

NBER WORKING PAPER SERIES

ESTIMATING AND TESTING MODELS WITH MANY TREATMENT LEVELS  
AND LIMITED INSTRUMENTS

Lance Lochner  
Enrico Moretti

Working Paper 17039  
<http://www.nber.org/papers/w17039>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
May 2011

Previously circulated as "Estimating and Testing Non-Linear Models Using Instrumental Variables." We thank Josh Angrist, David Card, Pedro Carneiro, Jim Heckman, Guido Imbens, the editor, two anonymous referees and seminar participants at the 2008 UM/MSU/UWO Summer Labor Conference, UCSD, and Stanford for their suggestions. We also thank Matias Cattaneo and Javier Cano Urbina for their excellent research assistance and comments, as well as Martijn van Hasselt and Youngki Shin for their many comments and suggestions. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2011 by Lance Lochner and Enrico Moretti. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Estimating and Testing Models with Many Treatment Levels and Limited Instruments  
Lance Lochner and Enrico Moretti  
NBER Working Paper No. 17039  
May 2011, Revised July 2014  
JEL No. C01,J0

**ABSTRACT**

Many empirical microeconomic studies estimate econometric models that assume a single finite-valued discrete endogenous regressor (for example: different levels of schooling), exogenous regressors that are additively separable and enter the equation linearly; and coefficients (including per-unit treatment effects) that are homogeneous in the population. Empirical researchers interested in the causal effect of the endogenous regressor often use instrumental variables. When few valid instruments are available, researchers typically estimate restricted specifications that impose uniform per-unit treatment effects, even when these effects are likely to vary depending on the treatment level. In these cases, ordinary least squares (OLS) and instrumental variables (IV) estimators identify different weighted averages of all per-unit effects, so the traditional Hausman test (based on the restricted specification) is uninformative about endogeneity. Addressing this concern, we develop a new exogeneity test that compares the IV estimate from the restricted model with an appropriately weighted average of all per-unit effects estimated from the more general model using OLS. Notably, our test works even when the true model cannot be estimated using IV methods as long as a single valid instrument is available (e.g. a single binary instrument). We re-visit three recent empirical examples that examine the role of educational attainment on various outcomes to demonstrate the practical value of our test.

Lance Lochner  
Department of Economics, Faculty of Social Science  
University of Western Ontario  
1151 Richmond Street, North  
London, ON N6A 5C2  
CANADA  
and NBER  
llochner@uwo.ca

Enrico Moretti  
University of California, Berkeley  
Department of Economics  
549 Evans Hall  
Berkeley, CA 94720-3880  
and NBER  
moretti@econ.berkeley.edu

# 1 Introduction

Many recent empirical papers seek to estimate causal relationships using instrumental variables (IV), including two-stage least squares (2SLS) estimators, when concerns about causality arise. A model frequently estimated in practice has the following form:

$$y_i = s_i\beta^L + x_i'\gamma^L + \nu_i, \tag{1}$$

where  $y_i$  is the outcome of individual  $i$ ;  $x_i$  is a  $k \times 1$  vector of exogenous covariates (including an intercept); and  $s_i$  is the potentially endogenous regressor. For example, the variable  $s_i$  might reflect different treatment levels of a government training program or different dosage levels for a new drug treatment. In our empirical examples and much of our discussion below,  $s_i$  reflects years of completed schooling.

Conclusions about exogeneity of  $s_i$  and consistency of the ordinary least squares (OLS) estimator are typically based on a comparison of OLS and IV estimates of  $\beta^L$ . When a standard Hausman test (Hausman 1978) indicates a significant difference between OLS and IV estimates, it is common to conclude that endogeneity of  $s_i$  plays an important confounding role in OLS.

Yet, in many economics applications, the true relationship between  $y_i$  and  $s_i$  is unlikely to be linear. In particular, suppose that the endogenous regressor  $s_i \in \{0, 1, 2, 3, \dots, S\}$  is discrete and the true model has the form:

$$y_i = \sum_{j=1}^S D_{ij}\beta_j + x_i'\gamma + \varepsilon_i, \tag{2}$$

where  $D_{ij} = 1[s_i \geq j]$  reflects a dummy variable equal to one if  $s_i \geq j$  and zero otherwise,  $E(\varepsilon_i) = 0$ , and  $E(\varepsilon_i x_i) = 0$ . When  $s_i$  reflects years of schooling, the  $\beta_j$  represent grade-specific effects of moving from  $j - 1$  to  $j$  years of schooling.

The difference between the models in equations (1) and (2) is that the former assumes a uniform per-unit or marginal effect of  $s_i$  across all levels of  $s_i$  while the latter does not. For example, in the classic case of the return to education, the model in (1) assumes that the effect of an extra year of elementary school is identical to the effect of the last years of high school and college, while the model in (2) allows for sheepskin effects and other non-linearities that are likely to arise in practice.

While variable per-unit treatment effects are likely to be important in many applications, relatively few studies have focused on their practical implications when instrumental variables may be

needed.<sup>1</sup> The difficulty in estimating a specification like equation (2) when endogeneity concerns arise is that there may be many  $\beta_j$  parameters to estimate, while researchers typically have very few valid instruments. In theory, a single continuous instrument may be sufficient for identification. In practice, there is often insufficient variation in the instrument to precisely estimate all per-unit effects. Discrete-valued instruments are also common in the literature. As a consequence, empirical studies commonly estimate models like equation (1) even when there is no theoretical reason to do so and in some cases there is prima-facie evidence of important non-linearities between  $y_i$  and  $s_i$ .

We demonstrate that when the per-unit effects of changes in  $s_i$  vary over the range of  $s_i$  as in equation (2) but the estimated model assumes that all per-unit effects are the same as in equation (1), OLS and IV methods estimate different weighted averages of all per-unit effects. Building on this insight, we develop a new exogeneity test that only requires a single (even binary) instrument and is useful when per-unit treatment effects vary across treatment levels.

We stress that our results do not apply to all non-linear models, but only to the specific case described in equations (1) and (2). In particular, we assume: (i) a single finite-valued discrete endogenous regressor; (ii) exogenous regressors are additively separable and enter the equation linearly; and (iii) all coefficients (including per-unit treatment effects) are homogeneous in the population. While these assumptions are strong, they are common in the applied microeconomics literature.

We are not the first to point out that estimates from a mis-specified linear model (i.e. constant marginal or per-unit treatment effects) yield weighted averages of each marginal/per-unit effect. Yitzhaki (1996) derives these weights in the context of OLS, while Angrist and Imbens (1995) and Heckman, Urzua, and Vytlačil (2006) derive weights for IV estimators in the presence of both variable multi-valued treatment effects and parameter heterogeneity. Angrist and Imbens (1995) show conditions under which 2SLS estimates a local average treatment effect (LATE).<sup>2</sup> In a very general setting, Heckman, Urzua, and Vytlačil (2006) discuss ordered and unordered choice models with unobserved heterogeneity and nonlinearity, developing weights for treatment effects using general instruments. Heckman and Vytlačil (2005) emphasize that in the presence of parameter heterogeneity, there is no single ‘effect’ of the regressor on an outcome, and different estimation

---

<sup>1</sup>Angrist, Graddy, and Imbens (2000), Lochner and Moretti (2001), and Mogstad and Wiswall (2010) are notable exceptions.

<sup>2</sup>Intuitively, the LATE reflects the effect of a regressor on outcomes for individuals induced to change their behavior in response to a change in the value of the instrument.

strategies provide estimates of different ‘parameters of interest’ or different ‘average effects’. While many studies focus on parameter heterogeneity across individuals with a uniform marginal effect over values of  $s_i$  (i.e.  $y_i$  is linear in  $s_i$ ), we consider the opposite case, assuming a non-linear relationship between  $y_i$  and  $s_i$  that is the same for all individuals.<sup>3</sup> Our setting is a special case of that used by Heckman, Urzua, and Vytlacil (2006); however, our emphasis on varying per-unit treatment effects and the endogeneity test are novel.

We begin by showing that inappropriately assuming model (1) when per-unit treatment effects vary across treatment levels will generally yield different OLS and IV/2SLS estimates even in the absence of endogeneity, since these estimators can be written as weighted averages of causal responses to each marginal change in the regressor, where the sets of weights differ for the estimators.<sup>4</sup> An appealing feature of our setting is that the weights have an intuitive interpretation, are functions of observable quantities, and can be easily estimated under very general assumptions. Therefore, it is possible to directly compare the OLS and IV weights.

This insight leads to our main contribution: a new exogeneity test that can be used to determine consistency of the OLS estimator for equation (2). Before describing our test, first note that the standard Hausman test is of limited applicability in this context. Since OLS and IV/2SLS identify different weighted averages of all per-unit effects, the Hausman test applied to equation (1) is uninformative about endogeneity of the regressor when per-unit treatment effects vary across treatment levels. It may reject equality of OLS and IV/2SLS estimates even when the regressor is exogenous, and it may fail to reject equality when the regressor is endogenous. Alternatively, in order to implement the Hausman test for equation (2), one would need to estimate all  $\beta_j$  parameters using IV methods. In practice, this is often impossible when there are many treatment levels, since researchers often have access to only a few valid instruments with limited variation. Rarely would researchers have instruments capable of identifying, for example, 20 different grade-specific  $\beta_j$  parameters associated with all potential schooling levels.

---

<sup>3</sup>Studies focused on parameter heterogeneity include Imbens and Angrist (1994), Wooldridge (1997), Heckman and Vytlacil (1998, 1999, 2005), Card (1999), Kling (2000), Moffitt (2009), and Carneiro, Heckman and Vytlacil (2010).

<sup>4</sup>Relative to the existing literature, our models are closer to those typically estimated in practice. Angrist and Imbens (1995) only consider discrete regressors that are indicators that place observations into mutually exclusive categories, and they interact their instrument (also assumed to be discrete) with each of these regressors to create a large set of effective instruments. The Heckman, Urzua, and Vytlacil (2006) discussion of instrumental variables estimation in ordered choice models is left implicit on all covariates affecting the outcome variable.

The test that we propose can be thought of as a generalization of the standard Hausman test and is informative about the consistency of OLS estimates for all  $\beta_j$  effects in equation (2). Our test re-weights OLS estimates of the  $\beta_j$ 's from equation (2) using estimated IV/2SLS weights and compares this with the corresponding IV/2SLS estimator of  $\beta^L$  in equation (1). Under fairly general conditions, our test can be implemented even when only a single valid (binary) instrument is available.<sup>5</sup>

Our proposed test has both strengths and weaknesses. The fact that our test requires only a single instrument should make it attractive to empirical researchers. In many contexts, researchers can easily use OLS to estimate models like equation (2) (e.g. regressing log wages on a set of 20 schooling dummies), yet they often have very few valid instruments with limited variation at their disposal. A researcher can use our test to establish whether the OLS estimates are consistent without having to estimate the more general equation (2) using IV/2SLS. If our test fails to reject exogeneity, researchers can have some confidence in their OLS estimates. However, if our test rejects, it does not help in estimating the true model. Our test, therefore, offers only a partial solution to the problem of estimating multiple per-unit treatment effects with limited instruments.

Three additional limitations are worth highlighting. First, it is important to note that we test whether the weighted average of all OLS  $\hat{\beta}_j$  asymptotic biases equals zero. Therefore, our test has no power against the possibility that some OLS  $\hat{\beta}_j$  estimates are asymptotically biased upwards and others downwards in such a way as to exactly cancel each other when averaged using the IV/2SLS weights. Still, rejection of the null implies that OLS estimates are inconsistent. Furthermore, we discuss conditions under which all  $\hat{\beta}_j$  asymptotic biases would be of the same sign, in which case our test is equivalent to testing whether all OLS  $\hat{\beta}_j$  estimates are consistent. In many applications, economic theory can be informative about the likely sign of any biases. For example, in the case of returns to schooling, most models of investment in human capital predict that OLS estimates of  $\hat{\beta}_j$  are all asymptotically upward biased.

Second, even if exogeneity cannot be rejected, researchers should exercise caution when conducting inference using OLS estimates of equation (2) when the instruments are not sufficiently

---

<sup>5</sup>Lochner and Moretti (2001) and Mogstad and Wiswall (2010) suggest that comparing re-weighted OLS estimates with IV/2SLS estimates may be a useful heuristic approach for assessing the importance of non-linearities. In this paper, we develop a formal econometric test for exogeneity based on this insight. Our test differs conceptually and practically from the omnibus specification tests developed by White (1981), which essentially compare different weighted generalized least squares estimators for a general nonlinear function.

strong. Like the Hausman test, our test does not have much power when instruments are weak. As Wong (1997) and Guggenberger (2010) demonstrate, this can cause size problems with inference in a two-stage approach where the Hausman test is used to determine exogeneity in a first stage, and OLS estimates are used in a second stage when exogeneity cannot be rejected.<sup>6</sup> Monte Carlo simulations confirm that similar inference problems can arise when using our test with insufficiently strong instruments.

Third, our approach assumes that equation (2) reflects the true model. Mis-specification due to, for example, non-separabilities between  $s_i$  and  $x_i$  or due to individual-level parameter heterogeneity would likely invalidate our test, since this would alter the relationship between OLS and IV estimators in unaccounted-for ways.

In the last part of the paper, we demonstrate the practical usefulness of our test by re-examining three recent empirical papers in which estimated 2SLS effects differ from OLS effects. In one example, our test suggests that schooling is exogenous for incarceration among white men. As we discuss below, this is empirically useful, since it lends credibility to OLS estimates that suggest a highly non-linear relationship between educational attainment and the probability of imprisonment. In contrast, our test strongly rejects exogeneity of schooling for incarceration among black men, while the standard Hausman test does not. In this case, the endogeneity of schooling is obscured when non-linearities between schooling and imprisonment are ignored. Our other examples produce greater concordance between the standard Hausman test and our exogeneity test; however, for different reasons.

The rest of the paper is organized as follows. In Section 2, we show conditions under which OLS, IV and 2SLS estimates of  $\beta^L$  in equation (1) can be written as weighted averages of the true underlying  $\beta_j$  parameters in the more general model given by equation (2). For expositional purposes, we will refer to  $s_i$  as years of schooling, so the  $\beta_j$  reflect grade-specific marginal or per-unit effects. Section 3 develops an exogeneity test that can be used to determine consistency of the OLS estimator for equation (2). Section 4 presents the results from three previous empirical examples, and Section 5 concludes.

---

<sup>6</sup>Specifically, Wong (1997) and Guggenberger (2010) provide simulation evidence (in a linear regression model like equation 1) for the null rejection probability of a simple hypothesis test conditional on a standard Hausman pretest for exogeneity not rejecting. Their findings indicate that when regressor endogeneity is small, the null rejection probability of the hypothesis test may be substantially higher than the nominal size if the instruments are not sufficiently strong.

## 2 Estimating Weighted Average Per-Unit Treatment Effects

In this section, we consider IV/2SLS and OLS estimators when equation (1) is estimated, but the true model is described by equation (2). We show conditions under which these estimators converge to a weighted average of each grade-specific  $\beta_j$  effect and discuss the weights. We assume throughout our analysis that all observations are independent across  $i = 1, \dots, N$  individuals and that standard conditions for the weak law of large numbers and central limit theorems apply.<sup>7</sup>

### 2.1 IV Estimation with a Single Instrument

We first consider IV estimation with a single instrument, discussing OLS as a special case. We study the case where the potentially endogenous variable  $s_i$  is discrete.<sup>8</sup> Throughout the paper, we assume  $\varepsilon_i$  is independent across individuals with  $E(\varepsilon_i) = 0$ ,  $x_i$  is distributed with density  $F_x(\cdot)$ , and  $E(\varepsilon_i x_i) = 0$ . The following decomposition is also useful:  $s_i = x_i' \delta_s + \eta_i$ , where  $\delta_s = [E(x_i x_i')]^{-1} E(x_i s_i)$  by construction and  $E(x_i \eta_i) = 0$ .

The following IV assumption is standard.

**Assumption 1.** *The instrument is uncorrelated with the error in the outcome equation,  $E(\varepsilon_i z_i) = 0$ , and correlated with  $s_i$  after linearly controlling for  $x_i$ ,  $E(\eta_i z_i) \neq 0$ .*

Let  $M_x = I - x(x'x)^{-1}x'$  and  $\tilde{s} = M_x s$  for any variable  $s$ . (We drop the  $i$  subscripts when we refer to the vector or matrix version of a variable that vertically stacks all individual-specific values.) With a single instrument, 2SLS estimation of equation (1) is equivalent to the following IV estimator:

$$\begin{aligned} \hat{\beta}_{IV}^L &= (z' M_x s)^{-1} z' M_x y \\ &= (\tilde{z}' \tilde{s})^{-1} \tilde{z}' \left( \sum_{j=1}^S D_j \beta_j \right) + (\tilde{z}' \tilde{s})^{-1} \tilde{z}' \varepsilon \\ &= \sum_{j=1}^S \hat{\omega}_j^{IV} \beta_j + (\tilde{z}' \tilde{s})^{-1} \tilde{z}' \varepsilon \end{aligned}$$

---

<sup>7</sup>For example, assume all random variables are independent and have finite first, second, and third moments. Finite third moments enable application of central limit theorems based on independent but not necessarily identically distributed random variables (e.g. Liapounov).

