

IZA DP No. 5216

## Linearity in Instrumental Variables Estimation: Problems and Solutions

Magne Mogstad  
Matthew Wiswall

September 2010

# Linearity in Instrumental Variables Estimation: Problems and Solutions

**Magne Mogstad**

*Statistics Norway  
and IZA*

**Matthew Wiswall**

*New York University*

Discussion Paper No. 5216  
September 2010

IZA

P.O. Box 7240  
53072 Bonn  
Germany

Phone: +49-228-3894-0

Fax: +49-228-3894-180

E-mail: [iza@iza.org](mailto:iza@iza.org)

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

## ABSTRACT

### Linearity in Instrumental Variables Estimation: Problems and Solutions<sup>\*</sup>

The linear IV estimator, in which the dependent variable is a linear function of a potentially endogenous regressor, is a major workhorse in empirical economics. When this regressor takes on multiple values, the linear specification restricts the marginal effects to be constant across all margins. This paper investigates the problems caused by the linearity restriction in IV estimation, and discusses possible remedies. We first examine the biases due to nonlinearity in the commonly used tests for non-zero treatment effects, selection bias, and instrument validity. Next, we consider three applications where theory suggests a nonlinear relationship, yet previous research has used linear IV estimators. We find that relaxing the linearity restriction in the IV estimation changes the qualitative conclusions about the relevant economic theory and the effectiveness of different policies.

JEL Classification: C31, C14

Keywords: linear model, variable treatment intensity, nonlinearity, instrumental variables

Corresponding author:

Magne Mogstad  
Research Department  
Statistics Norway  
P.O. Box 8131 Dep.  
N-0033 Oslo  
Norway  
E-mail: [magne.mogstad@ssb.no](mailto:magne.mogstad@ssb.no)

---

<sup>\*</sup> We appreciate the helpful comments from Rolf Aaberge, Christian Brinch, Audun Langørgen, Terje Skjerpen and participants at a number of seminars and conferences.

# 1 Introduction

The linear instrumental variables (IV) estimator, in which the dependent variable is a linear function of a potentially endogenous regressor, is a major workhorse in empirical economics and gaining ground in other sciences. When this regressor takes on multiple values, so-called “variable treatment intensity,” the linear specification restricts the marginal effects to be constant across all margins. There are numerous applications with variable treatment intensity, including the economic return to schooling, class size and child development, family size and child quality, fertility and maternal labor supply, family income and child development, and maternal smoking and children’s birth weight.

In most applications with variable treatment intensity, the linear specification does not come from economic theory, but is rather chosen out of convenience or because there may not be enough instruments available (or they are too weak) to instrument for multiple endogenous regressors. In the spirit of Leamer’s (1983) seminal critique of empirical economics, the results from studies relying on such an arbitrarily chosen functional form should be interpreted with caution. The remedy Leamer proposed was sensitivity analysis, in which researchers examine how their results vary with changes in functional form.

Over the last few years, however, it has been argued that the linearity restriction gives little cause for concern. The argument is twofold. First, the relationship between the dependent variable and the regressor of interest might in many cases be approximately linear.<sup>2</sup> And second, even if the relationship is nonlinear, the weighted average of marginal effects produced by the linearly restricted estimator might still be a relevant parameter for testing theory or assessing policy. These arguments are advocated forcefully in the review article by Angrist and Pischke (2010), who state that: “an emerging grasp of the sense in which regression and two-stage least squares [IV] produce average effects when the underlying relationship is heterogeneous and/or nonlinear has made functional form concerns less central. The linear models that constitute the workhorse of contemporary empirical practice usually turn out to be remarkably robust” (p. 10). However, others have challenged this view of linear models, arguing instead that a careful exploration of possible nonlinearities is of central importance in empirical research.<sup>3</sup>

In this paper, we investigate the problems caused by the linearity restriction in IV estimation, and discuss possible remedies. The first part of the paper builds on previous research showing that the linear OLS and IV estimands can be decomposed into weighted

---

<sup>2</sup>A generalization of this argument is that for some transformation of the dependent variable or the regressor of interest, e.g. the log transformation, the relationship is approximately linear. We examine this issue in the context of a particular application.

<sup>3</sup>See e.g. Keane (2010a), Sims (2010) and Stock (2010). Their comments also discuss a number of other aspects of the so-called experimentalist approach to empirical economics, advocated by Angrist and Pischke (2010). See also the comments by Leamer (2010) and Nevo and Whinston (2010), as well as the recent discussions in Deaton (2010), Heckman and Urzua (2010), Imbens (2010), Keane (2010b), and Heckman (2010).

averages of marginal effect estimates, where the OLS and IV weights are in general different (see Angrist and Imbens, 1995; Yitzhaki, 1996; Angrist and Krueger, 1999; Heckman, Urzua, and Vytlacil, 2006). Using such decompositions, we examine the the biases due to nonlinearity in commonly used tests for treatment effects, selection bias, and instrument validity. The insights from our analysis may be summarized in the following four conclusions:

(i) The linearly restricted IV estimator can lead to a conclusion of a zero treatment effect even if some (or all) marginal effects are non-zero. In particular, the bias in using linear models when the marginal effects are non-constant is one-sided, and can only lead to under-rejection of the null hypothesis of no treatment effect.

(ii) The linear IV estimator may assign negative weights to some marginal effects. An implication of the negative weights is that the linear IV estimator can be negative (positive) even if all the marginal effects are strictly positive (negative). Negative weights can be ruled out by assuming monotonicity in the direction the instrument affects treatment intensity across treatment levels, which is directly testable.

(iii) Because the OLS estimator and the IV estimator of the linearly restricted model generally assign different weights to the marginal effects, nonlinearities can lead to a conclusion of selection bias even when the regressor is exogenous, or lead to a conclusion of no selection bias even when the regressor is endogenous. We therefore propose a diagnostic tool designed to tell to what extent linear OLS and IV estimates differ because of different weighting of marginal effects.

(iv) Over-identification tests comparing the linear IV estimator using different instruments can be misleading, as nonlinearities can mask instrument invalidity or lead to an erroneous conclusion of instrument validity. The reason is that different instruments generally assign different weights to the marginal effects.

To emphasize the distinction between the issues related to functional form and other commonly cited issues with IV estimation, we show that the above results are solely due to nonlinearity, and not to heterogeneity in the marginal effects, instrument invalidity, or weak instruments.

To demonstrate that the implications of the linearity restriction in IV estimation are not merely theoretical peculiarities but real cause for concern, the second part of the paper considers three applications with variable treatment intensity: the economic return to schooling, the link between family size and child quality, and the effect of childhood family income on child development. The choice of applications is not incidental. In all three cases, economic theory and OLS results suggest a nonlinear relationship between the dependent variable and the regressor of interest, yet previous research has typically used linearly restricted IV estimators. It should be emphasized that our focus is not on instrument validity or weak instruments: Our point of departure is to illustrate the

implications of nonlinearities in IV estimation of linearly restricted models, using the same data and instruments as in already published papers.

From the applications, we have four empirical findings:

(i) In each of our applications, the OLS results (with controls for observable confounding factors) suggest highly nonlinear relationships, consistent with the theoretical predictions. This evidence of nonlinearities in the OLS estimates raises serious doubts of the appropriateness of the linearity restriction in the IV estimation.

(ii) From estimating the linear IV weights, we learn that even though the linear IV estimators produce some weighted average of underlying marginal effects, it is not clear how they should be interpreted and what policy question they could answer. On the one hand, the IV weights are quite different from the sample distribution of the regressor of interest. This indicates that the linear IV estimate will likely differ from the average marginal effect across the population distribution of the regressor, or some other typical policy-relevant treatment parameter.<sup>4</sup> On the other hand, the linear IV estimates do not identify specific marginal effects, since they tend to place non-trivial weight on several treatment margins. This suggests that the linearly restricted IV estimates are not interpretable as the effect on particular margins.

(iii) The OLS and IV estimates of the linearly restricted model assign different weights to the marginal effects. In fact, in one of the applications we consider, the linear IV estimator assigns negative weights to some marginal effects. The differences in weighting implies that linear IV estimates are not directly comparable to the linear OLS estimates. As a consequence, the commonly used tests for selection should be interpreted with caution. In one of the applications, for example, the proposed diagnostic tool tells us that about half of the difference between the linear OLS and IV estimates is a result of different weighting of marginal effects.

(iv) In each application, we find that the IV estimates are not robust to relaxing the linearity restriction. In particular, allowing for a more flexible functional form changes the qualitative conclusions about the relevant economic theory and the effectiveness of different policies.

The general lesson to be drawn from this paper is *not* that IV estimation of linearly restricted models is of little use in applications with variable treatment intensity. Rather, our findings echo the argument of Leamer (1983) that functional form concerns are inescapable in empirical research. In fact, IV estimation may exacerbate the sensitivity of the results to the linearity restriction: While the linear IV weights depend on the particular instrument chosen and the correlation of that instrument with the various treatment margins, the OLS weights solely depend on the sample distribution of the regressor of

---

<sup>4</sup>The average marginal effect is sometimes referred to as the average partial effect. See Heckman and Vytlacil (2001) for a discussion of policy-relevant treatment parameters.

interest. The ideal remedy is non-parametric IV estimation, or at least some flexible form of parametric estimation to show how the results vary with changes in functional form assumptions.<sup>5</sup> In some situations, however, such analysis might be infeasible because of small sample size or not enough (or too weak) instruments. In any case, one can compute the weights to learn what the linear estimators actually identify and interpret the results in view of that. And if the weighting differs across linear estimates, the commonly used tests for selection bias and instrument validity should be interpreted with caution.

The remainder of the paper unfolds as follows. Section 2 reviews what linear OLS and IV estimators identify, and Section 3 examines the implications of nonlinearities for the inference drawn from these linear estimators. Section 4 reports the empirical results from the three applications. Section 5 concludes.

## 2 What Linear Estimators Identify

This section reviews what linearly restricted OLS and IV estimators identify when the endogenous regressor takes on multiple values.

### 2.1 Potential outcomes, linearity, and marginal effects

Let  $f_i(s)$  denote the potential (or latent) outcome that individual  $i$  would receive with treatment intensity  $s$ . In the return to schooling application, for example,  $f_i(s)$  describes the potential earnings that individual  $i$  would receive with  $s$  years of schooling. In general,  $f_i(s)$  is determined by aspects of individual behavior and market forces. Without an explicit theoretical model for  $f_i(s)$ , we can think of this function as describing the outcome that individual  $i$  would achieve if he or she was randomly assigned treatment intensity  $s$ . The observed treatment intensity received by individual  $i$  is denoted  $S_i$ .

The standard regression framework used in the previous literature to link the potential outcome to the observed outcome,  $y_i$ , has the following form

$$y_i = f_i(S_i) = \mu + \beta S_i + \epsilon_i, \quad (1)$$

where  $\epsilon_i$  is a mean-zero error term. This linearly restricted model forms the basis for most previous studies using IV strategies to address the concern for selection bias in regression analysis. Our point of departure is to relax the linearity assumption, allowing instead the marginal treatment effects to vary with the treatment level. Let the treatment variable

---

<sup>5</sup>See e.g. Horowitz (2009) and Newey and Powell (2003) for discussions of non-parametric IV estimation. In all of the applications we consider, the support of the endogenous regressor is either discrete and the non-parametric model has a finite number of parameters, or we estimate less restrictive parametric models. We therefore avoid the “ill posed inverse” problem of non-parametric IV estimation. See also Abadie (2003) for a discussion of IV estimation of nonlinear models with a binary endogenous regressor.

take on values in the finite set  $S_i \in \{0, 1, \dots, \bar{s}\}$ . Using dummy variables constructed as  $d_{si} = 1\{S_i \geq s\}$ , we can specify the following non-parametric model

$$y_i = \mu + \sum_{s=1}^{\bar{s}} \gamma_s d_{si} + v_i, \quad (2)$$

where the  $\gamma_s$  coefficient represents the marginal effect of an increase in treatment intensity from  $s - 1$  to  $s$ , and  $v_i$  is a mean-zero error term. The non-parametric model (2) nests the linear model (1), which restricts the marginal treatment effects to be independent of the level of treatment:  $\gamma_s = \beta$  for all  $s > 0$ .

Note that both (1) and (2) assumes that the marginal effects are homogenous across individuals: The only individual-specific component is the error terms, which capture unobserved factors determining the outcome. We assume homogenous marginal effects to focus attention on the implications of nonlinearity in linearly restricted models. In contrast, much of the discussion on IV estimation has been focused on population heterogeneity in treatment effects in situations with a binary treatment variable. We simply reverse the focus. And even if there is heterogeneity in the marginal effects, which could be represented in (2) as random coefficients on the  $d_{si}$  indicators,  $\gamma_s$  is still interpretable as the *average* marginal effect of increasing treatment intensity from  $s - 1$  to  $s$ . However, with heterogeneity in the marginal effects, the linear IV estimator should be interpreted as a weighted average of the marginal local average treatment effects (LATE). The marginal LATEs are the average marginal effects for the particular groups whose treatment status is shifted by the instrument.<sup>6</sup> In our empirical analysis, we discuss the issue of population heterogeneity in the marginal effects.

## 2.2 IV Decomposition

Angrist and Imbens (1995) provide a decomposition of the linear IV estimand in the case of a binary instrument and under an assumption of monotonicity. We offer an alternative decomposition result, and moreover, discuss how the linear IV estimator can assign *negative* weights to some marginal effects if monotonicity does not hold.<sup>7</sup>

Consider the case of a scalar (binary or multi-valued) instrument  $z_i$ . We discuss the case of multiple instruments below. Suppose that the standard IV assumptions hold:

Assumption A1 (Instrument Uncorrelated with Error Term):  $Cov(v_i, z_i) = 0$ .

---

<sup>6</sup>See Imbens and Angrist (1994), Heckman, Urzua, and Vytlacil (2006), and Angrist and Imbens (1995) for discussions of the LATE interpretation of IV estimates. While the former two studies focus on the case of a binary treatment variable, the latter study considers a situation with variable treatment intensity.

<sup>7</sup>In the return to schooling application of Angrist and Krueger (1991), for example, monotonicity means that people born in the fourth quarter (who are obliged to stay in school longer due to compulsory schooling laws) get at least as much schooling as they would have if they had been born in the first quarter. See also Heckman and Vytlacil (2005), and Barua and Lang (2009) for a critical discussion of the monotonicity assumption, and examples of situations where it is likely to fail.

Assumption A2 (Instrument Correlated with Endogenous Regressor):  $Cov(S_i, z_i) \neq 0$ .

Assumption A1 implies that the instrument is uncorrelated with the unobserved factors determining the outcome. Note that the stronger assumption of the instrument being independent of all potential outcomes and potential treatment intensities, as in Angrist and Imbens (1995), would imply Assumption A1. Assumption A2 is a standard condition of the existence of an instrument, implying that the instrument affects treatment intensity.

Under Assumptions A1 and A2, the IV estimand for  $\beta$  in (1) is  $\beta(z) = Cov(y_i, z_i)/Cov(S_i, z_i)$ . To decompose  $\beta(z)$  into a weighted average of the marginal effects  $\gamma_s$ , substitute for  $y_i$  in (2), which yields

$$\begin{aligned} Cov(y_i, z_i) &= Cov((\mu + \gamma_1 d_{1i} + \cdots + \gamma_{\bar{s}} d_{\bar{s}i} + v_i), z_i) \\ &= \gamma_1 Cov(d_{1i}, z_i) + \cdots + \gamma_{\bar{s}} Cov(d_{\bar{s}i}, z_i) + Cov(v_i, z_i). \end{aligned}$$

Under Assumptions A1 and A2, we then have:

$$\beta(z) = \sum_{s=1}^{\bar{s}} w_s(z) \gamma_s, \quad (3)$$

where

$$w_s(z) = \frac{Cov(d_{si}, z_i)}{Cov(S_i, z_i)}.$$

These weights sum to one,<sup>8</sup> and can be directly estimated using the sample analog of the expression above.

