Dragon Children: Identifying the Causal Effect of the First Child on Female Labor Supply with the Chinese Lunar Calendar

James P. Vere

University of Hong Kong

26 March 2004

Abstract

Instrumental variables (IV) estimates of the effect of fertility on female labor supply have only been able to identify the causal effect of second and higher-parity children. This study uses exogenous variation in fertility caused by the Chinese lunar calendar to identify the effect of the first child. Additionally, weighting formulas are presented to interpret IV estimates as weighted average treatment effects in the case of multiple endogenous variables, which are useful when children vary in intensity by both number and age. The effect of the first child is found to be much greater than that of other children. (*JEL* J13, J22)

1 Introduction

The relationship between fertility and female labor supply has been of interest to economists and demographers for decades. Mincer's (1962) early model was the first of many to establish a strong, negative correlation between the two, and the subsequent debate on causality has continued unchecked. If childbirth has a causal effect on women's labor supply decisions, the implications are substantial. For one, it may be that children have a detrimental effect on women's careers, since they induce their mothers to leave the labor force (Gronau 1988). In the reverse direction, it is possible that a decline in fertility may cause an increase in female labor supply, which is one theory put forward for the rising levels of women's labor force participation in the United States since World War II (Coleman and Pencavel 1993). But because fertility and labor supply decisions are jointly determined, cause and effect relationships between the two are difficult to extract.

One way of addressing the endogeneity problem has been to impose a parametric structure on the labor supply and fertility decisions and then estimate the resulting model within a simultaneous equations framework (e.g., Moffitt (1984); Hotz and Miller (1988); Carrasco (2001)). However, the results of this literature have been varied, and fertility is generally not viewed as a causal factor in its own right, but merely a mechanism through which the effects of other causal factors might be transmitted to the labor supply decision (for example, changes in a woman's market wage, the efficacy and availability of contraceptive methods, or biological constraints on childbearing).

More recently, instrumental variables methods have been put to use to draw inferences on the causal link between fertility and female labor supply. The search for exogenous variation in fertility, however, has yielded few candidate instruments. Bronars and Grogger (1994) use the incidence of twins in the 1970 and 1980 U.S. Census Public Use Microdata Samples to estimate the effect of an unplanned second child on labor force participation for unwed mothers. Angrist and Evans (1998) study the labor supply of married women with at least two children, using both twins and the gender mix of the first two children as instruments. Since parents tend to prefer having a mix of genders among their children, gender mix operates as an instrument because couples with two children of the same sex are more likely to have a third child than couples with one boy and one girl.

The gap in the literature is that though the gender mix instrument can identify the causal effect of a third- or higher-order birth, and the twins instrument can identify the effect of birth of second order or higher, inference about the effect of the first birth is impossible with these instruments. Instrumental variables estimates of the marginal effects of higher-order births are still conditional on a woman's decision to have a first child. The marginal effect of the first birth, therefore, is of critical importance in determining the total effect of fertility on female labor supply. Moreover, if there are economies of scale to be realized in raising children, the marginal effect of the first birth may be much higher than the effects of subsequent births.

A further issue with existing instrumental variables estimates is that many rely on constant-coefficient linear regression models to identify the effect of interest. Additionally, the implicit conditional expectation function of the dependent variable in such models is not appropriate for examining measures of female labor supply, which are often binary and virtually always nonnegative. It might be tempting to surmise that ignoring these limitations and proceeding with two stage least squares (2SLS) estimation anyway can still yield an average population effect, much like ordinary regression yields the best linear predictor regardless of the underlying conditional expectations function, but this supposition is false (Angrist 2001). However, 2SLS estimation when the model is fully saturated yields weighted average population effects that fully incorporate innate restrictions on the dependent variable. This is because the probability limits of saturated 2SLS estimates, as weighted averages of individual causal effects, must fall within the smallest convex set containing the causal effects themselves. The weights are the variance in the predicted values of the endogenous variable in the second stage, conditional on the covariates (Angrist and Imbens 1995). Moreover, unlike nonlinear estimation strategies, inference of causal effects from a saturated 2SLS model does not depend on correctly specifying the relationship between the endogenous variable, the dependent variable, and a continuously distributed latent index.

One contribution of this paper is to identify the marginal effect of the first birth using a unique instrumental variables strategy that uses the Chinese lunar calendar as a source in exogenous variation in childbirth among Hong Kong women. In Chinese astrology, children born in certain years are perceived as being more blessed by fate than children born in other years, and corresponding increases in fertility during auspicious years are widely observed in Chinese populations. A woman's usual language spoken at home is used to determine if she is a member of the Chinese subpopulation in Hong Kong. Since one's cultural background is virtually exogenous, an indicator variable for this characteristic is a plausible instrument for fertility. Perhaps surprisingly, an overidentification test does not yield any evidence that, after conditioning on age and education, a woman's first language affects her labor supply decision apart from its effect on fertility. However, this finding is consistent with sociologists' observations that the social norms of

Hong Kong Chinese with regard to work and relationships within the nuclear family are remarkably similar to Western social norms (Podmore and Chaney 1974). This unusual confluence of highly Western attitudes toward work and household decisions and highly non-Western views on the optimal timing of children makes Hong Kong a unique environment in which to examine the relationship between fertility and female labor supply.

An additional contribution of this paper is to show that a saturated 2SLS model, in the case of multiple endogenous variables, can be used to estimate average causal effects without imposing functional form restrictions on the relationship between the instruments, the outcomes and the treatment. This is an extension of Angrist and Imbens's (1995) result for one endogenous variable; the extension is simple, but useful because it permits the intensity of the treatment to vary along more than one dimension. In this case the "treatment" comprises children that vary in both age and number. When the number of children is recast as a set of mutually orthogonal indicators, exclusive of an indicator for zero children, the coefficient on the ith indicator is a weighted average of the effects of ith children of varying ages, and the weights are proportional to the probability that the instruments induce the birth of a ith child of that age. Without parameterizing the relationship between childrens' ages and their effects on female labor supply, the weighting function permits formalizing the intuition that different instruments, since they induce the presence of children drawn from different age distributions, may yield different IV estimates of the causal effect of fertility.

The second section of this paper discusses the data and the instruments. The third section presents the methodology and the main empirical results, and the fourth section concludes.

2 Data and Instruments

2.1 Data

The empirical work for this paper uses information on women's labor supply, demographic characteristics and birth histories from the 5 percent sample data sets of the Hong Kong Population Censuses of 1991, 1996, and 2001. During each census, six-sevenths of the population is simply enumerated to provide basic demographic information. The remaining seventh is given a detailed survey on demographic and socioeconomic characteristics, including labor force activity, migration patterns and relationships to other members of the household. Households included in the 5 percent sample data sets are drawn exclusively from those households selected to complete the long survey, in a manner such that each household in Hong Kong has an equal probability of being included in the 5 percent sample. The 5 percent samples for 1991, 1996, and 2001 in total contain long survey records for 926,452 people.

