Estimating and Testing Non-Linear Models Using Instrumental Variables

by

Lance Lochner and Enrico Moretti

Working Paper # 2011-2 **June 2011**

CIBC Working Paper Series

Department of Economics
Social Science Centre
The University of Western Ontario
London, Ontario, N6A 5C2
Canada

This working paper is available as a downloadable pdf file on our website
http://economics.uwo.ca/centres/cibc/

Estimating and Testing Non-Linear Models Using Instrumental Variables

Lance Lochner
University of Western Ontario
NBER

Enrico Moretti*
University of California–Berkeley
NBER, CEPR and IZA

April 29, 2011

Abstract

In many empirical studies, researchers seek to estimate causal relationships using instrumental variables. When only one valid instrumental variable is available, researchers are limited to estimating linear models, even when the true model may be non-linear. In this case, ordinary least squares and instrumental variable estimators will identify different weighted averages of the underlying marginal causal effects even in the absence of endogeneity. As such, the traditional Hausman test for endogeneity is uninformative. We build on this insight to develop a new test for endogeneity that is robust to any form of non-linearity. Notably, our test works well *even when only a single valid instrument is available.* This has important practical applications, since it implies that researchers can estimate a completely unrestricted non-linear model by OLS, and then use our test to establish whether those OLS estimates are consistent. We re-visit a few recent empirical examples to show how the test can be used to shed new light on the role of non-linearity.

*We are very grateful to Matias Cattaneo and Javier Cano Urbina, who provided excellent research assistance and insightful substantive comments, as well as Martijn van Hasselt and Youngki Shin for their many comments and suggestions. We also thank Josh Angrist, David Card, Pedro Carneiro, Jim Heckman, Guido Imbens, and seminar participants at the 2008 UM/MSU/UWO Summer Labor Conference, UCSD, and Stanford.

1 Introduction

Many recent empirical papers seek to estimate causal relationships using instrumental variables or two-stage least squares estimators when concerns about causality arise. In many cases, only a single valid instrument is available, so researchers are often limited to estimating a *linear* relationship between the dependent and the potentially endogenous regressor.[1] Conclusions about consistency of the ordinary least squares estimator are then based on a comparison of OLS and 2SLS estimates. When a standard Hausman test (Hausman 1978) indicates that OLS estimates are sufficiently different from 2SLS estimates, endogeneity of the regressor is typically concluded to play an important confounding role in OLS. However, we demonstrate that when the true relationship is non-linear but the estimated model is linear, OLS and IV/2SLS estimate different weighted average effects and the standard Hausman test is uninformative about endogeneity. Based on this insight, we develop a new endogeneity specification test that is robust to general non-linear relationships and only requires a single (even binary) instrument.

While numerous empirical and econometric studies explore the implications of parameter heterogeneity for OLS and IV estimation, very few studies focus on the implications of non-linearity when the estimated model is assumed to be linear.[2] Angrist, Graddy, and Imbens (2000), Lochner and Moretti (2001), and Mogstad and Wiswall (2010) are notable exceptions. Yet, in many applications in economics, there is no particular reason to expect the true relationship to be linear. The empirical examples of Mogstad and Wiswall (2010) underscore this point. We discuss the consequences of this mis-specification for OLS and IV/2SLS estimators.

We begin by clarifying the interpretation of OLS and IV estimators under assumptions and specifications commonly employed in the empirical literature. We show that inappropriately assuming linearity will generally yield different OLS and IV/2SLS estimates even in the absence of endogeneity. The reason is that the OLS and 2SLS 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 two estimators. The weights have an intuitive interpretation, are functions of observable quantities, and can be estimated easily under very general assumptions.[3]

These insights motivate and guide our main contribution: a new specification test for endogene-

[1] More generally, the number of available instruments constrains the extent of non-linearity that can be estimated.

[2] Studies focusing on parameter heterogeneity include Imbens and Angrist (1994), Angrist and Imbens (1995), Yitzhaki (1996), Wooldridge (1997), Heckman and Vytlacil (1998), Card (1999), Heckman and Vytlacil (1999, 2005), Kling (2000), Heckman, Urzua and Vytlacil (2006), Moffitt (2009), and Carneiro, Heckman and Vytlacil (2010).

[3] See Heckman and Vytlacil (1999, 2005), Heckman, Urzua and Vytlacil (2006), Moffitt (2009), and Carneiro, Heckman and Vytlacil (2010) for estimation of marginal treatment effects and different average treatment effects under parameter heterogeneity.

ity in the presence of non-linearities. Because OLS and IV/2SLS applied to mis-specified linear models identify different weighted averages of marginal effects, the traditional Hausman test is uninformative about endogeneity of the regressor. It may reject equality of OLS and 2SLS estimates even when the regressor is exogenous, and it may fail to reject equality when the regressor is endogenous. We exploit the intuition underlying the failure of the standard Hausman test to develop a new test for whether OLS estimation of a general non-linear model produces consistent estimates of each unrestricted, level-specific marginal effect. This test can be thought of as a generalization of the standard Hausman test.[4]

Notably, our test works well *even when only a single valid instrument is available*. This is noteworthy, since in the presence of non-linearities, many parameters typically need to be estimated. Thus, one might expect to need at least as many instruments to test for endogeneity (as with a standard Hausman test).[5] We show that this is not necessary, since the general non-linear model need not be estimated via IV/2SLS to test for endogeneity.

The minimal requirements on instruments imply that our test has important practical implications for empirical researchers. Consider the common situation where only one valid instrument is available, but the true model may be non-linear. A researcher can use OLS to estimate a model that allows for a fully non-parametric relationship using a set of dummy variables for each level of the regressor. For example, the researcher might regress wages on a full set of 20 schooling dummies representing each year of potential schooling attainment. The researcher can then use our proposed test to establish whether the OLS estimates are consistent. Despite the fact that there are 20 OLS parameters of interest, the test only requires one instrument, which may be binary. Rather than using the instrument for direct estimation of the general causal relation of interest, the instrument is used here to determine whether OLS estimates of the non-linear model are consistent. Of course, in the case where they are not, our test does not help in estimating the true model. Thus, our test offers only a partial solution to the problem of non-linear models with few instruments.

To make things more concrete, consider a simple example where the true relationship between an outcome, y, and years of schooling, s, is non-linear. Let β_j represent the grade-specific effect of moving from $j-1$ to j years of schooling. For example, there are large empirical literatures focusing on the case where y measures wages, earnings, labor force participation, health, crime

[4]Note that 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.

[5]In theory, a single continuous instrument with broad support may enable identification. However, in practice, nonparametric or general nonlinear instrumental variable estimates obtained using few instruments are typically very imprecise.

or numerous other social and economic outcomes. It is common in these literatures to estimate this type of model by assuming a linear relationship between s and y, although in many of these cases there is evidence of substantial non-linearities.[6] The practical problem is that there are 20 potential schooling levels and β_j parameters to estimate, while researchers typically have very few valid instruments.

Our setting differs from the case most often discussed in the literature regarding heterogeneity in the regression parameter (e.g., Wooldridge 1997, Heckman and Vytlacil 1998, Card 1999), where everyone receives a constant marginal return to schooling regardless of their level of schooling (i.e. y_i is linear in s_i), but the constant marginal return is assumed to vary in the population. We focus on the opposite extreme, assuming a non-linear relationship between y_i and s_i that does not vary across individuals. In our analysis, the marginal return to schooling varies in the population, because the effects of schooling are non-linear and different individuals have different levels of schooling, but the marginal effect at each schooling level is assumed to be homogenous in the population.

We first show that in the absence of any endogeneity bias, the OLS estimate of the mis-specified linear model converges to a weighted average of the true grade-specific effects, β_j. IV and 2SLS estimates of the mis-specified linear model also converge to weighted averages of the true grade-specific effects, but the weights are different. The stronger the effect of instruments on a particular schooling transition, the greater the weight on the effect of that transition. Intuitively this means that the more people crossing the grade j barrier in response to a change in the instrument, the greater the weight placed on the marginal effect of finishing grade j, β_j. In general, different instruments yield estimates of different "weighted averages," even if the instruments are all valid. While OLS weights depend on the joint distribution of schooling and the controls, IV weights depend on the joint distribution of schooling, the controls and the instrument. As a consequence, IV and OLS estimates can be quite different even when schooling is exogenous. One appealing feature of our setting is that it is easy to empirically estimate the weights, and therefore, it is possible to directly compare the OLS and IV weights. One can obtain the grade-specific OLS weights by regressing indicators for whether schooling is above each grade on years of schooling. One can obtain the grade-specific IV weights by estimating the same set of models, instrumenting for the schooling indicators.