<sup>8</sup>While we study the case of a discrete endogenous regressor, OLS and IV estimators will also yield different weighted averages of marginal effects when the regressor is continuous. The insights of Yitzhaki (1996) might be used to develop weights and a related test specifically designed for the continuous regressor case.



where  $\hat{\omega}_j^{IV} = (\tilde{z}'\tilde{s})^{-1}\tilde{z}'D_j = \left(\frac{1}{N}\sum_{i=1}^N\tilde{z}_iD_{ij}\right) / \left(\frac{1}{N}\sum_{i=1}^N\tilde{z}_i\tilde{s}_i\right)$ . Since  $\sum_{j=1}^S D_{ij} = s_i$ , these  $\hat{\omega}_j^{IV}$  sum to one over  $j = 1, \dots, S$ . We refer to them as “weights” even though they may be negative for some  $j$ .<sup>9</sup>

One helpful assumption is monotonicity in the effects of the instrument on  $s_i$ . Although monotonicity is not necessary for deriving and estimating “weights”, it does help ensure that they are non-negative and simplifies their interpretation. When  $s_i$  reflects years of schooling, monotonicity implies that the instrument either causes everyone to weakly increase or causes everyone to weakly decrease their schooling. Without loss of generality, we assume that  $s_i$  is weakly increasing in  $z_i$ . Define  $s_i(\vartheta)$  to be the value of  $s_i$  for individual  $i$  when  $z_i = \vartheta$ .

**Assumption 2.** (*Monotonicity*) *The instrument does not decrease  $s_i$ :  $Pr[s_i(\vartheta) < s_i(\vartheta')] = 0$ , for all  $\vartheta > \vartheta'$ .*

To facilitate our analysis of  $\hat{\beta}_{IV}^L$ , it is useful to decompose  $z_i = x_i'\delta_z + \zeta_i$  where  $\delta_z = [E(x_i x_i')]^{-1}E(x_i z_i)$  and  $E(x_i \zeta_i) = 0$ .

**Proposition 1.** *If Assumption 1 holds, then  $\hat{\beta}_{IV}^L \xrightarrow{p} \sum_{j=1}^S \omega_j^{IV} \beta_j$ , where*

$$\omega_j^{IV} = \frac{Pr(s_i \geq j)E(\zeta_i | s_i \geq j)}{\sum_{k=1}^S [Pr(s_i \geq k)E(\zeta_i | s_i \geq k)]} \quad (3)$$

*sum to unity over all  $j = 1, \dots, S$ . Furthermore, if  $E(z_i | x_i) = x_i'\delta_z$  and Assumption 2 (*Monotonicity*) holds, then the weights are non-negative and can be written as*

$$\omega_j^{IV} = \frac{E\{Cov(z_i, D_{ij} | x_i)\}}{\sum_{k=1}^S E\{Cov(z_i, D_{ik} | x_i)\}} \geq 0. \quad (4)$$

Proof: See Online Appendix A.

This result shows that estimating the mis-specified linear-in-schooling model using IV yields a consistent estimate of a weighted average of all grade-specific  $\beta_j$  effects. The weights on all grade-specific effects are straightforward to estimate. From a 2SLS regression of  $D_{ij}$  on  $s_i$  and  $x_i$  using  $z_i$  as an instrument for  $s_i$ , the coefficient estimate on  $s_i$  equals  $\hat{\omega}_j^{IV}$ .

---

<sup>9</sup>When they cannot be shown to be non-negative, we use “weights” with quotation marks to distinguish them from cases when they are known to be proper weights that are both non-negative and sum to one.

When the instrument affects all persons in the same direction and its expectation conditional on  $x_i$  is linear (e.g.  $x$ 's are mutually exclusive and exhaustive categorical indicator variables), the weights are non-negative and depend on the strength of the covariance between the instrument and each schooling transition indicator conditional on other covariates. In general, different instruments yield estimates of different “weighted averages,” even if the instruments are all valid.

With Assumption 1,  $E(z_i|x_i) = x_i'\delta_z$ , and  $E(\varepsilon_i|x_i) = 0$ , it is straightforward to show that the IV estimator converges to a weighted average of all conditional (on  $x_i$ ) IV estimators, where the weights are proportional to the covariance between the instrument and schooling conditional on  $x_i$ :

$$\hat{\beta}_{IV}^L \xrightarrow{p} \int \beta_{IV}(\phi)h(\phi)dF_x(\phi),$$

where  $\beta_{IV}(\phi) = \frac{Cov(z_i, y_i|x_i=\phi)}{Cov(z_i, s_i|x_i=\phi)}$  is the population analogue of the IV estimator conditional on  $x_i = \phi$  and  $h(\phi) = \frac{Cov(z_i, s_i|x_i=\phi)}{\int Cov(z_i, s_i|x_i=a)dF(a)}$  is a weighting function that integrates to one for all  $x_i$  (with  $h(\cdot) \geq 0$  under Assumption 2). Notice that  $\beta_{IV}(\phi) = \sum_{j=1}^S \beta_j \omega_j^{IV}(\phi)$ , where  $\omega_j^{IV}(\phi) = \frac{Cov(z_i, D_{ij}|x_i=\phi)}{Cov(z_i, s_i|x_i=\phi)}$  are  $x$ -specific IV “weights” for each grade-specific effect,  $\beta_j$ . Each  $x$ -specific IV estimator is simply a weighted average of the grade-specific  $\beta_j$  effects, where the weights are proportional to the covariance between the instrument and  $D_{ij}$  conditional on  $x_i$ . Some re-arranging shows that the IV weights from equations (3) or (4) can be re-written as  $\omega_j^{IV} = \int \omega_j^{IV}(\phi)h(\phi)dF_x(\phi)$ .<sup>10</sup>

These results complement the IV/2SLS analyses of Angrist and Imbens (1995) and Heckman, Urzua, and Vytlacil (2006), who also consider parameter heterogeneity along with variable per-unit treatment effects. In order to ease interpretation in the presence of parameter heterogeneity, Angrist and Imbens (1995) make strong assumptions about the additional  $x_i$  covariates and how they enter in estimation. Specifically, they assume that the  $x_i$  regressors are indicator variables that place individuals into mutually exclusive categories and that the instrumental variable (also assumed to be discrete) is interacted with all of these additional covariates. By contrast, Heckman, Urzua, and Vytlacil (2006) consider a very general setting for ordered and unordered choice models; however, their discussion of IV estimation for these models implicitly conditions on all covariates  $x_i$  (deriving IV weights analogous to  $\omega_j^{IV}(\phi)$  in our setting). Results in this section could, therefore, be derived as a special case of their analysis. While our analysis ignores heterogeneity in the grade-specific effects, it considers estimation under common assumptions about covariates and the way they

---

<sup>10</sup>In Online Appendix A, we further show that with a binary instrument, the  $\omega_j^{IV}(\cdot)$  weights can be more easily interpreted along the lines of the LATE analysis of Angrist and Imbens (1995).

typically enter during estimation. We are not focused on finding an ‘economic interpretation’ for the IV estimator, since the weights we consider can easily be estimated. Instead, we are interested in empirically comparing the OLS and IV weights and deriving a test for whether the different weights can explain differences between the two estimators when per-unit treatment effects are incorrectly assumed to be uniform (i.e. linearity between  $y_i$  and  $s_i$ ).

Since OLS is a special case of IV estimation, in the absence of endogeneity, the OLS estimator for the linear-in- $s_i$  model (equation 1) also converges to a weighted average of the grade-specific effects,  $\beta_j$ , where the weights are non-negative and sum to one.

**Corollary 1.** *If  $E(\varepsilon_i s_i) = 0$  then*

$$\hat{\beta}_{OLS}^L \xrightarrow{p} \sum_{j=1}^S \omega_j^{OLS} \beta_j \quad (5)$$

where the

$$\omega_j^{OLS} = \frac{Pr(s_i \geq j)E(\eta_i | s_i \geq j)}{\sum_{k=1}^S Pr(s_i \geq k)E(\eta_i | s_i \geq k)} \geq 0 \quad (6)$$

sum to unity over all  $j = 1, \dots, S$ .

Proof: This result largely follows from Proposition 1 replacing  $z_i$  with  $s_i$ . Online Appendix A shows that the OLS weights are always non-negative.

The empirical counterpart to the OLS weights,  $\hat{\omega}_j^{OLS} \xrightarrow{p} \omega_j^{OLS}$ , is simply the coefficient estimate on  $s_i$  in an OLS regression of  $D_{ij}$  on  $s_i$  and  $x_i$ . Therefore, only data on  $x_i$  and  $s_i$  are needed to construct consistent estimates of the asymptotic weights. Of course, the weights implied by OLS estimation will not generally equal the weights implied by IV estimation.<sup>11</sup> In Section 4, we graph estimated OLS and IV weights in a few different empirical applications.

Researchers often estimate models like equation (1) rather than the more general equation (2), because they are limited in the instrumental variables at their disposal. Yet, even in the absence

---

<sup>11</sup>For example, consider the case with no  $x$  regressors (except an intercept). It is straightforward to show that  $\omega_{j+1}^{OLS} - \omega_j^{OLS} \propto (E(s_i) - j) \times Pr(s_i = j)$ , which is positive for  $j < E(s_i)$ , zero for  $j = E(s_i)$ , and negative when  $j > E(s_i)$ . This implies that OLS estimation of the linear specification places the most weight on grade-specific  $\beta_j$  effects near the mean schooling level. When schooling is uniformly distributed in the population, the weights decay symmetrically as one moves away from the mean in either direction. Contrast this with the IV weights in the case of a binary instrument  $z_i \in \{0, 1\}$  satisfying the monotonicity assumption. In this case, IV places all the weight on schooling margins that are affected by the instrument, while the underlying distribution of schooling in the population is irrelevant.

of endogeneity and individual-level parameter heterogeneity, there is no reason to expect OLS and IV estimators to be equal for a mis-specified linear-in- $s_i$  model that assumes uniform per-unit treatment effects. As a result, standard Hausman tests applied to the mis-specified linear-in- $s_i$  model may reject the null hypothesis of ‘exogenous  $s$ ’ due simply to variable per-unit treatment effects. Below, we develop a chi-square test for whether OLS estimation of equation (2) yields consistent estimates of the underlying  $\beta_j$  parameters (i.e. whether  $E(\varepsilon_i|s_i) = 0$ ) even when only a single valid instrumental variable is available. However, we first generalize our key results to the case of many instruments.

## 2.2 2SLS Estimation with Multiple Instruments

We now generalize the results to the case where we have  $I$  distinct instruments for schooling,  $z_i = (z_{i1} \dots z_{iI})'$ , but the researcher still estimates the linear-in-schooling model (1). Let  $s_i = x_i'\theta_x + z_i'\theta_z + \xi_i$ , with  $\hat{\theta}_x$  and  $\hat{\theta}_z$  reflecting the corresponding OLS estimates of  $\theta_x$  and  $\theta_z$ . Further define the predicted value of schooling conditional on  $x$  and  $z$ :  $\hat{s}_i = x_i'\hat{\theta}_x + z_i'\hat{\theta}_z$ . Then, 2SLS estimation of equation (1) yields

$$\hat{\beta}_{2SLS}^L = (\hat{s}'M_x\hat{s})^{-1}\hat{s}'M_x y = \sum_{j=1}^S \hat{\omega}_j \beta_j + (\hat{s}'M_x\hat{s})^{-1}\hat{s}'M_x \varepsilon \quad (7)$$

where the “weights”  $\hat{\omega}_j = (\hat{s}'M_x\hat{s})^{-1}\hat{s}'M_x D_j = (\hat{\theta}'_z z' M_x z \hat{\theta}_z)^{-1} \hat{\theta}'_z z' M_x D_j$  reflect consistent estimates of  $\omega_j$  from 2SLS estimation of

$$D_{ij} = s_i \omega_j + x_i' \alpha_j + \psi_{ij}, \quad \forall j \in \{1, \dots, S\}. \quad (8)$$

We assume that Assumption 1 holds for all  $z_{i\ell}$  instruments and that we have sufficient variation in  $z_i$  conditional on  $x_i$  for identification. Let  $\zeta_i = (\zeta_{i1}, \dots, \zeta_{iI})'$  be the  $I \times 1$  vector collecting all  $\zeta_{i\ell} = z_{i\ell} - x_i' \delta_{z\ell}$ , where  $\delta_{z\ell} = [E(x_i x_i')]^{-1} E(x_i z_{i\ell})$  was introduced above in the single-instrument case.<sup>12</sup>

**Assumption 3.** *The covariance matrix for  $z_i$  after partialling out  $x_i$ ,  $E(\zeta_i \zeta_i')$ , is full rank.*

As with the single-instrument IV estimator, we can show that the 2SLS estimator for  $\beta^L$  in equation (1) converges in probability to a “weighted” average of all grade-specific effects. Letting

---

<sup>12</sup>In the case of a single instrument, this analysis reduces to that for IV in the previous subsection with  $\hat{\beta}_{2SLS}^L = \hat{\beta}_{IV}^L$  and  $\hat{\omega}_j = \hat{\omega}_j^{IV}$  for all  $j$ .

$\omega_{j\ell}^{IV}$  reflect the grade  $j$  “weight” from the single-instrument IV estimator using  $z_{i\ell}$  as the instrument as defined by equation (3), the 2SLS estimator “weight” on any  $\beta_j$  is a weighted average of each of these single-instrument IV estimator “weights”.

**Proposition 2.** *Under Assumptions 1 and 3,  $\hat{\beta}_{2SLS}^L \xrightarrow{p} \sum_{j=1}^S \omega_j \beta_j$ , where  $\omega_j = \sum_{\ell=1}^I \Omega_\ell \omega_{j\ell}^{IV}$  sum to unity over all  $j = 1, \dots, S$  and*

$$\Omega_\ell = \frac{\theta_{z\ell} \sum_{k=1}^S \Pr(s_i \geq k) E(\zeta_{i\ell} | s_i \geq k)}{\sum_{m=1}^I \theta_{zm} \sum_{k=1}^S \Pr(s_i \geq k) E(\zeta_{im} | s_i \geq k)} \quad (9)$$

*sum to unity over all  $\ell = 1, \dots, I$ . Furthermore, if each instrument satisfies Assumption 2 and  $E(z_{i\ell} | x_i) = x_i \delta_{z\ell}$ , then all  $\omega_{j\ell}^{IV}$ ,  $\Omega_\ell$ , and  $\omega_j$  are non-negative.*

Proof: See Online Appendix A.

Not surprisingly, one can also show that the 2SLS estimator converges in probability to a weighted average of the probability limits of all single-instrument IV estimators, where the weights are given by  $\Omega_\ell$  in equation (9).<sup>13</sup>

### 3 A Wald Test for Consistent OLS Estimation of All $\beta_j$ 's

When at least one valid instrumental variable is available, the analysis of Section 2 suggests a practical test for whether OLS estimates of  $B \equiv (\beta_1, \dots, \beta_S)$  from equation (2),  $\hat{B}$ , are consistent.<sup>14</sup> We now develop a test that compares the 2SLS estimator from equation (1) with the weighted sum of the grade-specific OLS estimates of the  $\beta_j$ 's from equation (2), using the estimated 2SLS weights  $\hat{\omega} \equiv (\hat{\omega}_1, \dots, \hat{\omega}_S)'$ . Intuitively, if  $E(\varepsilon_i | s_i) = 0$  so the grade-specific OLS estimates are consistent, then the re-weighted sum of these OLS estimates (using the 2SLS weights) should asymptotically equal the 2SLS estimator from equation (1), i.e.  $\hat{\beta}_{2SLS}^L - \hat{\omega}' \hat{B} \xrightarrow{p} 0$ . This will not generally be true when  $E(\varepsilon_i D_{ij}) \neq 0$  for any  $j$ .

Applying 2SLS to equation (8) yields estimates  $\hat{\omega}_j$  and  $\hat{\alpha}_j$  for all  $j$ . In order to derive our test statistic, we frame estimation of  $\hat{B}$ ,  $\hat{\beta}_{2SLS}^L$ , and  $\hat{\omega}$  as a stacked generalized method of moments

<sup>13</sup>If we define  $\beta_{IV,\ell}^L = \text{plim } \hat{\beta}_{IV,\ell}^L$  where  $\hat{\beta}_{IV,\ell}^L$  is the single-instrument IV estimator using  $z_{i\ell}$  as an instrument for  $s_i$  in estimating equation (1), then  $\hat{\beta}_{2SLS}^L \xrightarrow{p} \sum_{\ell=1}^I \Omega_\ell \beta_{IV,\ell}^L$ , where  $\Omega_\ell$  is defined by equation (9).

<sup>14</sup>Formally,  $\hat{B} = (D' M_x D)^{-1} D' M_x y$ , where  $M_x$  and  $y$  are defined earlier and  $D$  reflects the stacked  $N \times S$  matrix of  $(D_{i1}, \dots, D_{iS})$  for all individuals.

(GMM) problem. This establishes joint normality of  $(\hat{B}, \hat{\beta}_{2SLS}^L, \hat{\omega})$  and facilitates estimation of the covariance matrix for all of these estimators. From this, a straightforward application of the delta-method yields the variance of  $\hat{\beta}_{2SLS}^L - \hat{\omega}'\hat{B}$ , which is used in developing a chi-square test statistic for the null hypothesis that  $\hat{T} \equiv \hat{\beta}_{2SLS}^L - \hat{\omega}'\hat{B} \xrightarrow{p} 0$ .