Since the Hong Kong census data do not include retrospective fertility information, it is necessary to construct birth histories based on the number and age of children present in the household. The method of doing so is similar to Angrist and Evans's (1999). Information on relationships within the household was used to match children to a universe of women consisting of those women between the ages of 15 and 40 who are reported as being either the head of household or the spouse of the head of household. Women over 40 were excluded because past this age it becomes less likely that their children are still included as part of the household register. But since very few women in Hong Kong start to have children before their early 20s[1], and very few children leave home before their early 20s, it is reasonable to use household composition to construct birth histories for women under 40.[2] In addition, any woman reported to have a child less than 15 years younger than herself was excluded from the sample.

Finally, the sample was limited to women with residency in Hong Kong of seven years or more. The purpose of this restriction is to exclude women who are temporarily in Hong Kong, since their residency is often contingent on their or their spouses' continuing employment. Very few temporary residents, however, stay for more than one two- or three-year employment contract. Further, people who have resided in Hong Kong for seven years acquire the right to permanent residency and are no longer subject to immigration conditions. Therefore, in light of the institutional characteristics of Hong Kong's labor force, the seven year criterion is sensible. In total, 82,028 women met the criteria for inclusion in the sample.

2.2 Fertility and the Dragon Year in Hong Kong

According to Chinese astrology, children born under different years of the lunar calendar have different characteristics, and some years are considered more auspicious for the birth of a child than others. In particular, the dragon year, which falls every twelfth year, is considered to be the most auspicious of all. Children born in these years are thought to have better fortune, do better in school, and be natural leaders.

[1] In fact, very few women even marry before their early 20s. In 2001, the median age at first marriage for women was 27.5.

[2] Children studying away from home, even those studying overseas, were included in the censuses as long as they were Hong Kong residents and the household remained their permanent address.

Whether one finds Chinese astrology credible or not, enough Chinese subscribe to these beliefs to generate measurable increases in the birth rate during dragon years in the countries where they live. For example, in 2000, the last such year, the general fertility rate in Taiwan increased by 6.7 percent over its 1999 level, and in China the crude birth rate was 6.8 percent higher. In Hong Kong, the number of births rose by nearly six percent, the first increase in the birth rate after six years of steady declines (Moi 2001).

This phenomenon is far from new. Sun, Lin and Freedman (1978) observed unusual increases in fertility in Taiwan in 1976, particularly during the last six months of that year. In Taiwan, the crude birth rate rose from 23.0 to 25.9, but they could explain only 21 percent of the increase through changes in Taiwan's age structure. They attributed the remainder of the increase in marital fertility to the dragon year superstition. Moreover, though marital fertility before 1976 had been increasing among women of all age and education levels, the 1976 rise in fertility was concentrated among young women of low education levels, lending support to the hypothesis that the superstition played a significant role in the 1976 fertility increase.

Goodkind (1993) further documents the 1976 baby boom, particularly the extent to which availability of modern contraceptive methods allowed women to engage in birth timing. He finds that use of birth control pills and other forms of contraception declined sharply during the conception window for dragon babies, and swiftly increased thereafter. Since previous dragon years were not accompanied by increases in the birth rate, he concludes that the availability of modern contraception methods was a decisive factor in causing the 1976 boom.

Aggregate birth statistics show that Hong Kong's experience has been similar. Figure 1 shows the number of births in Hong Kong during the previous three dragon years, 2000, 1988, and 1976, and the two years surrounding each of them. In 1976, there is no appreciable difference in the birth rate between that year and surrounding years. However, the numbers of births in 1988 and 2000 are both considerably higher than in surrounding years; the number of births in 1988 was 7.8 percent higher than the number of births in 1987, and the number of births in 2000 was 5.5 percent more than in 1999.[3] This may be because it took until 1988 for modern contraception methods to become widely used in Hong Kong. In any event, the statistics suggest that a considerable number of Chinese women timed their births to occur in 1988 and 2000. Table 1 lists the number of births in Hong Kong for all years between 1971 and 2002. Yip, Lee and Cheung (2002) make similar observations analyzing the pattern of total fertility rates over time.

A Cox proportional hazard model can yield further insight into the correlates of fertility for women in the sample. The proportional hazard framework asserts that the hazard rate for subject i at time t is

$$h\left(t \mid X_i\right) = h_0 \cdot \exp\left(X_i \cdot \beta\right)$$

where X_i is a vector of covariates, h_0 is the baseline hazard and β is a vector of parameters to be estimated. There is no parametric assumption about the shape of the baseline hazard over time.

Column I of Table 2 presents estimates of a simple, reduced-form Cox model

[3] The effect may have been particularly pronounced in 1988 because '8' is a lucky number in its own right. Economic conditions were also considerably better in 1988 than in 2000.

of fertility where time is defined as the woman's age, the event is the birth of a child, and the covariates are her years of education and a set of dummy variables for each year of the Chinese zodiac exclusive of the dragon year. The model also includes a full set of interaction terms between the zodiac dummy variables and an indicator variable for non-Chinese, equal to one when a woman reports speaking a language other than a Chinese dialect as her usual language at home.[4] In the table, the zodiac variables are listed in the order in which they occur in the Chinese calendar; the dragon year occurs between the year of the rabbit and the year of the snake. Standard errors are in parentheses. Marital status is excluded as a covariate because marital status and fertility are often chosen jointly.

The estimates in column I of Table 2 indicate that, holding number of children, age, and years of education constant, Chinese women are relatively less likely to give birth in most non-dragon years, particularly those years immediately before the dragon year. For example, giving birth in the year of the rabbit is more than 10 percent less likely than giving birth in the year of the dragon, when all other factors are held constant. This makes intuitive sense if women are more likely to wait for the auspicious year when that year is less far away. The only coefficient that is positive and significant is that for the year of the pig ($p = 0.021$). Moreover, though the zodiac dummy variables are highly significant for Chinese ($p \ll 0.001$), the hypothesis that the zodiac has no effect on fertility for the non-Chinese population cannot be rejected ($p = 0.45$). Therefore, since the pattern of fertility is exactly as Chinese astrology would suggest, and utterly absent among non-Chinese, it seems

[4] The results are not qualitiatively different when self-reported ethnicity is used in place of usual language spoken at home to define "Chinese." However, self-reported ethnicity is only available for the 2001 data.

compelling that the seasonality of births is caused by superstitious beliefs held by a significant part of the population.

Figure 2 shows these results graphically. The relative probabilities are obtained by exponentiating the estimated parameters for each zodiac year, and they hold when other factors in the model are held constant. A column for the dragon year is included for reference.

Column II of Table 2 shows estimates of a model that additionally includes interaction terms between years of education and the zodiac variables. The interaction terms permit the effect of different years of the zodiac to vary depending on the woman's education level. Although many of the zodiac variables are now significant and positive, these results do not contradict the results in Column I because most of the identification is from women with more than zero years of education. The median number of years of education for women in the sample is 11; the 10th and 90th percentiles are 6 and 16 years, respectively. With the exception of the year of the pig, which is no longer significantly different from the year of the dragon, the interaction terms are all sufficiently negative so that the probability of giving birth in a non-dragon year, when evaluated at a number of years of education typical for women in the sample, is relatively less than that for a dragon year. The interaction terms show that the influence of the zodiac is in fact more pronounced for more educated women, which could reflect greater awareness of fertility timing methods. Another possibility is that because more educated women have fewer children, and because it is difficult to time more than one child in this way, a greater proportion of their births are timed than are births to less educated women.

