Case A, Wednesday.

April 18, 2012

1 Introduction

Adverse birth outcomes have large costs with respect to direct medical costs aswell as long-term developmental consequences. Maternal health behaviors at conception and during pregnancy are important determinants of fetal growth and child development. Maternal smoking is one of the most commonly studied behavioral risk factors that affect fetal/child development and is often considered the single most important, modifiable factor affecting birth outcomes, see Kramer (1987). One way of measuring a birth outcome is the birthweight that is considered as a leading indicator of infant health. For instance it has been shown that babies with low birthweight have higher mortality rates, are more likely to have cognition and attention problems, and are more likely to be unemployed and earn lower wages later in life.

The literature on evaluting the effects of smoking on birthweight is vast and several different statistical and econometric methods have been applied. Endogenous maternal selection into smoking and biased reporting of smoking behaviors complicate the estimation of the causal effects of smoking on birth outcomes. Specifically, mothers who smoke during pregnancy are also likely to self-select into smoking based on their preferences for health and risk taking and their perceptions of fetal health endowments. These factors, typically unobserved in available data samples, are related to fetal health through other pathways besides smoking. For example, women who smoke during pregnancy may adopt other unhealthy behaviors that may also have adverse effects on the fetus (e.g. poorer nutrition or reduced prenatal care), but may also be less likely to have a family history of poor birth outcomes. In light of these endogeneity issues, IV estimation methods have been applied in several analyses, see e.g. Permutt and Hebel (1989)and Evans and Ringel (1999). Studies applying IV estimation generally suffers from recovering a good instrument. For instance Evans and Ringel (1999) apply the state tax on tobacco as an instrument. This variable suffers from low variation, and the authors find that that the estimated effect of smoking on birthweight is larger (although not significantly) when applying IV estimation (350-600g) compared to OLS (200-250g). Issues regarding identification will be discussed in the next section.

As an alternative to IV estimation Abrevaya (2006) applies fixed effects estimation on panel data. This approach allows for the identification of the smoking effect of women who change their

smoking behaviour from one pregnancy to another, and unobserved heterogeneity can be controlled for. The drawback of this approach is that the identification strategy imlies that the variation in data used to estimate the effect of smoking on birth weight only stems from mothers who have changed their smoking behavior between different births. It does not provide an estimate of the effect of smoking on birth weight for life-time-smoking mothers, and we thus suspect that the estimated effect only is a lower bound of the true effect for the full population. Abrevaya (2006) estimates the effect to be 100-150g.

Quantile regression is yet another econometric method previously used in the literature on birthweight. This method is motivated by the fact that both social and health costs associated with birthweight have been found to exist primarily in the low end of the birthweight distribution. In contrast to other methods, where low birthweight is based on a common unconditional threshold for low birthweight for the entire sample, quantile regression focuses on a particularly chosen quantile of the conditional birthweight distribution. As a result, the quantile regression is a convenient method for determining how different factors affect birthweight at different parts of the distribution. Abrevaya and Dahl (2008) estimates that maternal smoking reduces birthweight by around 160-185g between the 10% and 90% quantiles, based on cross-sectional estimation, whereas the effect is found to be around 50-80g when applying fixed effects estimation. Moreover, there is no significant effect on the 10% quantile.

Thus, what characterizes all the studies mentioned above in general, is that maternal smoking seem to have a statistically significant negative effect on birthweight. IV estimation tends to find a larger negative effect of smoking whereas panel data methods uncover a smaller negative effect. The results from the quantile regression results are not directly, only qualitatively, comparable to IV and panel data results.

The next section discusses the problem of identifying the effect of smoking on birthweight. Section 3 describes the data and variable selection in addition to present a summary statistics to motivate the empirical analysis. The applied methodology is explained in section 4. The results are presented and discussed in section 5 and finally section 6 concludes.

2 Identification

Unobserved heterogeneity is the main issue of identifying the effect of morthers' smoking behavior on birthweight. That is, unobservables that affect birthweight are likely to be correlated with the mothers decision to smoke, but are not possible to controlled for. In example, mothers who smoke are more likely to adopt other behaviors such as drinking and poor nutrition, which may be expected to have an additional effect on birthweight. Given the data provided, traditional econometric methods relying on panel data to control for unobserved heterogenity are not feasible. Abrevaya (2006) creates a pseudo-panel by matching on observables but this approach levies all identification on matching and given that the information on many health-related variables is not available

we are skeptical about this approach. Instead one may consider instrument variable methods. However, no valid instruments are present in our data set. In order for an instrument to be valid it should only be correlated with birthweight through its correlation with smoking. Previous studies have used instruments such as state level tax on cigarettes, maternal and paternal schooling or genetic risk factors for smoking as instruments, see Evans and Ringel (1999), Mullahy (1997), and Wehby, Fletcher, Lehrer, Moreno, Murray, Wilcox, and Lie (2011), respectively. Variables representing the parents level of education are present in our data set. However, we do not consider these variables as valid instruments to determine the causal effect of smoking behavior on birthweight. It is well reconized that the level of education is correlated health specific choices, e.g., drinking, exercising, drug using and nutrion, which are unobserved. As a result, such variables cannot be used as valid instruments because they will not only correlate with birthweight through smoking, but also through the unobservables and hence the exclusion restriction is violated in a instrumental variable regression using parental education as an instrument for smoking.

