)	(-6.51)	(-6.54)	(-6.51)	(-6.54)	(-4.07)	(-4.08)
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Other Controls								
Age	-0.017	-0.005	-0.007	-0.002	-0.007	-0.002	0.002	0.001
Age Squared	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Age at marriage	-0.06***	0.02***	0.06***	0.02***	0.06***	0.02***	0.06***	0.02***
Married b. 1960	0.41**	0.14**	0.45***	0.15**	0.45***	0.15**	0.55***	0.18***
Number of Siblings	0.02*	0.01*	0.01	0.00	0.01	0.00	0.01	0.00
Has no religion	-0.18***	-0.05***	-0.15***	-0.04***	-0.15***	-0.04***	-0.13***	-0.04**
Met spouse in school	-0.60***	-0.14***	-0.62***	-0.14***	-0.62***	-0.14***	-0.65***	-0.14***
Male	0.18***	0.05***	0.17***	0.05***	0.17***	0.05***	0.13**	0.04**
Father is farmer			0.04	0.01	0.04	0.01	0.01	0.00
Father first born			-0.01	0.00	-0.01	0.00	0.00	0.00
Father number of siblings			0.04***	0.01***	0.04***	0.01***	0.04***	0.01***
Mother is Farmer			0.12	0.04	0.12	0.04	-0.04	-0.01
Mother is first born			0.04	0.01	0.04	0.01	0.03	0.01
Mother's number of siblings			0.04***	0.01***	0.04***	0.01***	0.03**	0.01**
HH's number of children							-0.03*	-0.01**
HH's has brickwall							0.31*	0.08**
HH's has access to information (tv, radio)							-0.14**	-0.04**
HH's agricultural income							-0.02	0.00
HH's Wage							-0.07***	-0.02***
Rural Dummy							0.26**	0.07***
Labor Market Opportunities$_1$	-0.23	-0.07	-0.19	-0.06	-0.19	-0.06	-0.01	0.00
Sex Ratio	-0.003	-0.001	-0.002	-0.001	-0.002	-0.001	-0.005	-0.002
Pseudo R-squared	0.144	0.144	0.151	0.151	0.151	0.151	0.17	0.17

$_1$Indicates the proportion in the Non-Farm sector at each year of marriage

Number of Observations: 3219

Robust p-values in brackets, Robust t-stat in parenthesis

***P<0.01, **P<0.05, *p<0.1

Table 3

Panel A: Instrumental Variable Estimates (Coefficients and Marginal Effects)

Dependent variable: Dummy for Arranged Marriage

	Control Function		IV-Probit		2SLS	GMM	CUE-GMM
	Coefficient	M.E.$_1$	Coefficient	M.E.	Coefficient	Coefficient	Coefficient
Years of Schooling	**-0.59*** **	**-0.17*** **	**-0.38*** **	**-0.13*** **	**-0.13*** **	**-0.13*** **	**-0.13*** **
	(-5.18)	(-5.21)	(-13.20)	(-8.79)	(-4.29)	(-4.40)	(-4.38)
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
First Stage Residuals	0.55***	0.16***					
Pseudo R-squared	0.18	0.18					
Diagnostics for Instruments							
F stat for Exclusion of Instruments					12.49	12.49	12.49
Hansen's J p value					0.811	0.811	0.811

Summary Results for First Stage :OLS

Birthplace Dummies	
F-Stat	18.04
P-Value	[0.00]
Cohort Dummy	
Coefficient	0.32**
t-stat	(2.05)
p-value	[0.04]
R-squared	0.38

Panel B: Finlay-Magnusson Weak IV Robust Inference, H_0: beta[Arranged: Years of Schooling]=0

	Linear IV			IV Probit (Test for Coefficients not for M.E.)		
Test	Statistic	p-value	95% C.I.	Statistic	p-value	95% C.I.
CLR_2	35.99	0.00	[-.20,-.08]	26.89	0.00	[-.49,-.36]
AR_3	36.52	0.00	[-.24,-.07]	29.28	0.00	[-.49,-.32]

Notes: (1) Control variables are the same as in Table 2 (2) Number of Observations: 3219

$_1$M.E. stands for Marginal Effects. M.E. are evaluated at the sample mean.

$_2$Conditional Likelihood Ratio Test

$_3$Anderson- Rubin Test

Robust p-values in brackets, Robust t-stat in parenthesis

***P<0.01, **P<0.05, *p<0.1

Table 5: Gender, Education and Arranged Marriage
Table 4: Bounds for the effect of education on Arranged Marriage

Support for possible values of $\theta = \theta_1 = \theta_2$	95% confidence interval	
	Lower	Upper
$\theta \in [-0.0001, +0.0001]$	-0.18	-0.07
$\theta \in [-0.0005, +0.0005]$	-0.18	-0.07
$\theta \in [-0.001, +0.001]$	-0.19	-0.07
$\theta \in [-0.005, +0.005]$	-0.19	-0.06
$\theta \in [-0.01, +0.01]$	-0.20	-0.06