We learn from (3) that  $\beta(z)$  is a weighted average of the marginal treatment effects. The weight  $w_s(z)$  attached to  $\gamma_s$  depends on the covariance between each  $d_{si}$  indicator and the instrument. Hence,  $\beta(z)$  assigns more weight to the marginal effects at treatment levels that are most affected by the particular instrument chosen.

From (3), it is also clear that some (but not all)  $w_s(z)$  might be negative. Negative weights occur in situations where the instrument increases treatment intensity at some level of treatment, but decreases it at other levels of treatment, so that  $Cov(d_{si}, z_i)$  varies in sign depending on  $s$ . The possibility of negative weights has some troubling implications. First,  $\beta(z)$  is not necessarily restricted to lie between the maximum and minimum  $\gamma_s$ , when the IV weights are negative. As a consequence,  $\beta(z)$  may not be representative of *any* particular marginal effect or subset of marginal effects. An additional implication of negative IV weights is that the linear IV estimator can be negative (positive) even if all

---

<sup>8</sup>To see that the  $w_s(z)$  weights sum to 1, note that  $S_i = \sum_{s=1}^{\bar{s}} d_{si}$ , so that  $Cov(S_i, z_i) = \sum_{s=1}^{\bar{s}} Cov(d_{si}, z_i)$ .

of the marginal effects are strictly positive (negative).

To ensure non-negative weights ( $w_s(z) \geq 0$  for all  $s$ ), one can assume monotonicity in the direction the instrument affects treatment intensity across treatment levels:

Assumption A3 (Treatment Margin Monotonicity): Either  $Cov(d_{si}, z_i) \geq 0$  for all  $s$  or  $Cov(d_{si}, z_i) \leq 0$  for all  $s$ .

Note that the stronger assumption of monotonicity for each individual (i.e. the instrument effects everyone in the same way, if at all), as in Angrist and Imbens (1995), or equivalently “uniformity” of response to treatment, as in Heckman and Vytlacil (2005), would imply Assumption A3. However, Assumption A3 is a weaker assumption requiring only that the sign of the covariance is the same across all margins, which is sufficient to ensure non-negative IV weighting across all margins.<sup>9</sup> Unlike individual level monotonicity, Assumption A3 is directly testable since we can estimate  $Cov(d_{si}, z_i)$  for all  $s$ . The reason is that Assumption A3 does not involve unobserved counterfactual responses to various instrument values, as with the individual level monotonicity assumption. In one of the applications we examine, Assumption A3 actually fails and the linear IV estimator assigns negative weights to some marginal effects.

It should finally be noted that with Assumption A3 and a binary instrument  $z_i \in \{0, 1\}$ , it follows straightforwardly that (3) coincides with the IV decomposition proposed by Angrist and Imbens (1995):

$$w_s(z) = \frac{pr(d_{si} = 1|z_i = 1) - pr(d_{si} = 1)}{\sum_{j=1}^{\bar{c}} pr(d_{ji} = 1|z_i = 1) - pr(d_{ji} = 1)},$$

since  $Cov(d_{si}, z_i) = pr(z_i = 1)[pr(d_{si} = 1|z_i = 1) - pr(d_{si} = 1)]$ .

### 2.3 OLS Decomposition

Next, we provide an analagous decomposition for the linear OLS estimator which allows us to compare the OLS weighting to the IV weighting. The OLS estimand for  $\beta$  in (1) is  $\beta(OLS) = Cov(y_i, S_i)/Var(S_i)$ . Substituting for  $y_i$  from (2), we can write the linear OLS estimand as

$$\beta(OLS) = \sum_{s=1}^{\bar{s}} \gamma_s \frac{Cov(d_{si}, S_i)}{Var(S_i)} + \psi, \tag{4}$$

where  $\psi = Cov(v_i, S_i)/Var(S_i)$ . This expression shows that the linear OLS estimand is a weighted average of the marginal effects  $\gamma_s$ , and a selection bias term,  $\psi$ .

---

<sup>9</sup>Even if individual level monotonicity is violated, and some individuals have a non-uniform response to an instrument, Assumption A3 could still hold. However, the stronger assumption of individual monotonicity would be required for a LATE interpretation of the IV estimand in a situation with heterogenous treatment effects.

We can further decompose the linear OLS estimand into a weighted sum of the non-parametric OLS estimands. The non-parametric OLS estimand for  $\gamma_s$  in (2) is given by

$$\gamma_s(OLS) = E[y_i|S_i = s] - E[y_i|S_i = s - 1].$$

Substituting for  $y_i$ , we have

$$\gamma_s(OLS) = \gamma_s + \Delta_s,$$

where  $\Delta_s = E[v_i|S_i = s] - E[v_i|S_i = s - 1]$  represents the selection bias in the OLS estimates of the marginal effects,  $\gamma_s(OLS)$ . To see the relationship between  $\psi$  and  $\Delta_s$ , note that we can write the linear projection of  $v_i$  on  $d_{1i}, \dots, d_{\bar{s},i}$  as

$$v_i = \Delta_1 d_{1i} + \dots + \Delta_{\bar{s}} d_{\bar{s},i} + \omega_i,$$

where  $Cov(\omega_i, d_{si}) = 0$  for all  $s$  by definition of the linear projection. Substituting this expression for  $v_i$ , and noting that  $Cov(\omega_i, d_{si}) = 0$  for all  $s$  implies  $Cov(\omega_i, S_i) = 0$ , we can write  $\psi$  as a weighted average of the selection bias in the OLS estimates of the marginal effects:

$$\psi = \sum_{s=1}^{\bar{s}} \Delta_s \frac{Cov(d_{si}, S_i)}{Var(S_i)}.$$

Finally, substituting this expression for  $\psi$  into (4) and re-arranging, we can write the linear OLS estimand as a weighted average of the non-parametric OLS estimands:

$$\beta(OLS) = \sum_{s=1}^{\bar{s}} \gamma_s(OLS) w_s(OLS), \tag{5}$$

where the associated OLS weight on  $\gamma_s(OLS)$  is defined as

$$w_s(OLS) = \frac{Cov(d_{si}, S_i)}{Var(S_i)}.$$

These OLS weights are simply the regression coefficient of  $d_{si} = 1\{S_i \geq s\}$  on  $S_i$ . The OLS weights sum to one,<sup>10</sup> and can be directly estimated using the sample analog of the above expressions. Note that unlike the linear IV weights, the linear OLS weights are strictly non-negative since  $Cov(d_{si}, S_i) \geq 0$  for all  $s$ .

In general, the linear IV weight assigned to a given marginal effect,  $w_s(z)$ , is different from the linear OLS weight assigned to the same marginal effect,  $w_s(OLS)$ . While the

---

<sup>10</sup>The weights sum to 1 since  $\sum_{s=1}^{\bar{s}} Cov(d_{si}, S_i) = Var(S_i)$ .

linear IV weights depend on the particular instrument chosen and the correlation of that instrument with the various treatment margins, the OLS weights solely depend on the sample distribution of the regressor  $S_i$ . To see this, we can re-write the  $Cov(d_{si}, S_i)$  terms as

$$\begin{aligned} Cov(d_{si}, S_i) &= E[d_{si}S_i] - pr(d_{si} = 1)E[S_i] \\ &= (E[S_i|d_{si} = 1] - E[S_i|d_{si} = 0])pr(d_{si} = 1)(1 - pr(d_{si} = 1)). \end{aligned}$$

This implies that the linear OLS weights can be expressed as

$$w_s(OLS) = \frac{q_s}{\sum_{j=1}^{\bar{s}} q_j},$$

with

$$q_s = (E[S_i|S_i \geq s] - E[S_i|S_i < s])(pr(S_i \geq s)(1 - pr(S_i \geq s))).$$

An equivalent result is derived in Angrist and Krueger (1999), drawing on results from Yitzhaki (1996).

There are two key issues with OLS estimation in situations with variable treatment intensity. As is well known, OLS estimates will be biased if observed treatment intensity is correlated with the unobserved factors determining the outcome. The second issue, which has received far less attention in empirical research, is that the linear OLS estimand has a particular weighting over the marginal effects, given by (5). Specifically, weight is given to each  $\gamma_s$  in proportion to the conditional mean of  $S_i$ , above and below  $s$ . More weight is also given to marginal effects close to the sample median of  $S_i$ , since this is where  $pr(S_i \geq s)(1 - pr(S_i \geq s))$  is maximized. If there are nonlinearities in the OLS estimates of the marginal effects ( $\gamma_s(OLS) \neq \gamma_{s'}(OLS)$  for  $s \neq s'$ ), then it follows from (5) that the linear OLS estimate depends on how it weights the marginal effects, and thereby the sample distribution of  $S_i$ . Depending on the weights, the linear OLS estimate ranges between the maximum and minimum  $\gamma_s(OLS)$ .

## 2.4 Decomposition with Covariates and Multiple Instruments

In most empirical work, the IV assumptions are assumed to hold conditional on a set of covariates. Moreover, many studies use several instruments in the IV estimation. In all the three applications we consider, for example, previous research have included covariates in a linear fashion, and used multiple instruments. In such situations, we need to modify the above decompositions to estimate the linear OLS and IV weights.

Consider first the case where a vector of exogenous covariates,  $X_i$ , is added to (1). The linearly restricted model is then given by

$$y_i = X_i' \delta + \beta S_i + \epsilon_i, \quad (6)$$

where  $X_i$  includes the intercept. In order to derive the linear OLS weights, we transform the model using residual regression. Specifically,  $S_i$  is regressed on  $X_i$ , producing the residual  $S_i^* = S_i - X_i' \hat{\chi}$ , where  $\hat{\chi}$  is the estimated regression coefficients. The linear OLS estimand is then  $\beta(OLS) = Cov(y_i, S_i^*) / Var(S_i^*)$ , and the associated weights are  $w_s(OLS) = Cov(d_{si}, S_i^*) / Var(S_i^*)$  for all  $s$ . Note that while the OLS weights in models without additional covariates are strictly non-negative since  $Cov(d_{si}, S_i) \geq 0$ , the OLS weights with covariates could be negative since it is possible that  $Cov(d_{si}, S_i^*) < 0$ .

Consider next the case where the linearly restricted model includes a vector of covariates, and multiple instruments are used in the IV estimation. We restrict our attention to the commonly used Two Stage Least Squares (2SLS) IV estimator. The second stage of the 2SLS estimator is given by (6), whereas the first stage can be expressed as

$$S_i = X_i' \kappa + Z_i' \lambda + \eta_i,$$

where  $Z_i$  is a vector of instruments. The linear 2SLS estimator can be formed by first regressing  $S_i$  on covariates  $X_i$  and instruments  $Z_i$ , to form the fitted value  $\hat{S}_i = X_i' \hat{\kappa} + Z_i' \hat{\lambda}$ , where  $\hat{\kappa}$  and  $\hat{\lambda}$  are the OLS estimates from the first stage regression. Next, we regress  $\hat{S}_i$  on covariates  $X_i$  alone, producing the residual  $\hat{S}_i^* = \hat{S}_i - X_i' \hat{\chi}$ , where  $\hat{\chi}$  is the estimated regression coefficients from  $\hat{S}_i$  on  $X_i$ . The linear IV estimate is then the sample analog of  $Cov(y_i, \hat{S}_i^*) / Cov(\hat{S}_i^*, S_i)$ , and the associated weights are  $w_s(\hat{S}_i^*) = Cov(\hat{S}_i^*, d_{si}) / Cov(\hat{S}_i^*, S_i)$ .

It should be noted that the IV weights in a linear model with covariates,  $w_s(\hat{S}_i^*)$ , can be quite different from the IV weights using a single instrument or ignoring covariates.  $w_s(\hat{S}_i^*)$  is the appropriate weight for the linear IV estimator that uses multiple instruments and includes covariates in a linear fashion, as  $\hat{S}_i^*$  is the implied instrument in this framework.

### 3 Testing for Treatment Effects, Selection Bias, and Instrument Validity

This section uses the above expressions to discuss how inference can be misleading if we use linear OLS and IV estimators when the true relationship is nonlinear. We consider testing for non-zero treatment effects, selection bias, and instrument validity. For each test, we formulate the implicit null hypothesis in terms of the marginal effects, and discuss the conditions under which using linear test statistics leads to under- or over-rejection of the null hypothesis. To illustrate, we use a Monte Carlo study. Finally, we propose a diagnostic tool designed to tell to what extent linear OLS and IV estimates differ because

of different weighting of marginal effects.

### 3.1 Simulation Example

Our Monte Carlo study is based on a simple example where the treatment variable/regressor takes on the values 0,1, and 2. The observed outcome is given by

$$y_i = \gamma_1 d_{1i} + \gamma_2 d_{2i} + v_i,$$

where  $v_i \sim N(0, \sigma_v^2)$ .

The level of treatment is given by

$$d_{1i} = 1\{\pi_{11}z_{1i} + \pi_{12}z_{2i} + \alpha_{1,v}v_i + \psi_{1i} \geq 0\},$$

$$d_{2i} = d_{1i}1\{\pi_{21}z_{1i} + \pi_{22}z_{2i} + \alpha_{2,v}v_i + \psi_{2i} \geq 0\},$$

$$S_i = d_{1i} + d_{2i}$$

where  $\psi_{si} \sim N(0, \sigma_{\psi_s}^2)$  and  $z_{si}$  is a binary instruments distributed as  $z_{si} = 1$  with probability 0.5 and  $z_{si} = 0$  otherwise, for  $s = 1, 2$ . The instruments are constructed to be independent,  $z_{1i} \perp z_{2i}$ , and both satisfy Assumptions A1-A3.

To illustrate our results, we simulate the model for two combinations of parameters.

Case 1, no selection bias:  $\gamma_1 = 1$ ,  $\sigma_v = 10$ ,  $\sigma_{\psi_1} = 1$ ,  $\sigma_{\psi_2} = 1$ ,  $\pi_{11} = 3$ ,  $\pi_{12} = 0$ ,  $\pi_{21} = 0$ ,  $\pi_{22} = 3$ ,  $\alpha_{1,v} = 0$ ,  $\alpha_{2,v} = 0$ .

Case 2, selection bias: Same parameters as Case 1 except that  $\alpha_{1,v} = 0.01$ ,  $\alpha_{2,v} = 0.01$ .

In each case, we simulate three different linear estimators:

1) Linear OLS:

$$\beta(OLS) = Cov(y_i, S_i)/Var(S_i).$$

2) Linear IV using  $z_{1i}$  as the instrument:

$$\beta(z_1) = Cov(y_i, z_{1i})/Cov(z_{1i}, S_i).$$

3) Linear IV using  $z_{2i}$  as the instrument:

$$\beta(z_2) = Cov(y_i, z_{2i})/Cov(z_{2i}, S_i).$$

In the simulations, we compute the average values of the linear estimates across 500 replications, each using 5000 observations from the data generating process. In every replication, we set  $\gamma_1 = 1$  and vary the other marginal effect  $\gamma_2$ . As we vary  $\gamma_2$  away

from 1, we increase the degree of nonlinearity. At  $\gamma_1 = \gamma_2 = 1$ , the marginal effects are constant and the linear model is correctly specified. In Case 1, we then have that  $\beta(OLS) = \beta(z_1) = \beta(z_2) = 1$ . In Case 2, however, the positive  $\alpha_{1,v}$  and  $\alpha_{2,v}$  creates an upward bias in the linear OLS estimate. For example, if  $\gamma_1 = \gamma_2 = 1$ , then  $\beta(OLS) = 1.63$  rather than 1.

Figure 1 illustrates Case 1. This figure shows that as we move  $\gamma_2$  away from  $\gamma_1$  and introduce nonlinearity in the treatment effects, the three estimators diverge from each other. The reason is that the estimators weigh the marginal effects differently. On the one hand, the linear OLS estimator places about 46 percent weight on the  $\gamma_1$  marginal effect and 54 percent weight on the  $\gamma_2$  marginal effect. On the other hand, the two IV estimators have substantially different weighting of the marginal effects, reflecting the strength of the instruments on each treatment margin. In our simulation,  $z_{1i}$  is constructed to affect mainly the first treatment margin, while instrument  $z_{2i}$  affects exclusively the second margin. Given these particular instruments, the linear IV estimator using instrument  $z_{1i}$  places 2/3 weight on  $\gamma_1$  and 1/3 weight on  $\gamma_2$ . In comparison, the linear IV estimator using instrument  $z_{2i}$  places 0 weight on  $\gamma_1$  and all weight on  $\gamma_2$ . For this reason  $\beta(z_1)$  is the flattest line in Figure 1 as this linear estimator has relatively little weight placed on  $\gamma_2$ , while  $\beta(z_2)$  is a much steeper line as this linear estimator weighs  $\gamma_2$  more heavily.