Since OLS and IV estimates can differ even when schooling is exogenous, an important practical issue in this context is how to appropriately test for endogeneity. To test whether all β_j parameters

[6] For example, in the classic case of returns to schooling—where y reflects log wages or earnings—Hungeford and Solon (1987), Jaeger and Page (1996), Park (1999), and Heckman, Lochner and Todd (2008) estimate significant non-linearities.

are consistent using a Hausman test, one would need at least 20 instruments in order to estimate the full model using IV or 2SLS. We show that this is not necessary. Instead, our proposed generalization of the Hausman test compares the 2SLS estimate for the linear specification with the weighted sum of the unrestricted, level-specific OLS estimates of β_j's, where the weights are the estimated IV or 2SLS weights.[7] The test statistic turns out to have an intuitive form and is easy to implement empirically. Rejection implies that OLS estimation of the general model is asymptotically biased (i.e. endogeneity bias is a problem).[8] As mentioned above, a researcher can estimate the grade-specific effects (i.e. β_j's) by OLS using a fully non-parametric model with dummies for each level of schooling, and then use our test to determine whether these estimates are consistent, even in the case where only a single instrument is available.

In the last part of the paper, we conduct a Monte Carlo study and revisit data from some recent empirical papers to illustrate the value of our approach. In the Monte Carlo simulation, we show how varying the degree of non-linearity can induce differences between the OLS and the 2SLS estimates, even in the absence of endogeneity bias. We base this analysis on the return to schooling model discussed in Card (1999). We then focus on three recent empirical papers in which estimated 2SLS effects differ from OLS effects. We find that in some cases the standard Hausman test would lead the researcher to incorrectly conclude that OLS estimates are consistent, while our test leads us to conclude the opposite. We also find that in some cases re-weighting the OLS β_j estimates by the 2SLS weights suggests that some of the discrepancy between the linear OLS and 2SLS estimators may be explained by non-linearity in the true relationship.

We are not the first to point out that estimates from a mis-specified linear model will yield weighted averages of each grade-specific effect. This point has been made by Angrist and Imbens (1995) and Heckman, Urzua, and Vytlacil (2006), who discuss weights from 2SLS in the presence of parameter heterogeneity. More recently, Mogstad and Wiswall (2010) also emphasize the importance of accounting for non-linearities in a number of empirical contexts. As discussed earlier, numerous studies discuss the weighting of OLS and/or 2SLS in the presence of parameter heterogeneity, showing that under some conditions 2SLS estimates a local average treatment effect (LATE), or the effect of a regressor on those individuals induced to change their behavior in response to a change in the value of the instrument. In addition, Heckman and Vytlacil (2005) emphasize that the interpretation of OLS and 2SLS estimators can be quite complicated in the presence of parameter heterogeneity. There is no single 'effect' of the regressor on the outcome,

[7] 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.

[8] Alternatively, it my also indicate the presence of individual parameter heterogeneity.

and different estimation strategies provide estimates of different 'parameters of interest' or different 'average effects'.[9]

Our paper complements the existing literature in two respects. First, unlike the existing literature, we propose a test for endogeneity. Second, relative to the existing literature, our models are a step closer to the models typically estimated by researchers in practice. Because both Angrist and Imbens (1995) and Heckman, Urzua, and Vytlacil (2006) focus attention on the role of parameter heterogeneity, their estimating equations differ from those commonly employed in empirical studies. Angrist and Imbens (1995) only consider regressors that are indicators that place observations into mutually exclusive categories, and they interact their instrument 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. By contrast, our model considers estimation under common assumptions about covariates and the way they enter estimation. In addition, our analysis is not centered on finding an 'economic interpretation' for the IV estimator, as in Angrist and Imbens (1995) or Heckman, Urzua, and Vytlacil (2006). Instead, we are primarily 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 linearity is incorrectly assumed.

The remainder of the paper is organized as follows. In Section 2 we compare the OLS and 2SLS estimators when the true model is non-linear, but a linear model is estimated. In Section 3 we develop a test of consistency of the OLS estimator. Section 4 presents the results from a simple Montecarlo study, while Section 5 focuses on three real world examples. Section 6 concludes.

2 Estimating Non-Linear Models Under Linearity Assumptions

In this section, we consider instrumental variable and OLS estimators when the estimated model is linear but the true data generating process need not be. Assume that an outcome, y_i, for person i is given by

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

where total years of schooling is represented by $s_i \in \{0, 1, 2, 3, ..., S\}$, $D_{ij} = 1[s_i \geq j]$ reflects a dummy variable equal to one if total years of schooling are at least j, and x_i is a $k \times 1$ vector of other exogenous covariates (including an intercept), and ε_i are iid error terms with $E(\varepsilon_i|x_i) = 0$.

[9] Heckman, Urzua, and Vytlacil (2006) also discuss 2SLS weights in general ordered and unordered multinomial choice models with parameter heterogeneity.

In this general model, the parameter β_j reflects the grade-specific effect of moving from $j-1$ to j years of schooling. The β_j effects are assumed to be identical for everyone in the population.[10]

Suppose that instead of estimating the general nonlinear model described above, a researcher estimates a mis-specified version that is linear in schooling:

$$y_i = s_i\beta^L + x_i'\gamma^L + \nu_i. \tag{2}$$

We are interested in estimates of β^L and how those estimates relate to the underlying β_j's. We assume a sample size N is available.

2.1 IV Estimation with a Single Instrument

We show that under standard assumptions – the instrument z_i is correlated with s_i after projecting on x_i and uncorrelated with ε_i – the IV estimator for β^L in equation (2) converges in probability to a "weighted average" of all grade-specific effects, β_j.

It is useful to decompose schooling in the population as $s_i = x_i'\delta_s + \eta_i$, where $\delta_s = [E(x_i x_i')]^{-1}E(x_i s_i)$ by construction and $E(x_i\eta_i) = 0$.

Assumption 1. *The instrument is uncorrelated with the error in the outcome equation,* $E(\varepsilon_i z_i) = 0$*, and correlated with schooling 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, two stage least squares (2SLS) estimation of the linear model (equation 2) is equivalent to the following IV estimator:

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

where

$$W_j^{IV} = (\tilde{z}'\tilde{s})^{-1}\tilde{z}'D_j = \frac{\frac{1}{N}\sum_{i=1}^{N}\tilde{z}_i D_{ij}}{\frac{1}{N}\sum_{i=1}^{N}\tilde{z}_i\tilde{s}_i}. \tag{3}$$

[10] For expositional purposes, it is assumed that there are no gaps in the schooling distribution, so the empirical density for schooling is strictly positive for all $S+1$ schooling levels. It is straightforward to generalize these results to account for such gaps.

Since $\sum_{j=1}^{S} D_{ij} = s_i$, these W_j^{IV} sum to one over $j = 1, ..., S$. We refer to them as "weights" even though they may be negative for some j.[11]

We show that the IV estimator of the mis-specified linear model converges to a "weighted average" of each grade-specific β_j effect. In general, the asymptotic "weights" sum to one but need not be non-negative; however, we discuss a set of conditions that yield more interpretable weights that are non-negative.

In terms of interpretation, one helpful assumption is monotonicity in the effects of the instrument on schooling. Though monotonicity is not necessary for deriving and estimating "weights", it does help ensure that they are non-negative and facilitates a more intuitive interpretation along the lines of the Local Average Treatment Effect (LATE) analysis of Angrist and Imbens (1995). 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 $s_i(z)$ to be the value of s_i for individual i when $z_i = z$.

Assumption 2. *(Monotonicity) The instrument does not decrease schooling:*

$$Pr[s_i(z) < s_i(z')] = 0 \text{ for all } z > z'.$$

To facilitate the discussion, decompose $z_i = x_i'\delta_z + \zeta_i$ where $\delta_z = [E(x_i x_i')]^{-1}E(x_i z_i)$ and $E(x_i\zeta_i) = 0$. We assume that x_i is distributed according to the density function $F(x)$.

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

$$\omega_j^{IV} = \frac{Pr(s_i \geq j)E(\zeta_i|s_i \geq j)}{\sum_{k=1}^{S}[Pr(s_i \geq k)E(\zeta_i|s_i \geq k)]} \tag{4}$$

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

$$\omega_j^{IV} = \frac{E\{Cov(z_i, D_{ij}|x_i)\}}{\sum_{k=1}^{S} E\{Cov(z_i, D_{ik}|x_i)\}} \geq 0. \tag{5}$$

Proof: It is straightforward to show that $W_j^{IV} \xrightarrow{p} \omega_j^{IV}$, since the numerator for W_j^{IV} equals $\frac{1}{N}\sum_{i=1}^{N} \tilde{z}_i D_{ij} \xrightarrow{p} E(D_{ij}\zeta_i) = Pr(s_i \geq j)E(\zeta_i|s_i \geq j)$, the denominator is $\frac{1}{N}\sum_{i=1}^{N} \tilde{z}_i\tilde{s}_i \xrightarrow{p} E(\eta_i z_i)$ which is assumed to be non-zero, and W_j^{IV} and ω_j^{IV} sum to one over $j = 1, ..., S$. The assumption that

[11] 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.

