Estimation of causal effects of fertility on economic wellbeing: Evidence from rural Vietnam

Arnstein Aassve
(ISER, University of Essex)

Bruno Arpino
(University of Florence)

ISER Working Paper
2007-24

Acknowledgement:

This paper forms part of the project "Poverty dynamics and fertility in developing countries", funded by the ESRC grant no. We are grateful for comments by Fabrizia Mealli, Stefano Mazzuco, Letizia Mencarini and Stephen Pudney. All errors and inconsistencies in the paper are our own.

Readers wishing to cite this document are asked to use the following form of words:
Aassve, Arnstein, Bruno Arpino (November 2007) 'Estimation of causal effects of fertility on economic wellbeing: Evidence from rural Vietnam', ISER Working Paper 2007-24. Colchester: University of Essex.

The on-line version of this working paper can be found at http://www.iser.essex.ac.uk/pubs/workpaps/

The Institute for Social and Economic Research (ISER) specialises in the production and analysis of longitudinal data. ISER incorporates

- MISOC (the ESRC Research Centre on Micro-social Change), an international centre for research into the lifecourse, and

- ULSC (the ESRC UK Longitudinal Studies Centre), a national resource centre to promote longitudinal surveys and longitudinal research.

The support of both the Economic and Social Research Council (ESRC) and the University of Essex is gratefully acknowledged. The work reported in this paper is part of the scientific programme of the Institute for Social and Economic Research.

Institute for Social and Economic Research, University of Essex, Wivenhoe Park,
Colchester. Essex CO4 3SQ UK
Telephone: +44 (0) 1206 872957 Fax: +44 (0) 1206 873151 E-mail: iser@essex.ac.uk
Website: http://www.iser.essex.ac.uk

ABSTRACT

Estimating the effects of demographic events on households' living standards introduces a range of statistical issues. In this paper we analyze this topic considering our observational study as a quasi-experiment in which the treatment is expressed by childbearing events between two time points and the outcome is the change in equivalized household consumption expenditure. Our main question concerns how one can best estimate causal effects of demographic events on households' economic wellbeing. We first provide a brief discussion of different methods for causal inference stressing their differences with respect to the underlying assumptions and data requirement. In particular, we contrast methods relying on the *Uncounfoundedness Assumption* (UNA), such as regressions and propensity score matching, with methods allowing for selection on unobservables, such as the Instrumental Variable (IV) estimators. We stress the fact that these methods are not equivalent in what they estimate. With Regressions and Propensity Score Matching (PSM) we can identify and estimate the Average Treatment Effect (ATE) and the Average Treatment effect on the Treated (ATT), while IV methods give the Local Average Treatment Effect (LATE). Since LATE is the average causal effect of the treatment on the sub-group of compliers, it is generally different from ATE and ATT. Moreover, different instruments identify the effect on different groups of compliers giving different estimates of LATE. A problem for policy making is that the compliers are in general an unobserved sub-group. However, IV methods estimate relevant policy parameter if the instrument itself is a potential policy variable. We demonstrate these issues with an application on data derived from the Vietnam Living Standard Measurement Study.

NON-TECHNICAL SUMMARY

Several approaches are available in order to estimate causal effects. The appropriateness and interpretations of these models depend on the application at hand, and importantly, the available instruments. In many cases methods relying on the unconfoundedness assumption (UNA) are chosen, simply because instruments are hard to come by. The various implications of these methodological choices are rarely considered in applied work, but, as we point out, the underlying assumptions are important, especially when there is interest in comparing estimates from different methods. We discuss these methods in light of an application where we consider the effect of fertility on changes in consumption expenditure. The issue is that childbearing events cannot be considered as an exogenous measure of fertility, especially when the outcome relates to economic wellbeing – in our case measured in terms of consumption expenditure. However, the discussion of the methods is general and applies to many other applications.

Using methods based on the UNA assumption, such as simple linear regression and propensity score matching, we find that those households having children between the recorded waves have considerably worse outcomes in terms of changes in consumption expenditure. The negative impact is however, highly heterogeneous, and varies substantially with education for instance. We then demonstrate how one can make an assessment of the potential effect from omitting relevant but unobserved variables without actually implementing an Instrumental Variable (IV) approach. This is a very useful tool in the sense that valid and relevant instruments are often hard to come by. In our application the estimates are robust with respect to unobserved omitted variables. We find that the estimated effect becomes non-significant only if the association between the omitted covariate, selection and the outcome is extremely (and unreasonable) large.

Despite the robustness of the UNA in our application we implement nevertheless the IV method using two different instruments. The first is a well-used instrument that relates to couples' preference for sons. In this case the instrument is a binary variable taking value 1 in those households that at the

first wave had no male children and 0 otherwise. The IV estimation is implemented for sub-sample of households with at least two children in the first wave. Since the instrument is close to being randomised, a simple Wald estimator can be used. The second instrument takes value 1 for households residing in a community where none of the contraceptive methods IUD and condom was available at the first wave and 0 otherwise (at least one was available). This instrument is not randomized and hence requires controlling for covariates. Whereas both instruments satisfy the standard tests for relevance and validity, they provide very different parameter estimates. The fact that the IV estimates the Local Average Treatment Effect, as opposed to the Average Treatment Effect and Average Treatment Effect on Treated, is the key reasons for these differences. Moreover, since the second instrument requires inclusion of covariates, it typically involves more stringent assumptions on the functional forms. Using the approach suggested by Frölich (2007), which overcomes many of these assumptions, demonstrates that they do matter for the parameter estimates.

The use of Instrumental Variable methods in our application illustrates that reasonable instruments can lead to estimates that differ from those of methods based on UNA but also differ among them. In fact, compliers for one instrument can be very different from compliers to another instrument and consequently if the treatment effect is heterogenous the estimated LATE in the two cases will necessarily differ. With the first instrument we estimated a negative impact of fertility on poverty with a magnitude not dramatically different from that obtained by method based on the UNA. This could be an effect of the fact that the preference for son is quite a general phenomenon in Vietnam not involving particular kinds of households. The estimated proportion of compliers in this case is actually quite high: 20%. The estimate with the second instrument, on the contrary, is much higher, in absolute value. The estimated proportion of compliers in this case is small: 1%. This small sub-population of households reacting to the availability of contraceptives is likely to be highly selected. These households live in areas where no contraceptives were available. Clearly their opportunity to control fertility through contraceptive practices is much reduced as it is unlikely that compliers are able to get contraceptives from elsewhere. In this sense these households have a higher exposure to childbearing. These communities are also likely to be more disadvantaged compared to others.

Whereas the estimates based on this instrument is very different compared to the one based on the sex preference, an advantage is that it does have direct policy relevance, simply because the instrument itself is a policy variable. The effect on this sub-population is high and importantly, much higher than what is estimated for the whole population through the ATT and ATE. However, the size of this sub-population is rather small, which is an equally important consideration for the policy maker.

1 Introduction

A common observation in developing countries is that large households with many children tend to be poorer. The traditional explanation behind this pattern is that poor households relying on primitive farming with no state welfare provision depend on children as a source of cheap labour and insurance against old age. These households are trapped in an equilibrium of high fertility and high poverty. However, with the onset of the fertility transition, followed by diffusion of contraception and economic development, the underlying links between fertility and poverty (or living standards) become less clear cut. Whereas descriptive analysis and simple regressions will in most cases confirm that fertility remains positively associated with poverty, the behavioural mechanisms becomes different. As households become less dependent on children as a source of cheap labour and modern contraception is becoming available, households will have greater control over their fertility choices and with higher income and education, the reasons for having children become different. Despite declining fertility rates and economic development, fertility remains an important policy issue. In so far fertility has detrimental impact on living standards it remains a target for policy. Whereas there is a clear positive association between fertility and poverty, it is not equally clear to what extent fertility actually *leads* to a worsened economic situation. This is of course a very different question, since we are in this case interested in the causal effect of fertility on poverty, which ultimately is what we would need in order to give sound policy advice. The key problem is that it is generally difficult to find an exogenous measure of women's fertility.

In this paper we review methods for causal inference within the framework of the *potential outcomes* approach, stressing the fact that they rely on very different assumptions and data requirements. As an empirical application we consider childbearing as a treatment and a change in consumption expenditure as the outcome. Each household has two potential outcomes: Y_1 if it experiences a childbearing event between the two waves (treated) and Y_0 otherwise (untreated or control). The validity and usefulness of these methods for policy making depends of course on the specific application at hand. However, our analysis is general and relevant to other observational studies where the interest lies in recovering the causal effect of a demographic variable on certain outcomes.

The key problem in recovering the causal effect of fertility on poverty or economic wellbeing is to find an exogenous measure of fertility. Typically, recorded childbearing events are used as a measure of fertility. But childbearing is, at least in part, down to individual choice, giving rise to *self-selection*: households that choose to have more children (self selected into the treatment) may be very different from households that choose to have fewer children. Hence, if we observe that the first group of households has on average lower per capita expenditure, we cannot necessarily assert that this is due to fertility. In fact the two groups of households are likely to be different in respect to many other characteristics, such as education, which in turn impacts both fertility decisions and economic outcomes. As a consequence a simple difference in the average consumption (or income) for the two groups of households will give biased estimates of the causal effect of interest.

Many methods are available to correct for the selection bias problem. In observational studies we distinguish between two contexts: *regular* and *irregular* assignment mechanism (Imbens, 2002). In the first case we can reasonably assume that all the variables that impact both fertility and poverty are observed. In these scenarios we can use methods relying on the so-called *Uncounfoundedness Assumption* (UNA). Here we would compare households of similar characteristics that differ only by the childbearing event. The observed difference in economic wellbeing, either measured in terms of poverty status or more directly by consumption expenditure, can then be reasonably assumed to be driven by differences in fertility. Both multiple regression and propensity score matching relies on UNA, and we show both methods in section 3.

We then consider the more interesting case of irregular assignment mechanisms, also known as *latent regular assignment*, whereby selection also depends on *unobserved* characteristics. Childbearing is in this case endogenous despite controlling for observed characteristics. Endogeneity arises from the simple fact that the unobserved, but relevant variable(s) are omitted (i.e. giving rise to an omitted variables bias). There are several reasons why childbearing might be endogenous with respect to economic wellbeing. One obvious reason is that it is determined by adults' labour earnings. Given unobserved ability levels, fertility decisions are endogenous with respect to women's work decisions and therefore their earnings (Kim and Aassve, 2006). Unobserved ability may also influence take-up of modern contraception.

The instrumental variable approach has been the workhorse in economics, but relies of course on the fact that valid and relevant instruments are available. But even if an instrument is found, passing the standard tests for validity and relevance, the meaning of the associated parameter estimates depends on the nature of the instrument. The key reason is that instrumental variable estimates refer only to the unobserved sub-sample of the population that reacts to the chosen instrument, typically referred to as *compliers* (Angrist and Imbens, 1994; Angrist, Imbens and Rubin, 1996). The corresponding parameter estimate is consequently the *Local Average Treatment Effect* (LATE) which in general, in the presence of heterogeneous treatment effects, is different from Average Treatment Effect (ATE) and the Average Treatment Effect for the Treated (ATT) that we get from PSM estimation. This is of course important for policy analysis, since only if the instrument coincides with a variable of real policy relevance, can we also say that the estimated LATE has direct policy relevance (Heckman, 1997).

The fact that the method of IV estimates the LATE whereas PSM estimates ATT and ATE is problematic for comparing estimates. Any difference between the estimates will depend on the nature of the instrument used and only when a *constant* treatment effect can be assumed are we sure that IV estimates are representative of the whole population (as opposed to a sub-sample of compliers). The assumption of a constant treatment effect, is however strong in many cases. We demonstrate this by using two instruments described below.