It is necessary to introduce some additional notation in order to define the test statistic. We first define the regressors for OLS estimation of equation (2),  $X_{1i} = (D_i' x_i')$ , and the regressors,  $X_{2i} = (s_i x_i')$ , and instruments,  $Z_{2i} = (z_i' x_i')$ , used in 2SLS estimation of equations (1) and (8). Denote the corresponding matrices for all individuals as  $X_1$ ,  $X_2$ , and  $Z_2$ , respectively. Next, let  $\Theta = (B' \gamma' \beta^L \gamma^{L'} \omega_1' \alpha_1' \dots \omega_S' \alpha_S')$  reflect the full set of parameters to be estimated. Finally, let  $\hat{\Theta}$  denote the corresponding vector of parameter estimates, where  $(B' \gamma')$  is estimated by OLS and  $(\beta^L \gamma^{L'})$  and all  $(\omega_j' \alpha_j')$  are estimated via 2SLS.

The variance of  $\Theta$  can be consistently estimated from

$$\hat{V} = \hat{A}\hat{\Lambda}\hat{A}', \quad (10)$$

where

$$\hat{A} = \begin{pmatrix} [X_1'X_1]^{-1} & \mathbf{0} \\ \mathbf{0} & I_2 \otimes [\hat{X}_2'\hat{X}_2]^{-1}\hat{\Gamma}_2' \end{pmatrix}, \quad (11)$$

$\hat{\Gamma}_2 = (Z_2'Z_2)^{-1}Z_2'X_2$ ,  $\hat{X}_2 = Z_2\hat{\Gamma}_2$ , and  $\mathbf{0}$  reflects conformable matrices of zeros.<sup>15</sup> Furthermore,

$$\hat{\Lambda} = \frac{1}{N} \sum_{i=1}^N \begin{pmatrix} \hat{\varepsilon}_i^2(X_{1i}'X_{1i}) & \hat{\varepsilon}_i\hat{\nu}_i(X_{1i}'Z_{2i}) & \hat{\varepsilon}_i\hat{\Psi}_i' \otimes (X_{1i}'Z_{2i}) \\ \hat{\varepsilon}_i\hat{\nu}_i(Z_{2i}'X_{1i}) & \hat{\nu}_i^2(Z_{2i}'Z_{2i}) & \hat{\nu}_i\hat{\Psi}_i' \otimes (Z_{2i}'Z_{2i}) \\ \hat{\varepsilon}_i\hat{\Psi}_i \otimes (Z_{2i}'X_{1i}) & \hat{\nu}_i\hat{\Psi}_i \otimes (Z_{2i}'Z_{2i}) & \hat{\Psi}_i\hat{\Psi}_i' \otimes (Z_{2i}'Z_{2i}) \end{pmatrix}, \quad (12)$$

where  $\hat{\varepsilon}_i = y_i - D_i'\hat{B} - x_i'\hat{\gamma}$ ,  $\hat{\nu}_i = y_i - s_i\hat{\beta}_{2SLS}^L - x_i'\hat{\gamma}^L$ , and  $\hat{\Psi}_i = (\hat{\psi}_{1i} \hat{\psi}_{2i} \dots \hat{\psi}_{Si})'$  with  $\hat{\psi}_{ij} = D_{ij} - s_i\hat{\omega}_j - \hat{\alpha}_j'x_i$ .

Finally, define  $\hat{T} \equiv T(\hat{\Theta}) = \hat{\beta}_{2SLS}^L - \hat{\omega}'\hat{B}$ , and let

$$\hat{G} \equiv \nabla\hat{T} = (-\hat{\omega}' \ 0_x' \ 1 \ 0_x' \ (-\hat{\beta}_1 \ 0_x') \ (-\hat{\beta}_2 \ 0_x') \ \dots \ (-\hat{\beta}_S \ 0_x')$$

represent the  $(2S + 1 + (S + 2)K) \times 1$  jacobian vector for  $T(\hat{\Theta})$  (where  $0_x$  is a  $K \times 1$  zero vector).

It is now possible to derive a chi-square test statistic.

---

<sup>15</sup>See the proof of Theorem 1 in Online Appendix A.

**Theorem 1.** *Under Assumptions 1 and 3, if  $E(\varepsilon_i|s_i) = 0$ , then*

$$W_N = N \left[ \frac{(\hat{\beta}_{2SLS}^L - \hat{\omega}'\hat{B})^2}{\hat{G}\hat{V}\hat{G}'} \right] \xrightarrow{d} \chi^2(1). \quad (13)$$

Proof: See Online Appendix A.

It is important to note that  $\hat{T} \xrightarrow{p} 0$  need not imply that  $\hat{B} \xrightarrow{p} B$  for two reasons. First, this test cannot tell us anything about whether  $\hat{\beta}_j \xrightarrow{p} \beta_j$  for some grade transition  $j$  if  $\omega_j = 0$ . The test only provides information about the effects of grade transitions that are affected by the instrument. Second, the  $\hat{\beta}_j$  OLS estimates may be asymptotically biased upward for some  $j$  and downward for others. When  $E(\varepsilon_i|s_i) \neq 0$ ,  $\hat{B} \xrightarrow{p} B^* \equiv B + \{E(D_i D_i') - E(D_i x_i') [E(x_i x_i')]^{-1} E(x_i D_i')\}^{-1} E(D_i \varepsilon_i)$ . Thus,  $\hat{T} \xrightarrow{p} 0$  for any  $B^*$  satisfying  $\omega'(B - B^*) = 0$ , where  $\omega \equiv (\omega_1, \dots, \omega_S)'$ . A test based on Theorem 1 would have no power against these alternatives; although, rejection of the null hypothesis would imply that  $\hat{B}$  does not consistently estimate  $B$ .

Under reasonable conditions,  $W_N$  can serve as a valid test statistic for the null hypothesis that  $\hat{B} \xrightarrow{p} B$ . If  $\omega_j > 0$  for all  $j$  (a testable assumption) and if  $E(\varepsilon_i D_{ij}) = E(\varepsilon_i | s_i \geq j) Pr(s_i \geq j)$  were either non-negative for all  $j$  or non-positive for all  $j$ , then all  $\hat{\beta}_j$  would be asymptotically biased in the same direction and  $B^* \neq B \Leftrightarrow \omega'(B - B^*) \neq 0$ . In this case, testing whether  $\hat{T} \xrightarrow{p} 0$  would be equivalent to testing for consistency of  $\hat{B}$ .<sup>16</sup>

To better understand these conditions, consider a standard latent index ordered choice model for schooling of the form:

$$s_i^* = \mu(z_i, x_i) + v_i \quad (14)$$

$$s_i = j \quad \text{if and only if } j \leq s_i^* < j + 1. \quad (15)$$

Assume that all  $x$  regressors and instruments  $z$  are independent of both errors:  $(\varepsilon_i, v_i) \perp\!\!\!\perp (z_i, x_i)$ . It is straightforward to show that if  $E(\varepsilon_i|v_i)$  is weakly monotonic in  $v_i$ , then  $E(\varepsilon_i|s_i \geq j)$  will be either non-positive or non-negative for all  $j$ .<sup>17</sup> Monotonicity of  $E(\varepsilon_i|v_i)$  is trivially satisfied by all joint elliptical distributions (e.g. bivariate normal or t distributions), which produce linear conditional expectation functions.

<sup>16</sup>In the case where some  $\omega_j = 0$ , the test would be equivalent to testing for consistency of all  $\beta_j$  with  $\omega_j > 0$ .

<sup>17</sup>Strictly speaking, weak monotonicity is only required over the range of  $v_i$  covered by  $j - \mu(z_i, x_i)$  (i.e. for  $v_i \in [1 - \mu(z_i, x_i), S - \mu(z_i, x_i)]$ ), so behavior in the tails of the distribution is irrelevant. See Online Appendix A for details.

In practice, one is only likely to fail to reject the null hypothesis of  $\hat{T} \xrightarrow{p} 0$  when  $B^* \neq B$  in cases where individuals with both high and low propensities for education (conditional on observable characteristics) have a higher (or lower) unobserved  $\varepsilon_i$  than individuals with an average propensity for schooling. In the case of an ordered choice model, this would imply a U-shaped (or inverted U-shaped) relationship for  $E(\varepsilon_i|v_i)$ . In many economic contexts, these perverse cases seem unlikely.

We also note that if more than one valid instrument are available, then those instruments can be used in different combinations to perform separate tests. Because each 2SLS estimator (distinguished by the set of instruments used) converges to a different weighted average of the true  $B$  parameters (i.e.  $\omega'_\Upsilon B$  where  $\Upsilon$  denotes the set of instruments used), it is unlikely that one would reject the null of  $\omega'_\Upsilon B = \omega'_\Upsilon B^*$  for all sets of instruments unless  $B = B^*$ .<sup>18</sup>

To demonstrate the extent to which varying per-unit treatment effects can induce differences between OLS and IV estimates that our new exogeneity test can account for (while standard Hausman or Durbin-Wu-Hausman tests applied to equation (1) cannot), we perform a Monte Carlo simulation exercise based on Card’s (1995) log earnings – schooling model. In this framework, varying per-unit treatment effects is equivalent to a non-linear relationship between log earnings and schooling. These results are discussed in detail in Online Appendix B; however, we note here that our test (see Theorem 1) performs well in two important respects. First, the test has nearly identical performance to the standard Hausman test (applied to equation (1)) when all grade-specific effects are the same. Thus, there is no ‘cost’ to using our test rather than the more traditional Hausman test that assumes a linear relationship between log earnings and schooling. Second, our test has very similar properties regardless of the extent of non-linearity between log earnings and schooling, rejecting equality of the re-weighted OLS and IV estimates at noticeably higher rates for even small deviations from exogeneity as long as the instruments are sufficiently strong.

Of course, when the instruments are relatively weak, our test (like the standard Hausman test) has little power to detect endogeneity since the IV estimates tend to have large standard errors. In these cases, negligible amounts of endogeneity may be difficult to detect with our test. This can lead to poor size properties when conducting inference using OLS estimates of the  $\beta_j$  parameters as discussed by Wong (1997) and Guggenberger (2010) who study this issue in the context of linear models and use of the Hausman test to determine exogeneity. Monte Carlo results presented in

---

<sup>18</sup>Because these test statistics are not generally independent, the critical values for this type of joint testing procedure are likely to be quite complicated. We do not address this issue here.



Online Appendix B suggest caution when using OLS estimates for inference – even if our test fails to reject exogeneity – if the instruments are relatively weak. This is particularly true when the IV and re-weighted OLS estimates are quite different but the IV estimates are very imprecise.

Another important limitation to keep in mind is that our test is valid only if equation (2) represents the true model. This model assumes that the regressors are additively separable and that the coefficients are the same for all individuals. In the case of non-separability or individual heterogeneity in the model’s coefficients, our model would be mis-specified and our test invalid.

## 4 Practical Use of our Test and Three Empirical Examples

To demonstrate the practical value of our test, we re-examine three empirical papers on the effects of individual and maternal schooling which estimated 2SLS effects that differ non-trivially from their corresponding OLS estimates.<sup>19</sup> In all cases, the econometric specification assumed a linear relationship between the outcome of interest and educational attainment as in equation (1).<sup>20</sup> Of course, if the true relationship is non-linear so grade-specific effects differ, then differences between OLS and 2SLS weights may explain at least some of the difference between the two estimates. For each of the three cases, we examine the extent to which re-weighting the OLS estimates of the  $\beta_j$ ’s helps reconcile the difference between the potentially mis-specified OLS and 2SLS estimates that assume uniform grade-specific effects. We then test whether schooling is exogenous using both the standard Hausman test and our proposed test.

Results are reported in Table 1.<sup>21</sup> Columns 1 and 2 reproduce OLS and 2SLS estimates using the same models and similar data used in the original papers. For example, the first row indicates that using the Lochner and Moretti (2004) data for white men, a regression of an indicator for

---

<sup>19</sup>The instruments used in these examples have been employed in numerous studies examining a wide array of outcomes. See, e.g. Lochner (2011).

<sup>20</sup>In two of the applications we consider (Lochner and Moretti 2004, Currie and Moretti 2003), the outcome variables are binary and a linear probability model is assumed by the authors. Heteroskedasticity of errors does not pose any problems for our test; however, our assumption of separability between all regressors and measures of schooling is questionable in more general binary choice models for well-known reasons. We simply follow the specifications employed in the earlier studies, assuming the data are consistent with a linear probability model. This may not be unreasonable in these applications given the limited range of predicted outcome probabilities across values of the regressors – assuming an index model based on equation (2), the density for the error may be (approximately) linear over the range of estimated index values.

<sup>21</sup>Details regarding samples and estimating specifications are reported in the bottom of Table 1.

incarceration on years of schooling and controls yields an OLS coefficient equal to -.0010, and a 2SLS coefficient equal to -.0011. The 2SLS estimates use as instrumental variables three indicators for different compulsory schooling ages. The difference between OLS and 2SLS is reported in column 3. The 2SLS estimate is about 10% larger than the OLS estimate (in absolute value), even though most reasonable explanations for the endogeneity of schooling suggest that the OLS estimate should overstate the importance of schooling. The corresponding OLS and 2SLS estimates for Blacks are -.0037 and -.0048, respectively.

There are several well-understood reasons why one might find a larger 2SLS estimate (relative to the OLS estimate), including the presence of measurement error and individual-level heterogeneity in the effects of schooling. It is also possible that non-linearity in the incarceration-schooling relationship may play a role. This seems particularly relevant here given the pattern of OLS estimates for the grade-specific effects  $\beta_j$  reported in Figures 1 and 2. If the assumption of uniform grade-specific effects were correct and these estimates were consistent, all of the estimated  $\hat{\beta}_j$  should be the same. Instead, the estimated  $\hat{\beta}_j$  suggest that the marginal effects of different grade transitions vary considerably across years of schooling. Unless there are much stronger biases for some grades than others, the figures suggest strong non-linearities in the relationship between imprisonment and schooling, with the strongest effect for high school graduation (moving from grade 11 to 12).<sup>22</sup> Based on these findings, Lochner and Moretti (2004) suggest that high school graduation is an important margin for incarceration among men, but they are hesitant to draw strong conclusions from these OLS estimates due to concerns about endogeneity.

The lines in Figures 1 and 2 report estimates of the OLS and 2SLS weights, as defined in Section 2. These weights are clearly very different for white men: the OLS weights are high for years of schooling between 12 and 16, while the 2SLS weights are highest at 12 years of schooling, implying that the effect of moving from 11 to 12 years of schooling figures prominently in the 2SLS estimates. This is not surprising, since the instruments adopted (compulsory schooling laws) are most effective at shifting schooling levels just before or at high school graduation. For black men, the effect of compulsory schooling is strong at earlier grades, so that the weights are more shifted to the left. In column 4 of Table 1, we re-weight the estimated grade-specific effects ( $\hat{\beta}_j$ ) using the

---

<sup>22</sup>Standard errors for the  $\beta_j$  estimates are all less than 0.001 (0.003) for whites (Blacks) except for the first two grade levels. Estimates along with their standard errors are reported in Online Appendix Table C1. For comparison, we also report average marginal effects from analogous logit specifications in Online Appendix Table C2. The pattern of effects is quite similar.

2SLS weights in Figure 1.<sup>23</sup> For whites, the re-weighted OLS estimates are 0.0012, larger than the 2SLS estimates. The re-weighted OLS estimates are larger, because 2SLS puts more weight on the large  $\beta_j$  associated with moving from 11 to 12 years of schooling. For blacks, the re-weighted OLS estimate is smaller, because the 2SLS weights are more shifted to the left and, therefore, put less weight on larger  $\beta_j$ .

The last three columns of Table 1 are the most important, since they report on different tests for the exogeneity of schooling. Column 5 presents test statistics and associated p-values for our proposed test of exogeneity (see Theorem 1), which is valid even when the effects of schooling differ across grades. Columns 6 and 7 present results from the standard Hausman test and the Durbin-Wu-Hausman test (applied to the linear-in-schooling specification), respectively, which are both incorrect when the grade-specific effects differ. For white men, our test fails to reject, which is quite important in practice, since it suggests that our OLS estimates of the  $\beta_j$  in Figure 1 are consistent. Given a high first stage F-statistic of 1000.3 and the fact that the re-weighted OLS estimate is very close to the 2SLS estimate (a difference of less than 10%), it seems reasonable to conclude from our OLS estimates of  $\beta_j$  that high school completion has the strongest effect on incarceration rates while college attendance has much weaker effects.<sup>24</sup> This is extremely useful, since with our limited set of instruments, it is impossible to estimate all 20  $\beta_j$  parameters using 2SLS. Indeed, 2SLS estimates from highly restricted two-parameter models that relax linearity in schooling are very imprecise. Fortunately, our test suggests that IV methods are not necessary in this case.

The case of incarceration for black men is different: our test strongly rejects the hypothesis that the re-weighted OLS and 2SLS estimates are the same (p-value of .0005), while the standard Hausman test fails to reject. Re-weighting the OLS estimates for the  $\beta_j$  parameters reveals that the OLS estimates are significantly biased towards zero, on average, since the re-weighted OLS estimate is -.0007 compared to the 2SLS estimate of -.0048. In this case, we cannot draw any strong conclusions about the relative importance of different grades due to these biases. These findings empirically demonstrate that when grade-specific effects may differ, the standard Hausman test

---

<sup>23</sup>The standard error for this re-weighted effect is derived using the delta-method and the estimated covariance matrix  $\hat{V}$  defined in Section 3.