2.3 Twins and Gender Mix

Of the sample data sets from 1991, 1996, and 2001, only the 1991 data include information on both month and year of birth. For the 1991 data, twins were defined as two children born to the same mother, in the same month and year. This definition yields a twinning rate of 5.1 per thousand children, which is in line with Hong Kong's twinning rate at birth, measured by Imaizumi (1998).

Since month of birth is not reported in the 1996 and 2001 data, it is natural to wonder about the consequences if one were to ignore this fact and consider any two children born to the same mother in the same year as twins. When only the year of birth is used to match children as twins in the 1991 data, however, the measured twinning rate rises to 7.5 per thousand births, a 47 percent increase over the twinning rate measured with both month and year of birth data. Therefore, since ignoring month of birth results in considerable error in imputing the birth of twins, the twins instrument will only be used in connection with the 1991 data.

Though it is not uncommon for Western couples to prefer a mix of genders among their children (as observed by Westoff, Potter and Sagi (1963)), Hong Kong couples with two daughters are much more likely to have a third child than those with two sons. Table 3 shows, for women in the sample, the probability of a woman having an additional child conditional on the gender of children already present in the household. At first, the results are not unlike Angrist and Evans's (1998) for American women in that the gender of the first child, for both Chinese and non-Chinese, makes little difference in the probability the couple decides to have a second child. However, Chinese couples with two girls are significantly more likely to have a third child than Chinese couples with two boys or one boy and one girl.

In 2001, more than one in six couples with two girls chose to have a third child, but less than one in eight couples with two boys did. For non-Chinese couples, the results are inconclusive; the probabilities of a third child conditional on the gender of the first two are never significantly different from one another, and though the point estimates for 1991 and 1996 suggest Western-style preferences for a mix of genders, the point estimates for 2001 are more in line with the Chinese pattern.

One issue that arises with the use of gender mix as an instrument in this context is that gender selection, either by selective abortion or by *in vitro* filtering methods, is a known issue in Chinese societies, particularly mainland China. To some extent, this occurs in Hong Kong, but it is only an issue for high-parity births. Wong and Ho (2001) examine the issue closely and find that the gender ratio of births to women in Hong Kong with one child is generally as biology would predict. However, the gender ratio of births to women with two children was skewed significantly, in favor of boys for women with two daughters, and to a lesser extent in favor of girls for women with one or two boys. But though this is a problem for high-parity births, the gender composition of the first two children is still essentially random.

3 Fertility and Labor Force Participation

3.1 Methodology

Angrist and Imbens (1995) show that, for models with variable treatment intensity, a fully saturated 2SLS strategy produces estimates that can be interpreted as average causal effects under very general conditions. The estimated parameter is a weighted per-unit treatment effect, where the effects of unit changes are weighted by the probability that the instrument induces that unit change, conditional on the instrument inducing any unit change. Therefore, 2SLS, using the incidence of twins as the instrument and the number of children as a measure of treatment intensity will produce a weighted average of the causal effects of second- and higher-order births, since it is impossible for the incidence of twins to induce a shift in the number of children from zero to one.

When indicator variables for the number of children are included as endogenous variables in a fully saturated 2SLS model, the marginal effect of each child can be estimated directly. To formalize this point, let a woman's labor supply decision be adequately represented by

$$Y = M_0 + \sum_{i=1}^{I} \sum_{j=1}^{J} M_{ij} \cdot C_{ij} \tag{1}$$

where C_{ij} is an indicator for the presence of a ith child of age j, up to at most I children of maximum age J, and the children are in descending order by age. M_{ij} is the marginal effect on Y of an ith child of age j. Note it is assumed that the M_{ij} are not dependent on the characteristics of any other children in the household. Further, though only one outcome is observed, it is assumed that a full set of Y exists for every woman and these potential outcomes are independent across women in the sample.

Since only one pair of labor supply and fertility outcomes is observable for each woman in the sample, the M_{ij} are not directly observable. However, it is possible to infer information about the distribution of M_{ij} in the population when observations of Y and fertility can be grouped independently of any relationship between the two. When Y and fertility are themselves independent, no grouping is necessary. When

Y and fertility are not independent, inference can still proceed if pairs of outcomes can be grouped independently through a suitable instrumental variable. This is the principle behind the Wald estimator (Angrist 1991).

Additionally, while the fertility effect of being assigned by the instrument to one group relative to another need not be constant across individuals, for any pair of groups the effect of being reassigned from one to another must be either positive or negative for everyone (not necessarily strictly, though of course the effect cannot be zero for everyone in the sample). This monotonicity condition, which Angrist and Imbens (1995) discuss in some detail, is a nonparametric version of the traditional instrumental variables assumption that the instrument must be correlated with the endogenous variable.

Let Z be a set of mutually orthogonal binary instruments, satisfying the monotonicity condition, that are independent of the endogenous variables after conditioning on a set of mutually orthogonal binary covariates, X. Since any set of discrete variables can be recast as a set of mutually orthogonal indicator variables, this is not an undue restriction on the structure of available instruments. Let $C_i = \sum_{j=1}^{J} C_{ij}$ be an indicator variable for the presence of a ith child. Consider a 2SLS regression of Y on X and a full set of indicator variables for the presence of children of varying parity, with X and a full set of interactions between X and Z as instruments. The probability limit of the estimate of the coefficient of C_i is

$$\beta_i = \frac{E\{Y \cdot (E[C_i \mid X, Z] - E[C_i \mid X, E[C \mid X, Z]])\}}{E\{C_i \cdot (E[C_i \mid X, Z] - E[C_i \mid X, E[C \mid X, Z]])\}} \quad (2)$$

$$= \frac{E\{\beta_i(X) \cdot \theta_i(X)\}}{E[\theta_i(X)]} \quad (3)$$

where C is the vector of endogenous variables exclusive of C_i itself,

$$\theta_i(X) = E\{E[C_i \mid X, Z] \cdot (E[C_i \mid X, Z] - E[C_i \mid X, E[C \mid X, Z]]) \mid X\}, \quad (4)$$

and

$$\beta_i(X) = \frac{E\{Y \cdot (E[C_i \mid X, Z] - E[C_i \mid X, E[C \mid X, Z]]) \mid X\}}{E\{C_i \cdot (E[C_i \mid X, Z] - E[C_i \mid X, E[C \mid X, Z]]) \mid X\}} \quad (5)$$

This result is a generalization of Angrist and Imbens's (1995) result for a single variable treatment, and the proof is very similar. Equation (2) is the definition of the 2SLS estimate of β_i with binary instruments. Steps (3) and (5) can be verified by iterating expectations. Step (3) states that the probability limit of the 2SLS estimate of β_i is a weighted average of the $\beta_i(X)$, which are the probability limits of 2SLS estimates in the subpopulations where X is fixed. The weights are proportional to $\theta_j(X)$, the variance of $E[C_i \mid X, Z]$ conditional on the covariates and $E[C \mid X, Z]$. Therefore, subpopulations where the instruments lead to more variance in C_i, after partial effects on C_i through other C have been netted out, receive more weight in β_i.