On the basis of these issues, we consider here an application where interest lies in estimating the causal effect of fertility on economic wellbeing using data from Vietnam. We present first methods based on the UNA (regression and propensity score matching). In general instruments are often difficult to find and as outlined above, the nature of the instrument will have an impact on the parameter estimates and its interpretation. It is therefore useful to be able to perform sensitivity tests for the presence of potential unobserved confounders without actually implementing an IV procedure. We show how this is done for our application. Finally we implement the instrumental variable approach demonstrating its benefits and drawbacks by using two instruments. Even if these instruments are valid in the standard sense, we expect LATE to be generally different to ATE and ATT. The first instrument is the sex composition of existing children. This is a widely used instrument (see e.g. Angrist and Evans, 1998; Chun and Oh, 2002; Gupta and Dubey, 2003) and is based on the fact that parents in Vietnam tend to have a strong preference for boys, especially in the North (Haughton and Haughton, 1995; Johansson, 1996 and 1998; Belanger, 2002). The second instrument is the availability of contraception at the community level. This is similar to other well-used instruments related to the availability of services in the neighbourhood or its distance from the dwelling (examples include McClellan et al. (1994) who use proximity to cardiac care centers; Card (1995) who uses college proximity). The difference is however, that the former instrument can be thought of as randomised, while for the latter control for covariates is needed. In order to accommodate for covariates, the standard IV imposes functional form and additive separability in the error term. These and other strong assumptions can be avoided by implementation of a non-parametric approach as suggested by Frölich (2007).

Another crucial difference between these instruments is that the preference for sons is a wide spread phenomenon among Vietnamese households. In other words, we can expect the proportion of compliers to be quite high. In the second case, in contrast, we expect the proportion to be much smaller because very few communities in Vietnam still lack availability of contraception (Nguyen-Dinh, 1997; Duy et al., 2001; Anh and Thang, 2002). Therefore, households whose childbearing events depend on the availability of contraception in the community where they reside will be a rather small and selected sub-population. However, an advantage of the second instrument is that it corresponds to a potential policy variable on which policy makers can act to both reduce fertility and, *through it*, impact positively on poverty. Of course, areas without availability of contraceptives are few, which mean that it cannot be considered as a *general* policy tool.

The paper is organised as follow. Section 2 provides a review of the literature concerning the relationship between poverty and fertility and the Vietnamese context. Section 3 explains the statistical methods. Section 4 presents the Vietnamese Living Standard Measurement Survey and results. Section 5 concludes.

2 Fertility and economic wellbeing and the Vietnamese context

The interrelationship between fertility and economic wellbeing has received considerable interest in the economics literature. The traditional micro-economic framework considers children as an essential part of the household's work force as they generate income, as well as providing insurance against old age. This is especially true for male children. In rural underdeveloped regions of the world, which rely largely on a low level of farming technology and where households have no or little access to state benefits, this argument makes a great deal of sense (Admassie, 2002). In this setting households will have a high demand for children. The down side is that a large number of children participating in household production hamper investment in human capital (Moav, 2005). There are of course important supply side considerations to be made in this regard: rural areas in developing countries have poor access to both education and contraceptives, both limiting the extent couples are able to make choices about fertility outcomes (Easterlin and Crimmins, 1985).

As households attain higher levels of income and wealth, they also have fewer children, either due to a quantity-quality trade-off as suggested by Becker and Lewis (1973) or due to an increase in the opportunity cost of women earning a higher income as suggested by Willis (1973). Expansion of female education, which reduces women's willingness to give up work for childbearing, is possibly the most important driver behind increased opportunity cost and fertility decline. Consequently, fertility reduction is often seen as a direct result of increased empowerment of women through education. Educational infrastructure and educational policies are clearly important as higher compulsory childhood schooling will delay the onset of a young adult's working life, thereby reducing child labour (Livi-Bacci 2000; Kabeer 2001). Lack of education opportunities for women reinforces social norms of women's role and position in society. In many traditional societies, men's status depends very much on their ability to foster a large family and household heads are often considered more successful if they have many children. Such perceptions are likely to be stronger in rural areas, where, households always show a stronger gender bias in favour of boys when deciding to send kids to school. The consequence is that women's roles tend to be limited to childrearing and other household chores. With economic progress and urbanisation, however, women gain in empowerment through higher education and independence (Drovandi and Salvini, 2004). Social norms become weaker, and traditional demographic patterns fade, which is reflected by the demographic transition. Moreover, economic progress reduces labour intensive technologies, and thereby reduces the demand for child labour.

The extent to which these theoretical concepts apply to Vietnam is less clear. An important aspect is that Vietnam has experienced a tremendous decline in fertility over the past two decades, and at present one can safely claim that the country has completed the fertility transition. The figures speak for themselves: in 1980 Total fertility Rate (TFR) was 5.0, in 2003 it was 1.9. Naturally, fertility levels in rural areas remain higher than in urban areas, but with a rural population of 75 percent, the overall TFR reflect in any case a substantial decline in fertility. Contraceptive availability and knowledge is widespread and family planning programs were initiated already in the 1960s (Scornet, 2007).

Vietnam also scores well in terms of educational infra-structure, which is reflected by low rates of child labour. In 2003 this figure stood at 2.3% (out of children aged between 10 and 14), a reduction from 22 percent in 1980. An important factor behind these impressive figures is the country's socialist tradition which certainly ensured strong provision of education and health care. But equally important is the "Doi Moi" policy[1] that was introduced in the late eighties. The main elements of Doi Moi were replacement of collective farms by allocation of land to individual households; legalisation of many forms of private economic activity; removal of price controls; and legalisation and encouragement of Foreign Development Investment (FDI). Since the introduction of Doi-Moi, the country embarked on a remarkable economic recovery, followed by a substantial poverty reduction (Glewwe *et al.,* 2002). The average annual GDP growth was at a staggering 7 percent and in the period covered by the Vietnam LSMS panel (i.e. from 1993 to 1998), the growth rate was even higher at 8.9 percent. School enrolment rates increased during the period both for boys and girls. In particular, upper secondary enrolment rates increased from 6 to 27 percent for girls, and from 8 percent to 30 percent for boys (World Bank, 2000). Access to public health centres, clean water and other infrastructure have all improved, as well as the ownership of consumer durables.

[1] "Doi-Moi" is translated as renewal.

The introduction of the Vietnam LSMS has sparked several poverty studies (examples include Justino and Litchfield, 2004; White and Masset, 2002 and 2003; Glewwe *et al*, 2002; Haughton *et al*, 2001). These studies suggest that female headed households, lack of education, rural households (or living in the Northern Uplands), households dependent upon agriculture, are associated with higher poverty. They also suggest that children, despite declining fertility rates, remain an important driver behind poverty. This is confirmed by Pudney and Aassve (2007a) who shows that childbearing is strongly associated with lower living standards also during the nineties – a period where poverty reduction was strong. Simple descriptive statistics from the Vietnam LSMS shows this point (Table 1).

Table 1: Average equivalized household consumption expenditure at the two waves and its growth by number of children born[1] between the two waves.

No of children born between the two waves	Observations	Average consumption in 1992	Average consumption in 1997	Average consumption growth in 1997-1992
0	1232	970	2436	1466
1	581	856	1892	1036
2	182	790	1755	965
3	28	571	1154	583
At least 1	791	832	1835	1004
Total	2023	916	2201	1285

Notes: We consider the number of children of all household members born between the two waves and still alive at the second wave. All consumption measures are valued in dongs and rescaled using prices in 1992.The 2023 households represented in the table are selected taking only households with at least one married woman aged between 15 and 40 in the first wave. Consumption is expressed in thousands of dongs.

Of course, simple descriptive statistics like those presented in Table 1 do not say anything about causality. Rather, they merely show a positive association between number of children and poverty. In order to draw causal conclusions, we have to consider that households with more children are different from households with fewer children with respect to a range of factors. We turn to this issue in the remainder of the paper.

3 Causal inference: the potential outcomes approach

We start by reviewing the potential outcomes approach in light of estimating causal effects of fertility on economic wellbeing. In observational studies the use of the potential outcomes approach give rise to a quasi-experimental method, where in this application childbearing events are considered as the treatment and change in expenditure as the outcome.

The approach was pioneered by Neyman (1923) and Fisher (1925) and extended by Rubin (1974, 1978) to observational studies and recently it has been adopted by many in both statistics and econometrics (E.g. Rosenbaum and Rubin, 1983; Heckman, 1992 and 1997; Imbens and Angrist, 1994; Angrist, Imbens and Rubin, 1996; Heckman, Ichimura and Todd, 1997).

To formalise the idea, suppose we have a sample of individual units under study (in our case households) indexed by $i = 1, 2, \ldots N$, a treatment indicator D that assumes the value 1 for treated units (in our case if there is at least another child) and 0 for untreated or the controls (that is households not having new children) and an outcome variable, which is the growth in household equivalized consumption between the two waves, here indicated by Y. In this setting each unit, i, has two potential outcomes depending on its assignment to the treatment levels: Y_{i1} if D_i=1 and Y_{i0} if D_i=0. In essence our interest lies in whether having at least one new child has a casual effect on consumption dynamics and, if so, interest lies in the magnitude of the effect.

The fact that potential outcomes variables are labelled only by i and $d=\{0,1\}$ corresponds to the "no interference among units" assumption of Cox (1958) which Rubin (1980) refers to as the Stable Unit Treatment Value Assumption (SUTVA). SUTVA consists of two components. The first states that the potential outcomes for any unit do not vary with the treatments assigned to any other units. In our application it means that having a child has an effect on household consumption growth independently of other households' fertility behaviours. This seems to be an adequate hypothesis in the context of our application. The second component requires that the treatment is the same for each treated unit. In our case this implies that the number and characteristics of new children (sex, weight, timing etc) are not relevant. As with many of the other assumption to be discussed, it is important to note that

SUTVA is not directly informed by the data. In other words, it is an un-testable assumption that stems from the scientist's assessment or knowledge. Whereas several approaches have been developed to relax the SUTVA assumption in settings where it seems not to hold[2], we maintain this assumption in the following discussion.

Following Rubin (1978) the true causal effect of the treatment for the specific unit *i* can be written as

$$\Delta_i = Y_{i1} - Y_{i0}. \tag{1}$$

It is obvious that the two potential outcomes in (1) are not observable for the same unit - a feature referred to by Holland (1986) as the "fundamental problem of causal inference". The following relationship makes clear that the observed outcome for unit *i*, which we indicate with Y_i^{obs} depends on the treatment indicator

$$Y_i^{obs} \equiv Y(D_i) = D_i * Y_{i1} + (1-Di) * Y_{i0} = Y_{i0} + D_i * (Y_{i1}-Y_{i0}) \tag{2}$$

The last equality in (2) states that the treatment "adds" a quantity (Y_{i1}-Y_{i0}) to the outcome with respect to the case of no treatment. In spite of the apparent non resolvability of the fundamental problem, several approaches are developed to overcome it. Since, in most cases, we are not interested in the estimation of a causal effect on a single unit but on the entire population or on a sub-group of it, then the statistical solution is to focus on the estimation of average causal effects.

By using the potential outcomes approach we can define several causal parameters and at the same time make clear the assumptions needed to identify them. The two most commonly used causal parameters of interest are the *Average Treatment Effect* (ATE) and the *Average Treatment Effect on the Treated* (ATT) which are defined as:

$$\text{ATE} = E(Y_{i1}-Y_{i0}) \tag{3}$$

$$\text{ATT} = E(Y_{i1}-Y_{i0} | D_i=1). \tag{4}$$

[2] An example is the evaluation of the effect of a vaccination campaign on a contagious disease (Halloran and Struchiner, 1995). In this context, in fact, the effect of vaccinating a unit changes according to the number of the other units being vaccinated.

ATE is the expected effect of the treatment on a randomly drawn unit from the population. The policy relevance of ATE is in some sense doubtful since it averages across the entire population and hence includes units who would be never eligible to the treatment (Heckman, 1997). ATT instead gives the expected effect of the treatment on a randomly drawn unit from the population of treated and is therefore more interesting. Of course, if there is interest in estimating the effects for certain sub-populations, conditional versions of ATE and ATT can be specified:

$$\text{ATE}(x) = E(\text{Y}_{\text{i1}}\text{-}\text{Y}_{\text{i0}} | X_i=x) \tag{5}$$

$$\text{ATT}(x) = E(Y_{i1}\text{-}Y_{i0} | X_i=x, D_i=1) \tag{6}$$

Identifying assumptions and estimation methods

Generally, we cannot draw valid causal conclusions without considering what makes some units receive a treatment whereas others do not. This is referred to as the *assignment mechanism* and there is a critical distinction between *randomized* and *observational studies*. The key difference is that in randomized settings the analyst can control assignment to treatment and the probabilities of getting the treatment are known. In observational studies, as ours, these conditions are unlikely to hold and the researcher can only estimate probabilities of assignment to treatment on the basis of the data available.