Notes: (1) Control variables are the same as in Table 2
(2) Bounds are estimated using approach due to Conley et al. (2008)
(3) Number of Observations: 3219

Dependent variable: Dummy for Arranged Marriage

Panel A: Female Sample

	Control Function		IV-Probit		2SLS	GMM	CUE-GMM
	Coefficient	M.E.$_1$	Coefficient	M.E.	Coefficient	Coefficient	Coefficient
Years of Schooling	**-0.74*** **	**-0.22*** **	**-0.42*** **	**-0.15*** **	**-0.17*** **	**-0.17*** **	**-0.17*** **
	(-5.16)	(-5.22)	(-14.14)	(-9.55)	(-3.82)	(-3.87)	(-3.86)
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
First Stage Residuals	0.70***	0.20***					
Pseudo R-squared	0.18	0.18					
Diagnostics for Instruments							
F stat for Exclusion of Instruments					8.76	8.76	8.76
Hansen's J p-value					0.91	0.91	0.91

Finlay-Magnusson Weak IV Robust Inference, H_0: beta[Arranged: Years of Schooling]=0

	Linear IV			IV Probit (Test for Coefficients not for M.E.)		
Test	Statistic	p-value	95% C.I.	Statistic	p-value	95% C.I.
CLR_2	34.61	0.00	[-.29,-.10]	26.47	0.00	[-.53,-.43]
AR_3	34.87	0.00	[-.34,-.08]	27.27	0.00	[-.53,-.34]

Number of Observations: 1716

Panel B: Male Sample

	Control Function		IV-Probit		2SLS	GMM	CUE-GMM
	Coefficient	M.E.$_1$	Coefficient	M.E.	Coefficient	Coefficient	Coefficient
Years of Schooling	**-0.27**	**-0.08**	**-0.28** **	**-0.09***	**-0.05**	**-0.06***	**-0.06***
	(-1.55)	(-1.55)	(-2.39)	(-1.86)	(-1.54)	(-1.78)	(-1.80)
	[0.12]	[0.12]	[0.06]	[0.12]	[0.12]	[0.08]	[0.07]
First Stage Residuals	0.22	0.06					
Pseudo R-squared	0.18	0.18					
Diagnostics for Instruments							
F stat for Exclusion of Instruments					4.32	4.32	4.32
Hansen's J p-value					0.57	0.57	0.58

Panel B2: Finlay-Magnusson Weak IV Robust Inference, H_0: beta[Arranged: Years of Schooling]=0

	Linear IV			IV Probit (Test for Coefficients not for M.E.)		
Test	Statistic	p-value	95% C.I.	Statistic	p-value	95% C.I.
CLR_2	3.89	0.07	[-.18,-.00]	2.39	0.15	[-.75,-.11]
AR_3	4.95	0.18	[-.19,-.02]	5.85	0.12	[-.75,-.11]

Number of Observations: 1503

$_1$M.E. stands for Marginal Effects. M.E. are evaluated at the sample mean.

$_2$Conditional Likelihood Ratio Test, $_3$Andrerson-Rubin Test

Appendix

Table A-1: Linear Probability Model Estimates

Dependent Variable: Dummy for arranged marriage	(1)	(2)	(3)	(4)
Years of schooling	**-0.02*** **	**-0.02*** **	**-0.02*** **	**-0.01*** **
	(-6.91)	(-6.67)	(-6.67)	(-4.22)
	[0.00]	[0.00]	[0.00]	[0.00]
Other Controls				
Age	0.00	0.00	0.00	0.00
Age Squared	0.00	0.00	0.00	0.00
Age at marriage	-0.02***	-0.02***	-0.02***	-0.02***
Married b. 1960	0.018***	0.020***	0.020***	0.021***
Number of Siblings	0.01*	0.00	0.00	0.00
Has no religion	-0.06***	-0.05***	-0.05***	-0.04***
Met spouse in school	-0.12***	-0.12***	-0.12***	-0.13***
Male	0.05***	0.04***	0.04***	0.03**
Father is farmer		0.01	0.01	0.00
Father first born		0.00	0.00	0.00
Father number of siblings		0.01***	0.01***	0.01***
Mother is Farmer		0.02	0.02	-0.02
Mother is first born		0.01	0.01	0.01
Mother's number of siblings		0.01***	0.01***	0.01***
HH's number of children				-0.01
HH's has brickwall				0.08**
HH's has access to information (tv, radio)				-0.04**
HH's agricultural income				-0.01
HH's Wage				-0.02***
Rural Dummy				0.06**
Labor Market Opportunities$_1$	-0.12	-0.13	-0.13	-0.08
Sex Ratio	0.00	0.00	0.00	0.00
Constant	0.82**	0.62**	0.62**	0.56
R-squared	0.17	0.18	0.18	0.19

$_1$Indicates the proportion in the Non-Farm sector at each year of marriage

Number of Observations: 3219

Robust p-values in brackets, Robust t-stat in parenthesis

***P<0.01, **P<0.05, *p<0.1