When Equation (1) characterizes women's labor supply decisions, $\beta_i(X)$ is a weighted average of the marginal effects of ith children of varying ages, where the weights are proportional to the probability that the instruments induce an ith child of that age. Provided that instrument-induced changes in C_i are not systematically related to the ages of the other children – i.e., that

$$E[E[C_{kj} \mid X, Z] \cdot (E[C_i \mid X, Z] - E[C_i \mid X, E[C \mid X, Z]])] = 0 \; \forall \; k \neq i \quad (6)$$

where C_{kj} is an indicator variable for the presence of a kth child of age j, $\beta_i(X)$ can then be expressed

$$\beta_i(X) = \frac{E\left\{\left(\sum_{j=1}^{A} M_{ij} \cdot E[C_{ij} \mid X, Z]\right) \cdot (E[C_i \mid X, Z] - E[C_i \mid X, E[C \mid X, Z]]) \mid X\right\}}{E\{C_i \cdot (E[C_i \mid X, Z] - E[C_i \mid X, E[C \mid X, Z]]) \mid X\}} \tag{7}$$

This is established by substituting for Y in (5) and using (6).[5] Therefore, the coefficient of C_i within the subpopulation is a weighted average of the marginal effects of ith children of varying ages, where the weights are proportional to the probability that the instruments induce the birth of an ith child of that age.

3.2 Empirical Results

Table 4 shows 2SLS estimates of β_i using the 1991 data, with age and education as exogenous covariates and the gender composition of the first two children, the mother's primary language spoken at home and the incidence of twins as instruments. Ordinary least squares (OLS) results are presented by way of comparison. A very small number of women with four children or more were excluded from the sample. The model is fully saturated as described in the preceding section, and there are two choices for the dependent variable: an indicator for labor force participation, and an indicator for employment. The 1991 data is the most suitable for analysis because it is the only year for which all three instruments are present. Although estimation could proceed mechanically with the 1996 and 2001 data, the

[5] An alternate condition to (6) is to assume that the marginal effects of the other children do not depend on their ages, i.e., that $M_{kj} = M_k \ \forall \ k \neq i$. In this case, the expression follows because $\sum_{i=1}^{C} \sum_{j=1}^{A} M_{ij} \cdot C_{ij}$ can be written as $\sum_{i=1}^{C} M_i \cdot C_i$ and $E[E[C_k \mid X, Z] \cdot (E[C_i \mid X, Z] - E[C_i \mid X, E[C \mid X, Z]])] = 0 \ \forall \ k \neq i$ is true by construction. To see this, note that $E[C_i \mid X, Z] - E[C_i \mid X, E[C \mid X, Z]]$ is the residual from a regression of $E[C_i \mid X, Z]$ on X and $E[C \mid X, Z]$, which includes $E[C_k \mid X, Z]$.

approach is questionable because identification of the three marginal effects would rely heavily on interactions between two instruments within which there is already very little variation.

When the second stage of the 2SLS regression is viewed as a structural model of female labor supply – which imposes, in particular, the assumption that the treatment effects in the model are constant and invariant to the covariates – it is possible to test directly whether the instruments are orthogonal to the second stage residuals through an overidentification test. Table 4 reports these test statistics when both labor force participation and actual employment are used as the dependent variable. The null hypothesis of orthogonality is not rejected at any conventional significance level. Under the constant-effects assumption, after conditioning on the covariates, there is no evidence that the instruments influence female labor supply other than through fertility.

An additional concern is that, since a saturated 2SLS strategy requires estimation of a regression model with hundreds of degrees of freedom, small-sample bias may be an issue even when the sample, as the 1991 sample does, comprises nearly 26,000 observations. Following the suggestion of Bound, Jaeger and Baker (1995), Table 5 presents F statistics and partial R-squared values of the identifying instruments for each reduced-form regression in the first stage. In the case of one endogenous variable and conditional homoskedastistic errors, the approximate finite-sample bias of IV relative to OLS is equal to $1/F$. Bound et al. (1995) suggest that empirical researchers should be concerned when the F statistic is close to one, implying that IV may be no less biased than OLS. Staiger and Stock (1997) propose a rule that instruments be deemed "weak" if the F statistic is less than ten, which Stock and Yogo (2002) subsequently interpret as a 5 percent significance test of the hypothesis that the maximum relative bias does not exceed 10 percent. It should be emphasized, however, that in case of multiple endogenous variables, sampling error in the estimation of the remaining reduced-form equations will distort this interpretation of an individual F statistic, so they should be examined with this caveat in mind.

In the case of more than one endogenous variable, Staiger and Stock (1997) show that while it is not possible to evaluate the relative bias of IV directly, it is possible to place an upper bound on it. Their "worst case" relative bias measure, which is equal to the inverse of the minimum squared sample correlation between the endogenous variables and the instruments after the exogenous covariates have been partialled out, multiplied by the ratio of instruments to observations, is equal to 0.272. A reasonable value for the likely bias of OLS, based on the results in Table 4 and other researchers' findings (e.g., Angrist and Evans (1998)), is on the order of 10 percent for the first child and 5 percent for second children and higher. This suggests that the worst-case finite-sample bias of the IV estimates is on the order of 2.7 percent for the causal effect of the first child and 1.4 percent for the causal effects of second and third children. Of course, since conditional homoskedasticity cannot hold in the context of a linear probability model, these statistics can be no more than rough guides to the reliability of the IV estimates presented. Nevertheless, they do contain valuable information and suggest that the potential distortion of the IV estimates due to finite-sample bias is small.

While viewing the second stage as a model of female labor supply is useful in that it provides for specification tests, it does not accommodate innate restrictions on the conditional expectations of the dependent variables, which have binary outcomes. When a saturated 2SLS estimation strategy is used to estimate average causal effects, however, these restrictions are incorporated since the probability limits of the estimates, as weighted averages of individual causal effects, must fall within the smallest convex set containing the causal effects themselves. Therefore, when the dependent variable is binary, the only possible induced changes in the dependent variable are -1, 1, and 0, and the probability limit of the estimates, as a weighted average of these, must fall within negative one and one. If an estimate falls outside this range, then either the instruments are invalid or there are not enough data for it to converge to a reasonable value.

Table 4 indicates that the marginal effect of the first child on female labor supply is far higher than that of a second or third child – while the probability of a woman participating in the labor force drops by 31.0 percent with the birth of a first child, second and third children reduce the probability of participation by only 12.4 percent and 11.8 percent, respectively. The effect on the probability of actual employment is very similar – a first child reduces the probability of employment by 29.0 percent, and the effect of second and third children are 12.1 percent and 11.8 percent.