In randomized experiments, treatment assignment D is statistically independent of potential outcomes Y_1 and Y_0:

$$Y_1, Y_0 \perp\!\!\!\perp D, \tag{7}$$

where $\perp\!\!\!\perp$ in the notation introduced by Dawid (1979) meaning independence. Randomization guaranties that the treatment assignment is ignorable (unrelated to potential outcomes) and that on average there is no difference in the characteristics of the treated and control units, at least in large samples. A consequence is that in these types of experiments ATE and ATT coincide and, more importantly, if we observe a difference in the average outcome between the two groups this can be attributed to the treatment effect. Formally, given condition (7) taking expectations of equation (2) yields $E(Y_i^{obs} \mid D = 1) = E(Y_{i1} \mid D_i = 1) = E(Y_{i1})$ and similarly $E(Y_i^{obs} \mid D_i = 0) = E(Y_{i0} \mid D_i = 0) = E(Y_{i0})$. Thus:

$$\text{ATE} = \text{ATT} = E(Y_i^{obs} \mid D_i = 1) - E(Y_i^{obs} \mid D_i = 0) \tag{8}$$

The right-hand side of (8) is easily estimated through its sample equivalent as the difference in the sample means of Y_i^{obs} in the two groups:

$$\hat{ATE} = \frac{\sum_{i:D_i=1} Y_i}{\#\{D_i = 1\}} - \frac{\sum_{i:D_i=0} Y_i}{\#\{D_i = 0\}} \tag{9}$$

The (9) gives an unbiased, consistent and asymptotically normal estimate of ATE.

The independence assumption on which the estimator in (9) relies is of course strong in non-randomized studies. In our setting we would have to assume that childbearing decisions were purely random or at least that on average the characteristics of households that decide to have a child are the same of households that did not. The more likely scenario, however, is that the two groups will differ quite substantially. If the characteristics that determine the childbearing decisions impact also on the consumption growth, which is very likely to hold, then the simple average difference in (9) will give a biased estimate of childbearing effect. This is the well known problem of *self selection bias*.

Among observational studies we distinguish two situations referred to as *regular* and *irregular* assignment mechanisms. The first concerns studies where the analyst can reasonably assume that characteristics driving self selection into treatment are all observed. Among irregular assignment mechanism the most important case is represented by the *latent regular assignment* where selection also depends on unobserved characteristics. Randomized experiments with non compliance, and by extension, instrumental variables estimation, belong to this setting. The presence of relevant unobservables is a critical problem for methods based on the UNA. However, as we show in section 4, sensitivity analysis can be performed to assess to what extent selection on unobservables represents a serious problem or not. This represents a useful exercise, especially in case where instruments are not available, which in turn is a necessary requirement for implementing the Instrumental Variable approach.

Regular assignment mechanisms

The unconfoundedness assumption (UNA) is the fundamental identifying assumption in observational studies with regular assignment[3]. The assumption can be stated as follows:

$$Y_1, Y_0 \perp D \mid X. \tag{10}$$

As stated earlier, assumption (10) implies that we are confident that all relevant variables influencing selection and the outcome are indeed observed. Regression and matching techniques, as well as stratification and weighting methods, all rely on this assumption. We focus on the first two. Another assumption, termed *overlap*, is also needed:

$$0 < P\ (D=1|X) < 1. \tag{11}$$

Assumption (11) implies equality in the support of X in the two groups of treated and controls (i.e. Support($X|D$=1) = Support($X|D$=0)) which guaranties that ATE is well defined (Heckman et al. 1997). Otherwise for some values of the covariates there would be some units in a group for which we could not find any comparable units in the other. It is important to note that identification of ATT require weaker versions of assumptions (10) and (11). In particular:

$$Y_0 \perp D \mid X, \tag{12}$$

$$P\ (D=1|X) < 1. \tag{13}$$

Regression versus propensity score matching

Whereas hypothesis (11) is familiar within the context of matching methods, it is less common in other settings such as regression analysis. The hypothesis is nevertheless necessary in an evaluation framework to guarantee that we are not comparing two groups (treated and controls) that are too different from each other. In parametric regression analysis the assumption is not needed in so far we can be sure to have the correct specification of the model. However, in this case comparability of treated and

[3] The uncounfoundedness assumption has been referred to also as the conditional independence, selection on observables or the exogeneity assumption (Imbens, 2004).

control units in some cells rely completely on extrapolation. The assumptions underlying the regression model are well known and outlined in common econometric text books (e.g. Green, 2002; Wooldridge, 2002). It is however of interest to clarify the regression model under the potential outcome framework introduced earlier. We specify a linear model for the two potential outcomes:

$$Y_{0i} = \beta X_i + e_i \tag{14}$$

$$Y_{1i} = \beta X_i + \Delta + e_i \tag{15}$$

$$\Delta = E(Y_{1i})\text{-}E(Y_{0i}) = \text{ATE}. \tag{16}$$

The model expressed by 14-16 assumes that the relationship between potential outcomes and covariates are linear and that there is no interaction of X with the treatment. In fact, the vectors of parameters in the two regressions are equal. Moreover, the treatment effect is assumed to be constant (in fact Δ is not indexed with i). Substituting the two models for potential outcomes (14 and 15) in the (2) we get the traditional multiple regression model

$$Y^{obs} = \beta X(1\text{-}D) + (\beta X + \Delta)\ (D) + e = \beta X + \Delta D + e \tag{17}$$

If the true model were non linear, the OLS estimates of the treatment would be in general biased (Goodman and Sianesi, 2005). Moreover if the effect of the treatment changes by unit characteristics (heterogeneous treatment effect) OLS will not in general recover the ATT. In fact, the constant treatment effect assumption implies that ATE coincides with ATT. Both these problems are exacerbated if some units fall outside of the common support of the observables, which would be the case if there are treatment units for whom there are no comparable non-treatment units. In this case, performing OLS might hide the fact that the analyst is actually comparing incomparable units by using the linear extrapolation. Of course, the problem can be circumvented by estimating the common support and running the regression conditioning on it. Moreover, heterogeneous treatment effect can be allowed by including a complete set of interactions between X and D. This give rise to the so called Fully Interacted Linear Model (FILM – see Goodman and Sianesi, 2005). Also the linearity assumption can be relaxed if we use a nonparametric method to estimate the effect of D on Y which is the case for matching methods.

Matching is an intuitive and appealing estimation method which consists of contrasting treated and control units that have the same characteristics X. With X being large, matching becomes computationally demanding. Instead it can be done on the basis of a univariate *Propensity Score* (PS), which is defined as the conditional probability of receiving a treatment given pre-treatment characteristics: $e(X) \equiv Pr\{D = 1|X\} = E\{D|X\}$. When the propensity scores are balanced across the treatment and control groups, the distribution of all covariates, X, are balanced in expectation across the two groups. Therefore, matching on the propensity score is equivalent of matching on X. However, the probability of observing a treated and a control unit with exactly the same value of the propensity score is, in principle, zero since $e(X)$ it is a continuous variable. Thus the matching methods must facilitate comparisons between treated and control units with a certain distance. Several methods of matching can be used. The most common ones are kernel (gaussian and epanechnikov), nearest neighbour, radius and stratification matching.

Latent regular assignment mechanisms

The assumption of unconfoundedness is violated if selection also occurs on unobserved variables. In the econometrics literature the issue is known as an omitted variable bias (see Wooldridge, 2002). In the potential outcomes approach the implication is that even after conditioning on X, there remains a certain degree of dependence among potential outcomes and treatment status. Only if we can also condition on the unobserved variable U will this dependency disappear. In other words, the assumption has to be relaxed so that:

$$Y_1, Y_0 \perp D \mid X, U \qquad (18)$$

When we suspect failure of the uncounfoundedness assumption two alternatives are available. The first is to use a method that does not rely on the hypothesis of selection on observables. The second option is to implement a sensitivity analysis to assess the severity of unobservable confounders. We explore both solutions. As a method that overcomes the uncounfoundedness assumption we explore the IV approach. We highlight its benefit and drawbacks using the potential outcome framework under which Angrist, Imbens and Rubin (1996) demonstrate that IV estimates are generally different from ATE and ATT which are usually the parameters of interest. The parameter estimated by IV, is instead the Local Average Treatment Effect (LATE). It is the average causal effect of the treatment on the sub-population of units that react to the instrument, also known as compliers. Importantly, these units are not identifiable by the data. For this reason it is often questioned to what extent LATE, or the IV estimates, can be considered as relevant policy parameter.

Randomised and conditionally randomised instruments

The way IV method is implemented depends on whether the available instrument can be thought of as randomised or not. The prototype of a randomised instrument is a situation where it is possible to randomise the *encouragement* to receive the treatment. In this situation the instrument is the assignment indicator, Z, which is different from the treatment indicator, D. We indicate by Z_i the assignment received by unit i, which in our application is a binary variable. We indicate with $D_i(\mathbf{Z})$ the binary treatment indicator for unit i, that depends on the vector of assignments $\mathbf{Z}$. Similarly, the potential outcomes for unit i are indicated as $Y_i(\mathbf{Z}, \mathbf{D})$.

In this case, where we do not need to control for covariates, the IV estimator is defined as the ratio of two sample covariances (Durbin, 1954):

$$\beta^{IV} = \frac{\operatorname{cov}(Y, Z)}{\operatorname{cov}(D, Z)} \tag{19}$$

When both Z and D are binary variables, as in our application, the ratio of the two covariances in (19) simplifies to the so-called Wald estimator (1940):

$$\beta^{IV} = \frac{E[Y \mid Z=1] - E[Y \mid Z=0]}{E[D \mid Z=1] - E[D \mid Z=0]} \tag{20}$$

It is useful to briefly review the definition of an instrument in the view of Angrist, Imbens and Rubin, 1996 (in the following AIR) to better understand the conditions for identification. The assumptions are as follows:

1) Stable Unit Treatment Value Assumption (SUTVA), which requires that the potential outcomes for each unit do not depend on assignment and treatment status for the other units. Foramlly, $Y_i(\mathbf{Z}, \mathbf{D}) = Y_i(Z_i, D_i)$ and $D_i(\mathbf{Z}) = D_i(Z_i)$.

2) Random Assignment of treatment which implies $P(Z = c) = P(Z' = c')$ for all c and c' such that $\iota T\ c = \iota T\ c'$, where ι is the N- dimensional column vector with all elements equal to one. In other words, if the number of units assigned to the treatment ($Z_i = 1$) is equal in Z and Z', these vectors have the same probability.

3) Exclusion Restriction: Z impacts on Y only through D: $\boldsymbol{Y}(\boldsymbol{Z}, \boldsymbol{D}) = \boldsymbol{Y}(\boldsymbol{Z'}, \boldsymbol{D})$ for all $\boldsymbol{Z}$, $\boldsymbol{Z'}$ and $\boldsymbol{D}$. By virtue of assumption 3, we can define potential outcomes $\boldsymbol{Y}(\boldsymbol{Z}, \boldsymbol{D})$ as a function of $\boldsymbol{D}$ alone: $\boldsymbol{Y}(\boldsymbol{Z}, \boldsymbol{D}) = \boldsymbol{Y}(\boldsymbol{D})$. By assumption 1 we can also write $Y_i(D_i)$ instead of $Y_i(\boldsymbol{D})$. This assumption ensures validity of the instrument.

4) Nonzero Average Causal Effect of Z on D. The average causal effect of Z on D, $E[D_{1i} - D_{0i}]$ is different from 0 and corresponds to the hypothesis of non zero correlation between the instrument and the endogenous variable usually stated in econometrics. In other the words of econometricians, the instrument needs to be relevant.