<sup>24</sup>See Online Appendix Table C1 for coefficient estimates and their standard errors. While it is possible that some  $\hat{\beta}_j$  are biased upwards and others downwards so as to perfectly offset when the 2SLS weights are applied, this seems highly unlikely given the economics of the problem (see, e.g., Lochner and Moretti (2004)).

can fail to detect an endogeneity problem when one exists.

In the second panel, we turn to estimates of the effect of maternal schooling on infant health from Currie and Moretti (2003). The instrument in this case is an indicator for college proximity. (First stage F-statistics for the instruments are 398.7.) In this case, the re-weighted OLS estimates (column 4) are generally quite similar to the OLS estimates (column 1). Looking at Figures 3 and 4, it is clear why: the OLS and 2SLS weights are nearly identical. Not surprisingly, our test and the standard Hausman test produce very similar test statistics and the same conclusions: exogeneity cannot be rejected for either child health outcome.

Finally, in the bottom panel, we turn to estimates of the private return to schooling using three dummies for compulsory schooling as instruments. While this analysis is based on that of Acemoglu and Angrist (2000), we consider the effects of schooling on log annual earnings rather than weekly wages for white men in their 40s. Figure 5 reports the OLS estimates of the  $\beta_j$  parameters as well as the OLS and 2SLS weights. OLS estimates indicate that an additional year of schooling translates into an 8.2% increase in annual earnings, while the 2SLS estimates suggest a much larger return. The re-weighted OLS estimates fall in between the OLS and 2SLS estimates, although they are much closer to the OLS estimates. The effect of re-weighting is minor despite substantially different OLS and 2SLS weights. Our test rejects the hypothesis that the re-weighted OLS and 2SLS estimates are equal, even though the instruments are not particularly strong in this application (the first stage F-statistic for the instruments is only 29.5).

## 5 Conclusions

In applied work, OLS and IV estimates often differ. In many cases, the sign of the difference is surprising given economic theory and plausible assumptions about the direction of endogeneity bias. Influential work by Angrist and Imbens (1994, 1995) and Heckman and Vytlacil (2005) has clarified the interpretation of IV estimates as a local average treatment effect when the regression parameter of interest varies across individuals. Our work complements the existing understanding of the differences between IV and OLS estimates when the model is mis-specified.

We consider a specific class of models with a single finite-valued discrete endogenous regressor, exogenous regressors that are additively separable and enter linearly, and coefficients that do not vary across individuals. Models of this type are widely used in empirical research to study the

effects of multi-valued program treatments, drug dosage levels, and schooling attainment. We focus attention on the possibility that per-unit treatment effects vary across levels of treatment.

The growing focus on identification of causal effects in economics has led many researchers to estimate models of this type using IV methods. Yet, due to the limited availability of valid instruments, it is common to estimate models that assume uniform per-unit treatment effects even when those effects are likely to vary across treatment levels as frequently suggested by more general specifications estimated using OLS. We show that, in this case, OLS and IV/2SLS estimators identify different weighted averages of all per-unit effects, which can lead to incorrect conclusions about endogeneity when using a standard Hausman test.<sup>25</sup>

The main contribution of this paper is to develop a simple generalization of the Hausman test to assess whether differential weighting and variable per-unit treatment effects can explain the difference between OLS and IV/2SLS estimators. Within the class of models under consideration, this serves as a specification test for exogeneity under reasonable conditions. Conveniently, this test only requires a single instrument, making it useful in many applications.

## References

Daron Acemoglu and Joshua D. Angrist. “How Large are Human-Capital Externalities? Evidence from Compulsory-Schooling Laws,” in *NBER Macroeconomics Annual 2000*, Vol. 15, 9–74, National Bureau of Economic Research, 2001.

Angrist, D. Joshua, Kathryn Graddy, and Guido W. Imbens. “The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Models with an Application to the Demand for Fish,” *Review of Economic Studies*, 67, 499–527, 2000.

Angrist, D. Joshua, and Guido W. Imbens. “Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity,” *JASA*, 90, 431–442, 1995.

Card, David. “Using Geographic Variation in College Proximity to Estimate the Return to Schooling” in “*Aspects of Labour Economics: Essays in Honour of John Vanderkamp*”, edited by Louis Christofides, E. Kenneth Grant and Robert Swindinsky. University of Toronto Press. 1995.

---

<sup>25</sup>Other important concerns include the strength and exogeneity of the instrument(s). Our approach, abstracts from instrument-related problems, instead addressing problems associated with mis-specification in the structural equation if at least one exogenous instrument is available.

Card, David. “The Causal Effect of Education on Earnings”. In Orley Ashenfelter and David Card, editors, *Handbook of Labor Economics Volume 3A*. Amsterdam: Elsevier, 1999.

Carneiro, Pedro, James Heckman and Edward Vytlačil. “Evaluating Marginal Policy Changes and the Average Effect of Treatment for Individuals at the Margin,” *Econometrica*, 78(1), 2010.

Currie, Janet, and Enrico Moretti. “Mother’s Education and the Intergenerational Transmission of Human Capital: Evidence from College Openings,” *Quarterly Journal of Economics*, 118(4), 2003.

Guggenberger, Patrik. “The Impact of a Hausman Pretest on the Asymptotic Size of a Hypothesis Test,” *Econometric Theory*, 26, 369–82, 2010.

Hausman, J.A. “Specification Tests in Econometrics”, *Econometrica*, 46(6), 1251-71, 1978.

Heckman, James J., and Edward Vytlačil, “Instrumental Variables Methods for the Correlated Random Coefficient Model,” *Journal of Human Resources*, 33(4), 1998.

Heckman, James J., and Edward Vytlačil. “Structural Equations, Treatment Effects, and Econometric Policy Evaluation,” *Econometrica*, *Econometric Society*, vol. 73(3), 669–738, 2005.

Heckman, James J., Lance Lochner, and Petra Todd. “Earnings Functions and Rates of Return,” *Journal of Human Capital*, 2(1), 2008.

Imbens, Guido, and Joshua Angrist. “Identification and Estimation of Local Average Treatment Effects,” *Econometrica*, 62(2), 1994.

Hungford, Thomas, and Gary Solon. “Sheepskin Effects in the Returns to Education,” *Review of Economics and Statistics*, 1987.

Jaeger, David, and Marianne Page. “Degrees matter: New Evidence on Sheepskin Effects in the Returns to Education”, *Review of Economics and Statistics*, 78(4), 1996.

Kling, Jeffrey. “Interpreting Instrumental Variables Estimates of the Returns to Schooling”, *Journal of Business and Statistics*, 2000.

Lochner, Lance. “Nonproduction Benefits of Education: Crime, Health, and Good Citizenship,” in E. Hanushek, S. Machin, and L. Woessmann (eds.), *Handbook of the Economics of Education*, Vol. 4, Ch. 2, Amsterdam: Elsevier Science, 2011.

Lochner, Lance, and Enrico Moretti. “The Effect of Education on Criminal Activity: Evidence from Prison Inmates, Arrests and Self-Reports”, NBER Working Paper No. 8605, 2001.

Lochner, Lance, and Enrico Moretti. “The Effect of Education on Criminal Activity: Evidence from Prison Inmates, Arrests and Self-Reports”, *American Economic Review*. 94(1), 2004.

Lochner, Lance, and Enrico Moretti. “Estimating and Testing Non-Linear Models Using Instrumental Variables” NBER Working Paper No. 17039, 2011.

Moffitt, Robert. “Estimating Marginal Treatment Effects in Heterogeneous Populations,” *Annales d’Economie et de Statistique*, Special Issue on Econometrics of Evaluation, Fall 2009.

Mogstad, Magne, and Matthew Wiswall. “Linearity in Instrumental Variables Estimation: Problems and Solutions,” Working Paper, 2010.

Park, Jin Heun. “Estimation of Sheepskin Effects Using the Old and New Measures of Educational Attainment in the CPS,” *Economic Letters* 62, 1999.

White, Halbert, “Consequences and Detection of Misspecified Nonlinear Regression Models,” *Journal of the American Statistical Association*, 76, 419–33, 1981.

Wong, Ka-fu. “Effects on Inference of Pretesting the Exogeneity of a Regressor,” *Economic Letters*, 56, 267–71, 1997.

Wooldridge, Jeffrey. “On Two Stage Least Squares Estimation of the Average Treatment Effect in Random Coefficient Models,” *Economics Letters*, 56, 1997.

Yitzhaki, Shlomo. “On Using Linear Regressions in Welfare Economics,” *Journal of Business and Economic Statistics*, 14, 478–486, 1996.

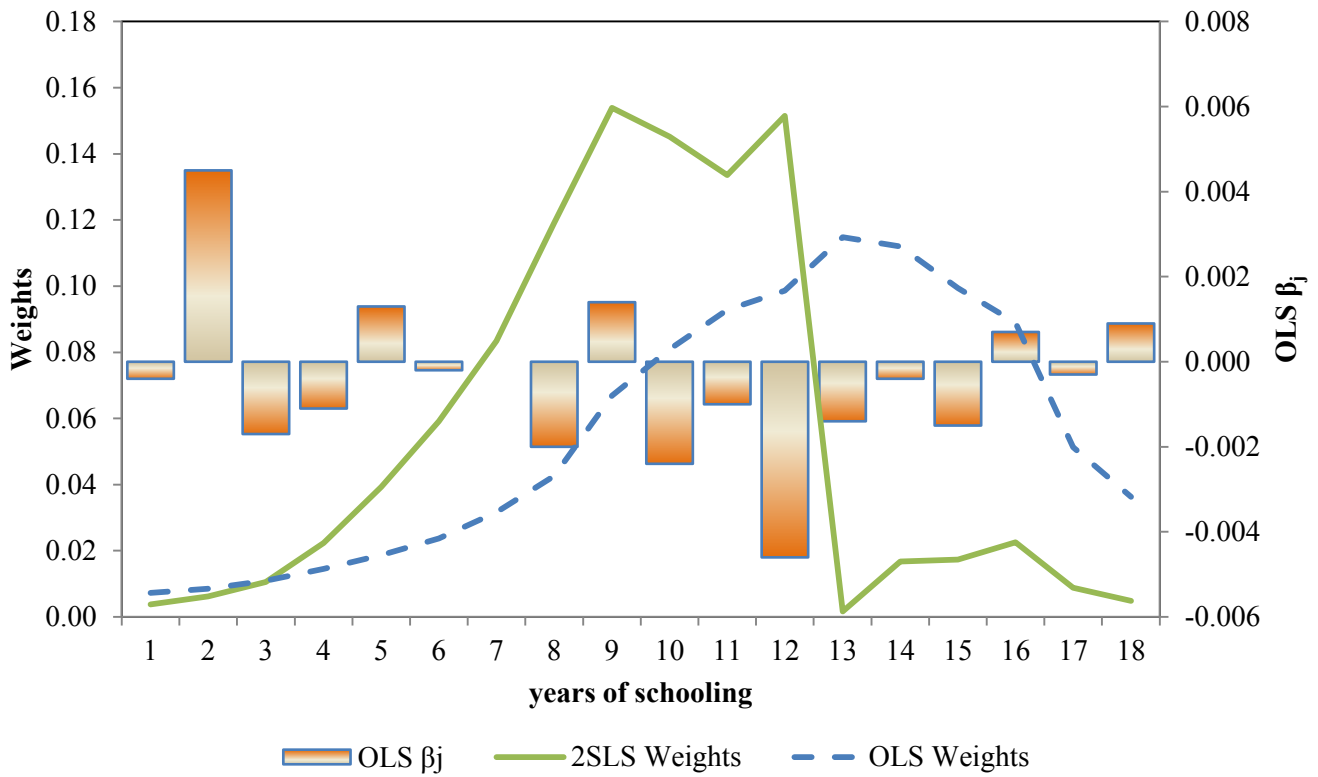
Table 1: Replication Results and Application of Wald Test for Endogeneity

	$\hat{\beta}_{OLS}^L$	$\hat{\beta}_{2SLS}^L$	$\hat{\beta}_{OLS}^L - \hat{\beta}_{2SLS}^L$	$\sum_j \hat{\omega}_j \hat{\beta}_j$	General Wald Test Statistic [p-value]	Hausman Test Statistic [p-value]	DWH Test Statistic [p-value]
<b>1. Lochner &amp; Moretti (2004):</b> Effect of Years of Schooling on Imprisonment							
White Males	-0.0010 (0.0000)	-0.0011 (0.0004)	-0.0002 (0.0004)	-0.0012 (0.0000)	0.0225 [0.8808]	0.2021 [0.6530]	0.1600 [0.6858]
Black Males	-0.0037 (0.0001)	-0.0048 (0.0012)	-0.0011 (0.0011)	-0.0007 (0.0002)	11.9441 [0.0005]	0.9757 [0.3233]	0.5154 [0.4728]
<b>2. Currie &amp; Moretti (2003):</b> Effect of Maternal Education on Infant Health							
Low birth weight	-0.0050 (0.0001)	-0.0098 (0.0038)	-0.0048 (0.0037)	-0.0053 (0.0002)	1.4376 [0.2305]	1.7022 [0.1920]	1.5566 [0.2122]
Pre-term birth	-0.0044 (0.0002)	-0.0104 (0.0044)	-0.0060 (0.0042)	-0.0046 (0.0002)	1.7639 [0.1841]	2.0472 [0.1525]	1.7749 [0.1828]
<b>3. Acemoglu &amp; Angrist (2001):</b> Private Returns to Schooling							
Annual Earnings	0.0822 (0.0003)	0.1442 (0.0256)	0.0620 (0.0253)	0.0832 (0.0017)	5.7093 [0.0169]	6.0028 [0.0143]	6.0218 [0.0141]

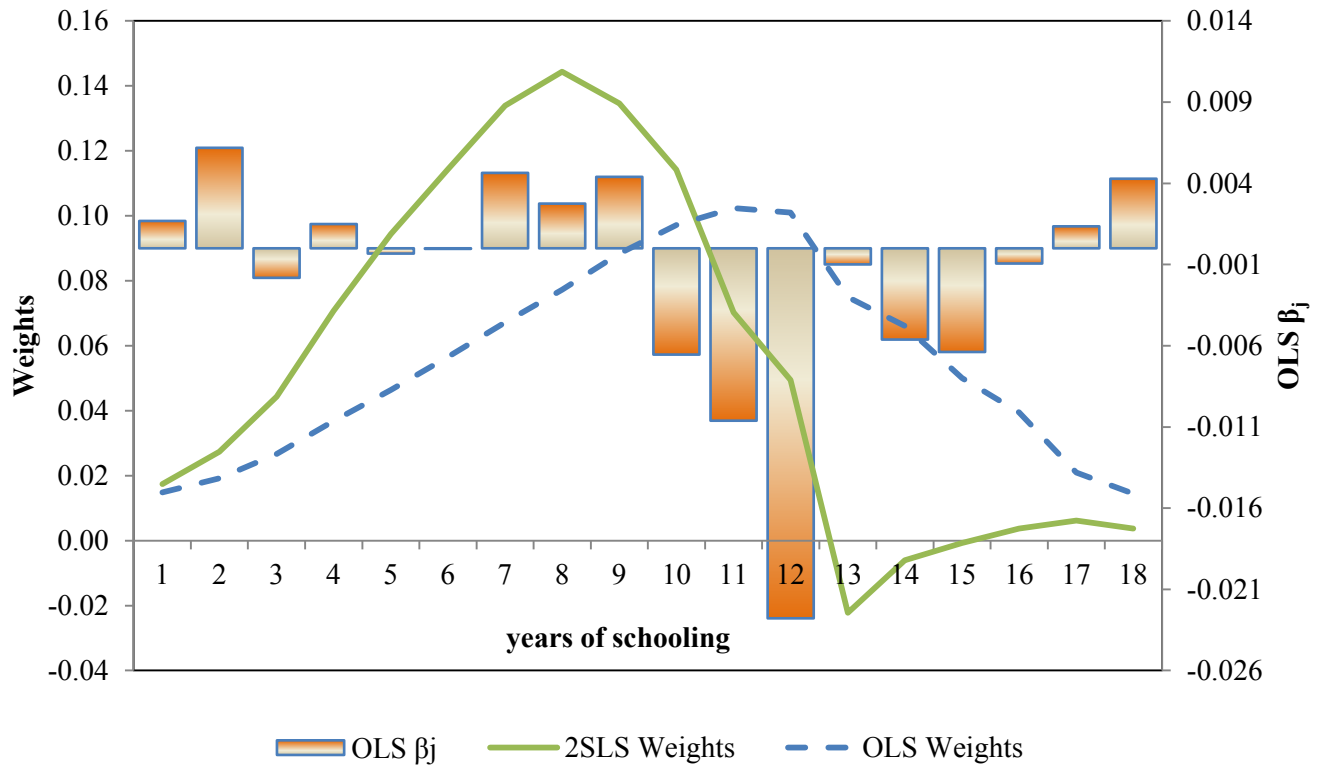
Notes: The first four columns report estimates for reported parameters with standard errors in parentheses. Columns for General Wald Test, Hausman, and Durbin-Wu-Hausman (DWH) Test report test statistics [p-values] for the null hypothesis of exogeneity. General Wald Test compares  $\hat{\beta}_{2SLS}^L$  and  $\sum_j \hat{\omega}_j \hat{\beta}_j$  as described in Theorem 1, while the Hausman and DWH Tests compare  $\hat{\beta}_{2SLS}^L$  and  $\hat{\beta}_L^{OLS}$ . Specifications for Lochner and Moretti (2004) use men ages 20-60 from the 1960-80 U.S. Censuses and include indicators for three-year age categories, year, state of birth, and state of residence. Specifications for blacks also include an indicator for whether the individual turned age 14 after 1957 and was born in the South. Specifications from Currie and Moretti (2003) use first-time white mothers ages 24-35 from Vital Statistics Natality records from 1970 to 1999 and include median county income, percent urban in county when the mother was 17, and indicators for ten-year birth cohorts, mother's age, and county-specific year of child's birth effects. Specifications for Acemoglu and Angrist (2001) results differ slightly from theirs, since we only use compulsory attendance indicators for instruments and do not estimate the 'social return' to schooling. Specifications use 40-49 year-old white men from the 1960-80 U.S. Censuses and include indicators for Census year, year of birth, state of birth, and state of residence.