The latter two figures are very similar in magnitude to existing estimates of the effect of a second or a third child on female labor supply in the United States. For example, Angrist and Evans (1998) find, using data from the 1980 U.S. census, that the 2SLS estimate of the effect of an indicator for having more than two children on paid employment is -12.5 percent when gender composition of the first two children is used as an instrument, and -7.9 percent when the incidence of a multiple second birth is used. They attribute differences in the estimates to age differences in the children induced by the instruments – a third child that is twin to a second child will necessarily be older than a third child born after an additional fertility and pregnancy cycle. Carrasco (2001) analyzes 1986-1989 data from the U.S. Panel Study of Income Dynamics with a switching probit model that endogenizes the fertility decision, using the panel nature of the data to control for unobserved heterogeneity among women in the sample and also incorporating the gender of previous children as an instrument. Her estimate of the average causal effect of an additional child on female labor force participation is -12.9 percent.

Since the estimates represent weighted averages of causal effects, the weights identifying these effects are critical to their interpretation. Figures 3, 4 and 5 show these weights for varying age-education cells in the data. In the calculation of the estimates in Table 4, measuring both age and education in years yielded 275 age-education cells with at least one observation in the data, and hence 275 age-education weights. Step (3) specifies the probability limits of these weights. For clarity of presentation in Figures 3, 4 and 5, however, the weights are aggregated into twenty age-education categories that cover five five-year age groups and four education stages. The weights can be estimated by regressing the predicted value of the endogenous variable in question on the predicted values of the remaining endogenous variables and the exogenous covariates, summing the squared residuals within age-education cells, and then dividing by the total sum of squared residuals.

Figure 3 indicates that most of the identification of the causal effect of a first child is from women 26 and over, particularly those who have completed an upper secondary education but did not continue to university. Figures 4 and 5 show that as the parity of the child increases, women over 30 become more important in the weighting. This makes sense because second and third children are born later in life. Additionally, typical education levels of the women identifying the estimates decline with the parity of the child, which is intuitive because more-educated women tend to have fewer children.

To summarize key features of the distributions shown in Figures 3, 4 and 5, Table 6 shows the average age and education of women identifying each of the marginal effects in Table 4. Differences in age among the three identifying populations turn out to be negligible, but the women identifying the estimates become distinctly less-educated as the parity of the child increases. However, since more-educated women incur greater opportunity costs in leaving the labor force, it is unlikely that this fact drives the finding that the causal effect of fertility declines with child parity.

Another reason that the estimated causal effects of first, second and third children may be different is that the children identifying these effects are drawn from different age distributions. The weights assigned to children of certain ages in this distribution are given by Equation (7). Since these weights are potentially different for each age-education cell, thirteen child ages from zero to twelve and 275 mother's age-education cells yield 3,575 age-age-education weights in total. They can be found by using the instruments and covariates to predict an indicator of the presence of an ith child of a particular age, multiplying the predicted indicator by the residuals from the regression used to calculate the age-education weights, summing the result within age-age-education cells, and dividing by the total sum of squared residuals from that regression.[6]

[6] In the special case of only one child age, this calculation reduces to the calculation of the age-education weights described earlier. In this case, the predicted indicator will differ from the residual by an age-education-specific constant. Since the residuals sum to zero within age-education cells, multiplying the indicator by the residual and summing within age-education cells is equivalent to

In this light, Figures 3, 4 and 5 are marginals of the distribution of age-age-education weights over wide age-education categories and lend insight into the characteristics of the mothers for whom the effects are identified. The marginals of this distribution within child age categories, which lends insight into the characteristics of the children identifying these effects, are shown in Figure 6. One could, of course, easily calculate any other marginal of interest.

Figure 6 indicates that the estimated effect of the first child reflects a child drawn from a younger age distribution than those reflected in the estimated effects of second and third children. In particular, there is a large spike at the age of two, which is exactly what one would expect if a significant portion of the exogenous variation in fertility for the first child is due to women delaying births until the 1988 dragon year. Children born in 1988 were two at the start of 1991, and many would not have reached their third birthday by the time of the census in March.[7] It is likely that differences in the age distributions of children induced by the instruments form a significant part of the reason why the estimated effect of the first child is much more negative than the estimated effects of second and third children.

One characteristic of the age distributions in Figure 6 is that the weights are negative for some children – in particular, they are negative for first eleven and twelve year olds. This occurs because the instruments may affect the timing of children as well as their number. For example, if, after considering the lunar calendar, a woman decides to delay a birth she otherwise would have had by three years, the effect will be to add one two-year-old to the distribution of first children induced by

summing the squared residuals with age-education cells and produces numerically identical age-education weights.

[7] In fact, as the date of the start of the lunar new year was February 17th, many children born before mid-March would not have been dragon children at all.

the instruments and subtract one five-year-old. Therefore, though the density will still sum to one across ages, it may not be positive for all of them, and the fact that it is actually negative, even slightly, for eleven- and twelve-year-olds suggests that birth timing is quite prevalent.

4 Conclusions

Economic models of household labor supply indicate that, if there are economies of scale in raising children, the causal effect of fertility on female labor supply should decline with child parity. A convincing test of this proposition, however, requires instruments capable of identifying the causal effect of the first child, which have proven elusive in practice. The results reported here are unique in that variation in fertility due to the Chinese lunar calendar affects the number of children born of any parity, permitting such identification.

The IV estimates presented are consistent with the idea that first children have a much greater causal effect on female labor supply than higher-parity children, though a partial explanation of this result may well be the fact that very young children received considerably more weight in the estimate of the first child's effect than in those of second and third children. An additional advantage of the IV procedure used is that interpretation of the estimates as average causal effects does not hinge on the correctness of a particular model of female labor supply. The weighting formulas presented extend Angrist and Imbens's (1995) result interpreting 2SLS estimates as weighted average causal effects to the case of multiple endogenous variables, and are helpful in interpreting estimated treatment effects when the intensity of the treatment varies along more than one dimension – in this case, children that vary in both age and number. Accordingly, the weights in Figure 6 formalize the intuition that if a significant part of the lunar calendar effect is to cause women to time their births for the 1988 dragon year, much of the identification of the causal effect of the first child should be from two-year-olds when 1991 data are used.

Finally, although the results were estimated from Hong Kong data, there is a surprising similarity between these results and those other researchers have obtained for the United States (e.g., Angrist and Evans (1998); Carrasco (2001)). This suggests that the labor supply behavior of Hong Kong women may not be all that different from the behavior of women in other developed countries, and the results here may be of wider relevance to the U.S. and other places outside Hong Kong.

References

Angrist, Joshua D., "Grouped Data Estimation and Testing in Simple Labor-Supply Models," *Journal of Econometrics*, February-March 1991, *47* (2-3), 243–66.

———, "Estimation of Limited Dependent Variable Models with Dummy Endogenous Regressors: Simple Strategies for Empirical Practice," *Journal of Business and Economic Statistics*, January 2001, *19* (1), 2–16.