### 3.2 Testing for Treatment Effects

Consider using IV estimation to test for whether the treatment has any effect on the outcome. The relevant null hypothesis is that each marginal effect is zero:  $\gamma_s = 0$  for all  $s$ . Suppose we follow the previous literature in using the linear IV estimator  $\beta(z)$  from (3) to test this null hypothesis.

There are two cases. In the first case, the null hypothesis is false and at least one of the marginal effects are non-zero:  $\gamma_s \neq 0$  for some  $s$ . Under-rejection of the false null hypothesis occurs when  $\beta(z) = 0$  and we fail to reject the false null hypothesis of no treatment effect. One possibility for under-rejection is when the linear IV estimands of the marginal effects are non-monotonic and cancel each other out. In fact, the linear IV estimate can suggest no treatment effect even if *all* marginal effects are non-zero. Figure 1 presents an example of this case. For the linear IV estimator using  $z_1$  as the instrument,  $\beta(z_1) = 0$  when  $\gamma_2 = -2$ . In this case, the linear estimator is zero, even though both marginal treatment effects are non-zero.

Another possibility for under-rejection occurs when the IV weights associated with non-zero marginal effects are zero. Considering again the example in Figure 1. When  $\gamma_2 = 0$ , the linear IV estimator using  $z_2$  indicates a zero treatment effect, despite the non-zero marginal effect at the 0 to 1 margin. The reason is that the instrument  $z_2$  only affects the second treatment margin,  $\gamma_2$ .

In the second case, the null hypothesis is true and over-rejection would have occurred if  $\beta(z) \neq 0$ . However, under the true null hypothesis, the linear IV estimator correctly imposes the constant marginal treatment effects restriction. As a consequence, the bias in using linear IV estimators when the marginal effects are non-constant is one-sided, and can only lead to under-rejection of the null hypothesis of no treatment effect.

### 3.3 Testing for Selection Bias

Following Hausman (1978), a standard test of selection bias is to compare the linear OLS and IV estimates. The idea is that if the regressor of interest is exogenous, the OLS and IV estimates would differ only by sampling error. For example, Caceres-Delpiano's (2006) study of family size effects and children's outcomes concludes that "the two-stage least-squares estimates are statistically distinguishable from OLS estimates, indicating an omitted variables bias in the single equation model" (p. 738). In the general case of treatment effect heterogeneity, the linear OLS and IV estimators can differ even if there is no selection bias, because the LATE identifies the average treatment effect of the subgroup of individuals affected by the particular instrument chosen. As shown below, however, even with homogeneous treatment effects the Hausman test for selection bias can be misleading if the marginal effects are non-constant.

Consider testing the null hypothesis that there is no selection bias, which we can write as  $\Delta_s = 0$  for all  $s$ . As before, suppose we use linear OLS and IV estimators to test the null hypothesis. We reject the null hypothesis if  $\beta(OLS) \neq \beta(z)$ , and fail to reject otherwise. The difference between the linear estimand (5) and the linear IV estimand (3) is

$$\beta(OLS) - \beta(z) = \sum_{s=1}^{\bar{s}} w_s(OLS)\gamma_s(OLS) - w_s(z)\gamma_s. \quad (7)$$

There are two cases. In the first case, the null hypothesis is true and  $\gamma_s(OLS) = \gamma_s$  for all  $s$ . Over-rejection of the true null hypothesis occurs when  $\beta(OLS) - \beta(z) \neq 0$ . If the marginal effects are constant ( $\gamma_s = \beta$  for all  $s$ ) or the OLS and IV weights assigned to non-constant marginal effects are the same ( $w_s(OLS) - w_s(z) = 0$ ), then  $\beta(OLS) = \beta(z)$  and the Hausman test is a valid test of selection bias. However, in the general case of non-constant marginal effects,  $\beta(OLS)$  may differ from  $\beta(z)$  even without selection bias if the OLS and IV weights are different:  $w_s(OLS) - w_s(z) \neq 0$  for some  $s$ . Intuitively, the Hausman test over-rejects the null hypothesis because it confuses selection bias with differences in the linear OLS and IV estimators due to different weighting of non-constant marginal effects.

Figure 1 presents an example of this case. Recall that this figure is constructed from a simulation imposing the null hypothesis of no selection bias. When the linear model is correct ( $\gamma_1 = \gamma_2 = 1$ ), the linear OLS and IV estimators provide the same estimate. However, as we introduce nonlinearities, these estimators diverge because of different weighting of

the marginal effects. To provide a sense of the over-rejection that is possible given non-linear treatment effects, we estimate for each instrument the p-value of the Hausman test for selection bias. Figure 2 plots the average p-values across the 500 replications of the data generating process, where we have normalized the p-value when the linear model is correct at 1.<sup>11</sup> This figure illustrates that the p-value for the Hausman selection bias test falls as we increase the degree of nonlinearity. As we move away from the linear model, the fall in the p-value indicates that we are over-rejecting the true null hypothesis of no selection bias. The graph of p-values suggests that even modest degrees of nonlinearity can change the probability of rejection of this commonly used test for selection bias.

Next, consider a second case in which the null hypothesis is false and there is selection bias:  $\gamma_s(OLS) \neq \gamma_s$  for at least one  $s$ . Under-rejection occurs when  $\beta(OLS) = \beta(z)$ , which means that selection bias at different treatment margins offset each other. Figure 3 presents an example of this case, graphing the three linear estimators under selection bias:  $Cov(d_{si}, v_i) \neq 0$  for  $s = 1, 2$ . As described above, we have constructed positive selection bias which shifts the linear OLS estimate up. In Figure 3, the linear OLS estimator intersects with each of the linear IV estimators. For this simulation, the linear IV estimator using instrument  $z_1$  is equal to the linear OLS estimator,  $\beta(z_1) = \beta(OLS)$ , when  $\gamma_2 = -2$ . In comparison,  $\beta(z_2) = \beta(OLS)$  when  $\gamma_2 = 2.5$ . These two intersection points indicate under-rejection of the null hypothesis, as the linear OLS and linear IV estimators are equal even though the level of treatment is endogenously determined.

In Figure 4, we provide the relative p-values for this simulation. Because the two linear IV estimators are equal to the linear OLS estimate at different degrees of nonlinearity, the peak of the p-value graphs is in different locations on the  $\gamma_2$  axis. Notice that for the linear estimator using  $z_1$ , values of  $\gamma_2 < 1$  lead to under-rejection of the false null hypothesis and higher p-values than at the point where the linear model is correct. An interesting aspect of this figure is that the p-value for the selection bias test falls below the level for the linear model at some points along the  $\gamma_2$  axis. This is because at some values of  $\gamma_2$ , both the selection bias and the nonlinearities push the linear IV estimators away from the linear OLS estimators.

To judge to what extent linear OLS and IV estimates differ because of different weighting of marginal effects, we propose a diagnostic tool that compares an IV-weighted OLS estimator, defined as  $\beta(OLS, z) = \sum_{s=1}^{\bar{s}} w_s(z)\gamma_s(OLS)$ , to the linear IV estimator. The difference between the IV-weighted OLS estimand and the IV estimand is given by

$$\beta(OLS, z) - \beta(z) = \sum_{s=1}^{\bar{s}} w_s(z)\Delta_s. \quad (8)$$

If the null hypothesis of no selection bias holds, i.e.  $\Delta_s = 0$  for all  $s$ , then  $\beta(OLS, z) =$

---

<sup>11</sup>The level of the p-value is not informative for this simulation example since it can be manipulated by changing the sample size or the degree of dispersion in the data.

$\beta(z)$ . Hence, by re-weighting the OLS estimator, we no longer confuse selection bias with differences in the OLS and IV estimators due to different weighting of non-constant marginal effects. An attractive aspect of this diagnostic tool is that it does not involve IV estimation of marginal effects, which may be difficult in practice because of too few or too weak instruments. All that is needed is to compute the weights underlying the linear IV estimator and estimate the marginal effects using OLS.

### 3.4 Testing for Instrument Invalidity

If the researcher has more instruments than included endogenous regressors, Hausman (1978) proposed an over-identification test for whether the additional instruments are valid (uncorrelated with the error term). A simple form of this test is to compare the linear IV estimator using different instruments.<sup>12</sup> The idea is that if all instruments are valid, the estimates should differ only as a result of sampling error.

A difficulty in interpreting the over-identification test is that varying the instrument shifts the weights assigned to different marginal effects. The difference between the IV estimand using instrument  $z$  and the IV estimand using instrument  $z'$  can be expressed as:

$$\beta(z) - \beta(z') = \sum_{s=1}^{\bar{s}} (w_s(z) - w_s(z')) \gamma_s + \Omega(z) - \Omega(z'), \quad (9)$$

where

$$\Omega(w) = \frac{Cov(v_i, w_i)}{Cov(w_i, s_i)}, w = z, z'.$$

Assume that instrument  $z_i$  is valid, implying that  $\Omega(z) = 0$ . Suppose that we want to test the null hypothesis that instrument  $z'_i$  is also valid,  $\Omega(z') = 0$ , and we use the linear IV estimators to test this hypothesis.

There are two cases. When the null hypothesis is true, over-rejection occurs when  $\beta(z) - \beta(z') \neq 0$ . Figure 3 presents an example of this case. In this figure, instruments  $z_{1i}$  and  $z_{2i}$  are both valid as they are uncorrelated with  $v_i$  by construction. When the linear model is correct ( $\gamma_2 = 1$ ), the two linear estimates are equal:  $\beta(z_1) = \beta(z_2)$ . As we introduce nonlinearity ( $\gamma_2 \neq 1$ ), the two linear IV estimators diverge from each other. This illustrates that the nonlinearity can lead to the erroneous conclusion that the IV estimates are different, which in the spirit of the Hausman test indicates that (at least) one of the instruments is invalid.

Next, consider a situation where the null hypothesis is false. Figure 5 illustrates this case. To do so, we introduce a third instrument,  $z_{3i} = \eta_i + 0.01v_i$ , with  $\eta_i \sim N(0, 1)$ .

---

<sup>12</sup>One could also compare the IV estimator using all of the instruments versus the IV estimator omitting some instruments variables.

The  $z_{3i}$  instrument is constructed to be invalid since  $cov(z_{3i}, v_i) = 0.01Var(v_i) \neq 0$ . In comparison, instruments  $z_{1i}$  and  $z_{2i}$  are valid, constructed as in Figure 3 (Case 2, selection bias). When the linear model is correct ( $\gamma_1 = \gamma_2$ ), the valid linear IV estimators are equal:  $\beta(z_1) = \beta(z_2)$ , modulo sampling error. However, because  $z_{3i}$  is an invalid instrument, the linear IV estimator  $\beta(z_3)$  provides a different estimate when the linear model is correct. As we introduce nonlinearities ( $\gamma_2 \neq 1$ ), the linear IV estimator using the invalid instrument  $z_{3i}$  intersects with the linear IV estimators using the valid instruments  $z_{1i}$  and  $z_{2i}$ . This illustrates a case where the linearity restriction erroneously suggests that an instrument is valid when in fact it is invalid.

## 4 Empirical Analysis

This section consider three applications with variable treatment intensity, in order to illustrate the implications of nonlinearities in IV estimation of linear models. We use the same data and instrument as Angrist and Krueger (1991) and Aakvik et al. (2010) in the return to schooling application, Black et al. (2005a) in the family size application, and Løken (2010) in the family income application. We refer to these studies for detailed description of the data, the identification strategy, as well as of relevant institutional details.<sup>13</sup>

### 4.1 Returns to Schooling Application

Research on the labor market returns to human capital has a long history in economics. Mincer (1958, 1974) provided two theoretical motivations for a log-linear specification between earnings and schooling, one based on the principle of compensated differentials and a second based on an accounting identity model of human capital formation. The two models are economically distinct, but both lead to a log-linear specification of the earnings equation. Under the simplifying assumptions used by Mincer, the coefficient on the linear schooling term equals the internal rate of return to schooling. This is a fundamental economic parameter that is often used to understand individuals' choice of schooling and assess whether public expenditure on education should be increased or decreased.<sup>14</sup>

Unfortunately, data do not support the assumptions underlying the log-linear earnings function (see e.g. Heckman et al., 2008). In particular, there is a considerable body of evidence suggesting higher returns to degree completion years (see e.g. Jaeger and Page, 1996). A log-linear relationship between earnings and schooling is also at odds with more realistic models of earnings. For instance, high returns to degree completion years

---

<sup>13</sup>Our analysis builds on Mogstad and Wiswall (2010) in the family size application, and Løken et al. (2010) in the family income application.

<sup>14</sup>See Heckman et al. (2008) for a detailed discussion of the interpretation of return to schooling estimates

are consistent with sorting models of education (Weiss, 1983), in which employers use credentials to identify workers with desirable traits that cannot be directly observed, as well as human capital models where lumpiness in the learning process leads to more skill acquisition in degree years than in preceding years (Chiswick, 1973). Nevertheless, the log-linear earnings equation continues to be applied in recent work, especially in IV estimation where the focus has been on identifying exogenous sources of variation that can be used as instruments for schooling. Our aim is to assess the role of the linearity restriction in schooling for the conclusions drawn from IV estimation of the returns to schooling.

#### 4.1.1 Data and empirical strategy

We consider two different data sources. First, we use the same data and instruments as Angrist and Krueger (1991), who were among the first to perform IV estimation of the return to schooling. This study was also the starting point for later research on weak instruments and population heterogeneity in the marginal effects of schooling.<sup>15</sup> While these issues are important, our focus is on illustrating the importance of the linearity restriction in IV estimation of the return to schooling.<sup>16</sup> As in Angrist and Krueger (1991), our sample consists of 247,199 full-time employed males who were aged 40-49 in the 1970 US Census. In line with this study, we use log earnings as the dependent variable. As an exogenous source to variation in schooling, we follow Angrist and Krueger (1991) in using the individual's quarter of birth since individuals born later in a year are obliged to stay in school longer due to compulsory school laws. Specifically, we use their full set of instruments: 3 quarter of birth dummy variables interacted with 10 year of birth indicators (30 total instruments). We also use the same set of covariates as Angrist and Krueger, which includes a quadratic specification in age (measured in quarters) as well as dummy variables for race, marital status, region of residence, urban residence, and year of birth.

As a second example, we use the same data and instruments as Aakvik et al. (2010). This study explores the impact of a mandatory education reform on educational attainment and returns to schooling in Norway.<sup>17</sup> The reform extended the minimum period of mandatory education from seven to nine years, and was implemented during a 12-year period from 1960 to 1972. Two distinct school systems coexisted in this period, provid-

---

<sup>15</sup>See e.g. Staiger and Stock (1997) for a discussion of weak instruments, and Card (2001) for a discussion of how the instruments used have been relevant for the educational decisions of individuals with relatively high returns to schooling. The argument is that instruments commonly used, including college proximity and compulsory schooling laws, mostly affect the education decision of persons with relatively poor family backgrounds, and that the return to more schooling in this subset of the population is relatively high.

<sup>16</sup>We have performed the same analysis using the data and instruments in Card (1993). The results are very similar.