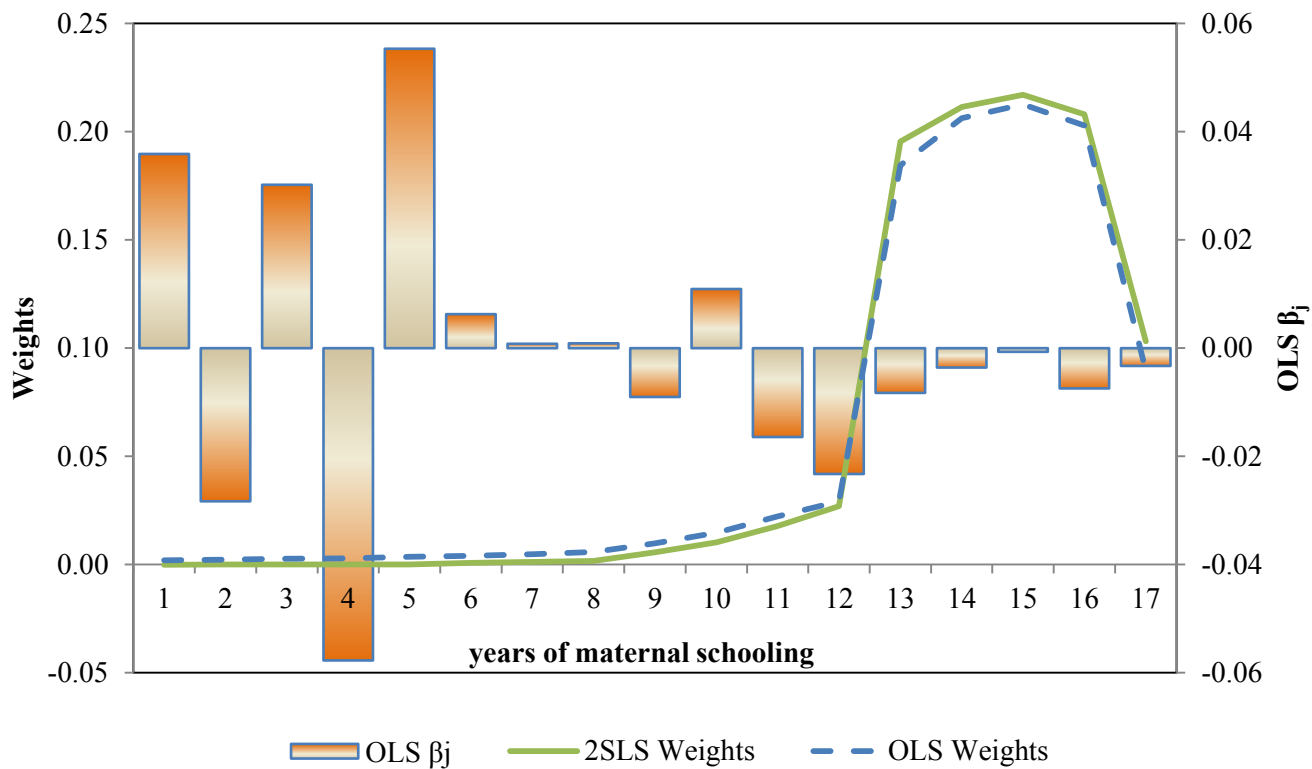
**Figure 1: Effects of Schooling on the Probability of Incarceration for White Males (OLS Estimates and Weights)**



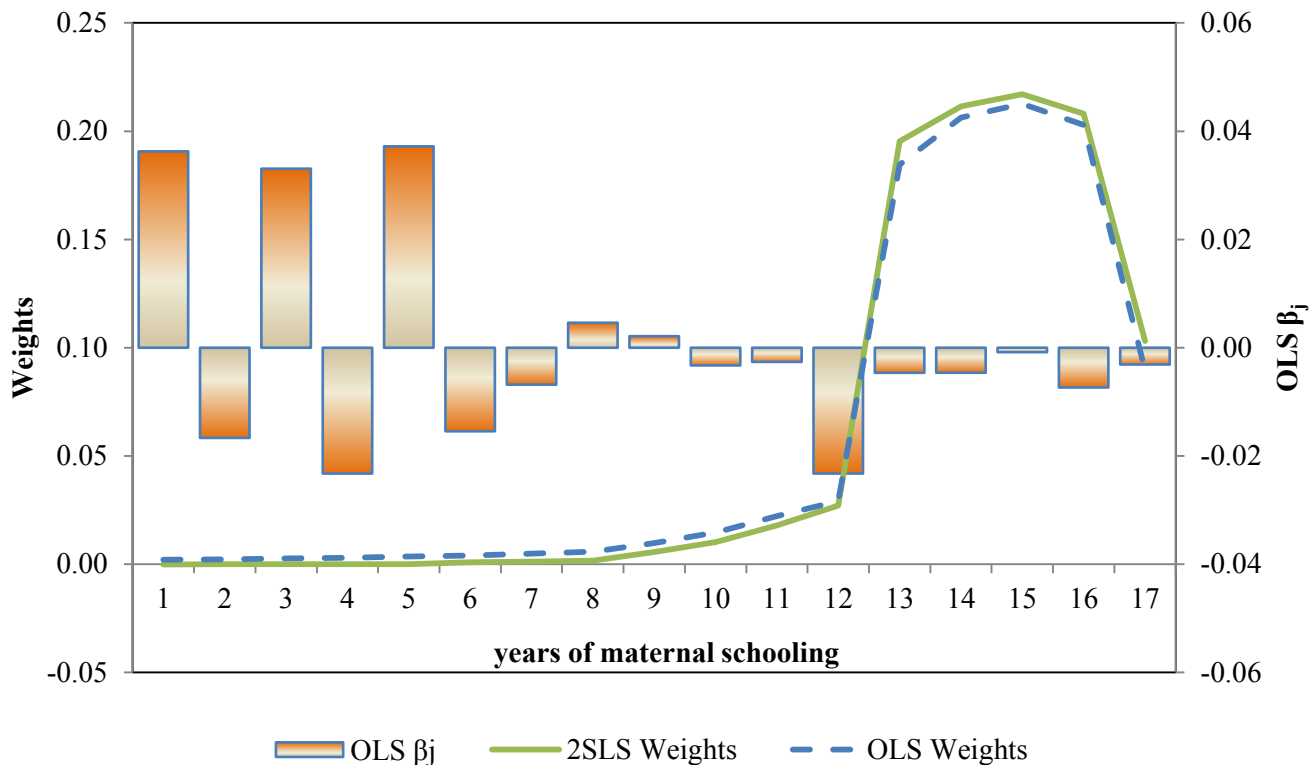
**Figure 2: Effects of Schooling on the Probability of Incarceration for Black Males (OLS Estimates and Weights)**



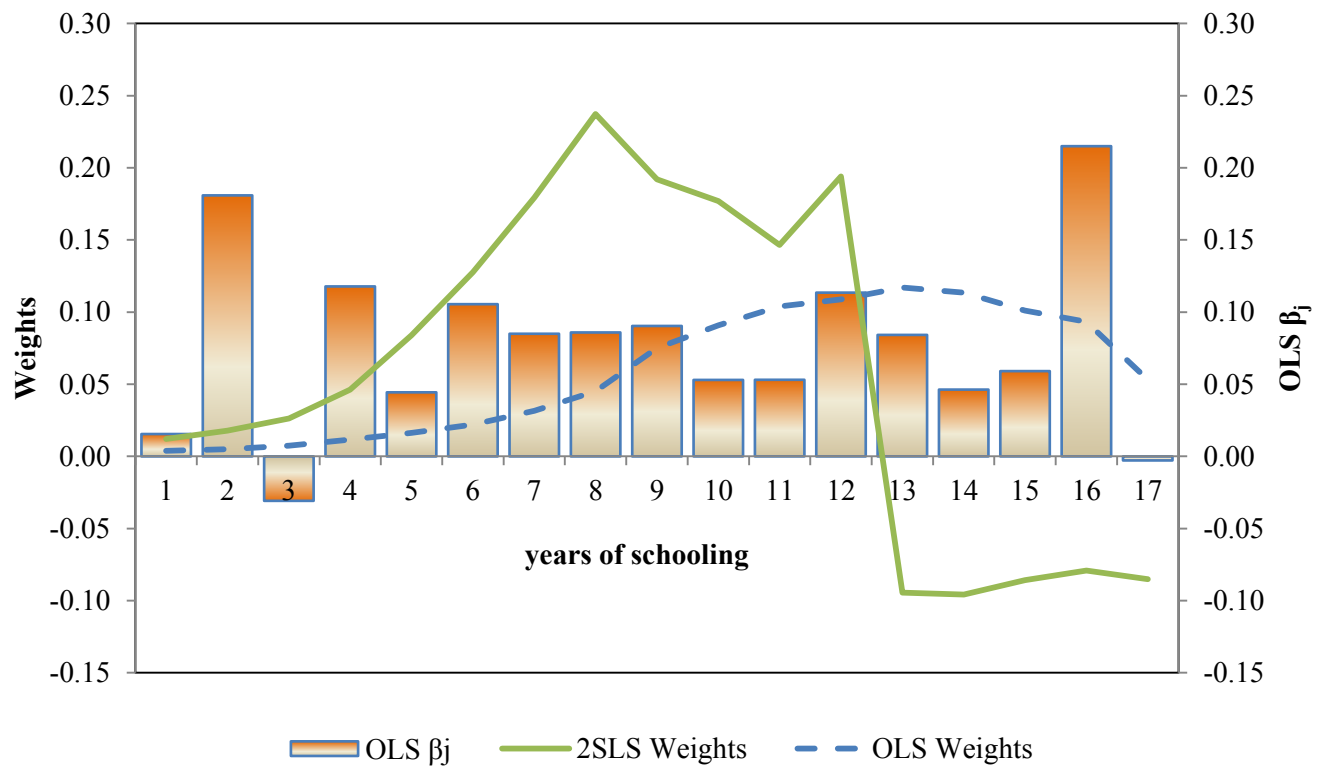
**Figure 3: Effects of Maternal Schooling on the Probability of Low Birth Weight (OLS Estimates and Weights)**



**Figure 4: Effects of Maternal Schooling on the Probability of Pre-Term Birth (OLS Estimates and Weights)**



**Figure 5: Effects of Schooling on Log Annual Earnings for Men  
(OLS Estimates and Weights)**



## Online Appendix A: Proofs and Technical Results

This is an online appendix for Lochner and Moretti (2012) that provides proofs of key propositions and theorems along with a few other technical results discussed in the paper.

### Proof of Proposition 1

It is straightforward to show that  $\hat{\omega}_j^{IV} \xrightarrow{P} \omega_j^{IV}$ , since  $\frac{1}{N} \sum_{i=1}^N \tilde{z}_i D_{ij} \xrightarrow{P} E(D_{ij} \zeta_i) = Pr(s_i \geq j) E(\zeta_i | s_i \geq j)$ ,  $\frac{1}{N} \sum_{i=1}^N \tilde{z}_i \tilde{s}_i \xrightarrow{P} E(\eta_i z_i)$  which is assumed to be non-zero, and  $\hat{\omega}_j^{IV}$  and  $\omega_j^{IV}$  sum to one over  $j = 1, \dots, S$ . The assumptions  $E(\varepsilon_i z_i) = 0$  and  $E(\varepsilon_i x_i) = 0$  imply that  $\frac{1}{N}(\tilde{z}'\varepsilon) \xrightarrow{P} 0$ . This proves the first part of the result.

To prove the second part of the result, note that the assumption  $E(z_i | x_i) = x_i' \delta_z$  implies

$$\frac{1}{N} \sum_{i=1}^N \tilde{z}_i D_{ij} = \frac{1}{N} \sum_{i=1}^N [z_i - x_i \hat{\delta}_z] D_{ij} \xrightarrow{P} E[(z_i - E(z_i | x_i)) D_{ij}] = E\{Cov(z_i, D_{ij} | x_i)\},$$

where  $\hat{\delta}_z = (x'x)^{-1}x'z \xrightarrow{P} \delta_z$ . Denoting the density function for  $z_i$  conditional on  $x_i$  by  $F_{z|x}(\cdot)$ , the  $Cov(z_i, D_{ij} | x_i) = \int [\varphi - E(\varphi | x)] Pr(D_{ij} = 1 | z_i = \varphi, x_i) dF_{z|x}(\varphi | x_i)$  is non-negative for all  $x_i$  and  $j$  if  $\partial Pr(D_{ij} = 1 | z_i, x_i) / \partial z \geq 0$  for all  $x_i$  and  $j$ . This is ensured by Assumption 2. Using the fact that the weights sum to one concludes the proof.

QED

### More Interpretable Weights

With a binary instrument, the  $\omega_j^{IV}(x_i)$  weights can be more easily interpreted along the lines of the LATE analysis of Angrist and Imbens (1995). For  $z_i \in \{0, 1\}$  and  $\pi(x_i) \equiv Pr(z_i = 1 | x_i)$ ,

$$Cov(z_i, D_{ij} | x_i) = \pi(x_i)[1 - \pi(x_i)][Pr(D_{ij} = 1 | z_i = 1, x_i) - Pr(D_{ij} = 1 | z_i = 0, x_i)].$$

In this case, the  $x$ -specific weights simplify to

$$\omega_j^{IV}(x_i) = \frac{Pr(D_{ij} = 1 | z_i = 1, x_i) - Pr(D_{ij} = 1 | z_i = 0, x_i)}{\sum_{k=1}^S [Pr(D_{ik} = 1 | z_i = 1, x_i) - Pr(D_{ik} = 1 | z_i = 0, x_i)]}.$$

Thus,  $\beta_{IV}(x_i)$  weights each  $\beta_j$  based on the fraction of all grade increments induced by a change in the instrument (conditional on  $x_i$ ) that are due to persons switching from less than  $j$  to  $j$  or more years of school. The effects of grade transitions at schooling levels that are unaffected by the

instrument receive zero weight. The IV estimator for the full sample weights each of the  $x$ -specific estimators according to the relative covariance of schooling with the outcome measure conditional on the value of  $x_i$ .

Under Assumptions 1 and 2, if  $E(x_i|z_i) = E(x_i)$ , then the weights in equations (3) or (4) simplify considerably, becoming independent of  $x_i$ :

$$\omega_j^{IV} = \frac{Pr(D_{ij} = 1|z_i = 1) - Pr(D_{ij} = 1|z_i = 0)}{\sum_{k=1}^S [Pr(D_{ik} = 1|z_i = 1) - Pr(D_{ik} = 1|z_i = 0)]} = \frac{Pr[s_i(0) < j \leq s_i(1)]}{\sum_{k=1}^S Pr[s_i(0) < k \leq s_i(1)]}.^{26}$$

The additional mean independence assumption  $E(x_i|z_i) = E(x_i)$  may apply naturally to many ‘natural experiments’, making this simple expression useful in those contexts. The resulting weights reflect the fraction of all grade increments induced by a change in the instrument that are due to persons switching from less than  $j$  to  $j$  or more years of school. The IV estimator, therefore, identifies the average effect of an additional year of schooling, where the average is taken across all grade increments induced by the instrument. If individuals change schooling no more than one grade in response to a change in the value of the instrument, then the IV estimator reflects the average marginal effect of an additional year of school among individuals affected by the instrument.

### Proof that OLS Weights are Non-negative in Corollary 1

To see that the OLS weights are always non-negative, note that the numerator for  $\omega_j^{OLS}$  equals  $E(\eta_i D_{ij})$ . To see that this is non-negative, notice that

$$E(\eta_i) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \eta dF_{\eta|x}(\eta|\phi) dF_x(\phi) + \int_{-\infty}^{\infty} \int_{-\infty}^{j-\phi'\delta_s} \eta dF_{\eta|x}(\eta|\phi) dF_x(\phi), \quad (16)$$

where  $F_x(\cdot)$  reflects the density of  $x_i$  and  $F_{\eta|x}(\cdot|\cdot)$  the density of  $\eta_i$  conditional on  $x_i$ . Assuming  $x_i$  includes a constant term,  $E(\eta_i) = 0$ . Since the first term in equation (16) is clearly greater than or equal to the second term and their sum is zero, the first term must be non-negative. Of course, the first term equals  $E(\eta_i D_{ij})$ .