——— **and Guido W. Imbens**, "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity," *Journal of the American Statistical Association*, June 1995, *90* (430), 431–442.

——— **and William N. Evans**, "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size," *The American Economic*

Review, June 1998, *88* (3), 450–477.

—— **and** ——, "Schooling and Labor Market Consequences of the 1970 State Abortion Reforms," in Solomon W. Polachek, ed., *Research in Labor Economics*, Stamford, Conn.: JAI Press, 1999, pp. 75–113. Volume 18.

Bound, John, David A. Jaeger, and Regina M. Baker, "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak," *Journal of the American Statistical Association*, June 1995, *90* (430), 443–450.

Bronars, Stephen G. and Jeff Grogger, "The Economic Consequences of Unwed Motherhood: Using Twin Births as a Natural Experiment," *The American Economic Review*, 1994, *84* (5), 1141–1156.

Carrasco, Raquel, "Binary Choice with Binary Endogenous Regressors in Panel Data: Estimating the Effect of Fertility on Female Labor Force Participation," *Journal of Business and Economic Statistics*, October 2001, *19* (4), 385–394.

Coleman, Mary T. and John Pencavel, "Trends in Market Work Behavior of Women Since 1940," *Industrial and Labor Relations Review*, July 1993, *46* (4), 653–76.

Goodkind, Daniel M., "New Zodiacal Influences on Chinese Family Formation: Taiwan, 1976," *Demography*, May 1993, *30* (2), 127–142.

Gronau, Reuben, "Sex-Related Wage Differentials and Women's Interrupted Labor Careers – The Chicken or the Egg," *Journal of Labor Economics*, July 1988, *6* (3), 277–301.

Hotz, V. Joseph and Robert A. Miller, "An Empirical Analysis of Life Cycle Fertility and Female Labor Supply," *Econometrica*, January 1988, *56* (1), 91–118.

Imaizumi, Y., "A Comparative Study of Twinning and Triplet Rates in 17 Countries, 1972-1996," *Acta Geneticae Medicae et Gemellologiae*, 1998, *47* (2), 101–14.

Mincer, Jacob, "Labor Force Participation of Married Women," in H. G. Lewis, ed., *Aspects of Labor Economics*, Princeton University Press Princeton, N.J. 1962.

Moffitt, Robert, "Profiles of Fertility, Labour Supply and Wages of Married Women : A Complete Life-Cycle Model," *The Review of Economic Studies*, April 1984, *51* (2), 263–278.

Moi, Patsy, "Dragon Brings Baby Boom; 3,000 More Born as Birthrate Rises for the First Time in Six Years," *South China Morning Post*, February 26 2001, p. 2.

Podmore, David and David Chaney, "Family Norms in a Rapidly Industrializing Society: Hong Kong," *Journal of Marriage and the Family*, May 1974, *36* (2), 400–407.

Rubin, Donald B., "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology*, October 1974, *66* (5), 688–701.

Staiger, Douglas and James H. Stock, "Instrumental Variables Regression with Weak Instruments," *Econometrica*, May 1997, *65* (3), 557–586.

Stock, James H. and Motohiro Yogo, "Testing for Weak Instruments in Linear IV Regression," October 2002. NBER Technical Working Paper No. 284.

Sun, Te-Hsiung, Hui-Sheng Lin, and Ronald Freedman, "Trends in Fertility, Family Size Preferences, and Family Planning Practice: Taiwan, 1961-76," *Studies in Family Planning*, April 1978, *9* (4), 54–70.

Westoff, Charles F., Robert G. Potter, and Philip C. Sagi, *The Third Child : A Study in the Prediction of Fertility*, Princeton, N.J.: Princeton University Press, 1963.

Wong, S. F. and L. C. Ho, "Sex Selection in Practice Among Hong Kong Chinese," *Social Science and Medicine*, August 2001, *53* (3), 393–97.

Yip, Paul S. F., Joseph Lee, and Y. B. Cheung, "The Influence of the Chinese Zodiac on Fertility in Hong Kong SAR," *Social Science and Medicine*, November 2002, *55* (10), 1803–1812.

Figure 1: Live Births in Hong Kong, Selected Years

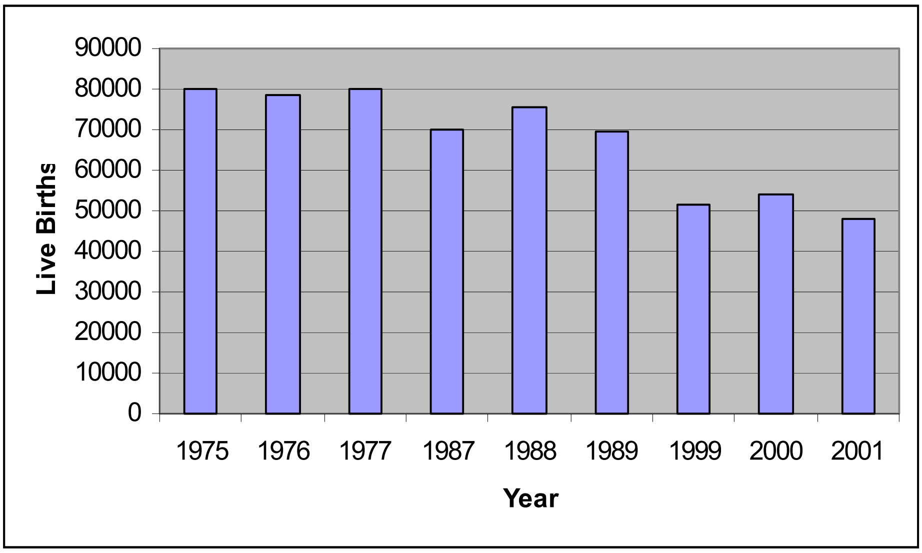

Figure 2: Relative Probability of Childbirth, by Zodiac Year (Dragon = 1)