As a consequence of the lack of panel data and of valid instruments, the present study only deals with estimation of correlations. From our point of view the available dataset is insufficient in order to recover a causal effect.

3 Data

The data is from the Nalality Data Sets from 1990 to 1998 and based on births records from every live birth that occurs in the United States. It contains 141,929 observations and is assumed to be representive of the population of interest.

3.1 Variable selection

We include the age of the mother as well as the age squared as it is well known that for older mothers, several complications may arise. We choose to include education dummies rather than years of education since we believe that completion of an educational level says more about the mother's type than simply the number of years. Also, including education dummies allows for non-linear effects of education. We include dummies for year of birth to take into account year fixed effects (for example, the average number of cigarettes smoked falls over the years). The region of the child's birth is included but not that of the mother's birth. Both variables are potentially relevant but in order to keep the analysis simple, we choose to only control for one and we believe that the current environment is more important than the past. Dummies for gender and race are included as proxies for other socio-economic variables that are not available and which may be relevant for healthiness and other relevant aspects of the mother's type. Marital status is included as it is relevant for the resources the mother has. Dummies for birth order 2, 3, 4 and 5 or above are included to capture parity fixed effects. All the relevant dummies for prenatal care and visits are included.

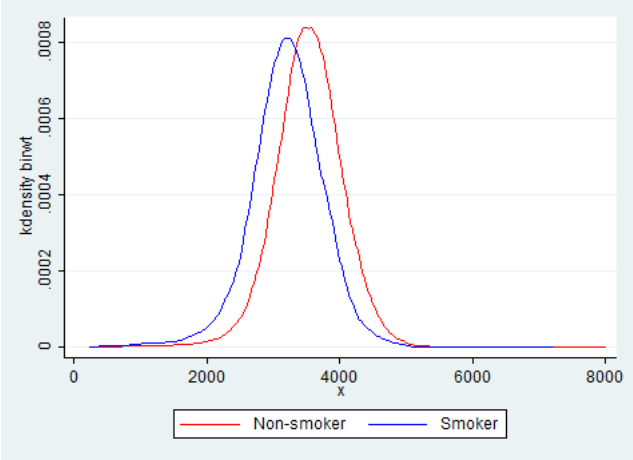
Table 1: Summary statistics

	All		Non-smokers		Smokers	
	mean	sd	mean	sd	mean	sd
birwt	3483.61	538.91	3525.05	523.25	3201.46	558.94
smoker	0.13	0.33				
cigs	1.65	5.13			12.85	7.85
age	29.62	5.37	30.01	5.24	26.98	5.47
hs	0.29	0.46	0.27	0.44	0.48	0.50
sc	0.23	0.42	0.24	0.43	0.17	0.37
cg	0.38	0.49	0.43	0.49	0.05	0.23
married	0.87	0.34	0.91	0.29	0.62	0.49
black	0.07	0.26	0.07	0.26	0.09	0.29
male	0.51	0.50	0.51	0.50	0.51	0.50
border2	0.25	0.43	0.25	0.43	0.30	0.46
border3	0.10	0.30	0.09	0.29	0.13	0.34
border4	0.03	0.18	0.03	0.18	0.05	0.21
border5	0.03	0.16	0.03	0.16	0.03	0.17
gest	39.17	2.13	39.19	2.05	38.98	2.58
nopnv	0.01	0.09	0.01	0.07	0.02	0.15
pnv2t	0.11	0.32	0.10	0.30	0.21	0.41
pnv3t	0.02	0.14	0.01	0.12	0.05	0.22
pnc_inter	0.17	0.38	0.16	0.36	0.27	0.44
pnc_inad	0.04	0.20	0.03	0.17	0.11	0.31
N	141175		123093		18082	

The variable `cigs` indicates how many cigarettes on average was smoked per day. The `smoke` variable is a dummy indicating whether `cigs` is non-zero. We choose to include both variables since the `smoke` dummy gives the effect on the extensive margin (the effect of smoking vs. not smoking) and `cigs` gives the effect on the intensive margin (the effect from the amount of smoking). This is in line with the analysis of Abrevaya (2006). 754 individuals have `cigs` missing and `smoke` equal to one meaning that these are smokers that have not reported how much they smoke. An issue could be that heavy smokers would be overrepresented among the individuals that cease to report due to bad conscience. In Appendix A we compare the group of smokers who report the number of cigarettes to the smokers who do not report, and we conclude that we do not face a selection issue when disregarding the smokers who do not report the number of cigarettes they smoke.