<sup>17</sup>This education reform has been used as an instrument for schooling in a number of other studies, such as Black et al. (2005b).

ing a natural experiment in which some municipalities provided nine years of mandatory schooling, while others provided only seven years. Aakvik et al. (2010) specify a generalized Roy model in the context of a control function framework with eight different types of education levels. To deal with selection bias, they use the following set of instruments: i) a reform indicator variable for the mandatory schooling reform, ii) interaction terms between the reform dummy and the pre-reform availability of other types of schools (vocational schools, upper secondary schools, regional colleges, and universities), and (iii) interaction terms between the reform dummy and family background variables (father's and mother's income in 1970 in quartiles, and the father's and mother's education represented by a dummy variable indicating whether they have college degrees). In our analysis, we use the same set of instruments. We also follow Aakvik et al. (2010) in controlling for cohort and municipality fixed effects as well as family background characteristics. The population of study consists of more than 220,000 males who were born in the years 1947-1958. Our dependent variable is log annual earnings in 1995 (excluding observations with zero earnings). Unlike Aakvik et al. (2010), we use a standard IV approach (rather than a control function framework), and define schooling as a continuous variable (instead of using the eight-level categorization). In doing so, we conform to most previous studies using an IV approach to address the issue of selection bias when estimating the returns to schooling.

#### 4.1.2 Results

We first replicate the results of Angrist and Krueger (1991), using their log-linear specification in earnings. The linearly restricted OLS estimate suggests a 7.01 percent return to schooling, and the linearly restricted IV estimate implies a 10.07 percent return to schooling. In both cases, the estimates are statistically significant from zero at the 5 percent confidence level.

To interpret the linear estimates, Figure 6 graphs the linear OLS and IV weights. These weights are constructed taking into account the covariates and multiple instruments, as discussed above. We see that the linear IV estimator places the most weight on high school grades, the margins from 11-12 and 12-13 years of school, mirroring that the source to exogenous variation in schooling is the compulsory schooling law. However, the instrument does not isolate just these particular margins. In fact, the quarter of birth instruments assign considerable weight to the marginal effects in the upper part of the schooling distribution. For instance, over 1/3 of the total weight is placed on the 13-14th and higher schooling margins, and 16 percent of the total weight is assigned to the 15-16th and higher schooling margins. Consequently, it is not clear what we can learn from the linear IV estimate, as it is a weighted average of marginal effects that may be of quite different magnitude.<sup>18</sup>

---

<sup>18</sup>While this weighting is in part due to the covariates used, the considerable weighting on higher

From Figure 6, we also see that the linear IV estimator assigns negative weights to a few of the low schooling margins (0-1, 1-2, and 2-3). The negative weighing suggests a violation of the monotonicity assumption. Even though the negative weights are of relatively little quantitative importance since the weights are small in absolute value, the results serves as an example of how linear IV estimators may assign negative weights to marginal effects.

A standard test of selection bias is to compare OLS and IV estimates from a linear model. In our case, the IV estimate is more than 40 percent higher than the OLS estimate (although the IV estimate is too imprecisely estimated to reject that they are the same). As discussed above, the difference between linear OLS and IV estimates might be attributable to different weighting of marginal effects, rather than selection bias. To examine this, it is useful to compare the linear OLS and IV weights displayed in Figure 6. We see that the distribution of weights are surprisingly similar. To judge more precisely to what extent linear the linear OLS and IV estimates differ because of different weighting of marginal effects, we compute the IV-weighted OLS estimate, defined by (8). The IV-weighted OLS estimate is equal to 0.0714, confirming that the difference between linear OLS and IV estimates is not driven by different weighting of marginal effects.

Next, we explore the possibility of nonlinearities in marginal effects by specifying a non-parametric model in schooling, as in (2). Specifically, we replace the linear specification of schooling with dummy variables for all 18 distinct schooling margins (0–1 year, 1–2 year, etc.), and re-estimate the OLS returns to schooling. Figure 7 plots the OLS estimates of the marginal effects, along with 95 percent confidence intervals. For sake of comparison, we also include the linearly restricted OLS estimate of the marginal effect, which is a flat line at 7 percent. With this large sample of close to a quarter million observations, we are able to precisely estimate most of the schooling margins after the 8-9 years of schooling margin. The pattern of estimates is bimodal, with high returns to degree completion years. The 13th year of schooling (high school completion) has a marginal return of 11 percent, and the marginal return to the 16th year is as high as 20 percent (college completion). In contrast, the marginal return to most other margins are around 5 percent or less. Our finding of strong nonlinearities in the OLS estimates conform to a considerable body of evidence suggesting higher returns to degree completion years

Next, we turn to examining how fragile the IV estimates are to relaxing the linearity restriction in schooling. For sake of comparison, the first row of Table 1 repeats the IV estimate from the log-linear model, whereas the next three rows reports the IV results

---

schooling margins may also give some cause for concern about the validity of the instrument. Angrist and Krueger (1991) examine this issue by using quarter of birth to predict college attendance and other later education outcomes. They argue that the weak and imprecise correlation between quarter of birth and these outcomes support the validity of the instrument. However, our results illustrate that the linear IV estimator places considerable weight on these higher margins. The concern for instrument validity revealed by this decomposition is that the weighting on the higher margins might be due to some correlation between the quarter of birth instruments and the probability of taking higher education.

from more flexible parametric models, all of which nest the log-linear model. In the second row, we specify earnings as a quadratic function of schooling, and instrument for both the linear and squared schooling variables. In this case, the IV estimates become too imprecise to reject the hypothesis that there is a zero return to schooling at the 5 percent confidence level. The same holds for a number of other models that nest the log-linear specification. While the third row specifies earnings as a cubic function of schooling, the fourth row allows for nonlinearities in the relationship between log earnings and schooling by adding dummy variables for high school and college completion. We have also replaced the linear specification in schooling with dummy variables for all 18 schooling margins. When performing IV estimation of this non-parametric model in schooling, we find that none of the estimated marginal effects are significantly different from zero at the 5 percent confidence level. The fragility of the linearly restricted IV estimator may not be surprising given the relatively weak correlation between the instruments and schooling, as previously noted. Nevertheless, our results illustrate the importance of the linearity restriction to IV estimates of the relationship between log earnings and schooling, since we are unable to rule out a *zero* return to schooling when we use more flexible parametric models.

In the remaining rows of Table 1, we report IV estimates from two models that restrict the relationship between log earnings and schooling to be captured by a single regressor, as in the log-linear specification. Although neither model nests the log-linear specification, they are both alternative approximations of the relationship between log earnings and schooling that preserve the efficiency gain from only having to instrument for one endogenous regressor. The fifth row replaces the linear schooling term with a squared schooling term, and the sixth row uses a cubic schooling term instead of a linear schooling term. The estimated coefficient on the squared schooling term is statistically significant at the 5 percent confidence level. This also holds true for the estimated coefficient on the cubic schooling term. Furthermore, it is instructive to note that both the model with a squared schooling term and the model with a cubic schooling term have greater R-square values than the log-linear model. Therefore, in terms of precision and model fit, there seems to be no reason to prefer a linearly restricted model in the IV estimation of the return to schooling.

Figure 8 graphs the IV estimates from the log-linear model, the model with a squared schooling term, and the model with a cubic schooling. As expected, we see that the estimated returns to another year of schooling vary substantially between the models. For example, the model with a squared schooling term suggests an 8 percent marginal return to the 11th year of schooling, and a nearly 15 percent marginal return for completing the 18th year of schooling. In comparison, the model with a cubic term implies a 7.3 percent marginal return to the 11th year of schooling, and almost 19.5 percent return for completing the 18th year of schooling. These results illustrate that the choice between different parametric models that generate sufficient precision in the IV estimates may

substantially affect the conclusions drawn about the returns to schooling.

To achieve sufficiently precise IV estimates of models that nest the log-linear specification, we next turn to a different data source and instrument set taken from Aakvik et al. (2010). Table 2 reports IV results from the alternative specifications of the earnings equations, whereas Figure 9 plots the IV estimates of the marginal effects.<sup>19</sup> The first row of this table shows the results from the log-linear specification, suggesting a 14.95 percent return to schooling. In the second row, we begin exploring for the possibility of nonlinearity in marginal effects by specifying log earnings as a quadratic function of schooling, and instrument for both the linear and squared schooling variables. In this case, the IV estimates are highly significant, suggesting a decreasing concave relationship. For example, the IV estimates suggest about 16 percent marginal return to the 11th year of schooling and as little as 3.4 percent marginal return for completing the 18th year of schooling. Next, the third row allows for nonlinearities in the relationship between log earnings and schooling by adding dummy variables for high school completion.<sup>20</sup> Also in this case, the IV estimates are highly significant. In particular, our results illustrate that taking into account the relatively large return to high school completion reduces the coefficient on the linear schooling term by as much as 40 percent. Table 2 also reveals that both the quadratic model and the model with the dummy variable at high school completion have greater R-square values than the log-linear model. Therefore, in terms of precision and model fit, there seems to be no reason to impose the linearity restriction in the IV estimation of the returns to schooling.

Admittedly, a more flexible specification than what is used in Table 2 would likely provide a better approximation of the empirical relationship between log earnings and schooling. To examine this, we have performed IV estimation of less restrictive parametric models. The estimates from these models are unfortunately too imprecise to draw any conclusions about nonlinearities. At the very least, the specifications used in the second and third row nests the specification used in the first row, and is therefore an improvement over the log-linear specification employed in most previous studies.

### 4.1.3 Discussion

Our results demonstrate that the linearity restriction in schooling plays an important role in the conclusion drawn about the returns to schooling. In line with previous studies, our OLS results indicate relatively high returns to degree completion years (high school

---

<sup>19</sup>The linear OLS estimate indicates a 5.74 percent return to schooling. The OLS estimates of a non-parametric model in schooling suggest a bimodal pattern, with much higher returns to degree completion (12 and 16 years). We are able to precisely estimate the marginal effects at all schooling margins (8-18 years). The first stage results are very strong for each model in Table 2, with F-values on the excluded instruments around 34.

<sup>20</sup>In a similar way, Harmon and Walker (1999) allow for a change in the returns to schooling around high school completion in their IV estimation. Their results suggest strong nonlinearities in the returns to schooling using data from the UK.

or college). And when allowing for a nonlinear relationship between log earnings and schooling in the IV estimation, we find that the estimated returns to schooling are quite sensitive to the parametric restrictions made. The evidence of substantially different marginal effects of an additional year of schooling depending on the choice of functional form underscores that careful exploration of possible nonlinearities is of central importance in empirical research that aims to understand individuals' choice of schooling.

Nonlinearities in the returns to schooling is also of interest from a policy perspective. In most developed countries, government subsidies and regulations vary systematically through the education system, from pre-school through secondary and post-secondary levels. To understand the marginal costs and benefits of these policies, it seems crucial to have some understanding of how the return to an additional year of schooling varies with the level of schooling. For instance, the evidence of relatively high return to high school completion would seem to suggest that there are great benefits to policies that reduce high school completion rates.

## 4.2 Family Size Application

Although the seminal quantity-quality model of fertility by Becker and Lewis (1973) assumes that family size and child quality are jointly determined, there is nothing in the theory which suggests that the relationship is linear. On the contrary, even with no heterogeneity in the production function of child quality, there can be nonlinearities in child quality from changes in family size (Mogstad and Wiswall, 2010). This is because parental preferences mediate changes in family size such that an increase in the number of children can have large or small, and negative or positive, effects on existing children. Hence, the relationship between family size and child outcome may not only be nonlinear but even non-monotonic.

The possibility of nonlinearities in the relationship between family size and child quality was pointed out by Rosenzweig and Wolpin (1980), but has received little attention in the subsequent empirical literature. Based on a linear model in family size, the pioneering study by Black et al. (2005a) concludes that “there is little if any family size effect on child education; this is true when we estimate the relationship with controls for birth order [OLS] or instrument family size with twin births” (p. 697). They therefore argue that we need to “revisit models of fertility and reconsider what should be included in the production function of children” (p. 698).<sup>21</sup> Motivated by the theoretical ambiguity about the magnitude and sign of the marginal effects on child quality of additional siblings, our point of departure is to allow for a nonlinear relationship between family size and child outcome in the IV estimation.

---

<sup>21</sup>Using data from the US, Caceres-Delpiano (2006) comes to a similar conclusion. Other recent studies reporting no effect of family size include Angrist et al. (2010) using data from Israel and Aaslund and Grønqvist (2010) using data from Sweden.

### 4.2.1 Data and empirical strategy

As in Black et al. (2005a), our data is based on administrative registers from Statistics Norway covering the entire resident population of Norway, who were between 16 and 74 years of age at some point during the period 1986-2000. Following Black et al. (2005a), the sample is restricted to children who were aged at least 25 in 2000 to make it likely that most individuals in our sample have completed their education. Conforming to Black et al. (2005a), our measure of family size is the number of children born to each mother. The only difference between our sample selection and that in Black et al. (2005a) is that we exclude a small number of families (less than 2 percent) with more than 6 children. Our reason for excluding large families is that the estimates of the marginal family size effects are unstable and imprecise for families with more than 6 children. The linear OLS and IV estimates barely move when including families with 6 or more children.

Following Black et al. (2005a), we use twin births as a source of exogenous variation in family size, conditional on the same set of controls for children's birth order and other characteristics.<sup>22</sup> The rationale for using twins as instruments is that for some families, twin births increase the number of siblings beyond the desired family size. In our sample, a twin birth occurs in around 1.4 percent of families.

The main empirical model used in the family size literature specifies outcomes for children as a linear function of their family size, as in (1). Our point of departure is to specify a non-parametric model in family size by including dummy variables for each number of siblings, as in (2). To avoid including the endogenously selected outcomes of children born after the twin birth as well as of twins themselves, we follow Black et al. (2005a) in restricting the sample to first born children in families with 2 or more children. In the IV estimation of the linear model, the instrument for family size is a dummy variable for whether the second birth was a twin birth (implying that the second and third born children are twins). When performing IV estimation of the non-parametric model in family size, there are four endogenous marginal family size variables that need to be instrumented for. Following Mogstad and Wiswall (2010), our strategy is to use twins at second, third, fourth, and fifth birth as instruments for the four endogenous dummy variables: 2 or more, 3 or more, 4 or more, and 5 or more siblings.

### 4.2.2 Results

As in Black et al. (2005a), OLS on the linear model yields a precisely estimated effect of family size on children's education that is very close to zero. However, the OLS estimates of the non-parametric model in family size reveal an inverse U-shaped pattern, with family size effects that are always statistically significant (at 5 percent level) and often

---

<sup>22</sup>Following Rosenzweig and Wolpin (1980), twin births have been frequently used as an exogenous shock to family size. See for example Black et al. (2005a) for results supporting the internal validity of twin birth as instrument for family size.

sizeable. Figure 10 graphs the predicted average child education from the OLS estimation of the linear and non-parametric model in family size. The figure presents educational attainment relative to only children, whose average educational attainment is normalized to 0. The slopes in this figure are the marginal effect estimates. Contrasting the estimates from the two types of models indicates the extent to which OLS estimation of the linear model approximates the relationship between child education and family size.

Examining Figure 10, we see a stark contrast between the estimates from the linear and non-parametric models in family size. While the linear OLS estimate is as small as 0.0001, the non-parametric results suggest that having one sibling increases educational attainment of first born by as much as 0.245 years. In contrast, adding a third sibling is estimated to reduce educational attainment of first born children by 0.086 years, adding a fourth sibling reduces education an additional 0.157 years, and adding a fifth sibling reduces education an additional 0.107 years.

Table 3 presents IV and OLS results for the linear model in family size in Panel I and the non-parametric models in family size in Panel II, when restricting the sample to first born children from families with 2 or more children. The strong first stage results are reported in Mogstad and Wiswall (2010). The first column of Panel I shows a linear effect of family size that is insignificant and close to zero, as in Black et al. (2005a). Turning attention to the IV results of the non-parametric model in family size, reported in the first column of Panel II in Table 3, the main finding is that there are significant and large marginal family size effects on children's education. Furthermore, the results indicate a non-monotonic causal relationship between family size and children's education. On the one hand, a third child is estimated to increase completed education by 0.15 years. On the other hand, changes in family size are estimated to reduce children's education by 0.47 years for a fourth child, another 0.8 years for a fifth child, and an additional 0.79 years for a sixth child. It should be noted that these marginal family size effects exceed the birth order effects that Black et al. (2005a) emphasize as large.