$E(\varepsilon_i z_i) = 0$ along with $E(\varepsilon_i|x_i) = 0$ implies 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 conditional on x by $F(z|x)$, the $Cov(z_i, D_{ij}|x) = \int[z - E(z|x)]Pr(D_{ij} = 1|z, x)dF(z|x)$ is non-negative for all x and j if $\partial Pr(D_{ij} = 1|z, x)/\partial z \geq 0$ for all x and j. This is ensured by Assumption 2. Using the fact that the weights sum to one concludes the proof.
QED

This result shows that estimating the mis-specified linear model using IV yields a consistent estimate of a weighted average of all grade-specific marginal effects. (This result is quite similar to that of Theorem 1 in Angrist and Imbens (1995). Their result allows for individual heterogeneity in β_j coefficients, but it assumes a binary instrument and does not consider additional covariates.) 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 W_j^{IV}.

These "weights" depend on the joint distribution of s, x, and z. 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 relationship between the instrument and each schooling transition indicator conditional on other covariates. The stronger the effect of the instrument on a particular schooling transition, the greater the weight on the effect of that transition. In general, different instruments yield estimates of different "weighted averages," even if the instruments are all valid.

2.1.1 Weighting across different observable types

Under Assumption 1 and $E(z_i|x_i) = x_i\delta_z$, it is straightforward to show that

$$\hat{\beta}_{IV}^{L} \xrightarrow{p} \int \beta_{IV}(x)h(x)dF(x)$$

where $\beta_{IV}(x) = \frac{Cov(z_i,y_i|x)}{Cov(z_i,s_i|x)}$ is the population analogue of the IV estimator conditional on $x_i = x$ and $h(x) = \frac{Cov(z_i,s_i|x)}{\int Cov(z_i,s_i|a)dF(a)}$ is a weighting function (that integrates to one) for different x. (The $h(x)$ weights are non-negative under Assumption 2.) Thus, the IV estimator converges to a weighted average of all conditional (on x) IV estimators, where the $h(x)$ weights are proportional to the covariance between the instrument and schooling conditional on x. $\beta_{IV}(x)$ estimators for those

types whose schooling is affected most by the instrument receive the greatest weight in calculating the average affect of schooling on y. Further, notice that $\beta_{IV}(x) = \sum_{j=1}^{S} \beta_j \omega_j^{IV}(x)$, where $\omega_j^{IV}(x) = \frac{Cov(z_i, D_{ij}|x_i)}{Cov(z_i, s_i|x)}$ are x-specific IV "weights" (i.e. they sum to one over all j) for each grade-specific effect, β_j.[12] So, each x-specific IV estimator is simply a weighted average of the grade-specific β_j effects, where the weights are proportional to the covariance between the instrument and D_{ij} conditional on x. Some re-arranging shows that we can write the IV weights from equations (4) or (5) as $\omega_j^{IV} = \int \omega_j^{IV}(x) h(x) dF(x)$.

With a binary instrument, the $\omega_j^{IV}(x)$ 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 $\pi(x) \equiv Pr(z_i = 1|x)$,

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

In this case, the x-specific weights simplify to

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

Thus, $\beta_{IV}(x)$ weights each β_j based on the fraction of all grade increments (for $x_i = x$ individuals) 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 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 x.

Under Assumptions 1 and 2, if $E(x_i|z_i) = E(x_i)$, then the weights in equations (4) or (5) simplify considerably, becoming independent of x_i:

$$\omega_j^{IV} = \frac{Pr(D_{ij} = 1|z = 1) - Pr(D_{ij} = 1|z = 0)}{\sum_{k=1}^{S} [Pr(D_{ik} = 1|z = 1) - Pr(D_{ik} = 1|z = 0)]} = \frac{Pr[s_i(0) < j \leq s_i(1)]}{\sum_{k=1}^{S} Pr[s_i(0) < k \leq s_i(1)]}.^{13} \quad (6)$$

The additional mean independence assumption $E(x|z) = E(x)$ 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

[12] These $\omega_j^{IV}(x)$ weights are non-negative under Assumption 2.

[13] See the Appendix for a proof of this result.

grade in response to a change in z, then the IV estimator reflects the average marginal effect of an additional year of school among individuals affected by the instrument.

Angrist and Imbens (1995) and Heckman, Urzua, and Vytlacil (2006) derive very similar weights on local average grade-specific effects when the β_j's vary across individuals. However, in order to ease interpretation, they make strong assumptions about the additional x_i covariates and how they enter the estimation procedure. For example, Angrist and Imbens (1995) assume that the x_i regressors are indicator variables that place individuals into mutually exclusive categories and that the instrumental variable is interacted with all of these additional covariates. Heckman, Urzua, and Vytlacil (2006) explicitly condition their ordered choice analysis on all covariates. Our analysis ignores heterogeneity in the grade-specific effects; however, it considers estimation under common assumptions about covariates and the way they enter during estimation. We are not focused on finding an 'economic interpretation' for the IV estimator (as in Angrist and Imbens (1995), Heckman and Vytlacil 2005), 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 linearity is incorrectly assumed.

2.1.2 Special Case: OLS Estimation of the Linear Specification

Since OLS is a special case of IV estimation, it is clear that in the absence of endogeneity (i.e. $E(\varepsilon_i|s_i) = 0$), the OLS estimator for the linear model also converges to a weighted average of the grade-specific effects, β_j, where the weights are non-negative and sum to one.

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

$$\hat{\beta}^L_{OLS} \xrightarrow{p} \sum_{j=1}^{S} \omega_j^{OLS} \beta_j \tag{7}$$

where the

$$\omega_j^{OLS} = \frac{Pr(s_i \geq j)E(\eta_i|s_i \geq j)}{\sum_{k=1}^{S} Pr(s_i \geq k)E(\eta_i|s_i \geq k)} \geq 0 \tag{8}$$

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

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

The empirical counterpart to ω_j^{OLS}, W_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. Note that $W_j^{OLS} \xrightarrow{p} \omega_j^{OLS}$ even if $E(s_i|x_i) \neq x_i'\delta_s$ and some "weights" are negative.

Of course, the weights implied by OLS estimation will not generally equal the weights implied by IV estimation. For example, consider the case with no x regressors (except an intercept). In this case, 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 β_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. The weights first decline slowly, then decline faster the further one gets away from the mean generating an inverted-U shape.

Contrast this with the weights implied by equation (6) 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. In Section 5, we graph estimated OLS and IV weights in a few different empirical applications.

Researchers often estimate linear specifications rather than more general non-linear models, because they are limited in the instrumental variables at their disposal. Yet, there is no reason to expect OLS and IV estimators for a mis-specified linear model to be equal even in the absence of endogeneity (i.e. if s_i and z_i are both uncorrelated with ε_i) or individual-level parameter heterogeneity (i.e. all β_j parameters are the same for everyone). As a result, standard Hausman tests applied to the mis-specified linear model may reject the null hypothesis of 'exogenous s' due simply to non-linearity in the relationship between s and y. Below, we develop a chi-square test for whether OLS estimation of equation (1) yields consistent estimates of the underlying β_j parameters (i.e. whether $E(\varepsilon_i|s_i) = 0$) even when only a single valid instrumental variable is available. However, first, we generalize our key results to the case of many instruments.

2.2 2SLS Estimation with Multiple Instruments

In Section 2.1 we have focused on the case where only one instrumental variable for schooling is available. Here we generalize the results to the case where we have I distinct instruments for schooling, $z_i = (z_{i1} \ ... \ z_{iI})'$, but the researcher still estimates the linear-in-schooling model (2).