Table 1 contains the mean and standard deviation of the variables we have decided to use in our analysis. The figures are calculated for the entire dataset and for the two sub-datasets containing the observations for non-smokers and smokers. We first notice that the mean birthweight for the non-smokers is 3,525g and 3,201g for the smokers. Moreover, non-smokers seem to have a higher

Figure 1: Kernel density estimate of the infints' birth weight for smokers and non-smokers.



level of education, are more likely to be older, white, and married compared to smokers. Based on the birth dummies (`border2-border5`) smokers get more children. The variables for prenental visits indicate that non-smokers tend to make their first prenental visit later in the pregnancy and are more likely to receive inadequate prenatal care, suggesting that non-smokers may care less about the health status of the pregnancy. Moreover, smokers are more likely to receive intermediate prenantal care which indicates that complications in the pregnancy are more likely to appear for smokers.

Overall, smokers and non-smokers appear to have different characteristics with respect to variables that may influence the birthweight. For instance smokers are less educated which is known to be an important determinant of health-related behaviour. This observation speaks for that level of education is a poor instrument for isolating the effect of smoking from other health issues. Instead, the obvious differences in the two sub-datasets indicate that a regression with smoking dummy interaction terms may be a useful method to estimate the effect of maternal smoking on birthweight.

4 Methodology

In the describtion of the relationship between smoking behaviour and the distribution of the infants' birth weight, we consider two main issues; the overall level of the birth weight in the groups of smokers and non-smokers, and shape effects in the lower part of the distribution due to smoking behaviour. This choice os motivated from Figure 1, which shows a density plot of the birth weight seperated into the the group of smokers and non-smokers. It is oberved the level is different between the two groups and also that lower tail of the distribution for the smokers appears slightly thicker than for the non-smokers. However, the relation is only conditinal on smoking behavior, and many other variables may have an important impact on the distribution. The rest of this paper analyses such relation in more detail.

We benchmark our results with a least square regression similar to Abrevaya and Dahl (2008). The benchmark model is then extended with a second degree polynomial in average consumption of cigarettes. By doing so, we try to seperate out effects on the intensive and extensive margin. The inclusion of the number of consumed cigarettes complicates to some extend the interpretation of the coeffitions to the smoking dummy as a treatment effect. Still, we believe important information is captured by the intensive margin of cigarette consumption and choose this approach.

Next, the gestation variable is added to the regression to account for the birthweight effects directly associated with gestation. We do so with the reservation that including gestation might blur the policy relevant effect from smoking behaviour on the distribution of infints' birth weight, as the policy relevant measure consists of both direct and indirect effects. However, as a causal effect from smoking is not attainable with the present data, we choose instead a more rich description of the conditional destribution of the infants' birth weight. The importance of the gestation variable in relation to smoking is also analysed in the appendix of Abrevaya (2006).

If the overall level effect from smoking on infants' birth weight is not the same over the whole conditional distribution, the estimates of the least squares regressions could be misleading for the overall level effects. In addition, as the many of the economically relevant consequences of infants' birth weight are related to the lower tail of the distribution, a more focused modelling of that part of the distribution is in itself of interest. As a consequence, we make quantiles regressions for the quantiles 1, 10, 25, 50, 75, and 90. Our main interest is wether the relation between smoking behaviour and infants' birthweight changes over the quantiles of the conditional distribution. Again, we consider the consequences of including the gestation variable in the regressions.

Finally, we consider interaction effects between smoking and control variables. This is done only for least squares regression with the aim to capture the different marginal correlations from smoking behaviour conditionally on the other controlling variables. This can be seen as another way of modelling changing relations between smoking behaviour and the birthweight over the conditional distribution.

5 Statistical analysis

5.1 OLS

Table 2 shows three OLS regressions. Before interpreting them, we note that they suffer from several misspecifications the most important being the following two. Firstly, we omit unobserved mother-specific heterogeneity which we are unable to handle since we do not have panel data. Secondly, we assume linearity, that is, we assume that the effect of smoking is homogenous. This we will deal with later.

The Column (1) of table 2 is the benchmark OLS estimation which shows an estimate of the effect of smoking on birth weight of -269.8 which is very close to the -250 typically found in the

literature. In column (2) we control for the number of cigarettes and this reduces the coefficient on smoking to -172.1. Further controlling for gestation reduces the estimate down to -147.8. For both cigarettes and gestation we note that the effect on the birth weight is highly nonlinear. We have tried with longer polynomials and found significant effects indicating more complicated effects but we believe that we have captured the main effects with the quadratic function. The results are in line with those found by Abrevaya (2006, Table IV) in the regressions where the number of cigarettes included. We note that the fixed effects estimates found there are substantially lower in magnitude (around 60–80g) and we expect that we would have found the same if we had had panel data available.