Both the OLS and IV results suggest that the linear model does a poor job approximating the relationship between child education and family size. However, as discussed above, a linear model will nevertheless produce some weighted average of the underlying marginal effects. The question is how the linear estimates should be interpreted, and what policy questions they could answer. For instance, if having twins on second birth only affects the probability of having 3 instead of 2 children, then IV estimation of the linearly restricted model would identify solely the marginal effect of moving from 2 to 3 children, which is readily interpretable. However, this is not the case. Figure 11 displays the linear IV weights. As expected, using twins at second birth as the instrument weights the 2 to 3 children margin most heavily (76 percent), but also places considerable weight on the marginal effects at higher parities (24 percent). Consequently, it is not clear what we can learn from the linear IV estimates, as they are arbitrarily weighted averages of

several marginal effects, which vary in sign and magnitude.

Figure 11 also shows that the distribution of OLS and IV weights are quite different. While the linear OLS estimator places 44 percent weight on the 1-2 child margin, the linear IV estimator places 76 percent of the weight on this marginal effect. As discussed above, a standard test of selection bias is to compare OLS and IV estimates from a linear model. By comparing the linear IV estimate (-0.0036) to the linear OLS estimate (-0.063), we see that the point estimates differ, although the IV estimate is too imprecisely estimated to reject that they are the same. By comparing the IV-weighted OLS estimate (-0.032) to the linear IV estimate (-0.0036), we learn that as much as half of the difference between the OLS and IV estimates is a result of different weighting of the marginal effects.

### 4.2.3 Discussion

Our results demonstrate that the conclusion of no family size effect is an artifact of a linear specification in family size, masking substantial marginal family size effects. This is true when we perform OLS estimation with controls for confounding characteristics such as birth order, or instrument family size with twin birth. In terms of the quantity-quality model, the estimated marginal effects can be interpreted as suggesting a tradeoff between quantity and quality in large families, and (strong) complementarities between quantity and quality in small families.

The finding of a nonlinear relationship between family size and children's outcomes is also of interest from a policy perspective. Most developed countries have a range of policies affecting fertility decisions, including publicly provided or subsidized child care as well as welfare and tax policies, such as maternity leave laws, family allowances, single parent benefits, and family tax credits. In fact, families with children receive special treatment under the tax and transfer provisions in twenty-eight of the thirty OECD countries (OECD, 2002). Many of these policies are designed such that they reduce the cost of having one child more than the cost of having additional children, in effect promoting smaller families.<sup>23</sup> If a policy goal is to slow or reverse the unprecedented fertility decline most developed countries have experienced over the last 30 years, the effects of family size on children's outcomes become ever more important. Accepting the recent findings of no effect of family size suggests that there is no need to be concerned with the externalities on the human capital development of existing children when designing policies promoting larger families. Our findings run counter to this conclusion. The

---

<sup>23</sup>For example, welfare benefits are, in many cases, reduced or even cut off after reaching a certain number of children. In the US, a recipient of the Earned Income Tax Credit program could in 2007 receive a maximum credit of USD 2,900 if he or she had one qualifying child; for two or more qualifying children, the maximum credit was only USD 4,700. In addition, a number of US states have implemented family cap policies, providing little or no increase in cash benefits when a child is born to a mother who is on welfare. Some developing countries have implemented far more radical policies to promote smaller families, such as China's One Child Policy and an aggressive public promotion of family planning in Mexico and Indonesia.

evidence of an inverse U-shaped pattern suggests that an efficient policy might be to target incentives for higher fertility to small families, and discourage larger families from having more children.

### 4.3 Family Income Application

The seminal model of Becker and Tomes (1979, 1986) predicts an increasing concave relationship between family income and child outcome. This is because parents with low family income are more likely to be credit constrained and will therefore under-invest in their children’s human capital.<sup>24</sup> The marginal return on human capital will therefore exceed that on assets in credit constrained families. This implies that poor parents will invest more of an increase in family income in children’s human capital compared to rich parents who would give more as bequests. The theoretical prediction of a concave relationship is supported by the OLS estimates reported in Blau (1999) and Duncan et al. (2009), suggesting a much stronger association between childhood family income and child development in poor families than in rich families, both in the US and in Norway. Yet, when performing IV estimation, child outcomes are typically restricted to be a linear function of family income.<sup>25</sup> Motivated by theoretical predictions and OLS results suggesting a nonlinear relationship, we depart from previous studies in allowing the marginal effects on children’s outcomes of an increase in family income to vary across the family income distribution.

#### 4.3.1 Data and empirical strategy

As in Løken (2010), our data is based on administrative registers from Statistics Norway covering the entire resident population of Norway. We also follow Løken (2010) in our definition of family income. Income is taken from tax registers, and includes all market income, from wages and self-employment, as well as (taxable) cash benefits such as unemployment benefits, disability benefits, and sickness pay. In every year, we add the income of the child’s mother to her spouse’s income (if she has any spouse), to create one variable reflecting annual family income. This means that we measure family income as the total income of the family that the child lived in, regardless of whether the spouse of the mother is the child’s biological father. We then define family income during childhood as

---

<sup>24</sup>As emphasized by Cunha and Heckman (2007), there are two distinct types of credit constraints operating on the family and its children, which can produce a nonlinear relationship between family income and child outcome. The first constraint is the inability of parents to borrow against their children’s future income to finance investment in them. The second constraint is the inability of parents to borrow against their own income to finance investment in their children. See also Cunha and Heckman (2007) for a discussion of evidence suggesting that credit constraints operating in childhood actually affect children’s ability and educational outcome as adults.

<sup>25</sup>See Shea (2000), Oreopoulos et al. (2005), Dahl and Lochner (2008), Milligan and Stabile (2007), and Løken (2010).

the average annual family income from age 1 until age 11. We use three different measures for children's outcomes: years of education, whether the individual is a high school drop out, and an IQ test score. Years of education is defined as the number of completed years of education in 2006, whereas high school drop out is defined as not obtaining a three year high school diploma by 2006. Unlike these two outcome measures, the IQ test score is only available for males, as it is collected from military records. Since military service is mandatory for men, we have information on the IQ tests for the vast majority of males.

Conforming to Løken (2010), we use time and regional variation in the economic boom that followed the initial (offshore) discovery of oil in Norway as a source to plausibly exogenous variation in childhood family income. At the end of 1969, the first major oil discovery was made in the North Sea in Norway, and in June 1970, the public was informed of it. Eventually, the discovery of oil fueled the entire Norwegian economy, but Rogaland county was the first and most strongly influenced because the main oil production in the relevant period was located off the coast there. The large increase in labor demand from the oil industry gradually spilled over into higher wages also for other types of jobs. As in Løken (2010), our sample of children is restricted to cohorts who were born in the years prior to the initial discovery of oil in Norway, to avoid threats to the validity of the instrument from endogenous migration. In addition, the sample is restricted to children who were born in Rogaland county or in ten other counties that were geographically distant from the offshore oil fields, but with similar family and child characteristics as those in Rogaland. After these sample restrictions, we have data on more than 120,000 children.

The rationale for using the initial discovery of oil as a natural experiment is twofold. First, the extent to which childhood family income of children born in a given year are affected by the subsequent oil boom depends on the geographical proximity of their place of birth to the offshore oil fields. And second, for children born in the same place, the effect of the oil boom on childhood family income depended on their year of birth. The instrument for family income exploits the fact that the oil boom most strongly affected the childhood family income of children born in the years immediately before the discovery of oil, in the county located just off the coast of the offshore oil fields. The first and second stage specifications include fixed effects for birth cohort and county of birth, controlling for unobserved permanent differences between children born in different years as well as between children born in different areas. We refer to Løken (2010) and Løken et al. (2010) for descriptive statistics as well as an in-depth discussion of the identification strategy, including a battery of robustness checks.

The main empirical model used in the family income literature specifies children's outcomes as a linear function of family income, as in (1). Our point of departure is to relax the linearity restriction. Following Løken et al. (2010), we specify children's outcomes to be a quadratic function of family income. In both the linear and the quadratic

specification, we condition on the same set of controls as Løken (2010), which in addition to the fixed effects for birth cohort and county of birth includes dummy variables for child's birth order, number of siblings, gender, parent's age and college attendance. All the control variables are measured in the year the child is born, and therefore before our regressor of interest, childhood family income.

In the IV estimation of the linear family income model, we first use a single binary instrument, equal to 1 if the child is born in Rogaland in the years 1967-1969, and 0 otherwise. However, to identify the parameters of both the linear and the squared family income terms in the quadratic specification of family income, we need more than one instrument. To construct multiple instruments, we follow Løken et al. (2010) in interacting the above binary instrument discussed with five control variables: indicator for father attended college, indicator for mother attended college, father's age, mother's age, and an indicator for large family size (3 or more siblings). Next, we use the full set of instruments in the first stage estimations for the linear and the squared family income terms.<sup>26</sup> To provide a direct comparison between the IV results of the linear and quadratic model using the same set of instruments, we also estimate the linear model with the full set of instruments.

### 4.3.2 Results

Columns 1-3 of Table 4 report the linear OLS and IV estimates for our three outcome measures: years of education, high school drop out, and IQ. From Panel I, we see that our precise OLS results indicate a positive and sizable effect of family income on children's educational attainment and IQ as adults. To get a perspective on the magnitude of the parameter estimates, a standard deviation increase in family income (NOK 101,000 or USD 17,414) is estimated to produce slightly less than 0.5 additional years of education, a fall in high school drop out rates in the range of 5 to 6 percent, and a rise in the IQ test score of more than 0.125 of a standard deviation. In Panel II, we report linear IV results when using only the binary instrument in the first stage estimation. The IV estimates show no sign of significant effects of family income on child outcome, although they are too imprecisely estimated to rule out some effect.

As discussed above, the interpretation and comparison of the linear estimates reported

---

<sup>26</sup>A number of previous studies have used this strategy to improve the precision of their IV estimates. The strategy exploits the fact that if the main instrument is valid and the included covariates are exogenous (a necessary assumption for consistent IV estimation of the linear model), then any function of the main instrument and the included covariates are valid instruments. Our reason for not interacting the Rogaland dummy variable with all the included control variables is that such a procedure would introduce a large number of over-identifying restrictions, which could increase the small sample bias of the IV estimator (see e.g. Staiger and Stock, 1997). Even though our first stage results are strong, we will therefore interact the Rogaland dummy variable with only five control variables. The reason for choosing these control variables is that they generate the strongest first stage results. We have also verified that our results are robust to performing IV estimation using only subsets of these interactions, as well when interacting the Rogaland dummy variable with every included covariate.

in Table 4 depend on the relationship between family income and child outcomes being linear or that the weights attached to the marginal effects are similar across the linear estimators. To directly examine the weighting implicit in the linear estimators, Figure 13 displays the OLS and IV weights. In order to compute the weights, we have discretized the income distribution using intervals of the family income distribution of length NOK 25,000 (USD 4,310). We see that the linear IV estimate assigns relatively little weight to the marginal effects at the lower part of the family income distribution, illustrating that the oil boom did not do that much for the family income of poor families. In comparison, the linear OLS estimator weights the different margins more evenly than the linear IV estimator. In particular, the OLS estimator assigns several times more weight to marginal effects in the lower part of the family income distribution than the linear IV estimator. This evidence of substantially different weighting of the marginal effects of family income underscores that we need to be cautious when interpreting and comparing the results from these linear estimators.

To directly test whether functional form assumptions drive the conclusions drawn about the causal relationship between family income and child outcome, we specify children's outcomes as a quadratic function of family income, with a linear and a squared family income term. Table 4 reports OLS and IV results from the quadratic specification in family income. The strong first stage results are reported in Løken et al. (2010). Panel IV reports the OLS results, whereas Panel V reports IV results using the full set of instruments in the first stage estimations. Our main finding is that all of the results suggest a concave relationship between family income and children's outcomes, with large marginal effects at the lower parts of the family income distribution. This holds true for both the OLS and IV results.

Although the full set of instruments are valid under the same assumptions as the binary instrument, they will in general identify different LATEs. Panel III therefore reports IV results from the linear model, using the same instruments as used in the IV estimates of the quadratic model, reported in Panel V. Comparing the results from the linear and quadratic specification using the *same* set of instruments, it is clear the role of the linearity restriction in masking the family income effects. The linear IV estimates are never significantly different, and always insignificantly different from zero, whereas the quadratic IV estimates using the same set of instruments show large and quite precisely estimated effects of family income. Hence, we can conclude that for a given set of instruments, the second stage restriction plays an important role in the conclusion of the effects of family income on child outcome.

It should finally be noted that our quadratic specification is intended to achieve a reasonable tradeoff between flexibility in functional form and achieving sufficient precision. We have, however, performed OLS estimation with higher order polynomials in family income. As illustrated in Figure 12, adding a cubic family income term barely moves the

marginal effect estimates over the majority of the family income distribution. We have also performed the IV estimation with a cubic term in family income. The coefficients associated with the cubic term are always insignificant, indicating that the concave specification might not be too bad. However, we admittedly cannot rule out that an even more flexible specification would provide a better approximation of the empirical relationship. At the very least, our quadratic model nests the linear model, and is therefore an improvement over the linear specification used in previous studies.

To examine whether a simple transformation of the family income variable is sufficient to capture the nonlinear relationship, we have also performed OLS and IV estimation with child outcome specified as a linear function of log family income. We find, like in Dahl and Lochner (2008), that measuring family income in logs rather than levels does not change our main results: The linearly restricted OLS estimates stay significant and substantial, whereas the linear IV estimates remain insignificant and closer to zero.

### 4.3.3 Discussion

When allowing for a nonlinear relationship between family income and child outcome, our IV results support the predictions from the seminal model of Becker and Tomes (1979, 1986) of an increasing concave relationship. In fact, the estimated marginal effects at the lower part of the family income distribution are several orders of magnitude larger than the marginal effects at the middle and upper part of the family income distribution. In contrast, the linear IV estimate suggests little, if any, effect of family income, because they assign little weight to the large marginal effects at the lower part of the income distribution. This illustrates how the conclusions drawn about the empirical content of an economic theory may depend on the chosen functional form in the IV estimation.

Our results may also be of interest from a policy perspective. Most developed countries have a range of policies targeted at family income during childhood, such as family allowances, maternity benefits, single parent benefits, and family tax credits. While some of these policies are means-tested, others are more universal of nature. The evidence of a concave relationship between family income and child outcomes, with relatively large positive effect at the lower part of the family income distribution, indicates that policies targeting poor families may be quite effective in promoting child development.

## 5 Conclusion

Mainstream economic theory postulates that the behavior of economic agents are determined by comparing benefits and costs that result from marginal changes. Similarly, to determine the optimal level of a public policy, voters or government officials must compare the marginal benefits and marginal costs. Yet, when testing theory and assessing policy,

empirical studies often use a linear IV estimator, in which the dependent variable is specified as a linear function of a potentially endogenous regressor. When this regressor takes on multiple values, the linear specification restricts the marginal effects to be constant across all margins

In this paper, we examine the implications for inference from using a linear IV estimator when the true relationship is nonlinear. Previous research has shown that the linear OLS and IV estimands can be decomposed into weighted averages of marginal effects estimates, where the OLS and IV weights generally differ. Using such decompositions, the first part of the paper examines the biases due to nonlinearity in commonly used tests for treatment effects, selection bias, and instrument validity.

To demonstrate that the implications of the linearity restriction in IV estimation are not merely theoretical peculiarities but real cause for concern, the second part of the paper considers three applications with variable treatment intensity: the economic return to schooling, the link between family size and child quality, and the effect of childhood family income on child development. The choice of applications is not incidental. In all three cases, economic theory and OLS results suggest a nonlinear relationship between the dependent variable and the regressor of interest, yet previous research has typically used linearly restricted IV estimators. In each application, we find that the IV estimates are not robust to relaxing the linearity restriction. In particular, allowing for a more flexible functional form changes the qualitative conclusions about the relevant economic theory and the effectiveness of different policies.

A general lesson to be drawn from this paper is that the linearity restriction in IV estimation may drive the conclusions reached in applications with variable treatment intensity. This finding stands in stark contrast to Angrist and Pischke (2010), who argue that the linearity restriction gives little cause for concern. Rather, our findings echo the argument of Leamer (1983) and others that exploration of possible nonlinearities is of central importance in empirical research, especially when it comes to testing theory and assessing policy.