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 θ_x and

θ_z. Further define the predicted value of schooling conditional on x and z: $\hat{s} = x_i'\hat{\theta}_x + z_i'\hat{\theta}_z$. Then,

$$\begin{aligned}\hat{\beta}_{2SLS}^{L} &= (\hat{s}'M_x\hat{s})^{-1}\hat{s}'M_x y \\ &= \sum_{j=1}^{S} W_j\beta_j + (\hat{s}'M_x\hat{s})^{-1}\hat{s}'M_x\varepsilon,\end{aligned}$$

where the "weights" $W_j = (\hat{s}'M_x\hat{s})^{-1}\hat{s}'M_xD_j = (\hat{\theta}_z'z'M_xz\hat{\theta}_z)^{-1}\hat{\theta}_z'z'M_xD_j$ reflect consistent estimates of ω_j from 2SLS estimation of

$$D_{ij} = s_i\omega_j + x_i'\alpha_j + \psi_{ij}, \qquad \forall j \in \{1, ..., S\}. \tag{9}$$

We will 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 $\zeta_i = (\zeta_{i1}, ..., \zeta_{iI})'$ be the $I \times 1$ vector collecting all $\zeta_{i\ell} = z_{i\ell} - x_i'\delta_{z\ell}$, where $\delta_{z\ell} = [E(x_ix_i')]^{-1}E(x_iz_{i\ell})$ was introduced above in the single-instrument case.

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 linear 2SLS estimator converges in probability to a "weighted" average of all grade-specific effects. Letting $\omega_{j\ell}^{IV}$ reflect the grade j "weight" from the single-instrument IV estimator using $z_{i\ell}$ as the instrument as defined by equation (4), the 2SLS estimator "weight" on any β_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 $j = 1, ..., S$ and

$$\Omega_\ell = \frac{\theta_{z\ell}\sum_{k=1}^{S} Pr(s_i \geq k)E(\zeta_{i\ell}|s_i \geq k)}{\sum_{m=1}^{I}\theta_{zm}\sum_{k=1}^{S} Pr(s_i \geq k)E(\zeta_{im}|s_i \geq k)} \tag{10}$$

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

Proof: See the Appendix.

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 Ω_ℓ in equation (10).[14]

3 A Wald Test for Consistent OLS Estimation of All β_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 $B \equiv (\beta_1, ..., \beta_S)$ from equation (1), $\hat{B}$, are consistent.[15] We now develop a test that compares the 2SLS estimator for the linear model with the weighted sum of the unrestricted grade-specific OLS estimates of the β_j's, using the estimated 2SLS weights $W \equiv (W_1, ..., W_S)'$. Intuitively, if $E(\varepsilon_i|s_i) = 0$ so OLS estimates of equation (1) are consistent, then the re-weighted sum of these OLS estimates (using the 2SLS weights) should asymptotically equal the 2SLS estimator from the linear model, i.e. $\hat{\beta}^L_{2SLS} - W'\hat{B} \xrightarrow{p} 0$. This will not generally be true when $E(\varepsilon_i D_{ij}) \neq 0$ for any j.

Recall that $\hat{B}$ is given by OLS estimation of equation (1), while $\hat{\beta}^L_{2SLS}$ is given by 2SLS estimation of equation (2). Applying 2SLS to equation (9) yields estimates W_j and $\hat{\alpha}_j$ for all j. In order to derive our test statistic, we frame estimation of $\hat{B}$, $\hat{\beta}^L_{2SLS}$, and W as a stacked generalized method of moments (GMM) problem. This establishes joint normality of $(\hat{B}, \hat{\beta}^L_{2SLS}, W)$ 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 $\hat{\beta}^L_{2SLS} - W'\hat{B}$, which is used in developing a chi-square test statistic for the null hypothesis that $\hat{T} \equiv \hat{\beta}^L_{2SLS} - W'\hat{B} \xrightarrow{p} 0$.

While most details are relegated to the Appendix, 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 (1), $X_{1i} = (D_i' \; x_i')$, and the regressors, $X_{2i} = (s_i \; x_i')$, and instruments, $Z_{2i} = (z_i' \; x_i')$, used in 2SLS estimation of equations (2) and (9). Denote the corresponding matrices for all individuals as X_1, X_2, and Z_2, respectively. Next, let $\Theta = (B' \; \gamma' \; \beta^L \; \gamma^{L\prime} \; W_1' \; \alpha_1' \; ... \; W_S' \; \alpha_S')'$ 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 $(\beta^L \; \gamma^{L\prime})$ and all $(W_j' \; \alpha_j')$ are estimated via 2SLS.

As shown in the Appendix, the variance of Θ can be consistently estimated from

$$\hat{V} = \hat{A}\hat{\Lambda}\hat{A}', \tag{11}$$

[14] If we define $\beta^L_{IV,\ell} = plim \; \hat{\beta}^L_{IV,\ell}$ where $\hat{\beta}^L_{IV,\ell}$ is the single-instrument IV estimator using $z_{i\ell}$ as an instrument for s_i in estimating equation (2), then $\hat{\beta}^L_{2SLS} \xrightarrow{p} \sum_{\ell=1}^{I} \Omega_\ell \beta^L_{IV,\ell}$, where Ω_ℓ is defined by equation (10).

[15] Formally, $\hat{B} = (D'M_xD)^{-1}D'M_xy$, where M_x and y are defined earlier and D reflects the stacked $N \times S$ matrix of $(D_{i1}, ..., D_{iS})$ for all individuals.

where

$$\hat{A} = \begin{pmatrix} [X_1'X_1]^{-1} & \mathbf{0} \\ \mathbf{0} & I_2 \otimes [\hat{X}_2'\hat{X}_2]^{-1}\hat{\Gamma}_2' \end{pmatrix}, \tag{12}$$

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

$$\hat{\Lambda} = \frac{1}{N}\sum_{i=1}^{N} \begin{pmatrix} \hat{\varepsilon}_i^2(X_{1i}'X_{1i}) & \hat{\varepsilon}_i\hat{\nu}_i(X_{1i}'Z_{2i}) & \hat{\varepsilon}_i\hat{\Psi}_i' \otimes (X_{1i}'Z_{2i}) \\ \hat{\varepsilon}_i\hat{\nu}_i(Z_{2i}'X_{1i}) & \hat{\nu}_i^2(Z_{2i}'Z_{2i}) & \hat{\nu}_i\hat{\Psi}_i' \otimes (Z_{2i}'Z_{2i}) \\ \hat{\varepsilon}_i\hat{\Psi}_i \otimes (Z_{2i}'X_{1i}) & \hat{\nu}_i\hat{\Psi}_i \otimes (Z_{2i}'Z_{2i}) & \hat{\Psi}_i\hat{\Psi}_i' \otimes (Z_{2i}'Z_{2i}) \end{pmatrix}, \tag{13}$$

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 $\hat{\Psi}_i = (\hat{\psi}_{1i}\ \hat{\psi}_{2i}\ ...\ \hat{\psi}_{Si})'$ with $\hat{\psi}_{ij} = D_{ij} - s_iW_j - \hat{\alpha}_j'x_i$.

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

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

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.

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

$$W_N = N\left[\frac{(\hat{\beta}_{2SLS}^L - W'\hat{B})^2}{\hat{G}\hat{V}\hat{G}'}\right] \xrightarrow{d} \chi^2(1). \tag{14}$$

Proof: See the Appendix.

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 $\omega_j = 0$. In other words, 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. In general, $\hat{B} \xrightarrow{p} B^* \equiv B + \{E(D_iD_i') - E(D_ix_i')[E(x_ix_i')]^{-1}E(x_iD_i')\}^{-1}E(D_i\varepsilon_i)$. Thus, $\hat{T} \xrightarrow{p} 0$ for any B^* satisfying $\omega'(B - B^*) = 0$. 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 $E(\varepsilon_iD_{ij}) = E(\varepsilon_i|s_i \geq j)$ 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 $\hat{B}$.[16]

[16] In the case where some $\omega_j = 0$, the test would be equivalent to testing for consistency of all β_j with $\omega_j > 0$.

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

$$s_i^* = \mu(z_i, x_i) + v_i \tag{15}$$

$$s_i = j \quad \text{if and only if } j \leq s_i^* < j+1. \tag{16}$$

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 $E(\varepsilon|v)$ is weakly monotonic in v, then $E(\varepsilon_i|s_i \geq j)$ will be either non-positive or non-negative for all j.[17] Monotonicity of $E(\varepsilon|v)$ is trivially satisfied by all joint elliptical distributions (e.g. bivariate normal or t distributions), which produce linear conditional expectation functions.

Intuitively, 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 ε 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 $E(\varepsilon|v)$. In many economic contexts, these perverse cases seem unlikely.

Finally, we 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. $\omega_z' B$ where z denotes the set of instruments used), it is unlikely that one would reject the null of $\omega_z' B = \omega_z' B^*$ for all sets of instruments unless $B = B^*$.[18]

4 A Monte Carlo Study

In this section, we use a Monte Carlo simulation exercise to show how varying degree of non-linearity can induce differences between the OLS and the IV estimates, even in the absence of endogeneity bias. As a setting, we consider a modified version of Card (1995) model of investment in human capital. An individual choose 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 κ in earnings at an arbitrary schooling level J

[17] Strictly speaking, weak monotonicity is only required over the range of v covered by $j - \mu(z,x)$ (i.e. for $v \in [1 - \mu(z,x), S - \mu(z,x)]$), so behavior in the tails of the distribution is irrelevant. See the Appendix for details.