The OLS results tells us how the treatment (smoking) works at the mean. That is, it tells the story about the shifting mean in the birth weight distribution that we reported earlier. As Abrevaya and Dahl (2008) point out, however, much of the interest is in what happens with the low birth weight children. In other words, we are interested in whether there are heterogenous treatment effects of smoking on birth weight, and in particular we want to investigate the role of smoking in the story about the fat left tail that we saw in Figure 1.

5.2 Quantile regression

In order to investigate the tail of the conditional distribution of birth weight, we employ a quantile regression approach. In Abrevaya and Dahl (2008), the 10 (conditional) quantile is the lowest considered but given the large dataset available, we choose to emphasize the tail even further by using the 1 quantile instead, as argued earlier. The results are shown in table 3. In the following we focus our attention to the coefficient on `smoker`.

We first compare our results to those of Abrevaya and Dahl (2008). We see that our estimates are almost the same for the 10, 25 and 50 quantiles but lower for 75 and 90. We will not go into detail with the latter two but note that since we have included cigarettes which they have not, this may be part of why. For the 1 quantile, we see a surprisingly sizeable coefficient of -321.9 (s.e. 83.92). An interesting hypothesis to test is whether the effects are significantly different across quantiles. However, this requires a simultaneous estimation of the quantiles to take cross-correlations properly into account. We have attempted this using Stata but found that running on 10% of the sample would take more than 2 hours and since our interest is in the tails of the distribution, further subsampling would leave us with too few observations to make inference. We conclude informally that while it may not be strongly different from the other coefficients, the point estimate is twice as large and the very imprecisely estimated.

In order to satisfactorily assess the impact of smoking on the first quantile, we depart slightly from the literature by including gestation as an explanatory variable (as we did in the OLS analysis). As before, we expect this to isolate part of the effect of smoking and separate it from the direct effect of smoking. The results are shown in table 4. Most importantly, the coefficient on smoking drops

Table 2: OLS regressions

	(1) birwt	(2) birwt	(3) birwt
smoker	-269.8***	-172.1***	-147.8***
	(4.409)	(10.02)	(8.946)
cigs		-9.871***	-9.745***
		(1.137)	(1.016)
cigssq		0.124***	0.119***
		(0.0293)	(0.0262)
gest			606.1***
			(6.988)
gestsq			-6.678***
			(0.0920)
age	24.07***	24.93***	16.23***
	(2.581)	(2.581)	(2.305)
agesq	-0.367***	-0.379***	-0.216***
	(0.0424)	(0.0424)	(0.0379)
hs	64.96***	62.18***	54.73***
	(5.679)	(5.681)	(5.073)
sc	96.61***	93.05***	82.73***
	(6.266)	(6.270)	(5.599)
cg	106.6***	102.9***	85.78***
	(6.574)	(6.579)	(5.875)
black	-276.7***	-280.6***	-200.7***
	(5.490)	(5.496)	(4.926)
male	130.8***	130.9***	142.1***
	(2.726)	(2.725)	(2.434)
nopnv	-33.63	-31.21	43.98*
	(20.25)	(20.24)	(18.09)
pnv2t	74.15***	74.23***	41.69***
	(6.605)	(6.602)	(5.904)
pnv3t	141.0***	140.8***	73.02***
	(16.09)	(16.08)	(14.37)
pnc_inter	-80.60***	-80.30***	-49.97***
	(5.462)	(5.459)	(4.878)
pnc_inad	-163.1***	-162.1***	-102.3***
	(12.96)	(12.96)	(11.57)
border2	17.82***	18.53***	21.98***
	(3.322)	(3.321)	(2.966)
border3	32.51***	33.73***	38.99***
	(4.866)	(4.865)	(4.344)
border4	51.92***	53.51***	57.17***
	(7.904)	(7.902)	(7.056)
border5	139.1***	139.5***	116.1***
	(8.969)	(8.964)	(8.006)
N	141175	141175	141175

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We also control for region and year fixed effects.

`borderX` is a dummy for birth order (parity) `X`.

substantially in the first quantile. It also drops in all the other quantiles but it drops disproportionately more in the first quantile. Again, we stress that we are not formally testing the coefficients different from each other (due to time constraints).

This tells us that the fatter tail in Figure 1 for the smokers is more due to the channel that goes through the shorter gestation than the direct effect on birth weight. Hence, once gestation is controlled for, the effect of smoking appears to be much more homogenous than table 3 first indicated. This suggests that an OLS approach which appropriately handles complicated interaction effects between smoking and gestation might be able to tell the mean story accurately without suffering from bias from heterogenous effects in the lower tail.