## References

- AAKVIK, A., K. G. SALVANES, AND K. VAAGE (2010): “Measuring heterogeneity in the returns to education using an education reform,” *European Economic Review*, 54(4), 483–500.
- AASLUND, O., AND H. GRÖNQVIST (2010): “Family size and child outcomes: Is there really no trade-off?,” *Labour Economics*, 17(1), 130–139.
- ABADIE, A. (2003): “Semi-Parametric Instrumental Variable Estimation of Treatment Response Models,” *Journal of Econometrics*, 113, 231–63.
- ANGRIST, J. D., AND G. IMBENS (1995): “Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity,” *Journal of American Statistical Association*, 90(430), 431–442.
- ANGRIST, J. D., AND A. B. KRUEGER (1991): “Does Compulsory School Attendance Affect Schooling and Earnings?,” *The Quarterly Journal of Economics*, 106(4), 979–1014.
- ANGRIST, J. D., AND A. B. KRUEGER (1999): “Empirical Strategies in Labor Economics,” in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card.
- ANGRIST, J. D., V. LAVY, AND A. SCHLOSSER (2010): “Multiple Experiments for the Causal Link between the Quantity and Quality of Children,” *Forthcoming in Journal of Labor Economics*.
- ANGRIST, J. D., AND J.-S. PISCHKE (2010): “The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics,” *Journal of Economic Perspectives*, 24(2), 3–30.
- BARUA, R., AND K. LANG (2009): “School Entry, Educational Attainment and Quarter of Birth: A Cautionary Tale of LATE,” NBER Working Papers 15236, National Bureau of Economic Research, Inc.
- BECKER, G. S., AND H. G. LEWIS (1973): “On the Interaction between the Quantity and Quality of Children,” *Journal of Political Economy*, 81(2), 279–288.
- BECKER, G. S., AND N. TOMES (1979): “An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility,” *Journal of Political Economy*, 87, 1153–89.
- (1986): “Human Capital and the Rise and Fall of Families.,” *Journal of Labor Economics*, 4, 1–39.
- BLACK, S. E., P. J. DEVEREUX, AND K. G. SALVANES (2005a): “The More the Merrier? The Effects of Family Size and Birth Order on Children’s Education,” *Quarterly Journal of Economics*, 120, 669–700.
- (2005b): “Why the Apple Doesn’t Fall Far: Understanding Intergenerational Transmission of Human Capital,” *American Economic Review*, 95(1), 437–449.
- BLAU, D. M. (1999): “The Effect of Income on Child Development,” *The Review of Economics and Statistics*, 81, 261–276.

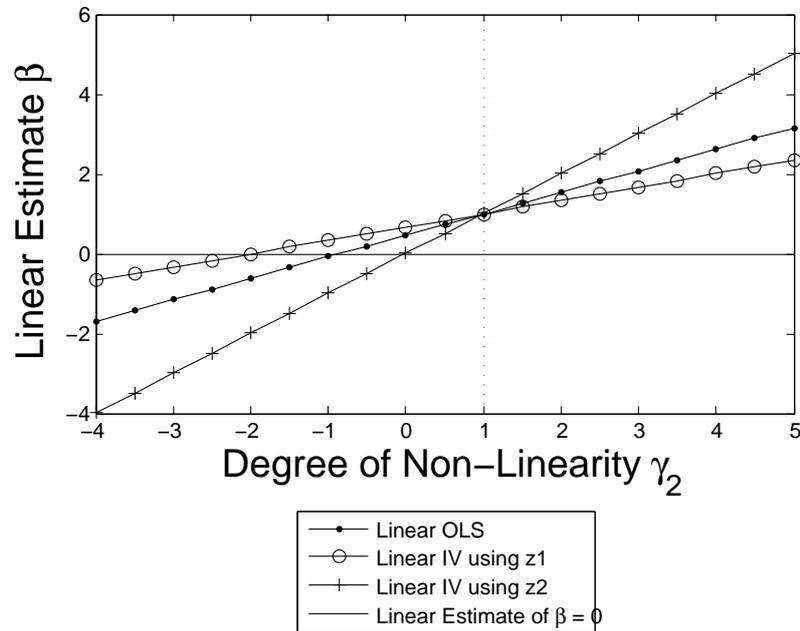
- CACERES-DELPINO, J. (2006): “The Impacts of Family Size On Investment in Child Quality,” *Journal of Human Resources*, 41(4), 738–754.
- CARD, D. (1993): “Using Geographic Variation in College Proximity to Estimate the Return to Schooling,” NBER Working Papers 4483, National Bureau of Economic Research, Inc.
- (2001): “Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems,” *Econometrica*, 69(5), 1127–60.
- CHISWICK, B. (1973): “Schooling, Screening, and Income,” in *Does College Matter?*, ed. by L. Solmon, and P. Taubman.
- CUNHA, F., AND J. HECKMAN (2007): “The Technology of Skill Formation,” *American Economic Review*, 97(2), 31–47.
- DAHL, G., AND L. LOCHNER (2008): “The Impact of Family Income on Child Achievement: Evidence from the Earned Income Tax Credit,” NBER Working Papers 14599.
- DEATON, A. (2010): “Instruments, Randomization, and Learning about Development,” *Journal of Economic Literature*, 48(2), 424–55.
- DUNCAN, G., K. TELLE, K. ZIOL-GUEST, AND A. KALIL (2009): “Economic Depreciation in Early Childhood and Adult Outcomes: Comparative Evidence from Norwegian Register Data and the US PSID,” in *Intergenerational Mobility Within and Across Nations*, ed. by M. J. R. Erikson, and T. Smeeding. Russell Sage Foundation, New York.
- HARMON, C., AND I. WALKER (1999): “The marginal and average returns to schooling in the UK,” *European Economic Review*, 43(4-6), 879–887.
- HAUSMAN, J. (1978): “Specification Tests in Econometrics,” *Econometrica*, 46, 1251–1271.
- HECKMAN, J., S. URZUA, AND E. VYTLACIL (2006): “Understanding Instrumental Variables in Models with Essential Heterogeneity,” *Review of Economics and Statistics*, 88(3), 389–432.
- HECKMAN, J., AND E. VYTLACIL (2001): “Policy-relevant treatment effects,” *American Economic Review*, 91, 107–111.
- HECKMAN, J. J. (2010): “Building Bridges between Structural and Program Evaluation Approaches to Evaluating Policy,” *Journal of Economic Literature*, 48(2), 356–98.
- HECKMAN, J. J., L. J. LOCHNER, AND P. E. TODD (2008): “Earnings Functions and Rates of Return,” *Journal of Human Capital*, 2(1), 1–31.
- HECKMAN, J. J., AND S. URZÚA (2010): “Comparing IV with structural models: What simple IV can and cannot identify,” *Journal of Econometrics*, 156(1), 27–37.
- HECKMAN, J. J., AND E. VYTLACIL (2005): “Structural Equations, Treatment Effects, and Econometric Policy Evaluation,” *Econometrica*, 73(3), 669–738.
- HOROWITZ, J. L. (2009): “Applied Nonparametric Instrumental Variables Estimation,” Discussion paper, Northwestern University.

- IMBENS, G. W. (2010): “Better LATE Than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009),” *Journal of Economic Literature*, 48(2), 399–423.
- IMBENS, G. W., AND J. D. ANGRIST (1994): “Identification and Estimation of Local Average Treatment Effects,” *Econometrica*, 62(2), 467–75.
- JAEGER, D. A., AND M. E. PAGE (1996): “Degrees Matter: New Evidence on Sheepskin Effects in the Returns to Education,” *The Review of Economics and Statistics*, 78(4), 733–40.
- KEANE, M. (2010a): “Structural vs. Atheoretic Approaches to Econometrics,” *Journal of Econometrics*, 156, 3–20.
- KEANE, M. P. (2010b): “A Structural Perspective on the Experimentalist School,” *Journal of Economic Perspectives*, 24(2), 47–58.
- LEAMER, E. (1983): “Let’s Take the Con out of Econometrics,” *American Economic Review*, 73, 31–43.
- LEAMER, E. E. (2010): “Tantalus on the Road to Asymptopia,” *Journal of Economic Perspectives*, 24(2), 31–46.
- LØKEN, K. V. (2010): “Family Income and Children’s Education: Using the Norwegian Oil Boom as a Natural Experiment,” *Labour Economics*, 17, 118–129.
- LØKEN, K. V., M. MOGSTAD, AND M. WISWALL (2010): “What Linear Estimators Miss: Re-Examining the Effects of Family Income on Child Outcomes,” IZA Discussion Papers 4971, Institute for the Study of Labor (IZA).
- MILLIGAN, K., AND M. STABILE (2007): “The Integration of Child Tax Credits and Welfare: Evidence from the Canadian National Child Benefit Program,” *Journal of Public Economics*, 91, 305–326.
- MINCER, J. (1958): “Investment in Human Capital and Personal Income Distribution,” *Journal of Political Economy*, 66, 281.
- (1974): *Schooling, Experience and Earnings*. National Bureau of Economic Research, New York.
- MOGSTAD, M., AND M. WISWALL (2010): “Testing the Quantity-Quality Model of Fertility: Linearity, Marginal Effects, and Total Effects,” Working paper, New York University.
- NEVO, A., AND M. D. WHINSTON (2010): “Taking the Dogma out of Econometrics: Structural Modeling and Credible Inference,” *Journal of Economic Perspectives*, 24(2), 69–82.
- NEWKEY, W. K., AND J. L. POWELL (2003): “Instrumental Variable Estimation of Nonparametric Models,” *Econometrica*, 71(5), 1565–1578.
- OECD (2002): *Taxing Wages: 2001 Edition*. Organization for Economic Co-operation and Development, Paris.

- OREOPOULOS, P., M. PAGE, AND A. STEVENS (2008): "The Intergenerational Effects of Worker Displacement," *Journal of Labor Economics*, 26, 455–483.
- ROSENZWEIG, M. R., AND K. I. WOLPIN (1980): "Testing the Quantity-Quality Fertility Model: The Use of Twins as a Natural Experiment," *Econometrica*, 48(1), 227–240.
- SHEA, J. (2000): "Does Parents' Money Matter?," *Journal of Public Economics*, 77, 155–184.
- SIMS, C. A. (2010): "But Economics Is Not an Experimental Science," *Journal of Economic Perspectives*, 24(2), 59–68.
- STAIGER, D., AND J. H. STOCK (1997): "Instrumental Variable Regression with Weak Instruments," *Econometrica*, 65(3), 557–86.
- STOCK, J. H. (2010): "The Other Transformation in Econometric Practice: Robust Tools for Inference," *Journal of Economic Perspectives*, 24(2), 83–94.
- WEISS, A. (1983): "A Sorting-cum-Learning Model of Education," *Journal of Political Economy*, 91(3), 420–42.
- YITZHAKI, S. (1996): "On Using Linear Regressions In Welfare Economics," *Journal of Business and Economics Statistics*, 14(4), 478–486.

## 6 Figures

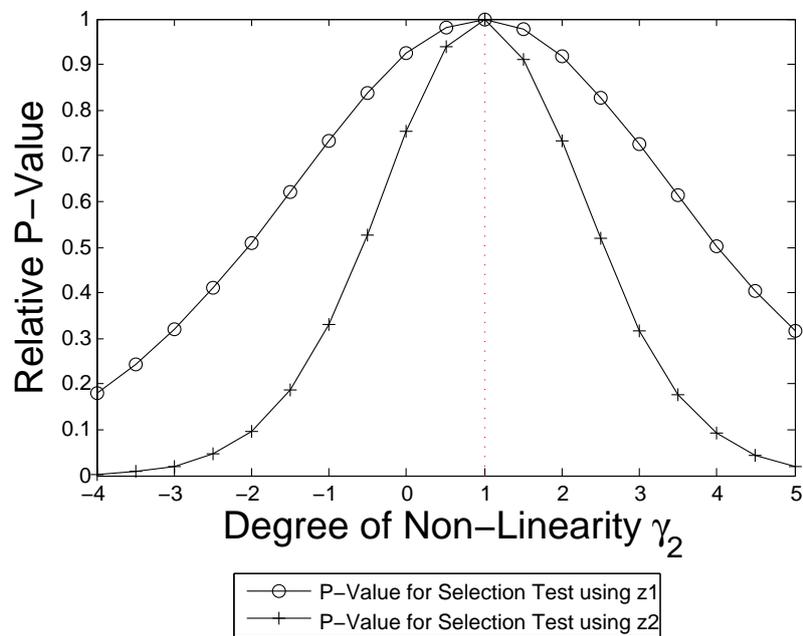
Figure 1: Linear IV and OLS Estimates in the Case without Selection Bias



Notes: This figure illustrates how nonlinearities affect OLS and IV estimates of a linear model. The underlying model is  $y_i = \gamma_1 d_{1i} + \gamma_2 d_{2i} + v_i$ , where the first marginal effect is fixed at  $\gamma_1 = 1$ . The vertical axis measures the level of one of three linear estimators: linear OLS,  $\beta(OLS)$ , linear IV using  $z_1$  as an instrument,  $\beta(z_1)$ , and linear IV using an  $z_2$ ,  $\beta(z_2)$ . The horizontal axis measures  $\gamma_2$ . At  $\gamma_2 = \gamma_1 = 1$ , the linear model is correct and the marginal effects are constant. This figure is drawn assuming no selection bias. Therefore, when the linear model is correct, all three estimators are equal:  $\beta(OLS) = \beta(z_1) = \beta(z_2)$ , modulo sampling error. As  $\gamma_2$  moves away from  $\gamma_2 = 1$ , the three estimators diverge from each other due to their different weighting of the marginal effects.

Source: Simulation from data generating process, described in Section 3.

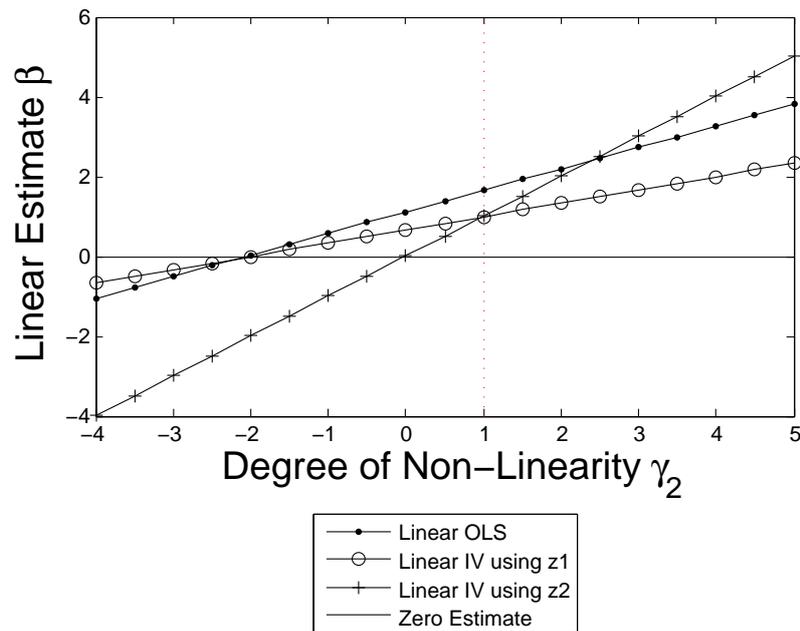
Figure 2: Relative P-Value for Test of No Selection Bias in the Case without Selection Bias



Notes: This figure is based on Figure 1. It provides the P-Value from two tests of no selection bias:  $H_0 : \beta(OLS) - \beta(z_1) = 0$  using  $z_1$  as the instrument;  $H_0 : \beta(OLS) - \beta(z_2) = 0$  using  $z_2$  as the instrument.

Source: Simulation from data generating process, described in Section 3.