5) Monotonicity, which implies that $D_{1i} \geq D_{0i}$ for all $i = 1, \ldots , N$. The assumption of monotonicity is critical when comparing the IV approach to the methods based on the UNA. To see how, characterize units by the way they might react to the assignment to treatment. A first group is termed *compliers* and defined by units that are induced to take the treatment by the assignment: $D_{1i} - D_{0i} = 1$. Other units may not change the treatment status from assignment. These units are defined as either *always-takers*, where $D_{1i} = D_{0i} = 1$ (they always take the treatment), or *never-takers*, if $D_{1i} = D_{0i} = 0$ (they always take the control). Finally, we might encounter *defiers*, who are units that do the *opposite* of their assignment status. The monotonicity assumption implies that there are no defiers. The assumption is crucial for identification since otherwise the treatment effect for those who shift from nonparticipation to participation when Z shift from 0 to 1 can be cancelled out by the treatment effect of those who shift from participation to nonparticipation (Angrist and Imbens, 1994). Of course, the monotoniticy assumption is untestable and its validity has to be evaluated in the context of the given application.

Under the assumptions 1-5 AIR demonstrate that the Wald estimator (19) identifies a particular causal parameter, namely:

$$\beta^{IV} = \frac{\text{cov}(Y, Z)}{\text{cov}(D, Z)} = \text{E}\left[Y_{1i}\text{-}Y_{0i} | D_{1i} \text{ - } D_{0i}\text{=}1\right] \tag{21}$$

which is the average causal effect calculated on the sub-population of units with $D_{1i} - D_{0i} = 1$ (i.e. for compliers). This is in other words the Local Average Treatment Effect (LATE) and is different from the ATE which is calculated on the whole population. A potentially serious drawback of the IV estimand given by (20) is that it refers to the subpopulation of compliers that is not identifiable by the data. Importantly, the ATE parameter is not identifiable unless we impose stronger assumptions. In particular, LATE coincide with ATE only if we can assume a constant treatment effect or if we suppose that the proportion of compliers in the population is equal to 1. However, both these assumption are very strong. It is also important to note that the effect identified by (20) depends on the instrument used. If several instruments are available LATE will differ and depend directly on the instrument used, simply because the instrument identifies the causal effect only for a specific sub-population.

In specific applications LATE can be an interesting parameter for policy. Let us suppose that the policy maker wants to know what is the (average) causal effect of *D* on *Y* but the only way to manipulate *D* is through *Z*. In this case interest lies in the (average) causal effect of *D* on *Y* for units that react to the policy intervention on *Z*. In this situation, however, the policy maker cannot identify which are the compliers, but can only estimate the dimension of this group. The presumption in such cases is that the average causal effect calculated on the subpopulation of units whose behavior was modified by assignment is likely to be informative about subpopulations that will comply in the future.

However, *Z* is not necessarily randomized in the way assumed above. In many applications *Z* itself is confounded with *D* or with *Y* or both. There are many examples. Suppose parental education is used as an instrument for the estimation of returns to schooling. However, parental education will be correlated with parent's profession, family income and wealth, which may directly affect the wage of their offspring. In these situations it is necessary to control for the confounder covariates *X* (examples include Card, 1995 and Angrist and Krueger, 1991). The conventional approach to accommodate covariates *X* in IV estimation consist of parametric or semi-parametric methods, 2SLS being the most common. However, the approach relies on functional form assumptions (see for ex. Angrist and Imbens, 1995; Heckman and Vytalic, 1999; Hirano, Imbens, Rubin and Zhou, 2000; Abadie, 2003). A serious drawback of these methods is that most of them impose additive separability in the error term that amounts to assume a constant treatment effect (see for ex. Newey, Powell and Vella, 1999; Das, 2000). In non-separable models identification require that the instrument is sufficiently powerful to move the value of D_i over the entire support of D (see for ex. Blundell and Powell, 2003; Florens, Heckman, Meghir and Vytalic, 2002; Imbens and Newey, 2003). However, assuming that such an instrument exists is difficult in many applications.

One approach that overcomes the strong assumptions used by the aforementioned IV methods is the non-parametric approach suggested by Frölich (2007). In this approach one estimates a conditional LATE which does not rely on the aforementioned strong assumptions. Frölich demonstrates that this estimator is asymptotically normal and efficient, and hence accommodation of X in the estimation does not give rise to curse of dimensionality. On the basis of assumptions 1 to 5, but conditioned on X, we can identify a conditional LATE[4]. This is the LATE defined for units with specific observed characteristics

$$LATE(x) = E[Y1 - Y0 \mid X = x, \tau = c] = \frac{E[Y \mid X = x, Z = 1] - E[Y \mid X = x, Z = 0]}{E[W \mid X = x, Z = 1] - E[W \mid X = x, Z = 0]} \quad (22)$$

Starting from this result Frölich (2007) demonstrates that the marginal LATE can be calculated as follow[5]:

$$LATE = E[Y_1 - Y_0 \mid \tau = c] = \frac{\int \left(E[Y \mid X = x, Z = 1] - E[Y \mid X = x, Z = 0]\right) f_x(x)dx}{\int \left(E[W \mid X = x, Z = 1] - E[W \mid X = x, Z = 0]\right) f_x(x)dx} \quad (23)$$

Formula (22) is similar to (20) but with both numerator and denominator conditional on X. When the number of covariates included in the set X is high, nonparametric estimation becomes difficult, especially in small samples. An alternative in these situations is to make use of the balancing property of the propensity score that allows us to substitute the high dimensional set X in (23) by a univariate variable. In fact, since adjusting for the distribution of X is equivalent to adjusting for the distribution of the propensity score $\pi(x) = P(Z=1|X=x)$ we can write

[4] Here we give only some intuition about the assumptions underlying this method. For a detailed and more formal discussion we refer to Frölich (2007).

[5] It is important to note that a common support assumption is needed as stated by Frölich: Supp(X/Z=1) = Supp(X/Z=0).

$$
\begin{aligned}
LATE &= \frac{\int \left(E[Y \mid X = x, Z = 1] - E[Y \mid X = x, Z = 0]\right) f_x(x) dx}{\int \left(E[W \mid X = x, Z = 1] - E[W \mid X = x, Z = 0]\right) f_x(x) dx} \\
&= \frac{\int \left(E[Y \mid \pi(X) = \pi, Z = 1] - E[Y \mid \pi(X) = \pi, Z = 0]\right) f_{\pi(x)}(\pi) d\pi}{\int \left(E[W \mid \pi(X) = \pi, Z = 1] - E[W \mid \pi(X) = \pi, Z = 0]\right) f_{\pi(x)}(\pi) d\pi}
\end{aligned} \tag{24}
$$

where $f_{\pi(x)}$ is the density function of $\pi(x)$ in the population.

An estimator of (24) can be obtained as the ratio of two propensity score matching estimator measuring the two intentions to treat effects of Z on Y (numerator) and of Z on D (denominator). Hence, to estimate the numerator (ITT_{ZY}) we consider the variable Z as the treatment and Y as the outcome. For the denominator (ITT_{ZD}) Z is still considered the treatment and D the outcome:

$$
\widehat{LATE} = \frac{\widehat{ITT}_{ZY}}{\widehat{ITT}_{ZD}} \tag{25}
$$

4 Estimating causal effects of fertility on economic wellbeing in Vietnam

In light of the discussion above we present here an application where we estimate the causal effect of fertility on economic wellbeing. Fertility is measured by childbearing events which are generally believed to be endogenous with respect to consumption expenditure, which is the most common way to measure a households' economic wellbeing. Economic wellbeing is measured in terms of the change in consumption expenditure between the waves. We proceed by presenting results from multiple regression and propensity score matching, both relying on the UNA. We then demonstrate the sensitivity analysis whereby we assess the extent bias is generated by imposing UNA. As discussed previously, IV is not expected to produce the same parameter estimates as the methods based on the UNA and we present how estimates differ, depending on instruments used. We explain these differences and assess to what extent these estimates can be used for policy making.

Data: The Vietnam LSMS

We use the Vietnam LSMS first surveyed in 1992/93 with a full follow up in 1997/98. It follows the LSMS format and includes rich information on education, employment, fertility and marital histories, together with rich information on household income and consumption expenditure. The overall quality of the panel is impressive with a very low attrition rate (Falaris, 2003). The Vietnam LSMS also provides detailed community information from a separate community questionnaire. Community level information is available for rural areas only and includes 120 communities, including information on health, schooling and main economic activities. The communities range in size from 8,000 inhabitants to 30,000.

We use as a measure of household's living standard the household's consumption expenditure, which requires detailed information on consumption behaviour and its expenditure pattern (Coudouel et al., 2002; Deaton and Zaidi, 2002). The expenditure variables are calculated by the World Bank procedure which is readily available with the Vietnam LSMS survey. It is clear that the distribution of consumption expenditure within the household is unlikely to be uniform across household members, and it is probable that children consume less than adults. The standard solution is to impose an assumption on inter-household resources allocation, and adjustments can be done by applying an equivalence scale that is consistent with the assumption made – producing a measure of expenditure per equivalent adult. Here we apply a simple equivalence scale similar to White and Masset (2001), giving to each child aged 0-14 in the household a weight of 0.65 relative to adults.

Our choice of variables is based mainly on dimensions which are important for both household's standard of living and fertility behaviour and hence are potentially confounders that have to be included in the conditioning set X to make the UNA more plausible. More specifically, we think that all these variables can theoretically have an impact on change in consumption expenditure and on the decision of have children. These are all variables measured at the first wave and hence can be viewed as pre-treatment variables.

Many of these variables are defined in terms of household ratios. That is, we include the number of household members that are engaged in gainful employment as a ratio of the total number of household members. We included also demographic characteristics of the household such the sex and the age of the household head, the household size and the presence of children. The effect of children is further distinguished by their age distribution, and again expressed as a ratio of the total number of household members. Other covariates are the ratio of male and female members aged 15-45, the ratio of male and female working members aged 15-45 out of the respective groups, an educational index, the level of equivalized consumption at the first wave and regional dummies. Importantly, we included two binary variables indicating, respectively, if the household is farmer or not and if the household head belong to the majority ethnic group (the Kinh) or not. The survey also includes rich information on the characteristics of the community where the household resides and we control for community differences through three indexes: 1) an index of economic development, 2) health facilities and 3) educational infrastructures.

Regression and propensity score results

Here we present the results of the estimation of the causal effect of childbearing on consumption expenditures. We consider the multiple regression model and the propensity score matching estimator. Our sample is restricted to households where in the first wave was one married woman aged between 15 and 40 years[6]. The selection is useful since it avoids units who are in effect incapable of childbearing.

Matching is based on the nearest neighbor method with replacement using the *nnmatch* module in *STATA* (Abadie et al, 2004)[7]. The results are presented in Table 2. We also report the results of the estimation of a simple regression of Y on D without any covariates. This is equivalent to use estimator (9) and can be obtained also from the Table 1. This would be an acceptable estimate of the ATE under the randomization of D. It is clear that selection is present and the estimate of fertility on expenditure is reduced by around 10% in the multiple regression. For the purpose of

[6] This sample selection criterion is part of the whole matching strategy since we avoid comparing households having a child with households who were essentially out of the risk set (here because there is no women of fecund age in the household). Obviously different selection strategies are possible. However, this selection criterion gives low attrition with respect to households having additional children. Moreover, we tried the following alternative selection criterion: 1) select households with at least one married woman aged 15-35 in the first wave; 2) select households where the head or its spouse is a married woman aged 15-40 in the first wave; 3) select households where the head or its spouse is a married woman aged 15-35 in the first wave. However, results are very similar to those presented here.