[18] 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.

$$log[y_i(s_i)] = a + bs_i + \kappa 1(s_i \geq J) + \varepsilon_i \tag{17}$$

where κ measures the degree of non-linearity between log earnings and schooling. A larger κ implies a stronger non-linearity. The individual-specific cost of schooling is assumed to be

$$C_i(s_i) = c + r_i s_i + \frac{k_2}{2} s_i^2 + \kappa 1(s_i \geq J), \tag{18}$$

where the inclusion of κ here ensures that the non-linearity in earnings does not affect schooling choices. This allows us to focus on the extent to which non-linearity in the outcome variable affects IV and OLS estimators and our exogeneity test given a fixed set of OLS and IV weights.[19] Finally, we assume that the instrumental variable z_i shifts the cost of schooling

$$r_i = dz_i + \eta_i, \tag{19}$$

and that individuals can only choose $s \in \{0, 1, 2, ..., S\}$.[20]

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.

We set the sample size for each Monte Carlo simulation equal to 1,000. For each independent observation, we randomly draw a binary instrument $z_i \in \{0, 1\}$ independently from bivariate normally distributed errors (ϵ_i, η_i). Given the value of the parameters, the level of schooling is determined and realized values of $log(y_i)$ are constructed. Given this information, point estimates and standard errors are computed and saved. For each choice of ρ and k, we use 10,000 simulated samples.

For each model, defined by a combination of endogeneity (ρ) and jump size (κ), we compute point estimates and standard errors for OLS estimator and IV estimator. Specifically, we estimate the model for all possible combinations of

$$\rho \in \{0, 0.05, 0.1, 0.15, 0.2, 0.3\} \text{ and } \kappa \in \{0, 0.1, 0.5, 1.0\}.$$

[19] Including κ 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 κ in our analysis, they will vary with the extent of 'endogeneity' as defined by ρ below.

[20] We have made two changes to Card's original model. First, Card allows for variation in b_i, while we set $b_i = b$ for all i. Second, in Card log earnings are quadratic in schooling. In our case, log earnings are non-linear, but non-linearity is parameterized with discrete jumps. This allows for an easier interpretation of the Monte Carlo estimates.

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

The estimation results for these Monte Carlo exercises are shown in Table 1. For each model, we report the average point estimates and their standard deviation from the simulation samples for OLS and 2SLS estimators from the mis-specified linear model, as well as the re-weighted OLS estimates from the non-linear model using the estimated 2SLS weights, $\sum_{j=1}^{S} W_j \hat{\beta}_j$. (Estimated OLS and 2SLS weights are shown in Figure 1.) 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 we reject the null of exogeneity based on the linear specification using the standard 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 κ). 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 between the linear OLS and IV estimators. 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. κ 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 OLS and IV estimates, it does not generally account for all of the difference. The greater the endogeneity (i.e. the higher is ρ), 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 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 OLS estimate is still smaller than the IV estimate, but the re-weighted OLS estimate is much closer. Indeed, it appears that the different weights and non-linearity explain roughly one-third of the difference between linear 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, but sharply increasing in the degree of endogeneity. It is also interesting to note that when the true underlying model is linear (i.e. $\kappa = 0$), 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 $\kappa = 0$.

5 Three Empirical Examples

In this section, we focus on three recent empirical papers in which estimated 2SLS effects are different from the OLS effects: estimates of the effect of schooling on the probability of incarceration, using compulsory schooling laws as instruments (Lochner and Moretti, 2004); estimates of the effect of mother schooling on child health at birth, using opening of new colleges as an instrument (Currie and Moretti, 2003); and estimates of the private return to schooling using compulsory schooling laws as instruments (Acemoglu and Angrist, 2001). In all cases, the econometric specification assumed linearity.[21] In the presence of non-linearities, 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 β_j's helps reconcile the difference between the linearly mis-specified OLS and 2SLS estimates. We then test whether schooling is exogenous using both the standard Hausman test and our proposed generalization that accounts for potential non-linearities.

Results are reported in Table 2. 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 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 3 dummies for different compulsory schooling ages. The difference between OLS and 2SLS is reported in column

[21] However, Lochner and Moretti (2001) explore the extent to which non-linearities may explain the difference between their 2SLS and OLS estimates.

3. The 2SLS estimate is about 10% larger than the OLS one in absolute value, even if reasonable assumptions on the endogeneity of schooling would suggest that the OLS estimate is likely to 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 heterogeneous effects.[22] It is possible that non-linearity in the incarceration-schooling relationship may also play a role. This is particularly true here, since non-linearities appear to be important. In the top panel of Figures 2 and 3, we plot OLS estimates of the grade-specific effect of moving from $j-1$ to j years of schooling — i.e. the OLS estimates of the β_j coefficients. If the linearity assumption were correct, all the β_j would be the same. Instead, the estimated β_j suggest that the grade-specific effect of moving from $j-1$ to j years of schooling varies considerably across years of schooling. Overall, the figures are consistent with strong non-linearities in the effect of schooling on imprisonment, with the strongest effect for high school graduation (11 to 12). 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 general OLS estimates due to concerns about endogeneity.

The bottom panels in Figures 2 and 3 report estimates of the OLS weights and the IV 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 exactly 12 years of schooling, implying that the effect of moving from 11 to 12 years of schooling figures prominently in the 2SLS estimates. This makes sense, given that 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 estimates the grade-specific effect of moving from $j-1$ to j years of schooling (β_j) using the 2SLS weights in the bottom panel of Figure 2. For whites, the re-weighted OLS estimates are 0.0012, larger than the IV estimates. Intuitively, the re-weighted OLS estimates are larger because the 2SLS weights put more weights on the large β_j that represent the effect of moving from 11 to 12 years of schooling. For blacks,

[22] With heterogenous effects, 2SLS estimates reflect the effects of schooling for those individuals whose schooling is affected by the instrument. For example, Kling (2001) shows that college proximity largely affects the schooling achievement of individuals from lower socioeconomic backgrounds. OLS will tend to reflect the impact for a broader population. If returns to schooling are higher for individuals from lower socioeconomic backgrounds, this could translate into a larger effect of schooling on criminal behavior for individuals most affected by the instrument. This may lead to larger 2SLS estimates relative to OLS estimates.

the re-weighted OLS estimate is smaller, because the 2SLS weights are more shifted to the left and therefore put less weight on β_j that are large.

The last three columns of Table 2 are the most important, since they report on different tests for the endogeneity of schooling. Column 5 presents test statistics and associated p-values for on our proposed test of endogeneity (see Theorem 1), which is valid in the presence of non-linearities. Columns 6 and 7 present results from the standard Hausman-based Wald test and the Durbin-Wu-Hausman test, respectively, which are both incorrect in the presence of non-linearities. For white men, our test fails to reject, which is quite important in practice. Based on this finding, we can confidently take the OLS estimates of the β_j in Figure 2 as consistent. This confirms the speculation by Lochner and Moretti (2004) that high school completion has the greatest effect on incarceration rates, and that college attendance has weaker effects. This is extremely useful, since with only three available instruments, it is impossible to estimate all 20 β_j parameters by 2SLS. Indeed, it is not possible to precisely estimate highly restricted two-parameter non-linear models. Fortunately, our test suggests that this is not necessary in this context.

The case of incarceration for black men is different: our test clearly rejects the hypothesis that the re-weighted OLS and 2SLS estimates are the same, with a p-value of .0005. Notably, the standard Hausman test fails to reject. This is particularly interesting, since it shows how the standard test may fail to detect an endogeneity problem when one exists if non-linearity is a problem. Our test, of course, correctly identifies the problem. Again, this is important in practice. A priori, one might have expected the endogeneity of schooling to produce OLS estimates that are too large (in absolute value); yet, a comparison of columns 1 and 2 suggests that there is little evidence of any bias. Despite the similarity of the linear OLS and 2SLS estimates, our test clearly implies that schooling is endogenous. Thus, in this case, a researcher would be wrong in concluding from the Hausman or Durbin-Wu-Hausman tests that schooling was exogenous. The unrestricted OLS coefficients reported in Table 2 are not consistent.