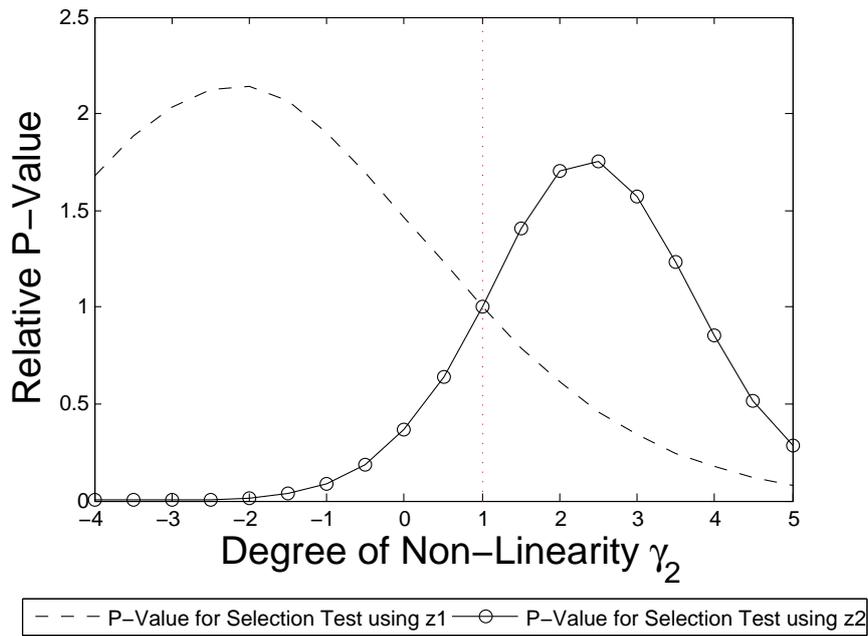
Figure 3: Linear IV and OLS Estimators in the Case with Selection Bias



Notes: This figure illustrates how nonlinearities affect OLS and IV estimates of a linear model. The underlying model is  $y_i = \gamma_1 d_{1i} + \gamma_2 d_{2i} + v_i$ , where the first marginal effect is fixed at  $\gamma_1 = 1$ . The vertical axis measures the level of one of three linear estimators: linear OLS,  $\beta(OLS)$ , linear IV using  $z_1$  as an instrument,  $\beta(z_1)$ , and linear IV using an  $z_2$ ,  $\beta(z_2)$ . The vertical axis measures  $\gamma_2$ . At  $\gamma_2 = \gamma_1 = 1$ , the linear model is correct and the marginal effects are constant. This figure assumes positive selection bias, and as a consequence  $\beta(OLS) > \beta(z)$ . However, because the treatment effects are constructed to be homogeneous, when the linear model is correct, the linear IV estimators are equal:  $\beta(z_1) = \beta(z_2)$ , modulo sampling error. As  $\gamma_2$  moves away from  $\gamma_2 = 1$ , the linear IV estimators diverge from each other due to their different weighting of the marginal effects.

Source: Simulation from data generating process, described in Section 3.

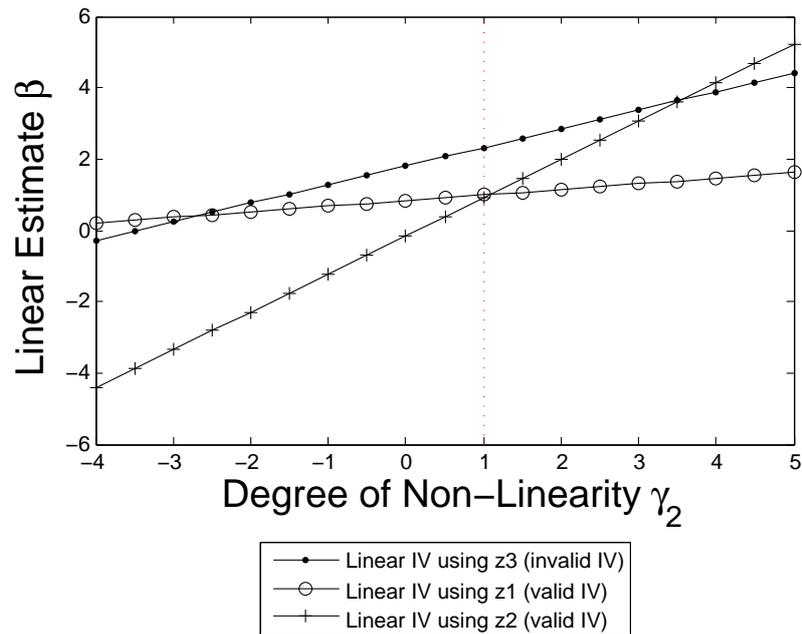
Figure 4: Relative P-Value for Test of No Selection Bias in the Case with Selection Bias



Notes: This figure is based on Figure 3. It provides the P-Value from two tests of no selection bias:  $H_0 : \beta(OLS) - \beta(z_1) = 0$  using  $z_1$  as the instrument;  $H_0 : \beta(OLS) - \beta(z_2) = 0$  using  $z_2$  as the instrument.

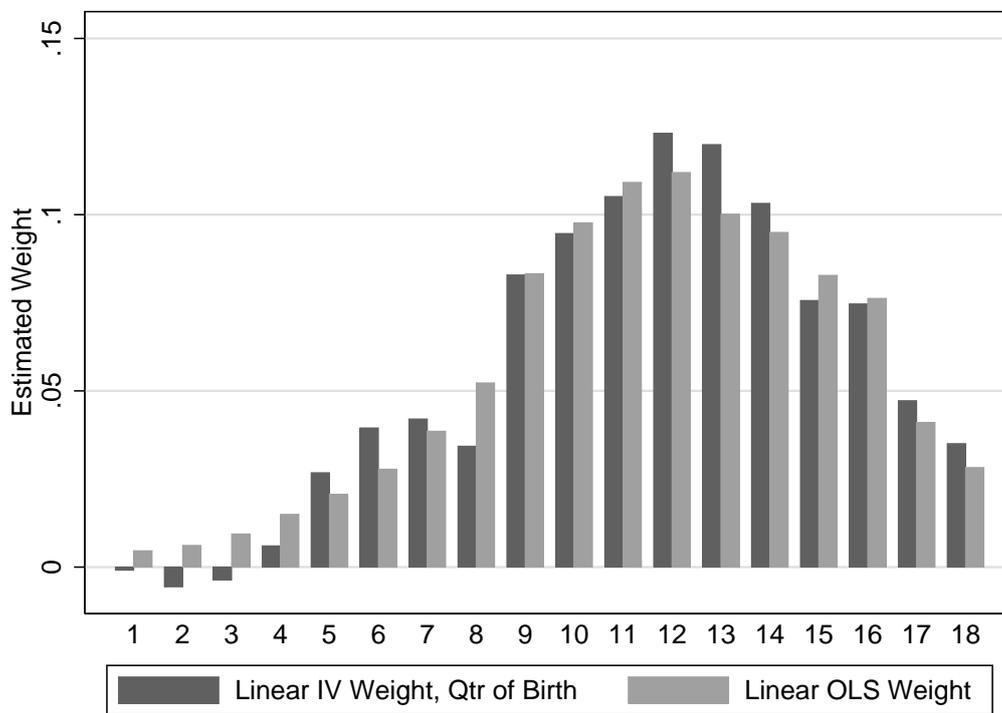
Source: Simulation from data generating process, described in Section 3.

Figure 5: Comparison of Linear IV Estimators ( $z_{1i}, z_{2i}$  valid,  $z_{3i}$  invalid)



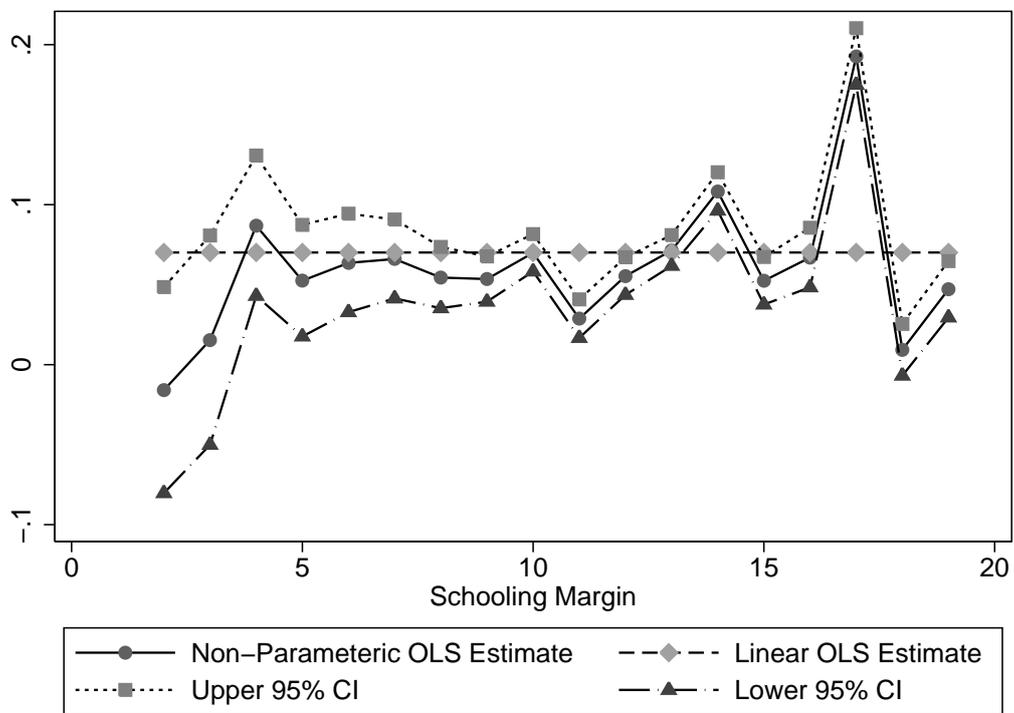
Notes: This figure illustrates how nonlinearities affect the over-identification test estimates of a linear model. We use the same model as in Figure 3, except for adding a third instrument constructed to be invalid:  $z_{3i} = \eta_i + 0.01v_i$ , with  $\eta_i \sim N(0, 1)$ . When the linear model is correct ( $\gamma_1 = \gamma_2$ ), the linear IV estimators using the valid instruments  $z_1$  and  $z_2$  are equal, but the linear IV estimator using the invalid instrument  $z_3$  is different:  $\beta(z_1) = \beta(z_2) \neq \beta(z_3)$ , modulo sampling error. As  $\gamma_2$  moves away from  $\gamma_2 = 1$ , the linear IV estimator using  $z_3$  intersects with the linear IV estimators using  $z_1$  and  $z_2$ .  
 Source: Simulation from data generating process, described in Section 3.

Figure 6: Returns to Schooling Application, Angrist and Krueger (1991): Linear OLS and IV Weights



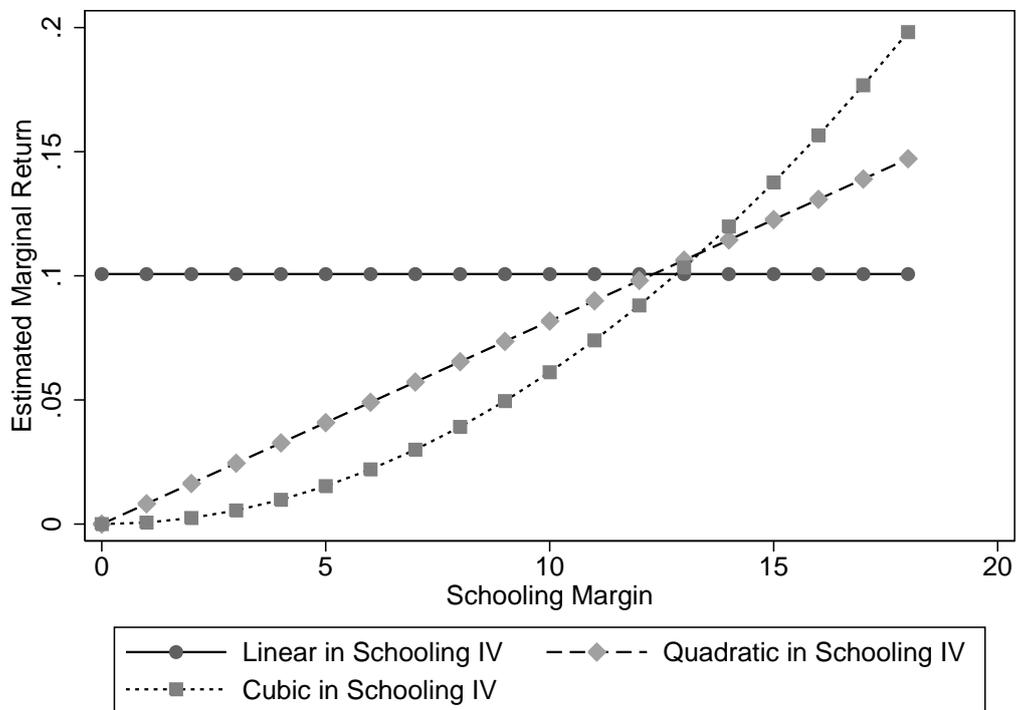
Notes: This table reports the weights for the linear OLS and IV estimates. To compute these weights, we use the decomposition described in Section 2. The decomposition adjusts for multiple instruments and covariates. The margins are labeled such that margin 1 is the schooling margin from 0 to 1 years of schooling, margin 2 is from 1 to 2 years of schooling, and so on.  
 Source: 1970 US Census data from Angrist and Krueger (1991).

Figure 7: Returns to Schooling Application, Angrist and Krueger (1991): OLS estimates of marginal effects



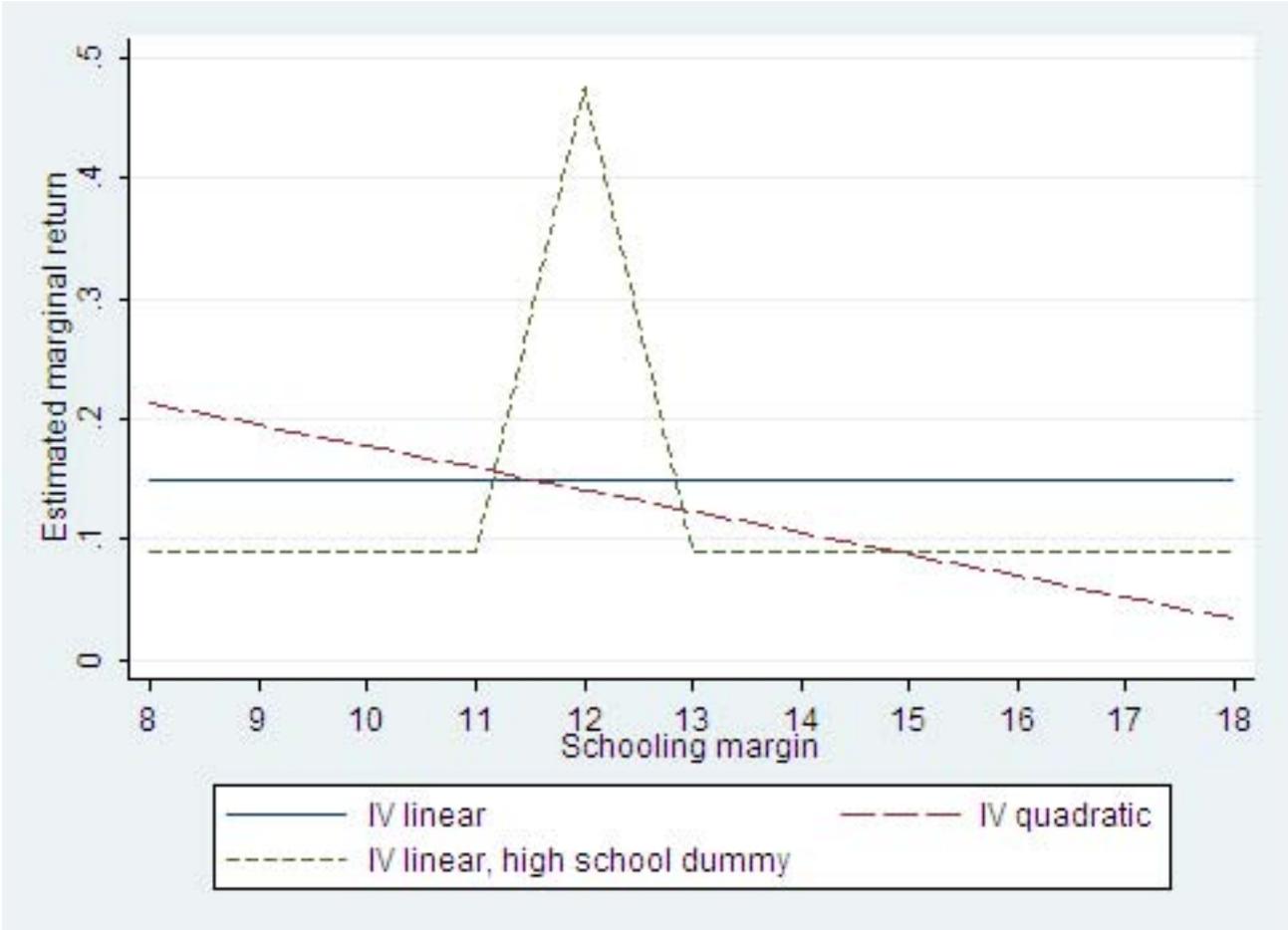
Notes: This figure graphs the estimated marginal effects from the log-linear and non-parametric in years of schooling, based on OLS regressions. Each model includes a full set of covariates.  
 Source: 1970 US Census data from Angrist and Krueger (1991).