[7] This software implements the estimators suggested by (Abadie and Imbens, 2002), permitting to obtain analytical standard errors which are robust to potential heteroschedasticity We preferred analytical to bootstrapped standard error because Abadie and Imbens (2006) show that bootstrap fails with nearest neighbor matching. This matching method, on the other hand, should be preferred since, ensuring the “best” matching, it reduces the bias with respect to other methods.

illustration we maintain the assumption that the treatment can be thought of as randomized after having controlled for covariates (UNA). What differ are the assumptions imposed for estimation. The standard multiple regression implicitly assume that the effect of childbearing on poverty is constant while FILM, including all interactions among *D* and covariates, allows it to change with covariate values. As a consequence, multiple regression does not distinguish between ATE and ATT since they coincide under a constant treatment effect. In contrast, FILM does and ATE and ATT will in general differ. FILM was implemented with and without conditioning on common support. Since results are very similar we show only FILM with the common support. FILM in this case requires a first stage estimation of the propensity score and the common support[8]. With FILM, multiple regression model is made as similar to PSM as possible. The difference of course, is that PSM does not impose any functional form for the relationship between poverty and fertility. Regression, in contrast imposes linearity. As we can see from Table 2 the estimate for ATE is similar in all this methods, while ATT is estimated a bit lower in the PSM. Thus, relaxing linearity matters and the PSM is to be preferred. Moreover, PSM permits us to assess in a simple way the underlying assumptions. From table 2 we note that using PSM, ATT is estimated to be different from ATE. In general, ATT and ATE differs if the distribution of covariates in the two groups of treated and control are different (this is expected due to likely potential self selection into treatment) and if the treatment interacts with covariates (treatment effect heterogeneity).

Table 2: Estimates from methods based on the Uncounfoundedness Assumption (robust standard error in parentheses)

REGRESSIONS				PROPENSITY SCORE MATCHING (nearest neighbour)	
SIMPLE	MULTIPLE WITH NO INTERACTIONS	FILM (conditioned on CS)			
(ATE=ATT)	(ATE=ATT)	ATE	ATT	ATE	ATT

[8] In principle, any standard probability model can be used to estimate the propensity score. For example, using the common logit or probit models, we can write Pr{Di = 1|Xi} = F(h(Xi)), where F(.) is, respectively, the normal or the logistic cumulative distribution and h(Xi) is a function of covariates with linear and higher order terms. The choice of which higher order terms to include, as well as interactions among covariates, is determined solely by the need to balance covariate distribution in the two treatment groups (see Dehejia,and Wahba, 2003). We used a logit specification including some interaction terms to achieve balancing. We avoided the inclusion of higher order terms because, as demonstrated by Zhao (2005) their inclusion could have some biasing effect (while the inclusion of interactions has not this drawback).

-462 (56.14)	-414 (61.78)	-421 (60.14)	-432 (59.13)	-356 (116.20)	-411 (87.47)

Notes: CS = common support. Figures are in thousands of dongs. Standard errors for regressions are robust to heteroschedasticity and correlation within communities. PSM standard errors are robust to heteroschedasticity.

In order to assess the magnitude of the estimate on consumption expenditure we compare the effect of the treatment variable *D* with other covariates in Table A.1 (See Appendix). The effect of *D* is strong and out of the various covariates it has the third strongest elasticity. Given that the average consumption growth between the two waves amounts to 1285 thousand dongs, the estimates ranging from -356 to -462 thousand dongs, as presented in Table 2, are clearly substantial. Moreover, the food poverty line in 1992 was estimated to 750 thousands of dongs (corresponding to 68 US$ in year 1992), which is another indication that the effects associated with childbearing events are potentially important for households' economic wellbeing[9]. For instance the amount needed in 1992 a quantity of rice giving 1000 calories (about 300 gr.) each day for one year, 215 thousand of dongs were needed[10].

An interesting, but not unexpected, result is that education is the most important confounder in the relationship between fertility and poverty. It has the strongest in effect in both models. It is interesting to note that the consumption level measured in 1993 little impact on the probability of childbearing events taking place between the two waves, and suggest that the issue of reversed causality seems to be not relevant in our application[11].

The previous estimate of ATT refers to the whole population of treated. However, for policy makers it would be of interest to assess if and how the treatment has different effects according to the specific characteristics of the treated households. Table 3 presents heterogeneity of the treatment effect. Bearing mind that the overall ATT is -356 with a standard error of 116.20, it is clear that there is substantial variation in the treatment effect for different groups[12]. First of all, we note huge

[9] For further insights on the goods composing the Vietnamese food basket and for more details about the construction of the Vietnamese food poverty line, see Tung (2004).
[10] These figures are derived by Molini (2006).
[11] As we have seen the effect of childbearing on consumption expenditure is significant. Its magnitude depends of course on the imposed equivalence scale, but qualitatively the differences between estimators persist. The sensitivity analysis is available from the authors on request.
[12] This is confirmed by the FILM model. After we ran it we test for the presence of heterogeneous effects of the treatment that is we tested the joint significance of all the interactions between *D* and *X*. The null hypothesis of no significant interaction is rejected (F = 1.94; Prob>F = 0.0120).

variations by regions. However, in some regions we have very few matched units hampering reliable causal effect estimation.

Interestingly, the distinction between farmers and non-farmer households give rise to clear heterogeneity. In particular, farmer households with an additional child are substantially more disadvantaged than non-farmers. This is also the case for non kinh versus kinh households but here the heterogeneity is smaller. The heterogeneity by education confirms the well known pattern. More highly educated individuals suffer more from a childbearing event, mainly due to differences in opportunity cost. That is, more highly educated women earn more and consequently suffer more from retracting from the labour market due to childbearing. We also find that households headed by older individuals fare worse from childbearing events. Interestingly, the percentage of children in the first wave seems less important, indicating that economies of scale are likely to be less strong.

Table 3 – Treatment effect heterogeneity

Sub-sample	Treated units			Untreated units			TOT units	ATT (standard error in parenthesis)
	MA	UN	TOT	MA	UN	TOT		
Regional variability								
Region 1	90	29	119	146	43	189	308	-433 (167.57)
Region 2	54	31	85	230	54	284	369	-52 (231.57)
Region 3	61	39	100	79	34	113	213	-512 (362.23)
Region 4	45	28	73	44	41	85	158	-1093 (380.74)
Region 5	2	35	37	1	18	19	56	-292 (na)
Region 6	11	34	45	17	53	70	115	-2239 (1384.04)
Region 7	90	17	107	186	22	208	315	-464 (211.91)
Farmer / non farmer								
Farmer	372	22	394	500	37	537	931	-720 (152.52)
Non Farmer	153	22	175	387	44	431	606	-339 (203.68)
Kinh / non kinh								
Kinh	409	33	442	796	24	820	1262	-328 (137.85)
Non Kinh	96	31	127	87	61	148	275	-407 (191.56)
Household Education Index								
Education – L	228	9	237	257	16	273	510	-355 (166.77)
Education – M	145	33	178	301	26	327	505	-334 (144.68)
Education – H	125	29	154	323	45	368	522	-497 (257.06)
Consumption in 1993								
Consumption in wave 1 – L	158	33	191	258	63	321	512	-278 (122.30)
Consumption in wave 1 – M	174	16	190	307	16	323	513	-317 (157.88)
Consumption in wave 1 – H	164	24	188	276	48	324	512	-451 (285.86)
Household size								
Household size – L	189	63	252	184	49	233	485	-408(164.03)
Household size – M	160	25	185	426	36	462	647	-333(175.14)
Household size – H	108	24	132	240	33	273	405	-377 (233.70)
% of kids in 1993								
% kids in wave 1 – L	138	21	159	389	64	453	612	-365 (172.03)
% kids in wave 1 – M	116	27	143	266	29	295	438	-368 (198.35)
% kids in wave 1 – H	215	52	267	193	27	220	487	-364 (223.43)
Age of household head								
Household head age – L	181	58	239	203	35	238	477	-169 (161.54)
Household head age – M	125	13	138	402	20	422	560	-294(205.08)
Household head age – H	188	4	192	253	55	308	500	-445(284.47)

Notes: Estimates of the ATT are based on the PSM with nearest neighbor. EDU is an educational index aggregating years of schooling for all household members and keeping into account the age; L, M, H attached to same variables means "low", "medium" and "high" groups obtained by splitting the sample in three parts of equal size; MA = observations matched; UN = observations unmatched; na = cannot be estimated.

Finally, the effect of household size forms an *U*-shape: the (negative) effect of childbearing is stronger for small and large households while medium sized households show a somewhat lower effect. The rather strong heterogeneity suggests of course that care is needed when implementing policy interventions. Clearly, policies related to fertility and economic wellbeing will have very different effects for different groups. As Crump et al (2006) notice, if there is strong evidence in favor of heterogeneous effects, one may be more reluctant to recommend extending the program to populations with different distributions of the covariates.

Sensitivity of the PSM estimator

The reliability of the PSM estimates depends on the balancing property being satisfied, the sensitivity to the imposition of common support and to the matching method used. All of these are checked with detailed analysis, and in general the estimates are robust[13]. However, the most critical requirement of the PSM is that it satisfies the UNA. It is therefore of interest to assess to what extent parameter estimates might be affected by any violation of the UNA.

We adopt the simulation-based approach suggested by Ichino, Mealli and Nannicini (2007 – in the following IMN). The underlying hypothesis[14] is that assignment to treatment may be confounded given the set of observables covariates X (i.e. the UNA does not hold) but it is unconfounded given *X and* an *unobservable* covariate U. Thus, $Y_0 \perp D \mid (X, U)$. The consequent idea behind the analysis is to simulate U which is added into the set of matching variables in order to assess the sensitivity of the estimates. It is assumed that U and the outcome are binary variables. In case of continuous outcomes, a transformation is needed so that the outcome takes the value 1 if it is above a certain threshold (the median for example) and 0 otherwise, alternatively one could consider other outcome variables such as poverty status which essentially is a dichotomous transformation of consumption expenditure[15]. The potential confounder can be specified in different ways. One alternative is a calibrated

---

[13] The balancing property has been checked comparing the distribution of covariates X before and after matching and calculating the reduction in the absolute bias. We calculated the ATE and ATT with different matching methods (nearest neighbor with/without imposition of caliper and with/without replacement, k- nearest neighbor, radius and kernels) in order to assess the sensitivity to methods different form the one presented here. The estimates reported in table 2 are obtained using the *nnmatch* module in *STATA* (Abadie et al, 2004). The other matching methods were implemented using the *psmatch2* module developed by Sianesi (2002). Finally, we found that very few units fall outside the common support calculated with the minima-maxima criterion. Using different methods for calculating CS, including the tick support, we get results similar to the ones presented in table 2. We also implemented matching on the underlying continuous index instead to the propensity score to control for the heavy – tails problem. Again results are stable. Material is available from authors on request.

[14] This assumption is used also in previous works but the approach used here is the only one that allows to assess point estimates sensitivity without relying on parametric estimation of the outcome.

[15] For technical details on the simulations, see Ichino, Mealli and Nannicini (2007) and Nannicini (2007) for details on the *STATA* module *sensatt*.

.

version where we make the distribution of U similar to the empirical distribution of important binary covariates in X. Another alternative is to specify a “Killer” confounder where the value of U is specified so that its association with the outcome and the treatment is increasingly high. We are particularly worried about the fact that the estimated ATT could become zero, or non significant, when including U. This is an interesting approach, because it gives us a measure of how large the association between U, Y_0 and D has to be in order to cancel out the ATT. The distribution of U is specified by four key parameters

$$p_{ij} = P\,(U{=}1|D{=}\,i,\ Y{=}\,j\,) = P\,(U{=}1|D{=}\,i,\ Y{=}\,j\,,X) \quad i, j = 0{,}1 \tag{26}$$

In (26) is made the hypothesis (used in the simulation) that U is independent to X conditional to D and Y. In order to choose the signs of the associations, IMN note that if $d = p_{01} - p_{00} > 0$ then U have a positive effect on Y_0 (conditioning on X) whereas if $s = p_{1.} - p_{0.} > 0$, where $p_{i.} = P(U{=}1|D{=}i)$, then U have a positive effect on D. If we set $p_u = P(U{=}1)$ and $d' = p_{11} - p_{10}$ the four parameters p_{ij} are univocally identified specifying the values of d and s. Hence we can fix the two quantities p_u and d’, which will not affect the baseline estimate, and change the values of d and s to assess the sensitivity of the estimates.

Table 4 shows the sensitivity analysis to some calibrated confounders. The key conclusion from this table is that for different simulated U, the resulting ATT is never substantially different from the ATT based on the UNA assumption. Thus, even if an unobservable variable (with the same distribution as those listed) was excluded from the conditioning set in the PSM the effect on the estimated ATT would be negligible.