In the second panel, we turn to estimates of the effect of maternal schooling on infant health and health inputs from Currie and Moretti (2003). The instrument in this case is a dummy for college proximity. In this case, the re-weighted OLS estimates (column 4) are generally similar to the OLS estimates (column 1). Looking at Figures 4 and 5, it is clear why: the OLS and 2SLS weights are nearly identical. Not surprisingly, our test and the standard Hausman test produce the same conclusion.

Finally, in the bottom panel, we turn to estimates of the private return to schooling using three dummies for compulsory schooling as instruments. While the original Acemoglu and Angrist paper includes estimates of the social return to schooling, we focus only on the more standard

private return to schooling, consistent with much of the literature. The dependent variable is log annual earnings. OLS estimates indicate that an additional year of schooling translates into 8.2% increase in annual earnings, while the 2SLS estimates suggest a much larger return. The re-weighted OLS estimates are in between, although the effect of re-weighting is minor despite the substantially different OLS and 2SLS weights (see Figure 6). Our test rejects the hypothesis that the re-weighted OLS and 2SLS estimates are equal.

6 Conclusions

In applied work, it is often the case that OLS and IV estimates differ, and sometimes the direction of the difference is not what one might expect based on economic theory and plausible assumptions on 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 is heterogenous. Our work complements the existing understanding of the differences between IV and OLS estimates when the model is mis-specified. We focus on the case where the true model is non-linear, but the researcher estimates a linear model. This case has become increasingly relevant as the growing emphasis on the validity of instruments has led many empirical researchers to estimate linear relationships with only a few instruments. Yet, in many instances the true relationship between the dependent and independent variables may be quite non-linear, as is frequently suggested by more general specifications estimated using OLS.

We develop a simple framework for thinking about the effects of nonlinearity when estimating mis-specified linear models using IV and OLS. In our setting, it is easy to compare IV estimates and OLS estimates and to interpret the difference. IV estimates and OLS estimates are both weighted averages of marginal effects, with different weights. For OLS, the marginal effects for levels near the average of the regressor tend to be weighted more heavily than marginal effects at low or high levels of the regressor. For IV, the stronger the effect of the instrument on a particular transition, the greater the weight on the effect of that transition. As a consequence, IV and OLS estimates may differ even in the absence of endogeneity. We show that it is easy to estimate these weights. The level-specific OLS weights can be obtained by regressing indicators for whether the regressor is above each level on the regressor. The IV weights can be estimated with a similar model, where the indicators for whether the regressor is above each level are instrumented for.

Building on these insights, the main contribution of this paper is to develop a simple generalization of the Hausman test to assess whether different weighting and non-linearity explains the

difference between linear IV/2SLS and OLS estimators. Under fairly weak conditions, this serves as a specification test for exogeneity of the regressor in a general non-linear context. A particularly appealing feature of our test is that it only requires a single instrument, the primary reason many researchers turn to linear models rather than estimate more general non-linear models that can be estimated using OLS. Our test offers researchers the ability to estimate general non-linear models using OLS and then easily test whether those estimates are consistent.

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.

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 Vytlacil. "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.

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

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

Heckman, James J., and Edward Vytlacil. "Structural Equations, Treatment Effects, and Econometric Policy Evaluation," Econometrica, Econometric Society, vol. 73(3), pages 669–738, 2005.

Heckman, James J., Lance Lochner, and Petra Todd. "Earnings Functions and Rates of Re-

turn," Journal of Human Capital, 2(1), 2008.

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

Hungeford, 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, 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.

Moffitt, Robert. "Estimating Marginal Treatment Effects in Heterogeneous Populations," Annales dEconomie 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.

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.

Appendix: Proofs and Technical Results

Derivation of Equation (6)

Equation (6) is easily verified using a slightly different decomposition of the empirical weights from the main text. Decompose $D_{ij} = x_i'\delta_{D_j} + \xi_{ij}$, where $\delta_{D_j} = [E(x_i x_i')]^{-1}E(x_i D_{ij})$ and $E(x_i\xi_{ij}) = 0$. With this, re-write

$$\omega_j^{IV} = \frac{E(z_i\xi_{ij})}{\sum_{k=1}^{S} E(z_i\xi_{ik})}.$$

Letting $\pi \equiv Pr(z_i = 1)$, observe that $E(z_i\xi_{ij}) = \pi(1-\pi)[E(\xi_{ij}|z_i = 1) - E(\xi_{ij}|z_i = 0)]$, so

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

Mean independence $E(x_i|z_i) = E(x_i)$, further simplifies the weights to

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

since $E(x_i'\delta_{D_j}|z_i) = E(x_i'|z_i)\delta_{D_j} = E(x_i')\delta_{D_j}$. Monotonicity of schooling in the instrument yields the final expression for ω_j^{IV} in the text.

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 ω_j^{OLS} equals $E(\eta_i D_{ij})$. To see that this is non-negative, notice that

$$E(\eta_i) = \int_{-\infty}^{\infty}\int_{j-x'\delta_s}^{\infty} \eta dF(\eta|x)dG(x) + \int_{-\infty}^{\infty}\int_{-\infty}^{j-x'\delta_s} \eta dF(\eta|x)dG(x), \tag{20}$$

where $G(x)$ reflects the density of x and $F(\eta|x)$ the conditional density of η conditional on x. Assuming x includes a constant term, $E(\eta_i) = 0$. Since the first term in equation (20) 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 $\hat{s}'M_x\hat{s} = s'M_xz(z'M_xz)^{-1}z'M_xs$. Since, $\frac{1}{N}s'M_xz \xrightarrow{p} E[(s_i - x_i'\delta_s)z_i'] = E(\eta_i z_i') \neq 0$ by Assumption 1 and $\frac{1}{N}z'M_xz \xrightarrow{p} E[z_i(z_i' - x_i'\delta_z)] = E(z_i\zeta_i') = E(\zeta_i\zeta_i')$, which is full rank by Assumption 3, the denominator for ω_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 W_j and ω_j sum to one. Now, consider the numerator for W_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 θ_z coefficient on $z_{i\ell}$. Since the ω_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, Ω_ℓ is given by equation (10).

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 Ω_ℓ, $\omega_{j\ell}^{IV}$, and ω_j are non-negative.
QED

Proof of Theorem 1

Proposition 2 shows that the linear 2SLS estimator converges to a "weighted average" of the true β_j's with the "weights", $\omega = (\omega_1, ..., \omega_S)'$, consistently estimated by 2SLS estimation of equation (9). That is, $W \xrightarrow{p} \omega$ and $\hat{\beta}^{2SLS} \xrightarrow{p} \omega' B$. If $E(\varepsilon_i | s_i) = 0$, then $\hat{B} \xrightarrow{p} B$, which implies that $\hat{\beta}_{2SLS}^{L} - W' \hat{B} \xrightarrow{p} 0$.

We write the estimation problems for equations (1), (2), and (9) in the form of a stacked linear GMM problem. (Note that equation (1) is estimated using OLS while the remaining equations are estimated using 2SLS.) This establishes joint normality of $(\hat{B}, \hat{\beta}_{2SLS}^{L}, W)$ in the limit and facilitates

estimation of their covariance matrix. A straightforward application of the delta-method yields the variance of $\hat{T} \equiv \hat{\beta}^L_{2SLS} - W'\hat{B}$, 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 $\Psi_i = (\psi_{1i} \ \psi_{2i} \ ... \ \psi_{Si})'$. Recall from Section 3 that $\Theta = (B' \ \gamma' \ \beta^L \ \gamma^{L\prime} \ W_1' \ \alpha_1' \ ... \ W_S' \ \alpha_S')'$ 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 (1), (2), and (9) (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 $\hat{\Theta} = \left[X'Z(Z'Z)^{-1}Z'X\right]^{-1} X'Z(Z'Z)^{-1}Z'Y$.

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 Ω is defined above.[23]

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

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

[23] Substituting in for C and Ω and simplifying yields

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

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 (11), (12) and (13) 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 (14).