5.3 OLS with interaction effects

In table 5 we present OLS estimates from a model where we interact the smoking dummy with all the demeaned covariates (except regional and year dummies). The quantile regression allows for differential associations between birthweight and smoking across the conditional distribution of birthweight. In this model, differential associations between birthweight and smoking are captured by the interactions of smoking with the demeaned covariates in this model. The estimates for the interaction terms in table 5 has the prefix `smoker_`.

First, we compare the estimate for the correlation between birthweight and smoking on the extensive margin between the quantile regressions (including gestation) and the estimate from the interaction model. The estimate from the latter regression is within two standard errors of the point estimates for the 0.10th to the 0.9th quantile in table 4, however we cannot formally test whether they are statistically different. The estimate for the 0.01th quantile is 31 grams higher than the estimate from the interacted model, but also has a very high standard error. Thus, using the standard error from the quantile regression, the estimate from the interacted model is within two standard errors of the 0.01th quantile as well.

Turning to the estimates for the interaction terms, we see that there does not appear to be a differential association between smoking and birthweight across the age of the mothers, their education, and ethnicity. Neither is there a differential association for girl- and boy-infants. However, the correlation between smoking and birthweight *given gestation* varies across the lenght of the gestation. The estimated partial effect of gestation on birthweight shows that a child born in week 40 instead of week 39 is 80 grams heavier if the mother is a non-smoker. If the mother is a smoker the child is instead only 73 grams heavier. Hence, it appears that smoking stunt the growth of the children, though we should keep in mind that this results cannot be interpreted as a causal impact.

We don't see very significant results for the interaction terms with the prenatal care dummies apart from the dummy indicating inadequate prenatal care. The latter estimate suggests that the negative impact of inadequate prenatal care on the birthweight is doubbled for smoking mothers compared with non-smoking mothers. The birthorder effects for non-smoking mothers are posi-

Table 3: Quantile regression — without gestation

	(1)	(2)	(3)	(4)	(5)	(6)
	1%	10%	25%	50%	75%	90%
smoker	-321.9***	-181.9***	-177.0***	-175.5***	-157.2***	-129.9***
	(83.92)	(17.53)	(12.68)	(11.24)	(13.25)	(15.39)
cigs	-5.221	-9.227***	-9.509***	-8.560***	-10.92***	-13.72***
	(8.401)	(1.903)	(1.473)	(1.242)	(1.418)	(1.593)
cigssq	0.167	0.104*	0.111**	0.0905**	0.156***	0.234***
	(0.183)	(0.0479)	(0.0388)	(0.0312)	(0.0338)	(0.0352)
age	132.7***	41.81***	26.40***	20.21***	15.10***	14.01***
	(24.79)	(4.725)	(3.240)	(2.949)	(3.471)	(4.019)
agesq	-2.330***	-0.694***	-0.407***	-0.297***	-0.200***	-0.163*
	(0.405)	(0.0773)	(0.0532)	(0.0484)	(0.0571)	(0.0660)
hs	101.4	73.88***	63.05***	57.91***	64.42***	57.23***
	(54.02)	(10.46)	(7.152)	(6.491)	(7.638)	(8.844)
sc	181.9**	115.7***	98.15***	90.72***	85.51***	74.04***
	(59.72)	(11.56)	(7.883)	(7.164)	(8.458)	(9.795)
cg	268.2***	143.7***	113.5***	99.66***	85.12***	66.08***
	(63.09)	(12.08)	(8.254)	(7.517)	(8.878)	(10.27)
black	-788.8***	-306.3***	-265.3***	-260.3***	-254.1***	-249.5***
	(52.39)	(9.959)	(6.843)	(6.281)	(7.457)	(8.700)
male	61.15*	106.1***	120.2***	134.7***	148.4***	158.3***
	(25.51)	(4.966)	(3.413)	(3.114)	(3.676)	(4.257)
nopnv	-256.4	-158.5***	6.580	-14.64	34.85	85.05**
	(184.8)	(36.59)	(25.24)	(23.13)	(27.37)	(31.79)
pnv2t	319.7***	110.5***	68.94***	63.72***	55.39***	55.81***
	(62.63)	(12.08)	(8.286)	(7.544)	(8.905)	(10.31)
pnv3t	629.2***	259.9***	179.9***	107.9***	75.95***	53.08*
	(147.2)	(29.07)	(20.03)	(18.38)	(21.81)	(25.32)
pnc_inter	-332.1***	-125.3***	-80.26***	-67.22***	-54.69***	-56.68***
	(52.56)	(10.04)	(6.878)	(6.238)	(7.319)	(8.431)
pnc_inad	-470.6***	-269.5***	-196.7***	-126.5***	-113.9***	-95.73***
	(118.0)	(23.36)	(16.08)	(14.81)	(17.63)	(20.47)
border2	11.94	20.36***	16.54***	20.29***	17.60***	20.82***
	(30.93)	(6.062)	(4.161)	(3.796)	(4.481)	(5.190)
border3	10.80	14.71	16.25**	36.13***	46.45***	44.26***
	(46.30)	(8.876)	(6.096)	(5.559)	(6.563)	(7.602)
border4	-25.90	33.59*	39.24***	59.26***	79.23***	68.76***
	(74.81)	(14.43)	(9.896)	(9.031)	(10.66)	(12.36)
border5	136.0	106.1***	114.7***	143.1***	145.5***	161.5***
	(85.48)	(16.45)	(11.28)	(10.24)	(12.05)	(13.91)
N	141175	141175	141175	141175	141175	141175