QED

### Proof of Proposition 2

First, note that  $\hat{s}'M_x\hat{s} = s'M_x z(z'M_x z)^{-1}z'M_x s$ . Since,  $\frac{1}{N}s'M_x z \xrightarrow{P} E[(s_i - x_i'\delta_s)z_i'] = E(\eta_i z_i') \neq 0$  by Assumption 1 and  $\frac{1}{N}z'M_x z \xrightarrow{P} E[z_i(z_i' - x_i'\delta_s)] = E(z_i\zeta_i') = E(\zeta_i\zeta_i')$ , which is full rank by

<sup>26</sup>See the Appendix of Locher and Moretti (2011) for a proof of this result.

Assumption 3, the denominator for  $\omega_j$  is non-zero.

Since  $\sum_{j=1}^S \hat{s}' M_x D_j = \hat{s}' M_x s = \hat{s}' M_x \hat{s}$ , both  $\hat{\omega}_j$  and  $\omega_j$  sum to one. Now, consider the numerator for  $\hat{\omega}_j$ :

$$\frac{1}{N} \hat{\theta}'_z z' M_x D_j \xrightarrow{p} \sum_{\ell=1}^I \theta_{z\ell} E(D_{ij} \zeta_{i\ell}),$$

where  $\theta_{z\ell}$  corresponds to the  $\theta_z$  coefficient on  $z_{i\ell}$ . Since the  $\omega_j$  sum to one, we can write

$$\begin{aligned} \omega_j &= \frac{\sum_{\ell=1}^I \theta_{z\ell} E(D_{ij} \zeta_{i\ell})}{\sum_{k=1}^S \sum_{m=1}^I \theta_{zm} E(D_{ik} \zeta_{im})} \\ &= \frac{\sum_{\ell=1}^I \theta_{z\ell} \left[ \omega_{j\ell}^{IV} \sum_{k=1}^S E(D_{ik} \zeta_{i\ell}) \right]}{\sum_{k=1}^S \sum_{m=1}^I \theta_{zm} E(D_{ik} \zeta_{im})} \\ &= \frac{\sum_{\ell=1}^I \omega_{j\ell}^{IV} \left[ \theta_{z\ell} \sum_{k=1}^S E(D_{ik} \zeta_{i\ell}) \right]}{\sum_{m=1}^I \theta_{zm} \sum_{k=1}^S E(D_{ik} \zeta_{im})} \\ &= \sum_{\ell=1}^I \Omega_\ell \omega_{j\ell}^{IV} \end{aligned}$$

where  $\omega_{j\ell}^{IV} = \frac{E(D_{ij} \zeta_{i\ell})}{\sum_{k=1}^S E(D_{ik} \zeta_{i\ell})}$  since  $E(D_{ij} \zeta_{i\ell}) = Pr(s_i \geq j) E(\zeta_{i\ell} | s_i \geq j)$ . Substituting the latter in

where it appears above,  $\Omega_\ell$  is given by equation (9).

Also, note that  $\frac{1}{N} \hat{s}' M_x \varepsilon \xrightarrow{p} \theta_z [E(z_i \varepsilon_i) + E(z_i x'_i) E(x_i x'_i) E(x_i \varepsilon_i)] = 0$ , since  $E(\varepsilon_i x_i) = 0$  and  $E(z_i \varepsilon_i) = 0$ . This implies that  $\hat{\beta}_{2SLS}^L \xrightarrow{p} \sum_{j=1}^S \omega_j \beta_j$ .

Finally, it is clear from the proof of Proposition 1 that if each instrument satisfies Assumption 2 and  $E(z_{i\ell} | x_i) = x'_i \delta_{z\ell}$ , then all  $\Omega_\ell$ ,  $\omega_{j\ell}^{IV}$ , and  $\omega_j$  are non-negative.

QED

## Proof of Theorem 1

Proposition 2 shows that the 2SLS estimator from equation 1 converges to a “weighted average” of the true  $\beta_j$ 's with the “weights”,  $\omega = (\omega_1, \dots, \omega_S)'$ , consistently estimated by 2SLS estimation of equation (8). That is,  $\hat{\omega} \xrightarrow{p} \omega$  and  $\hat{\beta}^{2SLS} \xrightarrow{p} \omega' B$ . If  $E(\varepsilon_i | s_i) = 0$ , then  $\hat{B} \xrightarrow{p} B$ , which implies that  $\hat{\beta}_{2SLS}^L - \hat{\omega}' \hat{B} \xrightarrow{p} 0$ .

We write the estimation problems for equations (2), (1), and (8) in the form of a stacked linear GMM problem. (Note that equation (2) is estimated using OLS while the remaining equations are estimated using 2SLS.) This establishes joint normality of  $(\hat{B}, \hat{\beta}_{2SLS}^L, \hat{\omega})$  in the limit and facilitates estimation of their covariance matrix. A straightforward application of the delta-method yields the variance of  $\hat{T} \equiv \hat{\beta}_{2SLS}^L - \hat{\omega}'\hat{B}$ , which is used in deriving a chi-square test statistic for the null hypothesis that  $\hat{T} \xrightarrow{p} 0$ .

Diagonally stack the regressor and instrument vectors for all equations as follows:

$$X_i = \begin{pmatrix} X_{1i} & \mathbf{0} \\ \mathbf{0} & I_2 \otimes X_{2i} \end{pmatrix} \quad \text{and} \quad Z_i = \begin{pmatrix} X_{1i} & \mathbf{0} \\ \mathbf{0} & I_2 \otimes Z_{2i} \end{pmatrix},$$

where  $I_2$  is an identity matrix of dimension  $S+1$  and  $\mathbf{0}$ 's reflect conformable vectors of zeros. Next, define  $Y_i = (y_i \ y_i \ D_i)'$  and  $U_i = (\varepsilon_i \ \nu_i \ \Psi_i)'$ , where  $\Psi_i = (\psi_{1i} \ \psi_{2i} \ \dots \ \psi_{Si})'$ . Recall from Section 3 that  $\Theta = (B' \ \gamma' \ \beta^L \ \gamma^{L'} \ \omega_1' \ \alpha_1' \ \dots \ \omega_S' \ \alpha_S)'$  is the full set of parameters to be estimated. ( $\hat{\Theta}$  reflects the corresponding vector of parameter estimates). Now, the three sets of estimating equations can be compactly re-written as:

$$Y_i = X_i\Theta + U_i.$$

Equation-by-equation estimation of (2), (1), and (8) (the first by OLS and the second and third by 2SLS) is mathematically equivalent to GMM estimation for this system:

$$\min_{\Theta} \left[ \sum_{i=1}^N Z_i'(Y_i - X_i\Theta) \right]' \hat{\Omega} \left[ \sum_{i=1}^N Z_i'(Y_i - X_i\Theta) \right],$$

using the weighting matrix  $\hat{\Omega} = \left[ \frac{1}{N} \sum_{i=1}^N Z_i'Z_i \right]^{-1} \xrightarrow{p} [E(Z_i'Z_i)]^{-1} \equiv \Omega$ . Stacking all individual-specific matrices into large matrices and using matrix notation, this system GMM estimator is  $\hat{\Theta} = [X'Z(Z'Z)^{-1}Z'X]^{-1} X'Z(Z'Z)^{-1}Z'Y$ .

Standard results in GMM estimation (under the assumptions specified in Theorem 1) imply that  $\sqrt{N}(\hat{\Theta} - \Theta) \xrightarrow{d} N(0, V)$  where

$$\begin{aligned} V &= (C'\Omega C)^{-1} C'\Omega \Lambda \Omega C (C'\Omega C)^{-1} \\ C &= E(Z_i'X_i) \\ \Lambda &= E(Z_i'U_iU_i'Z_i) \end{aligned}$$

and  $\Omega$  is defined above.<sup>27</sup>

<sup>27</sup>Substituting in for  $C$  and  $\Omega$  and simplifying yields

$$V = \{E(X_i'Z_i)[E(Z_i'Z_i)]^{-1}E(Z_i'X_i)\}^{-1} E(X_i'Z_i)[E(Z_i'Z_i)]^{-1}\Lambda[E(Z_i'Z_i)]^{-1}E(Z_i'X_i) \{E(X_i'Z_i)[E(Z_i'Z_i)]^{-1}E(Z_i'X_i)\}^{-1}.$$

Letting  $\hat{\Gamma} = (Z'Z)^{-1}Z'X$ ,  $\hat{X}_i = Z_i\hat{\Gamma}$ , and  $\hat{U}_i = Y_i - X_i\hat{\Theta}$ , the covariance matrix  $V$  can be consistently estimated by

$$\hat{V} = [\hat{X}'\hat{X}]^{-1}\hat{\Gamma}'\hat{\Lambda}\hat{\Gamma}[\hat{X}'\hat{X}]^{-1} \xrightarrow{p} V,$$

where

$$\hat{\Lambda} = \frac{1}{N} \sum_{i=1}^N (Z_i'\hat{U}_i\hat{U}_i'Z_i) \xrightarrow{p} \Lambda.$$

Due to the ‘diagonal’ structure of  $X_i$  and  $Z_i$ , it is possible to simplify the expressions for  $\hat{V}$ ,  $\hat{A}$ , and  $\hat{\Lambda}$  as provided in equations (10), (11) and (12) in the text.

Standard application of the delta-method implies that the variance of  $T(\hat{\Theta})$  can be estimated by  $\hat{G}\hat{V}\hat{G}'$ , where  $\hat{G}$  is the jacobian vector for  $T(\hat{\Theta})$  as defined in the text. With this, it is clear that

$$W_N = N\hat{T}'[\hat{G}\hat{V}\hat{G}']^{-1}\hat{T} \xrightarrow{d} \chi^2(1),$$

which can be more simply written as equation (13).

QED

## Ordered Choice Model

Assume schooling is determined by the ordered choice model defined by equations (14) and (15) in the paper. Then, the sign of the asymptotic bias for OLS estimation of any  $\beta_j$  in equation (2) depends on the sign of  $E(\varepsilon_i D_{ij}) = E(\varepsilon_i | s_i \geq j) = E(E[\varepsilon_i | v_i, z_i, x_i, v_i \geq j - \mu(z_i, x_i)])$ .

For illustrative purposes, consider the case in which the bias is non-negative. Clearly, if  $E(\varepsilon_i | z_i, x_i) = 0$  and  $\frac{\partial E(\varepsilon_i | v_i, z_i, x_i)}{\partial v} \geq 0$ , then  $E[\varepsilon_i | v_i, z_i, x_i, v_i \geq j - \mu(z_i, x_i)] \geq 0$  for any  $j$ . Furthermore, if  $(\varepsilon_i, v_i) \perp\!\!\!\perp (z_i, x_i)$ , then  $E(\varepsilon_i | v_i) = E(\varepsilon_i | v_i, z_i, x_i)$ . Altogether, if  $(\varepsilon_i, v_i) \perp\!\!\!\perp (z_i, x_i)$  and  $\frac{\partial E(\varepsilon_i | v_i)}{\partial v} \geq 0$ , then  $E(\varepsilon_i D_{ij}) \geq 0$  for all  $j$ . This implies that the asymptotic bias from OLS estimation will be non-negative for all  $\beta_j$  parameters.



## Online Appendix B: Monte Carlo Analysis

This is an online appendix for Lochner and Moretti (2012) that provides a Monte Carlo analysis of the test developed in that paper. We first show how varying the degree of non-linearity between  $y_i$  and  $s_i$  can induce differences between the OLS and the IV estimates, even in the absence of endogeneity bias. We further show that our exogeneity test accounts for this while the more standard Durbin-Wu-Hausman test (applied to IV and OLS estimates of  $\beta^L$  in equation 1) does not. Second, we discuss inference in a two stage procedure in which our exogeneity test is applied in a first stage and, if exogeneity cannot be rejected, OLS estimates from equation (2) are used to conduct inference in a second stage.

### General Setup

As a setting, we consider a modified version of Card's (1995) model of investment in human capital. An individual chooses schooling  $s_i$  to maximize  $V_i(s_i) = \log[y_i(s_i)] - C_i(s_i)$  where  $y_i(s_i)$  is earnings and  $C_i(s_i)$  is cost of schooling. We assume that the relation between log earnings and schooling is non-linear by allowing for jumps of size  $\kappa$  in earnings at an arbitrary schooling level  $J$

$$\log[y_i(s_i)] = a + bs_i + \kappa \mathbb{1}(s_i \geq J) + \varepsilon_i,$$

where  $\kappa$  measures the degree of non-linearity between log earnings and schooling and  $\mathbb{1}(\cdot)$  is an indicator function. A larger  $\kappa$  implies a stronger non-linearity with a greater discrepancy between the effect of moving from grade  $J - 1$  to  $J$  and the effects of all other grade transitions. The individual-specific cost of schooling is assumed to be

$$C_i(s_i) = c + r_i s_i + \frac{k_2}{2} s_i^2 + \kappa \mathbb{1}(s_i \geq J),$$

where the inclusion of  $\kappa$  here means that the non-linearity between schooling and earnings does not affect schooling choices. This allows us to focus on the extent to which variability in grade-specific effects influences IV and OLS estimators and our exogeneity test given a fixed set of OLS and IV weights.<sup>28</sup> Finally, we assume that the instrumental variable  $z_i$  shifts the cost of schooling

$$r_i = dz_i + \eta_i,$$

---

<sup>28</sup>Including  $\kappa$  in both the log earnings and cost functions is equivalent to assuming that individuals do not consider any non-linearities when making their schooling decisions. Although the IV and OLS weights will not vary with  $\kappa$  in our analysis, they will vary with the extent of 'endogeneity' as defined by  $\rho$  below.

and that individuals can only choose  $s \in \{0, 1, 2, \dots, S\}$ . The parameter  $d$  relative to the variance of  $\eta_i$  effectively controls the strength of the instrument.<sup>29</sup>

If we let

$$\begin{bmatrix} \varepsilon_i \\ \eta_i \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\varepsilon^2 & \sigma_{\varepsilon\eta} \\ \sigma_{\varepsilon\eta} & \sigma_\eta^2 \end{bmatrix} \right)$$

we can control the amount of ‘endogeneity’ by varying  $\rho = \frac{\sigma_{\varepsilon\eta}}{\sigma_\varepsilon\sigma_\eta}$ . Note that we naturally have monotonicity in the effects of  $z_i$  on schooling.

For each independent observation, we randomly draw a binary instrument  $z_i \in \{0, 1\}$  independently from bivariate normally distributed errors  $(\varepsilon_i, \eta_i)$ . Given the value of the parameters, the level of schooling is determined, and realized values of  $\log(y_i)$  are constructed.

In all of our simulations, we randomly draw  $z_i$  with probability  $Pr(z_i = 1) = 0.5$  and set other parameters of the model as follows:  $a = 1.5$ ;  $b = .04$ ;  $c = 0$ ;  $k_2 = .003$ ;  $\sigma_\varepsilon^2 = .25$ ;  $\sigma_\eta^2 = .00005$ ;  $J = 12$ ; and  $S = 20$ . With  $d = 0.01$  and  $\rho = \kappa = 0$ , this set of parameters generates a reasonable earnings and schooling distribution relative to recent Census years.

## Comparing IV and OLS Estimators and our Wald Test vs. the Durbin-Wu-Hausman Test

We begin by demonstrating the impacts of variable grade-specific effects on OLS and IV estimators when linearity in  $s_i$  is incorrectly assumed. We also consider the properties of our new exogeneity test along with those of the more standard Durbin-Wu-Hausman test (applied to a linear-in- $s_i$  specification) with varying degrees of endogeneity and non-linearity between  $y_i$  and  $s_i$ .

Here, we set the sample size  $N = 1,000$  for each Monte Carlo simulation and use 10,000 simulated samples. For each model, defined by a combination of endogeneity ( $\rho$ ) and non-linearity or ‘jump size’ ( $\kappa$ ), we compute point estimates and standard errors for the OLS and IV estimators. Specifically, we estimate the model for all possible combinations of

$$\rho \in \{0, 0.05, 0.1, 0.15, 0.2\} \quad \text{and} \quad \kappa \in \{0, 0.1, 0.5, 1\},$$

setting  $d = 0.01$  and other parameters as described above.

Estimation results for these Monte Carlo exercises are shown in Table B1. For each model, we report the average point estimates and their standard deviation from the simulation samples

---

<sup>29</sup>We have made two changes to Card’s original model. First, Card (1995) allows for variation in  $b_i$ , while we set  $b_i = b$  for all  $i$ . Second, Card (1995) assumes log earnings are quadratic in schooling. In our case, log earnings are non-linear in schooling, but non-linearity is parameterized with discrete jumps. This allows for an easier interpretation of the Monte Carlo estimates.

for OLS and 2SLS estimators from the mis-specified linear-in-schooling model, as well as the re-weighted OLS estimates (from the more general model that allows for grade-specific schooling effects) using the estimated 2SLS weights,  $\sum_{j=1}^S \hat{\omega}_j \hat{\beta}_j$ . We next report the fraction of cases where we reject the null hypothesis of equality between the IV and re-weighted OLS estimates using the general Wald test given in Theorem 1. Finally, we report the fraction of cases for which we reject the null of exogeneity based on the linear-in-schooling specification using the Durbin-Wu-Hausman (DWH) test. We use the critical value of 3.841 associated with a 0.05 significance level for both tests. Using our test, we should reject the null hypothesis that the re-weighted OLS estimates equal the IV estimates 5% of the time when schooling is exogenous (i.e.  $\rho = 0$ ) regardless of the amount of non-linearity (i.e. for any value of  $\kappa$ ). We only expect to reject the null 5% of the time using the DWH test when  $\rho = \kappa = 0$ .