The results for the “killer” confounder are given in Tables 5a to 5d. Here the parameters G and A are defined as:

$$G = \frac{P(Y=1|\,T=0\,,\,U=1\,,X)\,/\,P(Y=0|\,T=0\,,\,U=1\,,X)}{P(Y=1|\,T=0\,,\,U=0\,,X)\,/\,P(Y=0|\,T=0\,,\,U=0\,,X)}$$

and

$$A = \frac{P(T=1|U=1\,,X)\,/\,P(T=0|U=1\,,X\,)}{P(T=1|U=0\,,X)\,/\,P(T=0|U=0\,,X\,)}$$

The parameter G is the average odds ratio from the logit model of $P(Y=1 | D=0, U, X)$ calculated over 1000 iterations. It is in other words a measure of the effect of U on Y, and is in this sense an outcome effect. The parameter A refers to the average odds ratio from the logit model of $P(D=1|U, X)$. This is a measure of the effect of U on D, and is therefore a measure of the selection effect. We implement four separate sensitivity analysis according to the signs of the association between U and D and Y. As we can see from Tables 6a-6d the estimated effect is almost always negative and significant and not much different from the baseline estimate. Only if the associations of U with D and/or Y are strong the ATT becomes not significant. This happens in particular if the effects of U on D and on Y have opposite signs. It is interesting to note that a theoretical relevant omitted variables such as unobserved ability, has this characteristic since it is (potentially) positively associated with Y and negatively associated with D.

Table 4 – Sensitivity analysis to "calibrated" unobserved confounders

	Fraction $U=1$ By treatment / outcome				Outcome effect	Selection effect	ATT
	p_{11}	p_{10}	p_{01}	p_{00}			
No unobserved confounders	0.00	0.00	0.00	0.00	---	---	-356 (101.07)
Unobserved confounder similar to:							
Sex of household head	0.82	0.86	0.85	0.87	0.80	0.92	-406 (128.63)
Kinh ethnic group	0.86	0.73	0.79	0.78	2.46	0.64	-412 (132.41)
% children aged 0 to 4	0.65	0.53	0.41	0.41	1.16	2.58	-440 (146.24)
% women aged between 15 & 45	0.70	0.58	0.45	0.41	1.15	2.15	-445 (141.53)
Farmer	0.64	0.72	0.55	0.56	0.99	1.83	-440 (138.84)
Education	0.45	0.40	0.62	0.50	1.68	0.55	-421 (135.20)
Consumption expenditure wave 1	0.41	0.36	0.42	0.37	1.24	0.93	-419 (135.02)
Age of household head	0.38	0.35	0.37	0.36	1.05	1.02	-413 (133.52)

Note: Estimates based on the nearest neighbour matching. Standard errors are analytical and each reported ATT is the average over 1000 iterations.

Table 5a – Sensitivity analysis to confounders *G* < 1 ; A < 1

	s = -0.1	s = -0.2	s = -0.3	s = -0.4	s = -0.5
d= -0.1	-442 (142.80) G=0.73 A=0.48	-452 (157.74) G=0.71 A=0.29	-457 (172.87) G=0.67 A=0.17	-464 (193.93) G=0.66 A=0.09	-460 (217.86) G=0.62 A=0.04
d= -0.2	-446 (139.32) G=0.42 A=0.56	-456 (150.38) G=0.39 A=0.34	-463 (168.77) G=0.38 A=0.20	-466 (179.48) G=0.35 A=0.11	-471 (207.22) G=0.31 A=0.04
d= -0.3	-451 (135.91) G=0.25 A=0.64	-446 (146.52) G=0.22 A=0.40	-461 (161.10) G=0.20 A=0.23	-472 (173.45) G=0.18 A=0.12	-477 (196.51) G=0.14 A=0.05
d= -0.4	-446 (136.97) G=0.12 A=0.74	-441 (142.58) G=0.12 A=0.45	-470 (156.33) G=0.10 A=0.26	-469 (169.73) G=0.07 A=0.13	-480 (188.55) G=0.05 A=0.04
d= -0.5	-434 (135.45) G=0.07 A=0.86	-446 (140.16) G=0.05 A=0.52	-471 (152.22) G=0.04 A=0.29	-468 (161.82) G=0.03 A=0.15	***

Note: *** = combination resulting in inadmissible values of the parameters characterising the distribution of *U*.

Tab 5b – Sensitivity analysis to confounders G > 1 ; A < 0

	s = -0.1	s = -0.2	s = -0.3	s = -0.4	s = -0.5
d= +0.1	-445 (150.48) G=2.03 A=0.36	-448 (168.67) G=1.93 A=0.22	-449 (195.53) G=2.08 A=0.13	-446 (217.97) G=2.17 A=0.07	-435 (251.52) G=2.14 A=0.03
d= +0.2	-444 (154.12) G=3.33 A=0.31	-445 (177.78) G=3.27 A=0.18	-444 (202.73) G=3.77 A=0.11	-433 (238.97) G=3.79 A=0.06	-417 (290.73) G=4.28 A=0.02
d= +0.3	-443 (163.98) G=5.99 A=0.26	-435 (186.02) G=6.33 A=0.15	-437 (216.87) G=6.53 A=0.09	-415 (263.24) G=7.58 A=0.04	-402 (324.63) G=9.84 A=0.02
d= +0.4	-436 (172.83) G=11.57 A=0.21	-417 (197.69) G=13.01 A=0.12	-417 (238.73) G=13.29 A=0.07	-383 (308.06) G=18.17 A=0.03	-359 (435.28) G=31.28 A=0.01
d= +0.5	-420 (184.43) G=36.02 A=0.17	-398 (219.84) G=30.67 A=0.09	-369 (271.77) G=51.94 A=0.05	-338 (388.96) G=117.46 A=0.02	-238 (639.76) G=2128.61 A=0.01

Tab 5c – Sensitivity analysis to confounders G < 0 ; A > 0

	s = +0.1	s = +0.2	s = +0.3	s = +0.4	s = +0.5
d= -0.1	-431 (135.01) G=0.73 A=1.19	-440 (137.16) G=0.73 A=1.85	-445 (149.17) G=0.76 A=2.94	-448 (171.19) G=0.74 A=4.79	-450 (192.42) G=0.69 A=8.43
d= -0.2	-445 (135.33) G=0.45 A=1.39	-432 (140.47) G=0.43 A=2.16	-448 (156.82) G=0.43 A=3.43	-440 (175.76) G=0.43 A=5.69	-436 (200.39) G=0.38 A==10.03
d= -0.3	-441 (133.69) G=0.25 A=1.62	-436 (147.94) G=0.26 A=2.56	-441 (166.18) G=0.25 A=4.10	-437 (189.35) G=0.24 A=6.98	-441 (224.24) G=0.22 A=12.47
d= -0.4	-432 (136.73) G=0.14 A=1.90	-435 (149.66) G=0.15 A=3.05	-431 (172.88) G=0.14 A=4.99	-427 (201.40) G=0.13 A=8.73	-415 (247.32) G=0.11 A=16.41
d= -0.5	-429 (142.92) G=0.08 A=2.27	-430 (163.20) G=0.08 A=3.74	-422 (189.44) G=0.07 A=6.39	-402 (222.36) G=0.06 A=11.47	-389 (293.15) G=0.05 A=23.49

Tab 5d– Sensitivity analysis to confounders G + ; A +

	s = +0.1	s = +0.2	s = +0.3	s = +0.4	s = +0.5
d= +0.1	-428 (136.23) G=2.12 A=1.18	-451 (138.13) G=2.12 A=1.38	-441 (143.28) G=2.16 A=2.19	-460 (155.67) G=2.18 A=3.54	-459 (173.39) G=2.40 A=6.08
d= +0.2	-442 (134.12) G=3.29 A=1.16	-434 (137.07) G=3.85 A=1.18	-442 (136.95) G=4.17 A=1.88	-456 (150.22) G=4.49 A=3.08	-465 (169.44) G=5.28 A=5.33
d= +0.3	-441 (135.42) G=6.32 A=1.05	-416 (133.03) G=7.97 A=1.03	-447 (135.25) G=7.78 A=1.65	-447 (148.94) G=8.07 A=2.67	-461 (163.73) G=10.59 A=4.65
d= +0.4	-441 (136.97) G=13.09 A=1.05	-424 (134.08) G=15.94 A=1.08	-452 (135.74) G=40.14 A=1.41	-449 (145.73) G=24.41 A=2.34	-465 (155.62) G=49.47 A=4.14
d= +0.5	-433 (142.48) G=40.72 A=1.06	-437 (134.23) G=83.74 A=1.05	-435 (137.70) G=232.02 A=1.21	-450 (142.01) G=143.41 A=2.04	-465 (152.35) G=172.65 A=3.64

In order to assess how strong these effects are (measured by parameters *G* and A) we can compare the odds ratios *G* and *A* with those presented in the Appendix A.2 where we estimated two separate logit models and taking as outcomes *D* and the binary transformation of *Y*. For example, we can see from this table that very few covariates show odd ratios higher than 2 and lower than 0.7.

The conclusions we get from this sensitivity analysis is that ATT estimated through PSM is rather robust to the presence of potentially omitted variables. Only if the effect of this unobserved confounder would be unreasonably strong the estimated effect becomes insignificant.

Instrumental variable results

Availability of instruments is the key to avoid the UNA assumption. Whereas these are not always easy to come by we have (at least) two alternatives in our setting. The first is a variable that takes value 1 if the household has no male children in the first wave - 0 otherwise. As already said this kind of instrument is widely used (see for ex. Angrist and Evans, 1998; Chun and Oh, 2002; Gupta and Dubey, 2003). The argument is that couples have certain gender preferences for their children - in particular they tend to have a preference for having at least one son. In other words, couples are more likely to have another child if the previous ones were girls. In so far couples have a preference for boys, such a variable work well as an instrument since it is expected to have an impact on fertility but not a direct effect on poverty. Hence, the exclusion restriction seems to be reasonable. The strong preference for sons in Vietnam is confirmed by many studies (Haughton and Haughton, 1995; Johansson, 1996 and 1998; Belanger, 2002). Also monotonicity seems to be plausible with this instrument since the presence of defiers is unlikely. In fact, the no-defiers assumption implies that households who would have (at least) one child between the two waves if they had one or more male children in the first wave (so, no "encouraged" to have more children, $Z_i = 0$) would also have more children if they had no male children ("encouraged" to have more children, $Z_i = 1$). Moreover, the instrument can be thought of as being randomised, since households can clearly not choose the sex of their children[16]. To better highlight the preference for sons we selected only

[16] This is not completely true in those countries were the selective abortion are a current practice. It was found that this is the case for example for India where amniocentesis diagnoses are available and used for sex-selective abortions (Gupta and Dubey, 2003).

households that had at least 2 children in the first wave. Thus, having only girls in the first wave, can proxy an exogenous increase in fertility between the waves.

The second instrument is a variable equal to 1 if in the community where the couple reside, none contraceptive methods between IUD and condom is available and 0 otherwise[17]. Instruments based geographical variation in availability of services is not new (see for example: McClellan et al., 1994; Card, 1995). This variable works well as an instrument if households living in communities with no contraceptive facilities have higher risks of childbearing and if contraceptive availability in the community has no direct effect on consumption growth. However, it is not unlikely that community characteristics that impact on the availability of contraceptive can also have an effect on households' poverty. Therefore, we cannot necessarily assume this instrument to be completely randomised as we do for the first one. Randomisation can only be assumed in so far we are also conditioning on a set of background variables. Controlling for covariates are consequently important in this setting. Monotonicity seems plausible also in this case, as presence of defiers is unlikely. In this case, it implies that households who would have (at least) one child between the waves if they live in a community with available contraception (so, no "encouraged" to have more children, $Z_i = 0$) would also have one child if they live in a community without contraception ("encouraged" to have more children, $Z_i = 1$).

The fact that the first instrument is randomized means that we can apply the Wald estimator without covariates. The results indicate a strong and negative effect of new children on the consumption expenditure and comparing it with the Frölich estimate we can see that the results are similar, confirming that controlling for covariates do not make a huge difference[18]. This result can be seen as reinforcement of the fact that the instrument can be assumed as randomised.