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We also control for region and year fixed effects.

Table 4: Quantile regression — with gestation

	(1)	(2)	(3)	(4)	(5)	(6)
	1%	10%	25%	50%	75%	90%
smoker	-183.2***	-156.3***	-144.0***	-150.8***	-146.4***	-133.1***
	(38.77)	(13.72)	(10.94)	(10.08)	(12.27)	(15.74)
cigs	-4.157	-7.748***	-9.245***	-9.597***	-10.76***	-12.67***
	(3.845)	(1.456)	(1.235)	(1.144)	(1.313)	(1.619)
cigssq	-0.0676	0.0899*	0.101**	0.116***	0.155***	0.189***
	(0.0788)	(0.0352)	(0.0320)	(0.0295)	(0.0313)	(0.0354)
gest	1361.8***	1050.5***	926.9***	753.6***	513.3***	282.0***
	(19.71)	(8.136)	(7.246)	(7.874)	(10.60)	(11.45)
gestsq	-16.09***	-12.24***	-10.67***	-8.503***	-5.547***	-2.719***
	(0.277)	(0.112)	(0.0975)	(0.104)	(0.138)	(0.148)
age	39.24***	17.43***	16.87***	15.21***	11.93***	12.40**
	(10.54)	(3.667)	(2.869)	(2.596)	(3.211)	(4.126)
agesq	-0.687***	-0.255***	-0.232***	-0.195***	-0.135*	-0.121
	(0.172)	(0.0601)	(0.0471)	(0.0426)	(0.0528)	(0.0677)
hs	57.10*	62.14***	56.48***	51.63***	64.15***	50.05***
	(23.18)	(8.157)	(6.354)	(5.714)	(7.071)	(9.126)
sc	117.9***	93.96***	88.08***	79.05***	84.20***	65.07***
	(25.61)	(9.024)	(7.002)	(6.307)	(7.825)	(10.13)
cg	167.2***	117.4***	98.10***	84.14***	80.20***	49.79***
	(27.11)	(9.422)	(7.335)	(6.619)	(8.221)	(10.64)
black	-239.9***	-197.6***	-195.0***	-206.0***	-208.1***	-200.8***
	(22.06)	(7.836)	(6.134)	(5.548)	(6.905)	(8.862)
male	110.7***	126.4***	132.9***	145.9***	156.5***	157.3***
	(10.85)	(3.871)	(3.030)	(2.742)	(3.405)	(4.372)
nopnv	107.2	50.04	31.14	28.38	43.49	70.94*
	(79.19)	(28.64)	(22.46)	(20.37)	(25.34)	(32.49)
pnv2t	79.20**	21.95*	25.77***	33.79***	44.26***	47.29***
	(26.59)	(9.409)	(7.359)	(6.650)	(8.261)	(10.60)
pnv3t	172.3**	98.71***	57.46**	61.55***	50.22*	20.95
	(63.41)	(22.75)	(17.82)	(16.18)	(20.18)	(25.98)
pnc_inter	-112.9***	-47.25***	-41.60***	-41.43***	-46.02***	-47.36***
	(22.09)	(7.796)	(6.092)	(5.494)	(6.796)	(8.703)
pnc_inad	-242.4***	-140.2***	-102.2***	-77.05***	-84.29***	-71.88***
	(50.67)	(18.31)	(14.32)	(13.03)	(16.30)	(21.01)
border2	23.09	13.95**	21.77***	22.07***	23.87***	21.21***
	(13.00)	(4.715)	(3.695)	(3.341)	(4.147)	(5.324)
border3	-3.432	21.41**	28.43***	43.31***	46.98***	46.88***
	(19.33)	(6.917)	(5.406)	(4.894)	(6.075)	(7.806)
border4	18.22	40.90***	49.49***	61.80***	78.47***	56.47***
	(31.63)	(11.21)	(8.782)	(7.947)	(9.876)	(12.72)
border5	98.15**	103.6***	98.34***	112.1***	140.0***	140.7***
	(36.21)	(12.78)	(10.01)	(9.017)	(11.16)	(14.31)
N	141175	141175	141175	141175	141175	141175

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We also control for region and year fixed effects.

tive but they are off-set for smoking mothers.