The first row in Table 1 indicates that when the true relation between earnings and schooling is linear and there is no endogeneity, both OLS and 2SLS estimated returns to schooling are 4%. The next few rows (all with  $\rho = 0$ ) indicate that the difference between IV and OLS grows when we introduce increasingly large non-linearities in the relation between earnings and schooling; however, re-weighting the OLS estimates accounts for all of the difference. Thus, our test rejects the null only about 5% of the time as it should. The standard DWH test rejects the null about 5% of the time for small or no non-linearity (i.e.  $\kappa$  values of 0 and 0.1), but rejects much more frequently as non-linearity becomes a more important feature of the data. For  $\kappa = 1$ , the DWH test rejects over 40% of the time despite the fact that schooling is exogenous.

The remaining panels repeat the same exercise progressively increasing the amount of endogeneity. While re-weighting the OLS estimates using the IV weights often accounts for much of the difference between the linear-in-schooling OLS and IV estimates, it does not generally account for all of the difference. The greater the endogeneity (i.e. the higher is  $\rho$ ), the more the difference remains unexplained. Most importantly, our test begins to reject equality of the re-weighted OLS and IV estimates (i.e. exogeneity of schooling) at noticeably higher rates for even minor deviations from exogeneity (e.g.  $\rho = 0.05$ ). For  $\rho \geq 0.2$ , our test almost always rejects exogeneity. Consider, for example, the set of results with  $\rho = 0.2$ . In the linear-in-schooling model ( $\kappa = 0$ ), the IV estimate is basically 0.04; however, the OLS estimate is much lower at 0.012 due to the endogeneity of schooling. Re-weighting has a negligible effect on the OLS estimate, and we almost always reject the null of exogeneity. When  $\kappa = 1$ , the linear-in-schooling OLS estimate is still smaller than the IV

estimate, but the re-weighted OLS estimate is much closer. The different weights and non-linearity explain roughly one-third of the difference between linear-in-schooling OLS and IV estimates in this case. Still, our test correctly rejects the null in almost all cases. In general, the share of rejections is independent of the amount of non-linearity between log earnings and schooling, but it is sharply increasing in the degree of endogeneity. It is also important to note that when the true underlying model is linear in schooling, our more general test has very similar power to the DWH test: rejection rates for our test are typically less than 2% lower than for the DWH when  $\kappa = 0$ .

## Inference in a Two Stage Approach

We next consider the properties of inference in a two stage approach to estimation. In stage 1, our exogeneity test is applied. If exogeneity cannot be rejected, researchers may conduct inference in a second stage using OLS estimates of the  $\beta_j$  parameters based on equation (2) in the paper. In an analogous context with a linear relationship between the outcome variable and endogenous regressor, Wong (1997) and Guggenberger (2010) demonstrate potential problems with inference when using this type of two stage approach (they consider the standard Hausman test in the first stage).

Because inference problems arise when the instruments are insufficiently strong, we explore this two stage approach for different values of instrument ‘strength’ by letting the parameter  $d$  vary from 0.002 to 0.01. We consider values of  $\rho$  ranging from 0 to 0.1, values of  $\kappa$  equal to 0 and 0.1, and sample sizes of  $N = 1,000$  as above and  $N = 10,000$ .

Tables B2 ( $N = 1,000$ ) and B3 ( $N = 10,000$ ) report average first stage F-statistics (i.e. for the regression of  $s_i$  on  $z_i$ ) as a measure of the strength of the instruments for different values of  $d$ . We also report average point estimates and their standard deviations for OLS and 2SLS estimators from the mis-specified linear-in-schooling model, as well as the re-weighted OLS estimates from the more general model that allows for varying grade-specific effects. The next column reports the fraction of cases in which we reject the null hypothesis of equality between the IV and re-weighted OLS estimates using the exogeneity test given in Theorem 1 (.05 significance level). The last two columns consider inference about OLS estimates of the  $\beta_j$  parameters in cases where we fail to reject exogeneity. We report the fraction of cases that we reject linearity or equal grade-specific effects (i.e.  $\beta_1 = \beta_2 = \dots = \beta_S$ ) and the fraction of cases for which we reject that the OLS estimates are consistent (i.e.  $\text{plim } \hat{B} = B$ ) using a .05 significance level in both cases. When  $\rho = \kappa = 0$ ,

both tests should reject 5% of the time. When  $\rho = 0$  and  $\kappa > 0$ , the second test should have a 5% rejection rate, while the first should have a higher rejection rate.

We highlight a few main lessons from Tables B2 and B3. First, our second stage tests for non-linearity (i.e. variation in grade-specific effects) and consistency of the OLS estimators for all  $\beta_j$  parameters reject 5% of the time when they should (i.e. when  $\rho = \kappa = 0$  and  $\rho = 0, \kappa = 0.1$ ). As expected, rejection rates for the test for linearity (i.e. uniform grade-specific effects) are notably higher (15-16%) when  $\kappa = 0.1$ .

Second, when the first stage F-statistics are low, our exogeneity test has very little power to identify modest levels of endogeneity. For example, when  $N = 1,000$  and  $d = .003$ , we fail to reject exogeneity most of the time even for values of  $\rho$  as high as 0.1 yielding a sizeable difference between the re-weighted OLS estimate and the IV estimate. In these cases, the researcher is likely to conclude from our test that schooling is exogenous, which then suggests that the OLS estimates of the  $\beta_j$  parameters from equation (2) are consistent. However, as Wong (1997) and Guggenberger (2010) point out (when using the standard Hausman test in the first stage), inference using the OLS estimates is likely to have poor size properties. Even a negligible amount of endogeneity (too small to cause our test to reject exogeneity) is likely to lead one to reject  $plim\hat{B} = B$  when using the OLS estimates. This problem is reflected in the high rejection rates for the second stage test of  $plim\hat{B} = B$  even for cases with very low rejection rates with our exogeneity test (e.g.  $N = 1,000$ ,  $d = 0.003$  and  $\rho = 0.1$ ). These results suggest caution in using OLS estimates for inference when the instruments are not sufficiently strong even if our test fails to reject exogeneity. Researchers should be particularly cautious when the IV and re-weighted OLS estimates are quite different but the IV estimates are very imprecise. It is reassuring that these problems disappear as the instrument becomes stronger as observed for larger values of  $d$  and first stage F-statistics. It is also the case that our rejection rates for the test of linearity are largely independent of  $\rho$  (and  $d$ ) suggesting that endogeneity of  $s$  appears to bias all  $\beta_j$  estimates by a similar amount in our context.

Table B1: Monte Carlo Simulations for Card Model  
Our Exogeneity Test vs. Durbin-Wu-Hausman Test

$\rho$	$\kappa$	$\hat{\beta}_{OLS}^L$	$\hat{\beta}_{2SLS}^L$	$\sum_j \hat{\omega}_j \hat{\beta}_j$	General Wald Test (fraction reject, .05 sig. level)	Durbin-Wu-Hausman Test (fraction reject, .05 sig. level)
0	0	0.0399 (0.0054)	0.0399 (0.0096)	0.0399 (0.0056)	0.050	0.049
0	0.1	0.0540 (0.0054)	0.0557 (0.0095)	0.0556 (0.0056)	0.051	0.054
0	0.5	0.1099 (0.0057)	0.1180 (0.0100)	0.1179 (0.0063)	0.056	0.172
0	1	0.1801 (0.0063)	0.1961 (0.0111)	0.1960 (0.0080)	0.047	0.434
0.05	0	0.0330 (0.0055)	0.0399 (0.0096)	0.0332 (0.0057)	0.139	0.144
0.05	0.1	0.0470 (0.0054)	0.0556 (0.0095)	0.0489 (0.0056)	0.139	0.206
0.05	0.5	0.1030 (0.0056)	0.1179 (0.0099)	0.1112 (0.0063)	0.139	0.472
0.05	1	0.1729 (0.0063)	0.1960 (0.0111)	0.1892 (0.0080)	0.146	0.721
0.1	0	0.0260 (0.0054)	0.0402 (0.0094)	0.0265 (0.0056)	0.428	0.444
0.1	0.1	0.0399 (0.0055)	0.0557 (0.0095)	0.0420 (0.0057)	0.430	0.527
0.1	0.5	0.0959 (0.0057)	0.1179 (0.0100)	0.1044 (0.0064)	0.424	0.784
0.1	1	0.1659 (0.0063)	0.1960 (0.0112)	0.1823 (0.0081)	0.429	0.911
0.15	0	0.0191 (0.0055)	0.0402 (0.0096)	0.0197 (0.0057)	0.762	0.783
0.15	0.1	0.0331 (0.0055)	0.0558 (0.0096)	0.0354 (0.0057)	0.760	0.836
0.15	0.5	0.0890 (0.0057)	0.1180 (0.0100)	0.0977 (0.0064)	0.761	0.954
0.15	1	0.1590 (0.0063)	0.1962 (0.0113)	0.1757 (0.0080)	0.763	0.984
0.2	0	0.0119 (0.0053)	0.0401 (0.0096)	0.0129 (0.0055)	0.949	0.956
0.2	0.1	0.0261 (0.0054)	0.0558 (0.0095)	0.0286 (0.0056)	0.951	0.971
0.2	0.5	0.0820 (0.0057)	0.1182 (0.0101)	0.0910 (0.0064)	0.950	0.995
0.2	1	0.1519 (0.0063)	0.1958 (0.0112)	0.1688 (0.0081)	0.949	0.999

Notes: Columns for  $\hat{\beta}_{OLS}^L$ ,  $\hat{\beta}_{2SLS}^L$ , and  $\sum_j \hat{\omega}_j \hat{\beta}_j$  report the average (standard deviation) of estimates. Columns for General Wald Test and Durbin-Wu-Hausman (DWH) Test report the fraction of cases rejecting the null hypothesis of exogeneity using a .05 significance level. General Wald Test compares  $\hat{\beta}_{2SLS}^L$  and  $\sum_j \hat{\omega}_j \hat{\beta}_j$  as described in Theorem 1, while the DWH Test compares  $\hat{\beta}_{2SLS}^L$  and  $\hat{\beta}_{OLS}^L$ . A sample size of  $N = 1,000$  is used for all simulation samples.

Table B2: Monte Carlo Simulations for Card Model  
 Inference Using OLS Estimates of Nonlinear Model, N=1,000

d	$\rho$	$\kappa$	First-Stage F-Statistic	$\hat{\beta}_{OLS}^L$	$\hat{\beta}_{2SLS}^L$	$\sum_j \hat{\omega}_j \hat{\beta}_j$	Exogeneity Test (fraction reject, .05 sig. level)	Testing $\beta_j = \beta$ (fraction reject, .05 sig. level)	Testing $\hat{B} \stackrel{p}{\rightarrow} B$ (fraction reject, .05 sig. level)
0.002	0	0	20.75	0.0401 (0.0066)	0.0398 (0.0520)	0.0402 (0.0090)	0.025	0.051	0.052
		0.1	20.50	0.0538 (0.0066)	0.0538 (0.0529)	0.0538 (0.0093)	0.027	0.155	0.049
	0.01	0	20.70	0.0379 (0.0066)	0.0399 (0.0520)	0.0380 (0.0089)	0.027	0.052	0.051
		0.1	20.63	0.0518 (0.0067)	0.0546 (0.0530)	0.0519 (0.0095)	0.024	0.158	0.055
	0.03	0	20.69	0.0338 (0.0066)	0.0402 (0.0520)	0.0339 (0.0089)	0.027	0.051	0.071
		0.1	20.60	0.0477 (0.0066)	0.0543 (0.0524)	0.0478 (0.0093)	0.029	0.157	0.067
	0.05	0	20.86	0.0297 (0.0066)	0.0402 (0.0524)	0.0297 (0.0088)	0.030	0.052	0.111
		0.1	20.53	0.0435 (0.0066)	0.0544 (0.0525)	0.0435 (0.0095)	0.026	0.162	0.110
	0.1	0	20.77	0.0195 (0.0066)	0.0414 (0.0523)	0.0198 (0.0087)	0.040	0.049	0.364
		0.1	20.58	0.0332 (0.0066)	0.0550 (0.0534)	0.0335 (0.0094)	0.039	0.162	0.366
0.003	0	0	45.35	0.0400 (0.0066)	0.0404 (0.0334)	0.0400 (0.0075)	0.044	0.054	0.054
		0.1	45.47	0.0543 (0.0066)	0.0545 (0.0330)	0.0544 (0.0077)	0.039	0.166	0.049
	0.01	0	45.32	0.0380 (0.0065)	0.0402 (0.0327)	0.0381 (0.0074)	0.037	0.050	0.051
		0.1	45.18	0.0523 (0.0065)	0.0545 (0.0330)	0.0524 (0.0077)	0.040	0.161	0.051
	0.03	0	45.26	0.0339 (0.0066)	0.0399 (0.0326)	0.0340 (0.0075)	0.041	0.053	0.071
		0.1	45.42	0.0483 (0.0066)	0.0544 (0.0328)	0.0485 (0.0078)	0.044	0.157	0.072
	0.05	0	45.18	0.0298 (0.0065)	0.0405 (0.0335)	0.0299 (0.0075)	0.056	0.050	0.112
		0.1	45.44	0.0442 (0.0066)	0.0541 (0.0332)	0.0444 (0.0078)	0.049	0.163	0.108
	0.1	0	45.45	0.0200 (0.0066)	0.0401 (0.0332)	0.0202 (0.0075)	0.082	0.047	0.360
		0.1	45.46	0.0342 (0.0065)	0.0542 (0.0329)	0.0345 (0.0076)	0.078	0.166	0.361

Table B2 (continued): Monte Carlo Simulations for Card Model  
 Inference Using OLS Estimates of Nonlinear Model, N=1,000

d	$\rho$	$\kappa$	First-Stage F-Statistic	$\hat{\beta}_{OLS}^L$	$\hat{\beta}_{2SLS}^L$	$\sum_j \hat{\omega}_j \hat{\beta}_j$	Exogeneity Test (fraction reject, .05 sig. level)	Testing $\beta_j = \beta$ (fraction reject, .05 sig. level)	Testing $\hat{B} \xrightarrow{p} B$ (fraction reject, .05 sig. level)
0.005	0	0	124.42	0.0400	0.0400	0.0400	0.047	0.046	0.049
				(0.0063)	(0.0191)	(0.0066)			
	0	0.1	124.68	0.0548	0.0550	0.0552	0.046	0.162	0.052
				(0.0063)	(0.0193)	(0.0067)			
	0.01	0	124.32	0.0382	0.0402	0.0382	0.045	0.046	0.047
				(0.0063)	(0.0190)	(0.0066)			
	0.01	0.1	124.13	0.0529	0.0551	0.0534	0.048	0.153	0.053
				(0.0063)	(0.0194)	(0.0067)			
	0.03	0	124.52	0.0343	0.0401	0.0344	0.060	0.049	0.067
				(0.0063)	(0.0193)	(0.0066)			
0.03	0.1	124.31	0.0492	0.0554	0.0497	0.056	0.161	0.070	
			(0.0064)	(0.0193)	(0.0068)				
0.05	0	124.25	0.0306	0.0401	0.0307	0.081	0.055	0.101	
			(0.0063)	(0.0194)	(0.0066)				
0.05	0.1	124.55	0.0455	0.0554	0.0460	0.077	0.168	0.103	
			(0.0063)	(0.0193)	(0.0066)				
0.1	0	124.36	0.0214	0.0402	0.0216	0.176	0.050	0.324	
			(0.0062)	(0.0194)	(0.0065)				
0.1	0.1	124.53	0.0361	0.0553	0.0365	0.167	0.162	0.329	
			(0.0063)	(0.0192)	(0.0067)				
0.01	0	0	494.95	0.0400	0.0400	0.0401	0.052	0.053	0.049
				(0.0055)	(0.0097)	(0.0056)			
	0	0.1	494.93	0.0540	0.0557	0.0556	0.046	0.151	0.050
				(0.0054)	(0.0095)	(0.0056)			
	0.01	0	494.66	0.0385	0.0399	0.0385	0.053	0.050	0.052
				(0.0055)	(0.0095)	(0.0056)			
	0.01	0.1	494.76	0.0526	0.0557	0.0543	0.054	0.155	0.053
				(0.0055)	(0.0096)	(0.0057)			
	0.03	0	494.75	0.0359	0.0401	0.0360	0.082	0.048	0.056
				(0.0055)	(0.0094)	(0.0056)			
0.03	0.1	494.09	0.0498	0.0557	0.0516	0.083	0.160	0.057	
			(0.0055)	(0.0096)	(0.0057)				
0.05	0	494.53	0.0330	0.0399	0.0332	0.140	0.053	0.076	
			(0.0055)	(0.0095)	(0.0057)				
0.05	0.1	494.49	0.0470	0.0557	0.0488	0.148	0.168	0.077	
			(0.0054)	(0.0096)	(0.0056)				
0.1	0	495.21	0.0260	0.0401	0.0265	0.422	0.053	0.179	
			(0.0055)	(0.0096)	(0.0057)				
0.1	0.1	494.55	0.0399	0.0555	0.0420	0.421	0.184	0.180	
			(0.0055)	(0.0096)	(0.0057)				

Notes: First-Stage F-statistic reports the average F-statistic from all simulation samples from first-stage regressions of  $s_i$  on  $z_i$ . Columns for  $\hat{\beta}_{OLS}^L$ ,  $\hat{\beta}_{2SLS}^L$ , and  $\sum_j \hat{\omega}_j \hat{\beta}_j$  report the average (standard deviation) of estimates. Exogeneity Test compares  $\hat{\beta}_{2SLS}^L$  and  $\sum_j \hat{\omega}_j \hat{\beta}_j$  as described in Theorem 1. The final two columns report tests for equal grade-specific effects (i.e.  $\beta_j = \beta$  for all  $j$ ) and consistency of OLS estimates for all  $\beta_j$  in equation (2) using only subsamples for which we fail to reject exogeneity. The fraction of cases rejecting the null for all tests is reported using a 0.05 significance level. A sample size of  $N = 1,000$  is used for all simulated samples.