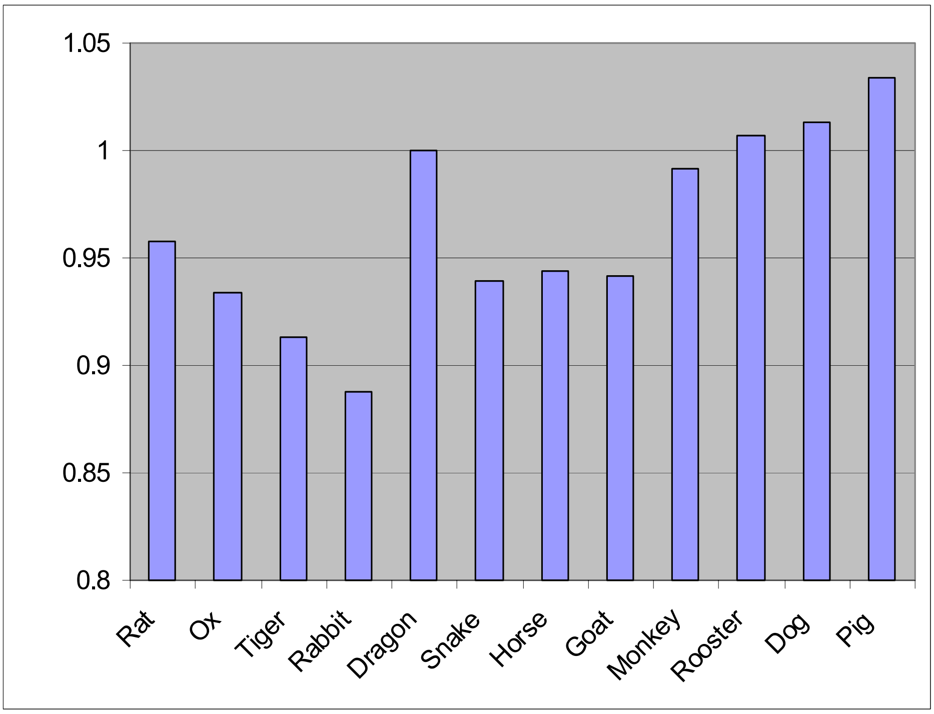

Figure 3: Weights, by Age and Education, Identifying the Causal Effect of a First Child

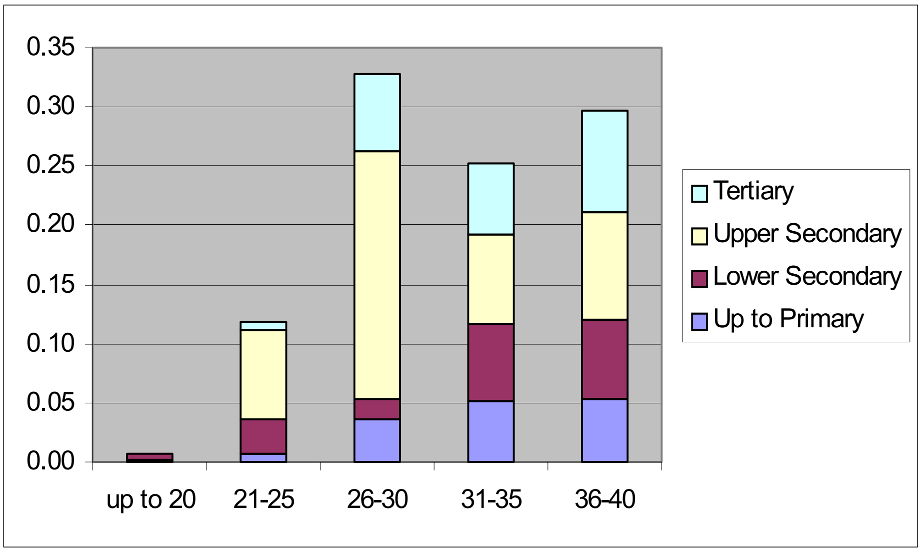

Figure 4: Weights, by Age and Education, Identifying the Causal Effect of a Second Child

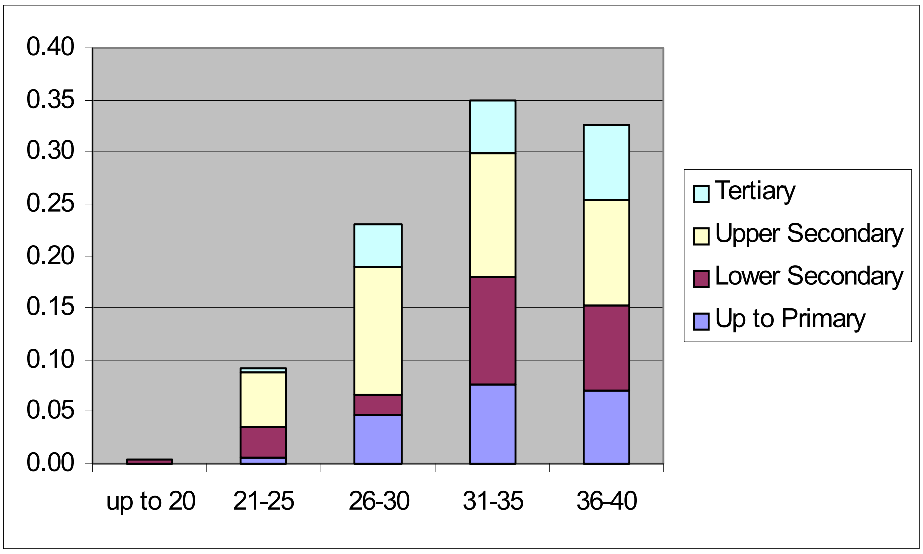

Figure 5: Weights, by Age and Education, Identifying the Causal Effect of a Third Child

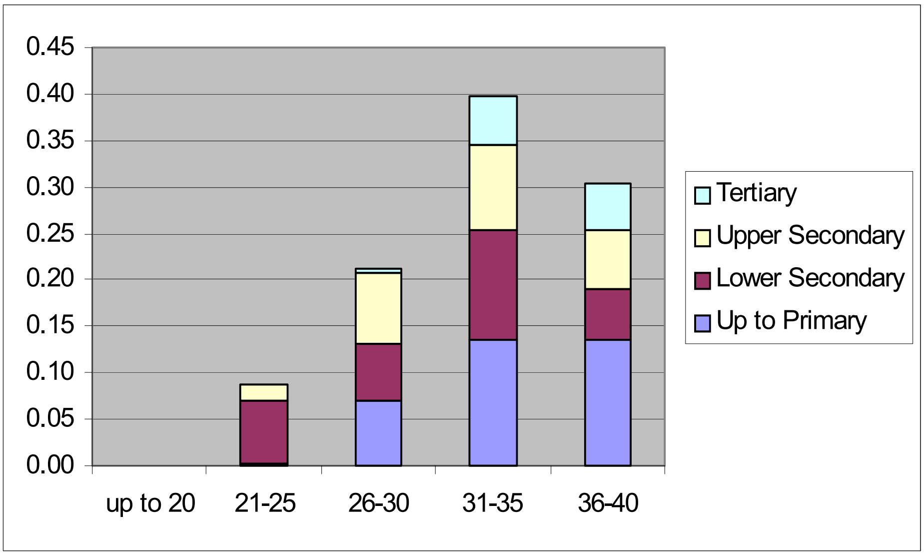

Figure 6: Age Distribution of Children Identifying Parity-Specific Causal Effects