QED

Ordered Choice Model

Assume schooling is determined by the ordered choice model defined by equations (15) and (16). Then, the sign of the asymptotic bias for OLS estimation of any β_j in equation (1) depends on the sign of

$$\begin{aligned} E(\varepsilon D_j) &= E(\varepsilon | s \geq j) \\ &= E(E[\varepsilon | v, z, x, v \geq j - \mu(z,x)]). \end{aligned}$$

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

Table 1: Monte Carlo Simulations for 'Card Model'

ρ	κ	Linear OLS	Linear IV	Re-weighted OLS	General Wald Test (fraction reject using .05 sig. level)	DWH Test (fraction reject using .05 sig. level)
0	0	0.0399	0.0399	0.0399	0.050	0.049
		(0.0054)	(0.0096)	(0.0056)		
0	0.1	0.0540	0.0557	0.0556	0.051	0.054
		(0.0054)	(0.0095)	(0.0056)		
0	0.5	0.1099	0.1180	0.1179	0.056	0.172
		(0.0057)	(0.0100)	(0.0063)		
0	1	0.1801	0.1961	0.1960	0.047	0.434
		(0.0063)	(0.0111)	(0.0080)		
0.05	0	0.0330	0.0399	0.0332	0.139	0.144
		(0.0055)	(0.0096)	(0.0057)		
0.05	0.1	0.0470	0.0556	0.0489	0.139	0.206
		(0.0054)	(0.0095)	(0.0056)		
0.05	0.5	0.1030	0.1179	0.1112	0.139	0.472
		(0.0056)	(0.0099)	(0.0063)		
0.05	1	0.1729	0.1960	0.1892	0.146	0.721
		(0.0063)	(0.0111)	(0.0080)		
0.1	0	0.0260	0.0402	0.0265	0.428	0.444
		(0.0054)	(0.0094)	(0.0056)		
0.1	0.1	0.0399	0.0557	0.0420	0.430	0.527
		(0.0055)	(0.0095)	(0.0057)		
0.1	0.5	0.0959	0.1179	0.1044	0.424	0.784
		(0.0057)	(0.0100)	(0.0064)		
0.1	1	0.1659	0.1960	0.1823	0.429	0.911
		(0.0063)	(0.0112)	(0.0081)		
0.15	0	0.0191	0.0402	0.0197	0.762	0.783
		(0.0055)	(0.0096)	(0.0057)		
0.15	0.1	0.0331	0.0558	0.0354	0.760	0.836
		(0.0055)	(0.0096)	(0.0057)		
0.15	0.5	0.0890	0.1180	0.0977	0.761	0.954
		(0.0057)	(0.0100)	(0.0064)		
0.15	1	0.1590	0.1962	0.1757	0.763	0.984
		(0.0063)	(0.0113)	(0.0080)		
0.2	0	0.0119	0.0401	0.0129	0.949	0.956
		(0.0053)	(0.0096)	(0.0055)		
0.2	0.1	0.0261	0.0558	0.0286	0.951	0.971
		(0.0054)	(0.0095)	(0.0056)		
0.2	0.5	0.0820	0.1182	0.0910	0.950	0.995
		(0.0057)	(0.0101)	(0.0064)		
0.2	1	0.1519	0.1958	0.1688	0.949	0.999
		(0.0063)	(0.0112)	(0.0081)		
0.3	0	-0.0021	0.0401	-0.0007	1.000	1.000
		(0.0053)	(0.0097)	(0.0055)		
0.3	0.1	0.0120	0.0557	0.0149	1.000	1.000
		(0.0052)	(0.0095)	(0.0054)		
0.3	0.5	0.0679	0.1181	0.0772	1.000	1.000
		(0.0055)	(0.0101)	(0.0063)		
0.3	1	0.1379	0.1959	0.1552	1.000	1.000
		(0.0062)	(0.0112)	(0.0081)		

Table 2: Replication Results and Application of Wald Tests for Endogeneity

	OLS	IV	IV - OLS	Re-weighted OLS[1]	Our Generalized Wald Test	Naïve Wald Test [2]	DWH Test [3]
	1	2	3	4	5	6	7
1. Lochner and Moretti (2004)	Effect of Years of Schooling on Imprisonment						
White Males	-0.0010	-0.0011	-0.0002	-0.0012	0.0225	0.2021	0.1600
	0.0000	0.0004	0.0004	0.0000	0.8808	0.6530	0.6858
Black Males	-0.0037	-0.0048	-0.0011	-0.0007	11.9441	0.9757	0.5154
	0.0001	0.0012	0.0011	0.0002	0.0005	0.3233	0.4728
2. Currie & Moretti (2003)	Effect of Maternal Education on Infant Health and Health Inputs						
Low birth weight	-0.0050	-0.0098	-0.0048	-0.0053	1.4376	1.7022	1.5566
	0.0001	0.0038	0.0037	0.0002	0.2305	0.1920	0.2122
Preterm birth	-0.0044	-0.0104	-0.0060	-0.0046	1.7639	2.0472	1.7749
	0.0002	0.0044	0.0042	0.0002	0.1841	0.1525	0.1828
3. Acemoglu & Angrist (2001)	Private Returns to Schooling						
Annual Earnings	0.0822	0.1442	0.0620	0.0832	5.7093	6.0028	6.0218
	0.0003	0.0256	0.0253	0.0017	0.0169	0.0143	0.0141

Notes: Re-weighted OLS reports the weighted average of all OLS βj estimates using the 2SLS weights. 'Our Generalized Wald Test' reports test statistics and p-values for the test developed in Theorem 1 of this paper. The 'Niave Wald Test' reports standard Hausman (1978) test stastics and p-values for the difference between the linear 2SLS and OLS estimates. 'DWH' reports test statistics and p-values for the Durbin-Wu-Hausman test using an augmented regression. Specifications for Lochner and Moretti (2004) use men ages 20-60 friom the 1960-80 U.S. Censuses and include indicators for three-year age categories, year, state of birth, and state of residence. Specifications from Currie and Moretti (2003) use first-time white mothers ages 24-35 from Vital Statistics Natality records from 1970-99 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: 2SLS and OLS weights for Monte Carlo Study (ρ=0)

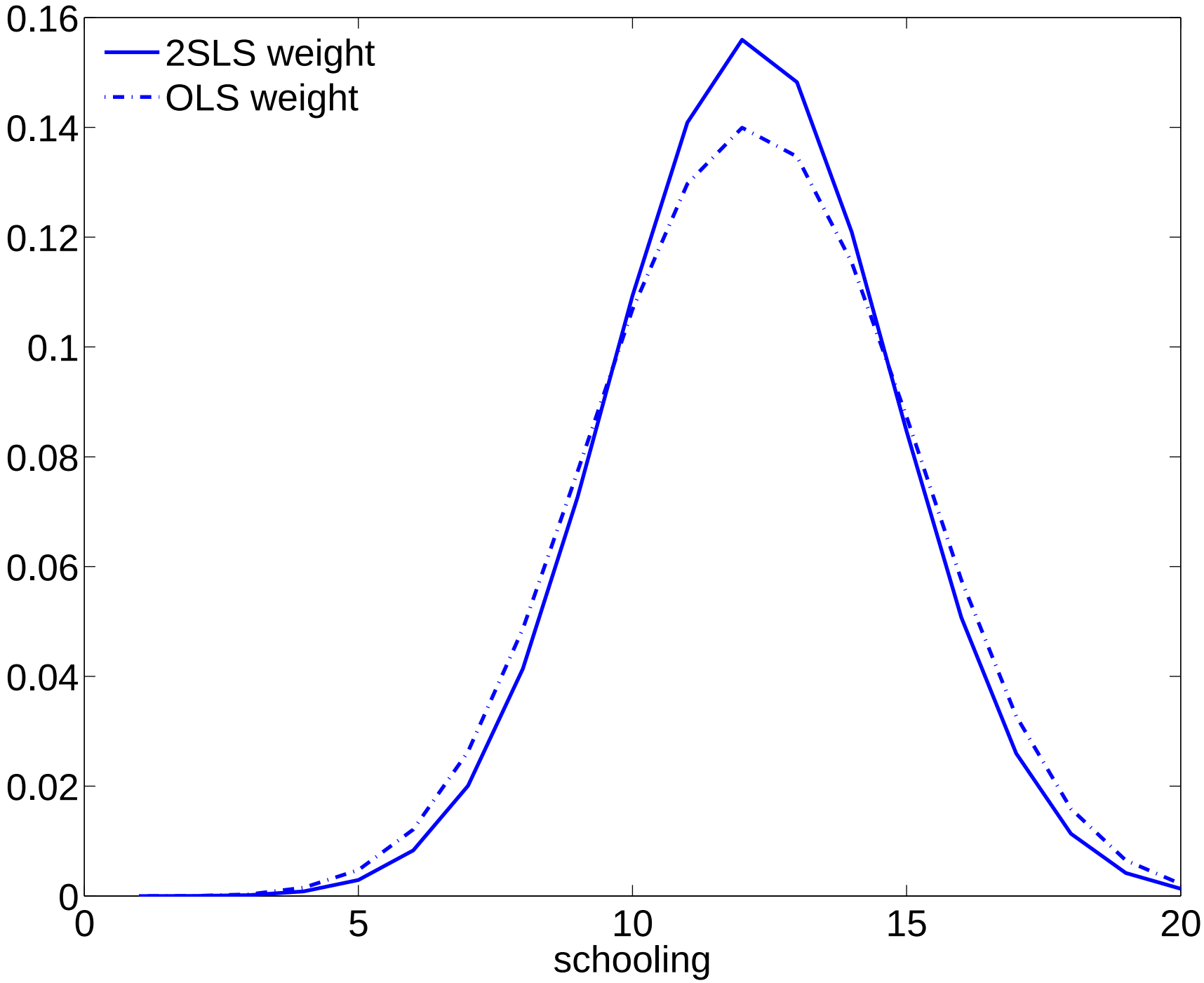

Figure 2: Effects of Schooling on Probability of Incarceration for White Males
(Estimated OLS Effects and Weights)