We have to do two important considerations here. First, since we selected households with at least 2 children these estimates refer properly only to this sub-population. Moreover, and more importantly, since IV estimates the LATE, these results are referred properly only to the latent sub-population of compliers who are those choosing to have another child, because they did not already have a male child.

---

[17] IUD and condom are the most available contraceptives method in Vietnam and the IUD is the most largely used (Anh and Thang, 2002).

[18] The Frölich estimates have been obtained as the ratio of two matching estimators. We used kernel based matching. The final point estimates and standard errors were obtained from bootstrapping over 1000 iterations.

The extent to which estimates can be compared with PSM estimates depends on the nature of the instrument. If we can assume that the average causal effect for "currently" non-compliers (always-takers and never-takers) is equal to the average causal effect for "currently" compliers (the LATE) then LATE and ATE coincide. This hypothesis is a-priori strong. However, in the case where we use the sex ratio of the children as the instrument, one may argue that LATE and ATE will indeed be quite similar. First we note that the estimated proportion of compliers equals 0.2, which is a quite high compared to many other studies (see for ex. Angrist and Evans, 1998). This proportion of compliers is equal to the estimated causal effect of the instrument on the treatment variable D, $E[D_{1i} - D_{0i}]$. Hence, there is evidence to suggest this instrument being strong. Moreover, the fact that male preference is likely to be a nationwide phenomenon in Vietnam implies that the estimated effect on the sub-group of compliers could be reasonably referred to the whole population. This argument is of course supported by the fact that the LATE from the IV estimation (-429) is in this case almost equal to the ATE from the PSM estimation (-414). In this sense the IV can be viewed as a robustness check for the estimates resulting from the methods based on the UNA.

Using the community level availability of contraception gives a very different picture. The first issue is that we cannot assume that this instrument is truly randomised, which means that controlling for other covariates is essential. As a confirmation of this fact we observe a huge difference between the Wald and the Frölich estimates. Frölich estimates, which are more reliable, show a strong and negative effect for the sub-population of households that are "encouraged" to have a child by the lack of contraception in the community. This LATE is not comparable with the LATE estimated with the first instrument (i.e. using the sex ratio) since the sub-population of compliers are very different. The estimated proportion of compliers for the second instrument is 0.01, hence the sub-population of households that "reacts" to this instrument is small. One could argue that since in Vietnam contraception is quite diffused compliers are in this case a rather selected group and, perhaps, constituted by less educated and generally marginalised households This also could contribute to explain the much stronger effect. The low proportion of compliers implies that this instrument is weak and hence care is needed in the interpretation of this result since, as noted by AIR, the sensitivity of IV estimator to violations of exclusion restriction and monotonicity is higher when the proportion of compliers is lower. We have to note that IV estimates are quite imprecise with high standard errors, especially for this second instrument.

Whereas these considerations would favour the sex ratio as an instrument over and above the community level availability of contraception, it is important to bear in mind that the latter has clear policy relevance. In fact, contraception availability in the communities is a variable on which policy makers could act. The Frölich estimates indicate that that the expected causal effect from a fertility reduction induced by raising contraception availability in the communities is quite high. However, the size of the sub-population reacting to this policy (compliers) is rather small. Obviously, policy makers should consider both aspects in order to calibrate efficient policies. In our case, the policy could have a huge effect on a small group.

The other estimates also indicate that fertility impact considerably on economic wellbeing but the policy implications are less direct. If policy makers worry about large households with many children, then targeted tax reductions and other benefits could help. In this way the existing gap between households with different number of children will be reduced. This policy clearly does not act on poverty through a fertility reduction and instead could influence a raise in fertility. However, our estimates do not generally suggest that the level of economic wellbeing will increase fertility levels (see Table A.1). In particular, the expenditure levels as recorded in the first wave do not seem to have any strong income effect on childbearing decisions.

Table 6 – Local average treatment effect estimates through Instrumental Variables

Estimates	Instrumental variable			
	son preference		contraception availability in the community	
	Wald estimator	Frölich estimator	Wald estimator	Frölich estimator
LATE (standard errors in parenthesis)	-429 (228)	-490 (630)	-8822 (8597)	-785 (1672)
Proportions of compliers	0.28	0.16	0.09	0.01

Notes: point estimates and standard errors for the Frölich's method have been obtained by bootstrapping over 1000 iterations.

6 Conclusions

Several approaches are available in order to estimate causal effects. The appropriateness and interpretations of these models depend on the application at hand, and importantly, the available instruments. In many cases methods relying on the UNA assumption are chosen, simply because instruments are hard to come by. The various implications of these methodological choices are rarely considered in applied work, but, as we point out, the underlying assumptions are important, especially when there is interest in comparing estimates from different methods. We discuss these methods in light of an application where we consider the effect of fertility on changes in consumption expenditure. The issue is that childbearing events cannot be considered as an exogenous measure of fertility, especially when the outcome relates to economic wellbeing – in our case measured in terms of consumption expenditure. However, the discussion of the methods is general and applies to many other applications.

Using methods based on the UNA assumption, such as simple linear regression and propensity score matching, we find that those households having children between the recorded waves have considerably worse outcomes in terms of changes in consumption expenditure. The negative impact is however, highly heterogeneous, and varies substantially with education for instance. We then demonstrate how one can make an assessment of the potential effect from omitting relevant but unobserved variables without actually implementing an Instrumental Variable approach. This is a very useful tool in the sense that valid and relevant instruments are often hard to come by. In our application the estimates are robust with respect to unobserved omitted variables. We find that the estimated effect becomes non-significant only if the association between the omitted covariate, selection and the outcome is extremely (and unreasonable) large.

Despite the robustness of the UNA in our application we implement nevertheless the IV method using two different instruments. The first is a well-used instrument that relates to couples' preference for sons. In this case the instrument is a binary variable taking value 1 in those households that at the first wave had no male children and 0 otherwise. The IV estimation is implemented for sub-sample of households with at least two children in the first wave. Since the instrument is close to being randomised, a simple Wald estimator can be used. The second instrument takes value 1 for households residing in a community where none of the contraceptive methods IUD and condom was available at the first wave and 0 otherwise (at least one was available). This instrument is not randomized and hence requires controlling for covariates. Whereas both instruments satisfy the standard tests for relevance and validity, they provide very different parameter estimates. The fact that the IV estimates the LATE, as opposed to the ATE and ATT, is the key reasons for these differences. Moreover, since the second instrument requires inclusion of covariates, it typically involves more stringent assumptions on the functional forms. Using the approach suggested by Frölich (2007), which overcomes many of these assumptions, demonstrates that they do matter for the parameter estimates.

The use of Instrumental Variable methods in our application illustrates that reasonable instruments can lead to estimates that differ from those of methods based on UNA but also differ among them. In fact, compliers for one instrument can be very different from compliers to another instrument and consequently if the treatment effect is heterogenous the estimated LATE in the two cases will necessarily differ. With the first instrument we estimated a negative impact of fertility on poverty with a magnitude not dramatically different from that obtained by method based on the UNA. This could be an effect of the fact that the preference for son is quite a general phenomenon in Vietnam not involving particular kinds of households. The estimated proportion of compliers in this case is actually quite high: 20%. The estimate with the second instrument, on the contrary, is much higher, in absolute value. The estimated proportion of compliers in this case is small: 1%. This small sub-population of households reacting to the availability of contraceptives is likely to be highly selected. These households live in areas where no contraceptives were available. Clearly their opportunity to control fertility through contraceptive practices is much reduced as it is unlikely that compliers are able to get contraceptives from elsewhere. In this sense these households have a higher exposure to childbearing. These communities are also likely to be more disadvantaged compared to others.

Whereas the estimates based on this instrument is very different compared to the one based on the sex preference, an advantage is that it does have direct policy relevance, simply because the instrument itself is a policy variable. The effect on this sub-population is high and importantly, much higher than what is estimated for the whole population through the ATT and ATE. However, the size of this sub-population is rather small, which is an equally important consideration for the policy maker.

References

Abadie, A. (2003) Semiparametric instrumental variable estimation of treatment response models. *Journal of Econometrics*, 113, 231–263.

Abadie, A., and Imbens, G. W. (2002) Simple and Bias-Corrected Matching Estimators for Average Treatment Effects. Technical Working Paper T0283, NBER.

Abadie, A., and Imbens. G. W. (2004) On the Failure of the Bootstrap for Matching Estimators. NBER Technical Working Paper T0325.

Abadie, A., Drukker, D., Leber Herr, J. and Imbens, G.W. (2004), Implementing Matching Estimators for Average Treatment Effects in Stata. Stata Journal, 4(3), 290-311.

Admassie, A. (2002) Explaining the High Incidence of Child Labour in Sub-Saharan Africa. *African Development Review,* 14(2): 251 – 275.

Angrist, J. D. and Evans, W. N. (1998) Children and their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size. *American Economic Review*. 88(3), 450–77.

Angrist, J. *D*., and Krueger, A. (1991) Does compulsory school attendance affect schooling and earnings? *Quarterly Journal of Economics*, 106, 979–1014.

Angrist, J., Imbens, G. (1995) Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of American Statistical Association*, 90, 431–442.

Angrist, J. *D*., Imbens, G. W. and Rubin, *D*. B. (1996) Identification of causal effects using instrumental variables. *Journal of the American Statistical Association,* 91, 444–472.

Anh, *D*. N. and Thang, N. M. (2002) *Accessibility and Use of Contraceptives in Vietnam*. *International Family Planning Perspectives*, 28(4), 214-219.

Becker G. S. and Lewis H.G. (1973) On the interaction between the quantity and quality of children, *Journal of Political Economy*, 81(2), S279-S288.

Belanger, *D*. (2002) Son preference in a rural village in North Vietnam. *Studies in Family Planning, 33*(4), 321–334.

Blundell, R. and Powell, J. (2003) Endogeneity in nonparametric and semiparametric regression models. In: Hansen, L., Dewatripont, M. and Turnovsky, S.J. (Eds.), *Advances in Economics and Econometrics*. Cambridge University Press, Cambridge, pp. 312–357.

Card, *D*. (1995) Using geographic variation in college proximity to estimate the return to schooling. In: Christofides, L., Grant, E., Swidinsky, R. (Eds.), *Aspects of Labor Market Behaviour: Essays in Honour of John Vanderkamp*. University of Toronto Press, Toronto, pp. 201–222.

Carneiro, P., Heckman, J. and Vytlacil E. (2005) Understanding What Instrumental Variables Estimate: Estimating Marginal and Average Returns to Education. Unpublished manuscript, University of Chicago, Department of Economics.

Chun, H. and Oh, J. (2002) An instrumental variable estimate of the effect of fertility on the labour force participation of married women. *Applied Economics Letters*, 9, 631-634.

Coudouel A., Hentschel J. and Wodon Q. (2002) Poverty Measurement and Analysis, Poverty Reduction Strategy Paper Sourcebook, World Bank, Washington *D*.C.

Cox, *D*. R. (1958) *Planning of experiment*. New York, Wiley.

Crump, R. K., Hotz, V. J., Imbens, G. W. and Mitnik, O. A., (2006) Nonparametric Tests for Treatment Effect Heterogeneity. IZA Discussion Paper No. 2091

Das, M. (2005) Instrumental variables estimators of nonparametric models with discrete endogenous regressors. *Journal of Econometrics* 124, 335–361.

Dawid, A. P. (1979) Conditional Independence in Statistical Theory, *Journal of the Royal Statistical Society*, Series B, 41, 1-31

Deaton, A. and Zaidi, S. (2002) Guidelines for Constructing Consumption Aggregates for Welfare Analysis, Living Standards Measurement Study Working Paper No. 135, The World Bank.

Dehejia, R., and Wahba, S. (1999) Causal effects in non-experimental studies: re-evaluating the evaluation of training programs. *Journal of the American Statistical Association*, 94, 448, 1053–1062.

Drovandi, S. and Salvini. S. (2004) Women's Autonomy and Demographic Behavior. *Population Review,* 43(2).

Durbin, J. (1954), Errors in Variables, *Review of the International Statistical Institute*, 22, 23-32.