Figure 8: Returns to Schooling Application, Angrist and Krueger (1991): IV estimates of marginal effects



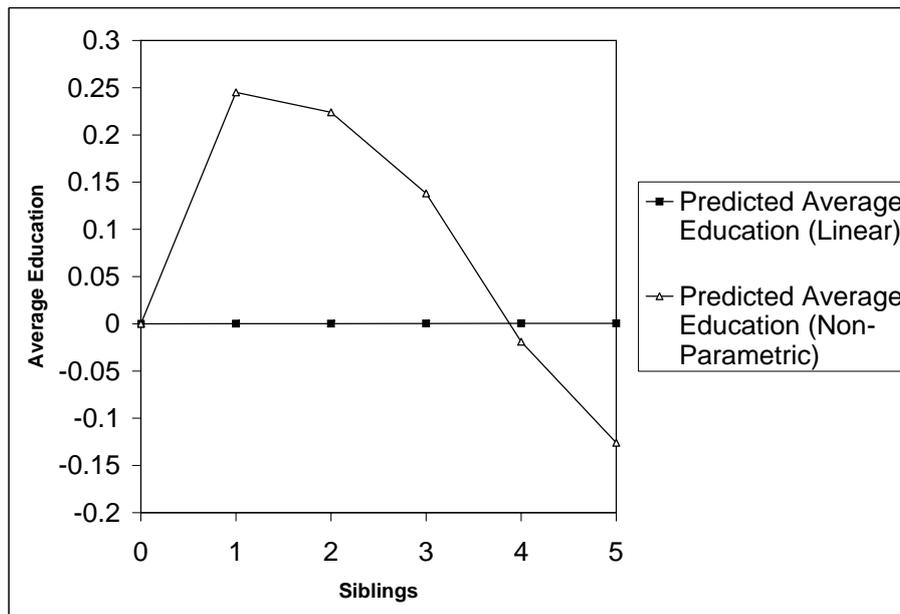
Notes: This figure graphs the estimated marginal effects based on the IV estimates reported in the first (log-linear), fifth (squared), and sixth (cubed) row of Table 1.  
 Source: 1970 US Census data from Angrist and Krueger (1991) including only men aged 40-49.

Figure 9: Returns to Schooling Application, Aakvik et al. (2010): IV estimates of marginal effects



Notes: This figure graphs the estimated marginal effects based on the IV estimates reported in the first (log-linear), second (quadratic), and third (Linear and dummy) row of Table2. Source: 1970 US Census data from Angrist and Krueger (1991) including only men aged 40-49.

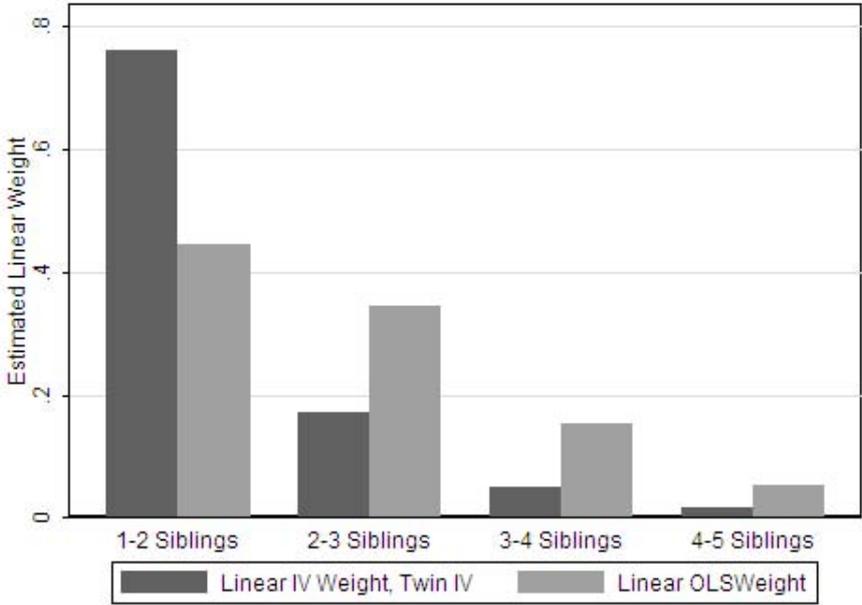
Figure 10: Family Size Application: Average Educational Attainment for First Born Children by Number of Siblings (Relative to Only Children)



Notes: This figure graphs the linear and non-parametric in family size predictions from OLS regressions. The sample consists of first born children in families with 2 or more children. Each model includes a full set of covariates. The values are graphed relative to only children (0 siblings), i.e. the education of only children is normalized to 0. The slopes in this figure provide the estimated marginal family size effects at each margin, where the linear model imposes constant slopes whereas the non-parametric model allows non-constant slopes. The marginal effect estimate from the linear model is close to zero (represented by a flat line), while the non-parametric estimate of the marginal effects indicates that they are non-monotonic. The linear prediction of total effect is  $\hat{y} = \hat{\beta} * S$  for  $S = 0, 1, \dots, 5$ , where  $S$  is number of siblings and  $\hat{\beta}$  is the OLS estimate from the linear model in family size. Non-parametric prediction of total effect is  $\hat{y} = \hat{\gamma}_1 * 1\{s \geq 1\} + \dots + \hat{\gamma}_5 * 1\{s = 6\}$ , where  $\hat{\gamma}_s$  are the OLS estimates from the non-parametric model in family size.

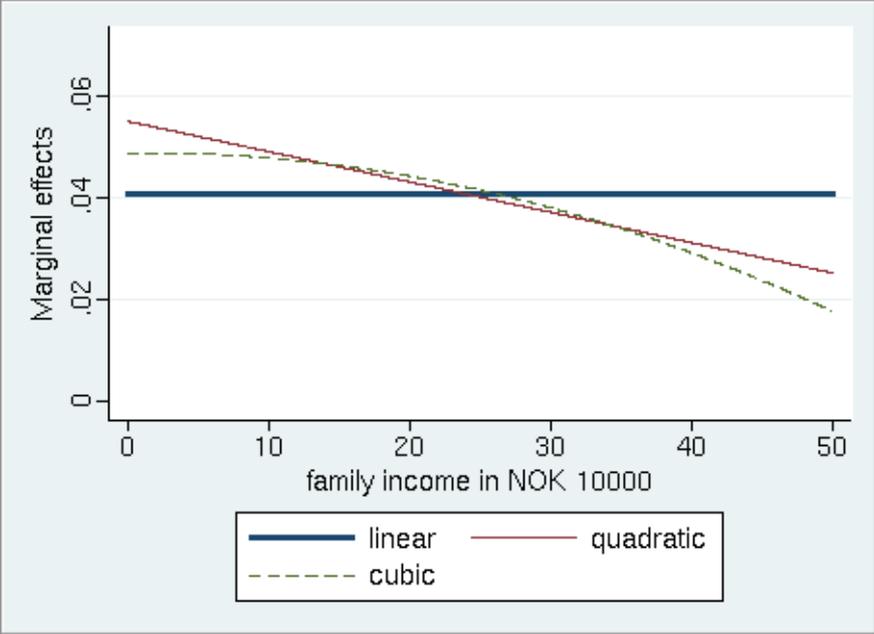
Source: Administrative registers from Statistics Norway.

Figure 11: Family Size Application: Linear OLS and IV Weights



Notes: This figure reports the weights for the linear OLS and IV estimates displayed in Table 3. To compute these weights, we use the decompositions described in Section 2.  
Source: Administrative registers from Statistics Norway.

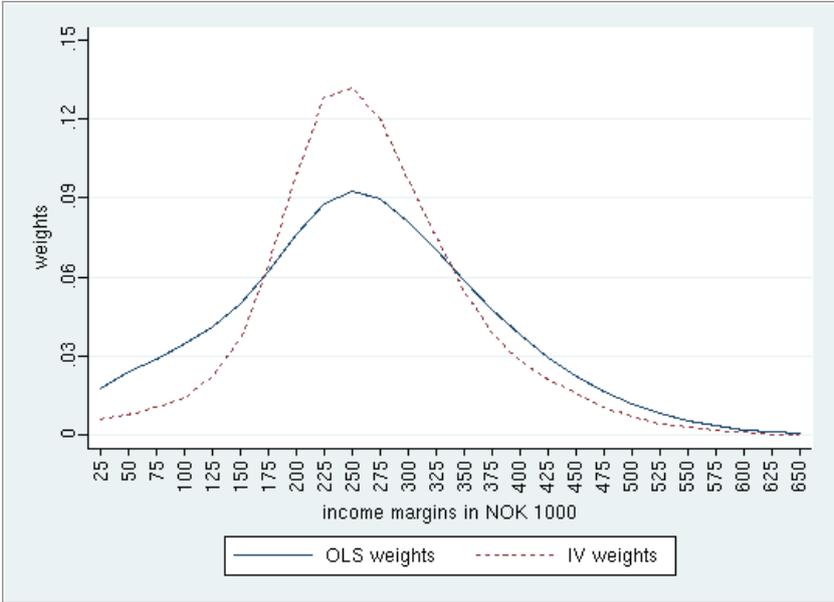
Figure 12: Family Income Application: OLS Estimates of Marginal Effects on Years of Education



Notes: This figure graphs OLS estimates of marginal effects from a linear specification in family income, a quadratic specification in family income, and cubic specification in family income. A full set of controls is used in all regressions.

Source: Administrative registers from Statistics Norway.

Figure 13: Family Income Application: Linear OLS and IV weights



Notes: This table reports the weights for the linear OLS and IV estimates displayed in Table4. To compute these weights, we use the decomposition described in Section 2.  
Source: Administrative registers from Statistics Norway.

## 7 Tables

Table 1: Return to Schooling Application, Angrist and Krueger (1991): IV estimates from Alternative Models

Model	Linear $s_i$	Squared $s_i^2$	Cubed $s_i^3$	High School $1\{s_i \geq 12\}$	College $1\{s_i \geq 16\}$	$R^2$
Log-Linear	0.1007 (0.0336)					0.2065
Quadratic	0.1080 (0.1366)	-0.00032 (0.0058)				0.2052
Cubic	0.9075 (0.4734)	-0.0848 (0.0478)	0.0027 (0.0015)			0.0506
Linear and dummies	0.1382 (0.0844)			-0.3439 (0.4154)	0.0646 (0.4715)	0.1766
Squared		0.0041 (0.0014)				0.2183
Cubed			0.00020 (0.000072)			0.2147

Notes: This table reports IV estimates of six different models of schooling. The first row specifies log earnings as a linear function of schooling. The next three rows nest the log-linear model. The second row specifies log earnings as a quadratic function of schooling, including a linear and a squared schooling term. The third row specifies log earnings as a cubic function of schooling, including linear, squared, and cubed schooling terms. The fourth row specifies log earnings as a linear function of schooling, with dummy variables at high school and college completion. The last two rows are non-nested alternatives to the log-linear model. The fifth row replaces the linear schooling term with a quadratic schooling term. The sixth row replaces the linear schooling term with a cubed schooling term. A full set of controls is used in all specifications. In each first stage, a full set of instruments is used (3 quarter of birth dummy variables interacted with 10 birth indicators). Standard errors in parentheses are heteroschedastic robust. Source: 1970 US Census data from Angrist and Krueger (1991).

Table 2: Return to Schooling Application, Aakvik et al. (2010): IV estimates from Alternative Models

Model	Linear $s_i$	Squared $s_i^2$	High School $1\{s_i \geq 12\}$	$R^2$
Log-Linear	0.1495 (0.0060)			0.0831
Quadratic	0.3567 (0.0830)	-0.0084 (0.0034)		0.1059
Linear and dummy	0.0893 (0.0189)		0.3867 (0.1310)	0.1161

Notes: This table reports IV estimates of three different models of schooling. The first row specifies log earnings as a linear function of schooling. The second row specifies log earnings as a quadratic function of schooling, including a linear and a squared schooling term. The third row specifies log earnings as a linear function of schooling, with a linear schooling term and a dummy variable at high school completion. A full set of controls is used in all specifications. In each first stage, a full set of instruments is used (reform indicator, interaction terms between reform indicator and the pre-reform availability of other types of school, and interaction terms between the reform indicator and family background variables). Standard errors in parentheses are heteroscedastic robust.

Source: Administrative registers from Statistics Norway.

Table 3: Family Size Application: OLS and IV Estimates from the Linear and Non-Parametric Models in Family Size

Panel I: Linear Estimates of Marginal Effects		
	IV	OLS
Numb. of Children	-0.0036 (0.0460)	-0.063 (0.004)
Panel II: Non-Parametric Estimates of Marginal Effects		
Marginal effect	IV	OLS
Siblings $\geq 2$	0.153 (0.063)	-0.0127 (0.008)
Siblings $\geq 3$	-0.474 (0.079)	-0.079 (0.011)
Siblings $\geq 4$	-0.800 (0.129)	-0.151 (0.019)
Siblings $\geq 5$	-0.787 (0.247)	-0.102 (0.033)

Notes: This table reports the OLS and IV estimates of marginal effects from the linear and non-parametric models in family size. The sample consists of first born children in families with 2 or more children. Each column is separate regression. Siblings  $\geq 2$  is the marginal effect from moving from 1 to 2 siblings, Siblings  $\geq 3$  is the marginal effect from moving from 2 to 3 siblings, and so on. All models include a full set of covariates, and twin births are used as the instrument for family size. Standard errors are heteroscedastic robust.

Source: Administrative registers from Statistics Norway.

Table 4: Family Income Application: OLS and IV Estimates of the Linear and Quadratic Models in Family Income

	Dependent variables		
	Education	Dropout	IQ (males only)
Panel I: Linear OLS			
Family income in 10000 NOK	.043 (.001)	-.006 (.000)	.028 (.001)
Panel II: Linear IV			
Binary instrument			
Family income in 10000 NOK	.022 (.057)	-.012 (.011)	.033 (.023)
Panel III: Linear IV			
Full set of instruments			
Family income in 10000 NOK	.026 (.020)	-.008 (.006)	-.017 (.013)
Panel IV: Quadratic OLS			
Family income in 10000 NOK	.051 (.002)	-.010 (.000)	.032 (.002)
Quadratic income (x 100)	-.016 (.005)	.008 (.001)	-.008 (.004)
Panel V: Quadratic IV			
Full set of instruments			
Family income in 10000 NOK	.180 (.087)	-.030 (.016)	.234 (.109)
Quadratic income (x 100)	-.302 (.164)	.042 (.021)	-.401 (.210)
N	121122	121122	57788

Notes: This table reports OLS and IV estimates of the linear and quadratic family income model. Panels I and IV display the OLS results, whereas Panels II, III, and V show the IV results. Panel II uses only the binary instrument (born in Rogaland in 1967-1969 dummy variable). Panels III and V use the full set of instruments (born in Rogaland in 1967-1969 dummy variable and interactions with family background variables). A full set of controls is used in all regressions. Standard errors in parentheses are heteroskedastic robust.

Source: Administrative registers from Statistics Norway.