Table B3: Monte Carlo Simulations for Card Model  
 Inference Using OLS Estimates of Nonlinear Model, N=10,000

d	$\rho$	$\kappa$	First-Stage F-Statistic	$\hat{\beta}_{OLS}^L$	$\hat{\beta}_{2SLS}^L$	$\sum_j \hat{\omega}_j \hat{\beta}_j$	Exogeneity Test (fraction reject, .05 sig. level)	Testing $\beta_j = \beta$ (fraction reject, .05 sig. level)	Testing $\hat{B} \stackrel{p}{\rightarrow} B$ (fraction reject, .05 sig. level)
0.002	0	0	197.96	0.0400 (0.0021)	0.0397 (0.0152)	0.0400 (0.0022)	0.051	0.046	0.047
		0.1	197.97	0.0538 (0.0021)	0.0537 (0.0151)	0.0538 (0.0023)	0.049	0.985	0.049
	0.01	0	198.61	0.0380 (0.0021)	0.0399 (0.0150)	0.0380 (0.0021)	0.046	0.050	0.070
		0.1	198.19	0.0517 (0.0021)	0.0535 (0.0150)	0.0517 (0.0023)	0.049	0.986	0.068
	0.03	0	197.88	0.0338 (0.0021)	0.0401 (0.0152)	0.0338 (0.0022)	0.065	0.048	0.334
		0.1	197.77	0.0476 (0.0021)	0.0539 (0.0152)	0.0477 (0.0023)	0.067	0.984	0.335
	0.05	0	198.51	0.0297 (0.0021)	0.0402 (0.0149)	0.0297 (0.0022)	0.101	0.048	0.814
		0.1	197.92	0.0435 (0.0021)	0.0541 (0.0149)	0.0435 (0.0023)	0.102	0.981	0.814
	0.1	0	197.95	0.0195 (0.0021)	0.0401 (0.0151)	0.0195 (0.0022)	0.277	0.053	1.000
		0.1	197.56	0.0332 (0.0021)	0.0540 (0.0152)	0.0333 (0.0023)	0.274	0.984	1.000
0.003	0	0	443.68	0.0400 (0.0021)	0.0401 (0.0101)	0.0400 (0.0021)	0.047	0.047	0.053
		0.1	444.48	0.0542 (0.0021)	0.0544 (0.0100)	0.0543 (0.0022)	0.048	0.984	0.053
	0.01	0	444.01	0.0380 (0.0021)	0.0398 (0.0101)	0.0380 (0.0021)	0.053	0.051	0.068
		0.1	444.92	0.0523 (0.0021)	0.0543 (0.0103)	0.0524 (0.0022)	0.060	0.985	0.069
	0.03	0	444.95	0.0340 (0.0021)	0.0399 (0.0100)	0.0340 (0.0021)	0.091	0.051	0.318
		0.1	444.17	0.0482 (0.0021)	0.0544 (0.0102)	0.0484 (0.0022)	0.095	0.984	0.324
	0.05	0	444.79	0.0300 (0.0021)	0.0400 (0.0101)	0.0300 (0.0021)	0.171	0.052	0.802
		0.1	444.10	0.0442 (0.0021)	0.0545 (0.0101)	0.0444 (0.0022)	0.176	0.985	0.805
	0.1	0	444.62	0.0200 (0.0021)	0.0400 (0.0100)	0.0200 (0.0021)	0.529	0.052	1.000
		0.1	444.86	0.0342 (0.0020)	0.0545 (0.0101)	0.0344 (0.0022)	0.533	0.988	1.000

Table B3 (continued): Monte Carlo Simulations for Card Model  
Inference Using OLS Estimates of Nonlinear Model, N=10,000

d	$\rho$	$\kappa$	First-Stage F-Statistic	$\hat{\beta}_{OLS}^L$	$\hat{\beta}_{2SLS}^L$	$\sum_j \hat{\omega}_j \hat{\beta}_j$	Exogeneity Test (fraction reject, .05 sig. level)	Testing $\beta_j = \beta$ (fraction reject, .05 sig. level)	Testing $\hat{B} \xrightarrow{p} B$ (fraction reject, .05 sig. level)
0.005	0	0	1232.11	0.0400 (0.0020)	0.0401 (0.0060)	0.0400 (0.0020)	0.049	0.049	0.052
			1233.34	0.0547 (0.0020)	0.0552 (0.0060)	0.0552 (0.0020)	0.048	0.985	0.054
	0.01	0	1232.69	0.0381 (0.0020)	0.0399 (0.0060)	0.0381 (0.0020)	0.061	0.051	0.066
			1234.08	0.0529 (0.0020)	0.0552 (0.0061)	0.0534 (0.0020)	0.067	0.985	0.066
	0.03	0	1234.14	0.0344 (0.0020)	0.0399 (0.0060)	0.0344 (0.0020)	0.161	0.047	0.280
			1232.78	0.0492 (0.0020)	0.0553 (0.0061)	0.0497 (0.0020)	0.169	0.987	0.295
	0.05	0	1233.70	0.0307 (0.0020)	0.0400 (0.0060)	0.0307 (0.0020)	0.372	0.052	0.750
			1232.57	0.0455 (0.0020)	0.0553 (0.0060)	0.0459 (0.0020)	0.380	0.986	0.753
	0.1	0	1233.38	0.0214 (0.0020)	0.0400 (0.0060)	0.0214 (0.0020)	0.912	0.050	1.000
			1234.14	0.0362 (0.0020)	0.0553 (0.0060)	0.0367 (0.0020)	0.909	0.984	1.000
0.01	0	0	4934.29	0.0400 (0.0017)	0.0400 (0.0030)	0.0400 (0.0018)	0.050	0.055	0.047
			4931.60	0.0540 (0.0017)	0.0557 (0.0030)	0.0556 (0.0018)	0.050	0.978	0.049
	0.01	0	4934.93	0.0386 (0.0017)	0.0400 (0.0030)	0.0386 (0.0018)	0.089	0.054	0.061
			4934.07	0.0526 (0.0017)	0.0556 (0.0030)	0.0542 (0.0018)	0.086	0.982	0.059
	0.03	0	4932.66	0.0358 (0.0017)	0.0400 (0.0030)	0.0359 (0.0018)	0.395	0.052	0.176
			4933.39	0.0498 (0.0017)	0.0556 (0.0030)	0.0515 (0.0018)	0.395	0.985	0.167
	0.05	0	4931.62	0.0330 (0.0017)	0.0401 (0.0030)	0.0332 (0.0018)	0.808	0.053	0.416
			4931.88	0.0470 (0.0017)	0.0556 (0.0030)	0.0488 (0.0018)	0.807	0.987	0.439
	0.1	0	4933.61	0.0260 (0.0017)	0.0401 (0.0030)	0.0264 (0.0017)	1.000	-	-
			4935.82	0.0400 (0.0017)	0.0556 (0.0031)	0.0419 (0.0018)	1.000	1.000	1.000

Notes: First-Stage F-statistic reports the average F-statistic from all simulation samples from first-stage regressions of  $s_i$  on  $z_i$ . Columns for  $\hat{\beta}_{OLS}^L$ ,  $\hat{\beta}_{2SLS}^L$ , and  $\sum_j \hat{\omega}_j \hat{\beta}_j$  report the average (standard deviation) of estimates. Exogeneity Test compares  $\hat{\beta}_{2SLS}^L$  and  $\sum_j \hat{\omega}_j \hat{\beta}_j$  as described in Theorem 1. The final two columns report tests for equal grade-specific effects (i.e.  $\beta_j = \beta$  for all  $j$ ) and consistency of OLS estimates for all  $\beta_j$  in equation (2) using only subsamples for which we fail to reject exogeneity. The fraction of cases rejecting the null for all tests is reported using a 0.05 significance level. A sample size of  $N = 10,000$  is used for all simulated samples.

## Online Appendix C: OLS and Logit Grade-Specific Estimates and Standard Errors

This is an online appendix for Lochner and Moretti (2012) that reports estimates and corresponding standard errors for all grade-specific effects discussed in that paper.

Table C1 reports OLS estimates from equation (2) based on the specifications used in Figures 1-5 of the paper.

In Table C2, we explore logit specifications for the binary outcome variables in Lochner and Moretti (2004) and Currie and Moretti (2003). These specifications are analogous to those from Figures 1-4 of the paper and Table C1. The specifications in Currie and Moretti (2003) include too many fixed effects for standard software packages, so we consider conditional logit specifications where we condition on county-year fixed effects. Table C2 reports the estimated grade-specific effect (standard error) when all regressors (except the county-year fixed effects) are set at their sample average and the county-year effect is set to the value that yields an estimated sample proportion for low birth weight or pre-term birth that equals the observed sample proportion. That is, the table reports the grade-specific effects (standard errors) given by :  $\hat{\beta}_j \times \left[ \frac{1}{N} \sum_{i=1}^N \xi \left( \sum_{j=1}^S \bar{D}_{ij} \hat{\beta}_j + \bar{x}' \hat{\gamma} + \tilde{\gamma}_0 \right) \right]$ , where  $\bar{D}_{ij}$  and  $\bar{x}$  are sample averages,  $\tilde{\gamma}_0$  solves  $\frac{1}{N} \sum_{i=1}^N \Xi \left( \sum_{j=1}^S \bar{D}_{ij} \hat{\beta}_j + \bar{x}' \hat{\gamma} + \tilde{\gamma}_0 \right) = \frac{1}{N} \sum_{i=1}^N y_i$ , and  $\xi(\cdot)$  and  $\Xi(\cdot)$  are the logit pdf and cdf, respectively.

**Figure C1: Estimated Grade-Specific Schooling Effects and Standard Errors (OLS)**

Years of Schooling	<b>1. Lochner and Moretti (2004)</b>				<b>2. Currie &amp; Moretti (2003)</b>				<b>3. Acemoglu &amp; Angrist (2001)</b>	
	Effect of Schooling on Imprisonment				Effect of Maternal Education on Infant Health and Health Inputs				Private Returns to Schooling	
	White Males		Black Males		Low birth weight		Preterm birth		Annual Earnings	
1	-0.0004	(0.0012)	0.0017	(0.0039)	0.0359	(0.0354)	0.0363	(0.0415)	0.0155	(0.0332)
2	0.0045	(0.0013)	0.0062	(0.0041)	-0.0283	(0.0412)	-0.0167	(0.0483)	0.1809	(0.0342)
3	-0.0017	(0.0009)	-0.0018	(0.0030)	0.0302	(0.0362)	0.0331	(0.0424)	-0.0308	(0.0228)
4	-0.0011	(0.0007)	0.0015	(0.0026)	-0.0577	(0.0319)	-0.0232	(0.0373)	0.1178	(0.0179)
5	0.0013	(0.0006)	-0.0003	(0.0024)	0.0553	(0.0260)	0.0372	(0.0304)	0.0444	(0.0155)
6	-0.0002	(0.0005)	0.0000	(0.0021)	0.0063	(0.0239)	-0.0154	(0.0280)	0.1055	(0.0122)
7	0.0000	(0.0004)	0.0046	(0.0019)	0.0008	(0.0175)	-0.0068	(0.0205)	0.0850	(0.0089)
8	-0.0020	(0.0003)	0.0028	(0.0016)	0.0009	(0.0125)	0.0046	(0.0147)	0.0859	(0.0064)
9	0.0014	(0.0002)	0.0044	(0.0014)	-0.0090	(0.0067)	0.0021	(0.0079)	0.0904	(0.0052)
10	-0.0024	(0.0002)	-0.0065	(0.0013)	0.0109	(0.0053)	-0.0033	(0.0062)	0.0530	(0.0054)
11	-0.0010	(0.0002)	-0.0106	(0.0012)	-0.0164	(0.0042)	-0.0026	(0.0049)	0.0531	(0.0054)
12	-0.0046	(0.0002)	-0.0228	(0.0010)	-0.0233	(0.0028)	-0.0232	(0.0033)	0.1135	(0.0043)
13	-0.0014	(0.0002)	-0.0010	(0.0012)	-0.0083	(0.0011)	-0.0046	(0.0013)	0.0842	(0.0043)
14	-0.0004	(0.0002)	-0.0056	(0.0015)	-0.0036	(0.0012)	-0.0046	(0.0014)	0.0463	(0.0053)
15	-0.0015	(0.0002)	-0.0064	(0.0018)	-0.0007	(0.0015)	-0.0008	(0.0017)	0.0591	(0.0065)
16	0.0007	(0.0002)	-0.0009	(0.0020)	-0.0074	(0.0014)	-0.0073	(0.0016)	0.2150	(0.0062)
17	-0.0003	(0.0002)	0.0014	(0.0028)	-0.0033	(0.0009)	-0.0031	(0.0010)	-0.0028	(0.0040)
18	0.0009	(0.0002)	0.0043	(0.0031)						

Notes: Table reports estimated coefficients (standard errors) on all schooling dummies from equation (2) in the paper. Specifications are the same as those in Figures 1-5 of the paper.

**Figure C2: Estimated Grade-Specific Schooling Effects and Standard Errors (Logits)**

Years of Schooling	1. Lochner and Moretti (2004)				2. Currie & Moretti (2003)			
	Effect of Schooling on Imprisonment				Effect of Maternal Education on Infant Health and Health Inputs			
	White Males		Black Males		Low birth weight		Preterm birth	
1	0.0005	(0.0009)	0.0025	(0.0054)	0.0243	(0.0274)	0.0333	(0.0372)
2	0.0022	(0.0009)	0.0122	(0.0056)	-0.0170	(0.0333)	-0.0131	(0.0441)
3	-0.0002	(0.0006)	-0.0021	(0.0037)	0.0190	(0.0301)	0.0253	(0.0379)
4	-0.0005	(0.0005)	0.0020	(0.0032)	-0.0452	(0.0269)	-0.0190	(0.0313)
5	0.0005	(0.0004)	-0.0004	(0.0027)	0.0409	(0.0225)	0.0274	(0.0248)
6	-0.0003	(0.0003)	-0.0013	(0.0023)	0.0045	(0.0174)	-0.0080	(0.0222)
7	-0.0001	(0.0003)	0.0032	(0.0019)	0.0007	(0.0127)	-0.0056	(0.0174)
8	-0.0009	(0.0002)	0.0005	(0.0014)	0.0012	(0.0091)	0.0032	(0.0126)
9	-0.0003	(0.0002)	-0.0013	(0.0011)	-0.0038	(0.0050)	0.0015	(0.0066)
10	-0.0012	(0.0002)	-0.0050	(0.0010)	0.0069	(0.0040)	-0.0018	(0.0052)
11	-0.0008	(0.0002)	-0.0078	(0.0010)	-0.0091	(0.0032)	-0.0020	(0.0042)
12	-0.0036	(0.0001)	-0.0200	(0.0009)	-0.0172	(0.0023)	-0.0188	(0.0029)
13	-0.0012	(0.0002)	0.0001	(0.0012)	-0.0077	(0.0011)	-0.0045	(0.0013)
14	-0.0009	(0.0002)	-0.0083	(0.0017)	-0.0036	(0.0012)	-0.0047	(0.0014)
15	-0.0024	(0.0003)	-0.0099	(0.0025)	-0.0006	(0.0015)	-0.0008	(0.0017)
16	-0.0010	(0.0004)	-0.0099	(0.0033)	-0.0081	(0.0014)	-0.0077	(0.0016)
17	-0.0018	(0.0006)	-0.0039	(0.0056)	-0.0029	(0.0010)	-0.0027	(0.0011)
18	0.0016	(0.0006)	-0.0017	(0.0067)				

Notes: Results for Lochner and Moretti (2004) reflect average grade-specific effects based on logit specifications analogous to those reported in Figures 1-4 of the paper and in Table C1. Results for Currie and Moretti (2003) are based on conditional logits where county-year fixed effects are conditioned out. Table reports marginal effects given the average level of all other covariates and the value of the fixed effect that yields a predicted share of observations with low birth weight or pre-term birth that matches the sample proportions. Standard errors are in parentheses.