6 Conclusion

To determine how maternal smoking during pregnancy affects the distribution of infants' birthweight, several issues must be addressed. Given the data provided, panel data and IV estimation methods have been disregarded. We argue that parent's education do not provide a valid instrument for smoking because it may be correlated with health-relating unobserved variables which may affect birthweight. Hence, the estimates shown in this paper should be interpreted as correlations rather than causal effects. The summary statistics show that there exist several differences between smokers and non-smokers and thus, we focus on regressions with and without interaction terms to highlight the association between smoking and birthweight. To analyse the lower tail of the distribution of birthweight we additionally estimate quantile regressions.

Our results are in line with the literature and suggest that smoking is in general associated negatively with infants' weight. However, we emphasise the link from smoking behaviour over gestation, which our analysis shows to be of great importance. When gestation is included in the quantile regressions, the heterogenous effects from smoking over the quantiles of the infants' birthweight revealed to be associated with the gestation mainly, and the direct relation between smoking and birthweight seems constant over the conditional distribution.

Finally we find by inclusion of interaction terms that the unborn infants of the smokers gain less weight than non-smokers unborn infants. It is tempting to conclude a clear negative consequence of smoking on birthweight, and that this effects is due to less weight gain over time for the smokers as well as shorter gestation. However, as we have no valid instrument variables or panel data it is problematic to draw causal interpretations from the analysis, which hope to investigate further tomorrow.

Appendix A

754 smoking mothers have not reported the number of cigarettes they smoke. Figure 1A compares the distribution of birthweigt for the smoking mothers that report the number of cigarettes compared to the mothers that do not report.

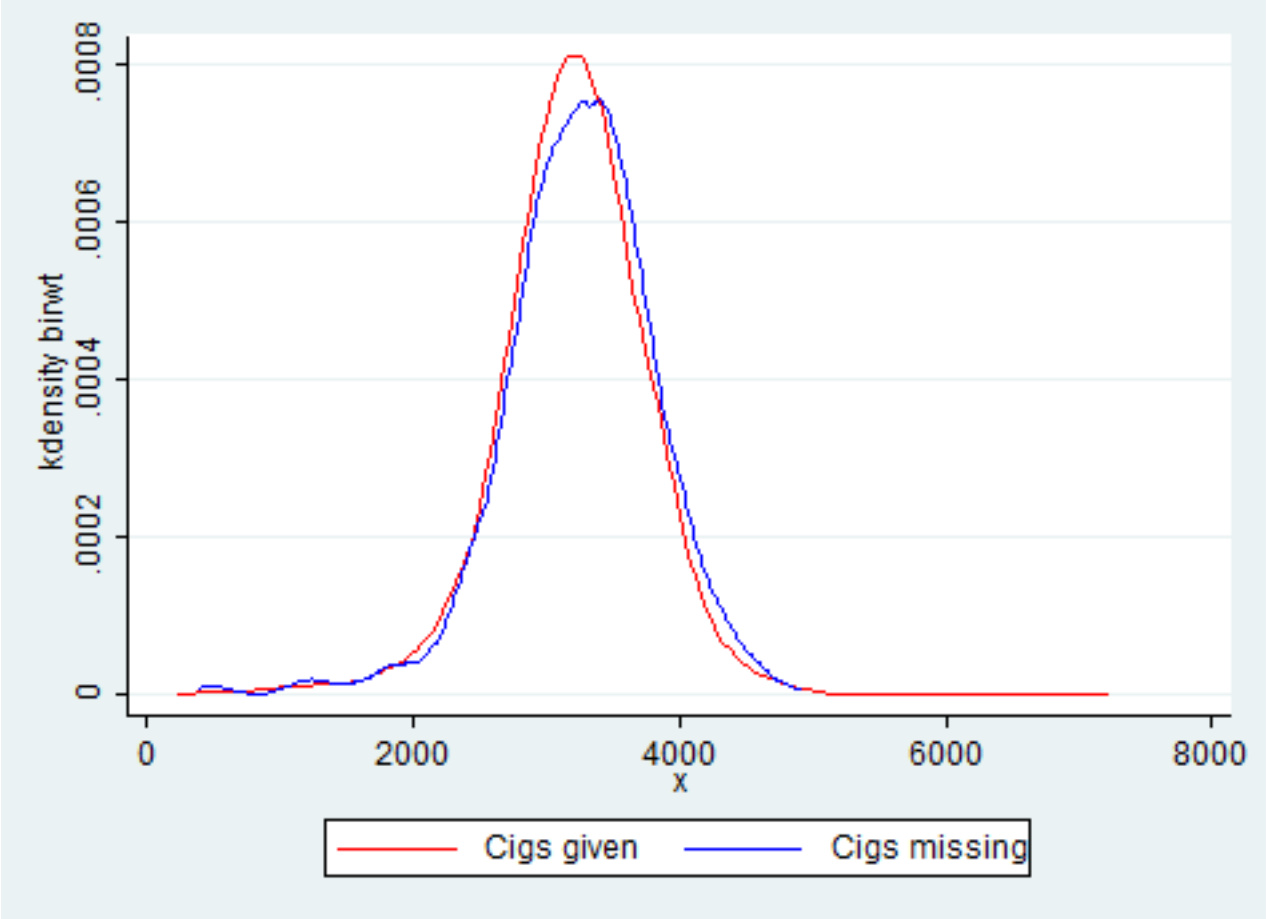
Table 5: OLS estimates with interaction effects