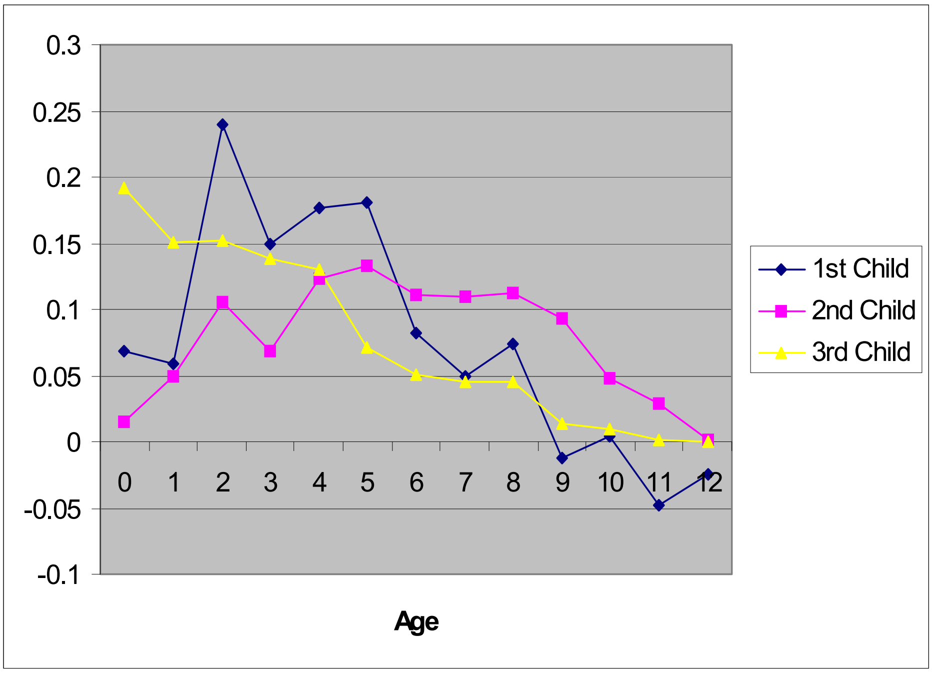

Table 1: Live Births in Hong Kong, by Year

Year	Live Births	Year	Live Births
1971	79,789	1987	69,958
1972	80,344	1988	75,412
1973	82,252	1989	69,621
1974	83,581	1990	67,731
1975	79,790	1991	68,281
1976	78,511	1992	70,949
1977	80,022	1993	70,451
1978	80,957	1994	71,646
1979	81,975	1995	68,637
1980	85,290	1996	63,291
1981	86,751	1997	59,250
1982	86,120	1998	52,997
1983	83,293	1999	51,281
1984	77,297	2000	54,134
1985	76,126	2001	48,219
1986	71,620	2002	48,200

Source: Hong Kong Census and Statistics Department. # indicates a provisional figure.

Table 2: Parameter Estimates of a Cox Proportional Hazards Model of Childbirth over Time, with Zodiac Dummy Variables

Variable	I	II
Years of Education	-0.086 (0.001)	-0.074 (0.003)
Rat	-0.044 (0.015)	0.081 (0.038)
Ox	-0.068 (0.015)	0.072 (0.038)
Tiger	-0.091 (0.015)	0.063 (0.037)
Rabbit	-0.120 (0.015)	-0.023 (0.038)
Snake	-0.063 (0.015)	0.039 (0.038)
Horse	-0.058 (0.015)	0.050 (0.037)
Goat	-0.060 (0.015)	0.141 (0.038)
Monkey	-0.009 (0.015)	0.180 (0.037)
Rooster	0.007 (0.015)	0.149 (0.037)
Dog	0.013 (0.014)	0.101 (0.037)
Pig	0.033 (0.009)	0.053 (0.037)
Rat · Years of Education	–	-0.014 (0.004)
Ox · Years of Education	–	-0.016 (0.004)
Tiger · Years of Education	–	-0.018 (0.004)
Rabbit · Years of Education	–	-0.011 (0.004)
Snake · Years of Education	–	-0.012 (0.004)
Horse · Years of Education	–	-0.012 (0.004)
Goat · Years of Education	–	-0.024 (0.004)
Monkey · Years of Education	–	-0.022 (0.004)
Rooster · Years of Education	–	-0.016 (0.004)
Dog · Years of Education	–	-0.010 (0.004)
Pig · Years of Education	–	-0.002 (0.004)

Standard errors are in parentheses. The models in column I and II also include interaction terms between the zodiac variables and an indicator for speaking a non-Chinese language as the primary language at home. The model in column II additionally includes interactions between this indicator and the zodiac-education interaction terms. The hypothesis that the zodiac exerts no influence on the fertility of non-Chinese cannot be rejected for either model ($p = 0.45$ and $p = 0.16$, respectively).

Table 3: Fraction of Women in Sample with Additional Children, by Gender of Initial Children

	Fraction with First Child		
	1991	1996	2001
Chinese Women	0.696 (0.003)	0.643 (0.003)	0.561 (0.003)
Observations	25,973	28,124	26,387
Non-Chinese Women	0.618 (0.028)	0.516 (0.020)	0.512 (0.020)
Observations	293	645	606
	Fraction with Second Child*		
	1991	1996	2001
Gender of first child			
Chinese Women			
Boy	0.561 (0.005)	0.495 (0.005)	0.457 (0.006)
Girl	0.560 (0.005)	0.493 (0.006)	0.445 (0.006)
Observations	18,087	18,090	14,797
Non-Chinese Women			
Boy	0.592 (0.050)	0.520 (0.038)	0.500 (0.039)
Girl	0.518 (0.054)	0.500 (0.040)	0.496 (0.042)
Observations	181	333	310
	Fraction with Third Child**		
	1991	1996	2001
Gender of first two children			
Chinese Women			
Two Boys	0.205 (0.008)	0.164 (0.008)	0.118 (0.008)
One Boy, One Girl	0.235 (0.006)	0.179 (0.006)	0.129 (0.006)
Two Girls	0.257 (0.009)	0.227 (0.009)	0.169 (0.009)
Observations	10,145	8,969	6,706
Non-Chinese Women			
Two Boys	0.344 (0.081)	0.333 (0.061)	0.146 (0.060)
One Boy, One Girl	0.245 (0.063)	0.152 (0.049)	0.173 (0.043)
Two Girls	0.375 (0.115)	0.359 (0.070)	0.258 (0.070)
Observations	101	170	154

Standard errors are in parentheses.

* Conditional on having at least one child.

** Conditional on having at least two children.

Table 4: Ordinary Least Squares and Instrumental Variables Estimates of the Marginal Effects of the First, Second and Third Child on Female Labor Supply

	Measure of Labor Supply			
	Labor Force Participation		Employment	
	OLS	IV	OLS	IV
First Child	-0.230 (0.007)	-0.310 (0.068)	-0.220 (0.007)	-0.290 (0.069)
Second Child	-0.168 (0.007)	-0.124 (0.033)	-0.161 (0.007)	-0.121 (0.034)
Third Child	-0.055 (0.011)	-0.118 (0.050)	-0.056 (0.011)	-0.118 (0.050)
$\chi^2\,(df)$	–	562.1 (530)	–	560.4 (530)
p	–	0.162	–	0.175

Table 5: First Stage F Statistics and Partial R-Squared Values of the Identifying Instruments in the Reduced-Form Equations for First, Second and Third Children

Endogenous Variable	F	Partial R^2
First Child	10.36	0.259
Second Child	23.57	0.443
Third Child	4.52	0.132

Table 6: Average Age and Education of Women Identifying the Marginal Effects of First, Second and Third Children

	Average Age	Average Education
First Child	31.5	10.5
Second Child	32.5	9.8
Third Child	32.7	8.0