Table 1: Data used in the recent literature

	Data	# of obs.
Evans and Ringel (1999)	NCHS (1989-1992)	10.5 million
Almond et al. (2005)	NCHS(1989-1991, PA only)	491,139
Abrevaya (2006)	matched panel constructed from NCHS (1989-1998)	296,218
Arellano and Bonhomme (2011)	matched panel #3 in Abrevaya (2006)	1,445
Jun et al. (2013)	matched panel #3 in Abrevaya (2006)	2,113
Hoderlein and Sasaki (2013)	random sample from NCHS (1989-1999)	100,000

and Sasaki (2013) took advantage of cigarette tax rates or tax increases.[13] The last strand takes a panel data approach. This approach isolates the effects of unobservables using data on mothers with multiple births and identifies the effect of smoking from the change in their smoking status from one pregnancy to another. To do this, Abrevaya (2006) constructed the panel data set with novel matching algorithms between women having multiple births and children on federal natality data. The panel data set constructed by Abrevaya (2006) has been used in other recent studies such as Arellano and Bonhomme (2011) and Jun et al. (2013). Jun et al. (2013) tested stochastic dominance between two marginal distributions of potential birth weight with and without smoking. Arellano and Bonhomme (2011) identified the distribution of smoking effects using the random coefficient panel data model.

To the best of my knowledge, the only existing study that examines the distribution of smoking effects is Arellano and Bonhomme (2012). While they point-identify the distribution of smoking effects, their approach presumes access to the panel data with individuals who changed their smoking status within their multiple births. Specifically, they use the following panel data model with random coefficients:

$$Y_{it} = \alpha_i + \beta_i D_{it} + X_{it}'\gamma + \varepsilon_{it}$$

where Y_{it} is infant birth weight and D_{it} is an indicator for woman i smoking before she had her t-th baby. Extending Kotlarski's deconvolution idea, they identify the *distribution* of $\beta_i = E\left[Y_{it}|D_{it}=1,\alpha_i,\beta_i\right] - E\left[Y_{it}|D_{it}=0,\alpha_i,\beta_i\right]$, which indicates the distribution of smoking effects in this example. For the identification, they assume strict exogeneity that mothers do not change their smoking behavior from their previous babies' birth weight. Furthermore, their estimation result is somewhat implausible. It is interpreted that smoking has a positive effect on infant birth weight for approximately 30% mothers. They conjecture that this might result from a misspecification problem such as the strict exogeneity condition, i.i.d. idiosyncratic shock, etc.

[13] Permutt and Hebel (1989), Evans and Ringel (1999) and Lien and Evans (2005) two-stage linear regression to estimate the average effect of smoking using an instrument. Hoderlein and Sasaki (2011) adopted the number of cigarettes as a continuous treatment, and identified and estimated the average marginal effect of a cigarette based on the nonseparable model with a triangular structure.