	(1)	
	birwt	
cigs	-8.993***	(1.025)
cigssq	0.118***	(0.0262)
smoker	-152.1***	(9.988)
gest	631.0***	(7.894)
gestsq	-6.977***	(0.104)
age	21.82***	(2.550)
agesq	-0.298***	(0.0415)
hs	61.60***	(6.284)
sc	86.97***	(6.663)
cg	88.06***	(6.871)
black	-199.0***	(5.334)
male	141.5***	(2.605)
nopnv	41.74	(22.50)
pnv2t	44.33***	(6.563)
pnv3t	53.95**	(17.31)
pnc_inter	-49.26***	(5.278)
pnc_inad	-79.13***	(13.73)
border2	27.29***	(3.182)
border3	46.40***	(4.715)
border4	69.78***	(7.730)
border5	131.2***	(8.597)
smoker_age	-8.595	(6.691)
smoker_agesq	0.0758	(0.116)
smoker_hs	-15.88	(10.84)
smoker_sc	-10.60	(13.55)
smoker_cg	24.66	(19.27)
smoker_black	23.35	(13.67)
smoker_male	4.743	(7.279)
smoker_gest	-104.1***	(17.04)
smoker_gestsq	1.226***	(0.226)
smoker_nopnv	24.56	(38.33)
smoker_pnv2t	-6.800	(15.23)
smoker_pnv3t	68.72*	(31.39)
smoker_pnc_inter	-7.641	(13.69)
smoker_pnc_inad	-78.63**	(26.03)
smoker_border2	-46.12***	(8.754)
smoker_border3	-58.72***	(12.12)
smoker_border4	-90.90***	(18.97)
smoker_border5	-129.4***	(23.58)
N	141175	

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure 1A



From Figure 1A, it does not appear as if the non-reporting mothers are heavy smokers that have children with particularly low birth weight. The two distributions appear to coincide more or less, taking into account that the group with missing observations is much smaller than the other group resulting in a less smooth distribution. In opposition to our intuition, the mothers with missing information have given birth the heavier children (3,201g on average) than those who report the number of cigarettes they smoke (3,252g). The difference is not large, and a t-test yields a p-value of 0.015 implying that the difference is only borderline significant at the one percent level. Because the non-reporting mothers give birth to heavier children and the difference is not too significant taking the large sample size into account, we conclude that we can disregard the non-reporting mothers from the analysis.

References

ABREVAYA, J. (2006): "Estimating the effect of smoking on birth outcomes using a matched panel data approach," *Journal of Applied Econometrics*, 21(4), 489–519.

ABREVAYA, J., AND C. DAHL (2008): "The effects of birth inputs on birthweight," *Journal of Business and Economic Statistics*, 26(4), 379–397.

EVANS, W. N., AND J. S. RINGEL (1999): "Can higher cigarette taxes improve birth outcomes?," *Journal of Public Economics*, 72(1), 135–154.

KRAMER, M. (1987): "Determinants of low birth weight: methodological assessment and meta-analysis.," *Bulletin of the World Health Organization*, 65(5), 663.

MULLAHY, J. (1997): "Instrumental-variable estimation of count data models: Applications to models of cigarette smoking behavior," *Review of Economics and Statistics*, 79(4), 586–593.

PERMUTT, T., AND J. HEBEL (1989): "Simultaneous-equation estimation in a clinical trial of the effect of smoking on birth weight," *Biometrics*, pp. 619–622.

WEHBY, G., J. FLETCHER, S. LEHRER, L. MORENO, J. MURRAY, A. WILCOX, AND R. LIE (2011): "A Genetic Instrumental Variables Analysis of the Effects of Prenatal Smoking on Birth Weight: Evidence from Two Samples," *Biodemography and Social Biology*, 57(1), 3–32.