Duy, L. V., Haughton *D*., Haughton J., Kiem *D*. A. and Ky L. *D*. (2001) Fertility decline. In: Haughton *D*., Haughton J., Phong N. (eds), *Living Standards during an Economic Boom. Vietnam 1993-1998*, Statistical Publishing House, Hanoi.

Easterlin, R.A. and Crimmins E.M. (1985) *The Fertility Revolution.* Chicago: University of Chicago Press.

Falaris, E.M. (2003) The effect of survey attrition in longitudinal surveys: evidence from Peru, Cote d'Ivoire, and Vietnam. *Journal of Development Economics*, 70, 133-157.

Fisher, R. A. (1925) *Statistical Methods for Research Workers*. 1st Edition. Oliver and Boyd, Edinburgh.

Florens, J., Heckman, J., Meghir, C. and Vytlacil, E. (2002) Instrumental variables, local instrumental variables and control functions. Cemmap Working Paper 15/02.

Frölich, M. (2007) Non parametric IV estimation of local average treatment effects with covariates, *Journal of Econometrics*, 139, 35-75.

Glewwe, P., Gragnolati, M. and Zaman, H. (2002), Who Gained from Vietnam's Boom in the 1990s? *Economic Development and Cultural Change*, 50(4), 773-792.

Goodman, A. and Sianesi B. (2005) Early Education and Children's Outcomes: How long the impacts last, *Fiscal Studies*, 26(4).

Green, W.H. (2002). *Econometric Analysis,* 5th edition, Prentice Hall.

Gupta, N. D. and Dubey A. (2003) Poverty and Fertility: - An Instrumental Variables Analysis on Indian Micro Data, Working Paper 03-11, Aarhus School of Business.

Halloran, M. E. and Struchiner, C. J.,(1995) Causal inference in infectious disease, *Epidemiology,* 6, 142-151.

Haughton, D., Haughton, J., Phong, N. (2001). *Living Standards during an Economic Boom. Vietnam 1993-1998*, Statistical Publishing House, Hanoi.

Haughton, J. and Haughton. *D*. (1995) Son preference in Vietnam. *Studies in Family Planning*, 26 (6), 325-337.

Heckman, J. J. (1992) Randomization and social program evaluation, in Evaluating welfare and training programs, In: Masnski, C. F. and Garfinkel, I. (eds.), *Evaluating welfare and training programs,* Cambridge, MA: Harward University Press, 201-230.

Heckman, J. J. (1997) Instrumental Variables: A study of implicit behavioural assumptions used in making program evaluations, *Journal of Human Resources*, 32, 441-462.

Heckman, J.J., Ichimura, H. and Todd, P. (1997) Matching As An Econometric Evaluation Estimator, *Review of Economic Studies*, 65, 261-294.

Heckman, J. and Vytlacil, E. (1999) Local instrumental variables and latent variable models for identifying and bounding treatment effects. Proceedings National Academic Sciences, USA Economic Sciences 96, 4730–4734.

Heckman, J. and Vytlacil, E. (2001) Local instrumental variables. In: Hsiao, C., Morimune, K. and Powell, J. (Eds.), *Nonlinear Statistical Inference: Essays in Honor of Takeshi Amemiya*. Cambridge University Press, Cambridge.

Hirano, K., Imbens, G., Rubin, *D*. and Zhou, *X*. (2000) Assessing the effect of an influenza vaccine in an encouragement design. *Biostatistics*, 1, 69–88.

Holland, P. (1986) Statistics and causal inference. *Journal of American Statistical Association*, 81, 945–970.

Ichino, A., Mealli, F. and Nannicini, T. (2007) From Temporary Help Jobs to Permanent Employment: What Can We Learn from Matching Estimators and their Sensitivity? *Journal of Applied Econometrics*, forthcoming.

Imbens, G. (2002) A Classification of Assignment Mechanisms. Mimeo, available at http://elsa.berkeley.edu/users/imbens/e244_f02/chap3.pdf.

Imbens, G. (2004) Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review, *Review of Economics and Statistics*, 86, 4-30.

Imbens, G. W. and Angrist, J. *D*. (1994) Identification and estimation of local average treatment effects. *Econometrica*, 62, 467–475.

Imbens, G. and Newey, W. (2003) Identification and estimation of triangular simultaneous equations models without additivity. Presented at the EC2 Conference, London, December 2003.

Johansson, A. (1996) Family planning in Vietnam - women's experiences and dilemma: a community study from the Red River Delta. *Journal of Psychosomatic Obstetrics and Gynecology*, 17, 59-67.

Johansson, A. (1998) Population policy, son preference and the use of IUDs in North Vietnam. *Reproductive Health Matters*, 6, 66-76.

Justino, P. and Litchfield, J. (2004) Welfare in Vietnam During the 1990s: Poverty, Inequality and Poverty Dynamics. *Journal of the Asian Pacific Economy*, 9(2), 145-169.

Kabeer, N. (2001) Deprivation, discrimination and delivery: competing explanations for child labour and educational failure in South Asia. Institute of Development Studies Working Paper 135, Sussex, Brighton, UK

Kim, J and Aassve A. (2006) Fertility and its Consequence on Family Labour Supply, Institute of Labour Studies (IZA) Discussion Paper No. 2162.

Leuven, E. and Sianesi, B. (2003) PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. Statistical Software Components S432001, Boston College Department of Economics.

Livi-Bacci, M. 2000. *A concise history of the world population*. Blackwell: Oxford.

Livi-Bacci, M. and De Santis, G. (1998) *Population and Poverty in the Developing World*. Clarendon Press: Oxford.

McClellan, M., McNeil, B. J. and Newhouse, J. P. (1994) Does More Intensive Treatment of Acute Myocardial Infarction in the Elderly Reduce Mortality? *Journal of the American Medical Association*. 272(11), 859 – 66.

Moav, O. (2005) Cheap Children and the Persistence of Poverty. *The Economic Journal,* 115, 88-110.

Molini, V. (2006) Food Security in Vietnam during the 1990s - The Empirical Evidence. United Nation University, Reaserch paper No. 2006/67.

Nannicini, T. (2007) Simulation-Based Sensitivity Analysis for Matching Estimators. *The Stata Journal*, 7(3), 334-350.

Newey, W., Powell, J., and Vella, F., (1999) Nonparametric estimation of triangular simultaneous equations models. *Econometrica,* 67, 565–603.

Neyman, J. (1923) On the application of probability theory to agricultural experiments: essay on principles, section 9. Translated in *Statistical Science*, 5(4), 465–480, (1990).

Nguyen-Dinh, H. (1997) A socioeconomic analysis of the determinants of fertility: The case of Vietnam. *Journal of Population Economics*, 10 (3), 251-271.

Pudney, S. and Aassve, A. (2007a) Poverty transitions in developing countries: the roles of economic and demographic change, *Working Paper of Institute for Social and Economic Research*, paper 2007-25. Colchester: University of Essex

Pudney, S. and Aassve A. (2007b) Endogenous fertility and its impact on poverty: Evidence from Vietnam, *Working Paper of Institute for Social and Economic Research*, paper 2007-26. Colchester: University of Essex

Rosenbaum, P. R. and Rubin, *D*. B. (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41–55.

Rubin, *D*. B. (1974) Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66, 688–701.

Rubin, *D*. B. (1978) Bayesian inference for causal effects: the role of randomization. *Annals of Statistics*, 6, 34–58.

Rubin, *D*., (1980) Discussion of Randomization Analysis of Experimental Data: The Fisher Randomization Test by *D*.Basu. *Journal of the American Statistical Association*, 75, 591-93.

Scornet, C. (2007) 1963-2003: Quarante ans de planification familiale au Viêt-Nam. In Adjamagbo, A., Msellati, P. and Vimard P. (eds.), *Santé de la reproduction et fécondité dans les pays du Sud. Nouveaux contextes et nouveaux comportements,* Academia Bruylant, Louvain la Neuve, pp. 142-171

Tung, P. D. (2004) Poverty line, poverty measurement, monitoring and assessment of MDG in Viet Nam. Report presented at the “2004 International Conference on Offical Poverty Statistics – Methodology and Comparability”, Manila October 2004.

Wald, A. (1940) The fitting of straight lines if both variables are subject to error. *Annals of Mathematical Statistics*, 11, 284-300.

White, H. and Masset, E. (2003) *Constructing the Poverty Profile: An Illustration of the Importance of Allowing for Household Size and Composition in the Case of Vietnam*. Young Lives Working Paper No. 3, London: Young Lives and Save the Children Fund UK.

White, H. and Masset, E. (2002) *Child poverty in Vietnam: using adult equivalence scales to estimate income-poverty for different age groups*. MPRA Paper 777, University Library of Munich, Germany.

Willis, R. J. (1973) A New Approach to the Economic Theory of Fertility Behavior. *Journal of Political Economy,* 81(2), part 2: S14-S64.

World Bank (2000) Vietnam: Attacking Poverty - Vietnam Development Report 2000, Joint Report of the Government-Donor-NGO Working Group, Hanoi.

Wooldridge, J.M. (2002) *Econometric Analysis of Cross Sectional Panel Data*, MIT Press.

Zhao, *Z*. (2005) Sensitivity of Propensity Score Methods to the Specifications. IZA Discussion Paper No. 1873

Appendix

Table A.1 – Estimates of the linear model for consumption growth, *Y*, and the logit model for selection into treatment.

Variables	Outcome equation (linear regression)			Selection equation (logistic regression)		
	Coef.	Robust Std. Err.	ey/ex (elasticity)	Coef.	Robust Std. Err.	ey/ex (elasticity)
DUMnewchild2	-413.58	66.07	-0.12			
Sexhhh	96.83	91.40	0.06	-0.02	0.18	-0.01
Kinh	173.39	61.71	0.11	-0.03	0.19	-0.02
perkids_04	-2.61	3.64	-0.03	0.01	0.01	0.14
perkids_59	-1.25	2.87	-0.02	-0.02	0.01	-0.28
perkids_1014	-8.45	2.95	-0.08	-0.03	0.01	-0.28
permale_1545	-0.73	3.74	-0.01	0.02	0.01	0.25
perfema~1545	-1.97	4.33	-0.04	0.02	0.01	0.28
Farm	-35.95	64.44	-0.02	0.11	0.13	0.05
Edu	10.00	1.63	0.42	-0.02	0.00	-0.62
rlpcex1	-0.12	0.08	-0.11	0.00	0.00	-0.01
region1	45.44	98.58	0.01	-0.03	0.24	0.00
region2	26.95	102.00	0.01	-0.49	0.25	-0.08
region3	-50.55	109.75	-0.01	0.56	0.25	0.05
region4	208.02	104.51	0.02	0.64	0.24	0.04
region5	400.58	147.09	0.01	1.61	0.37	0.04
region6	1163.51	198.20	0.07	0.30	0.27	0.01
Agehhh	4.22	4.22	0.13	0.00	0.01	-0.13
Hhsize	-0.65	16.13	0.00	-0.09	0.04	-0.35
peractm_1545	4.32	1.50	0.31	0.01	0.00	0.31
peractf_1545	0.55	1.42	0.04	0.00	0.00	0.22
IEI	12.81	15.22	0.05	-0.06	0.03	-0.20
EDI	-64.92	52.35	-0.14	0.14	0.13	0.27
HFI	22.52	18.80	0.11	-0.06	0.04	-0.23
Constant	1079.89	522.43		0.13	1.00	

Table A.2 – Estimated odds ratios in the models for D and the binary transformation of *Y*.

Covariates	Odds ratios	
	Y	*D*
D	0.41	---
Sexhhh	0.95	0.98
Kinh	1.71	0.97
Perkids_04	0.98	1.01
perkids_59	0.98	0.97
perkids_1014	0.97	0.96
permale_1545	0.99	1.02
perfema~1545	0.97	1.02
Farm	0.98	1.12
Edu	1.02	0.98
rlpcex1	0.99	0.99
region1	0.90	0.97
region2	0.85	0.61
region3	0.70	1.75
region4	1.43	1.89
region5	1.34	4.99
region6	5.65	1.35
agehhh	0.99	0.99
hhsize	0.97	0.91
peractm_1545	0.99	1.01
peractf_1545	0.99	1.00
IEI	1.07	0.95
EDI	0.91	1.16
HFI	1.02	0.95