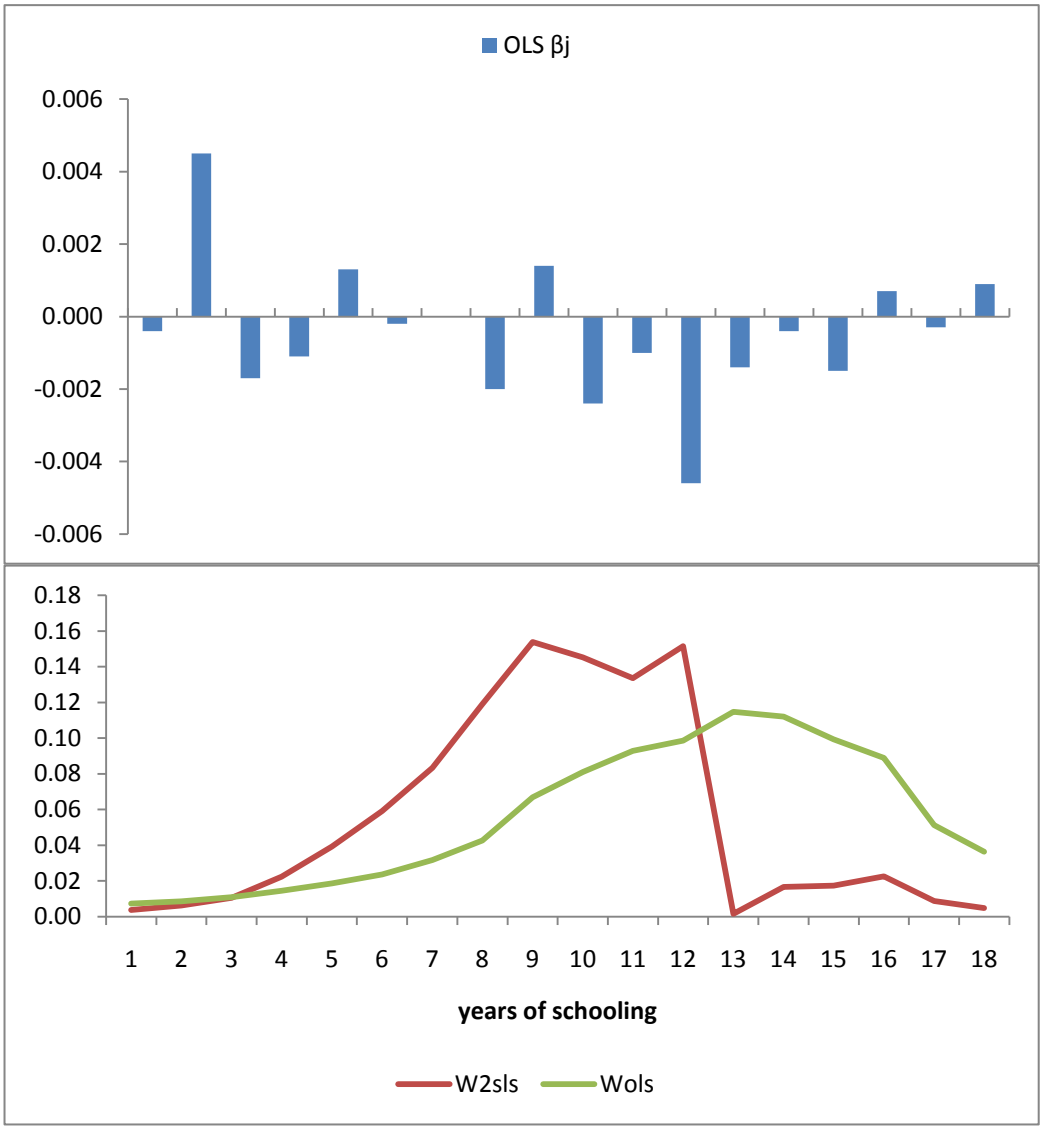

Figure 3: Effects of Schooling on Probability of Incarceration for Black Males
(Estimated OLS Effects and Weights)

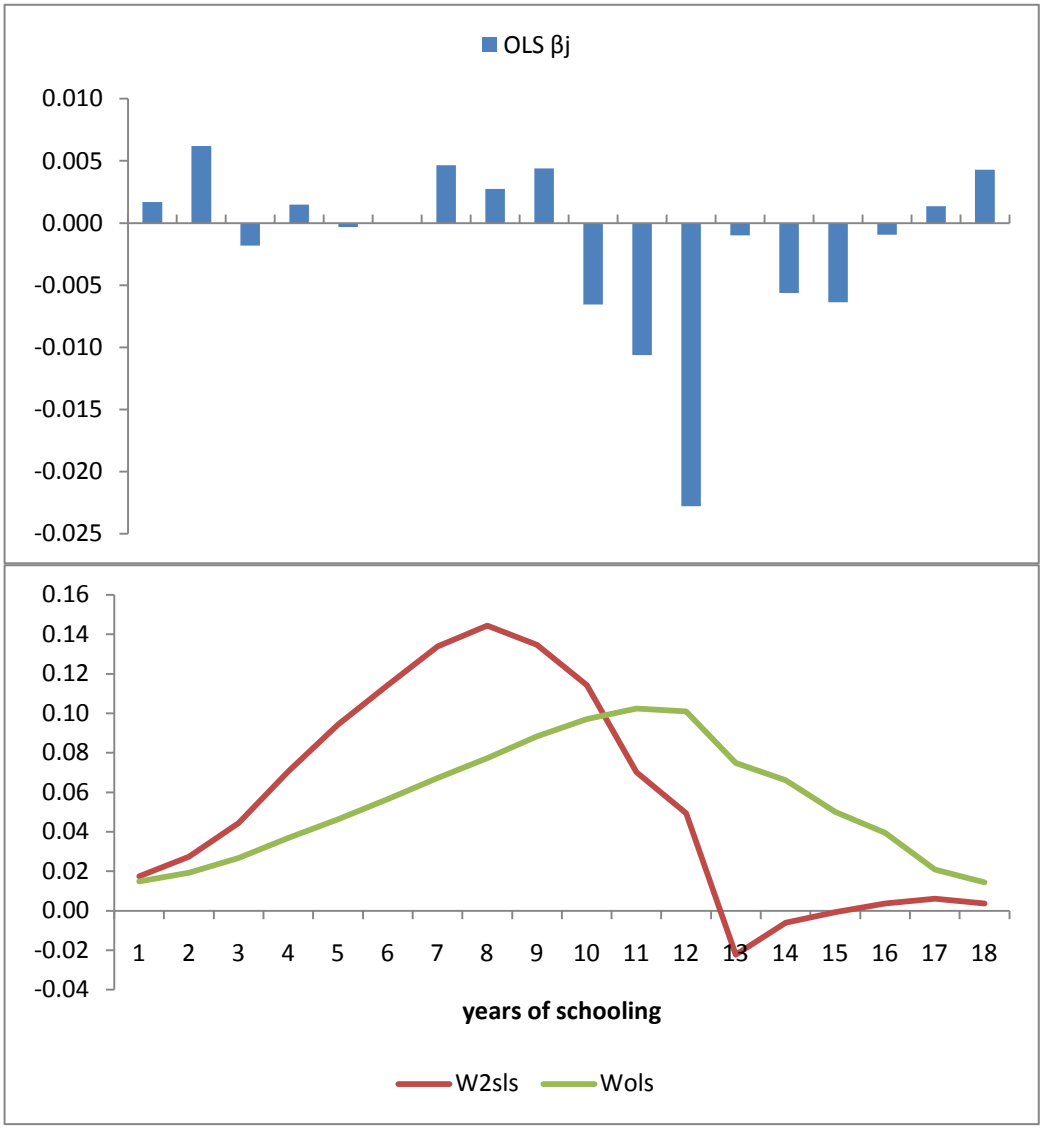

Figure 4: Effects of Maternal Schooling on Probability of Low Birth Weight
(Estimated OLS Effects and Weights)

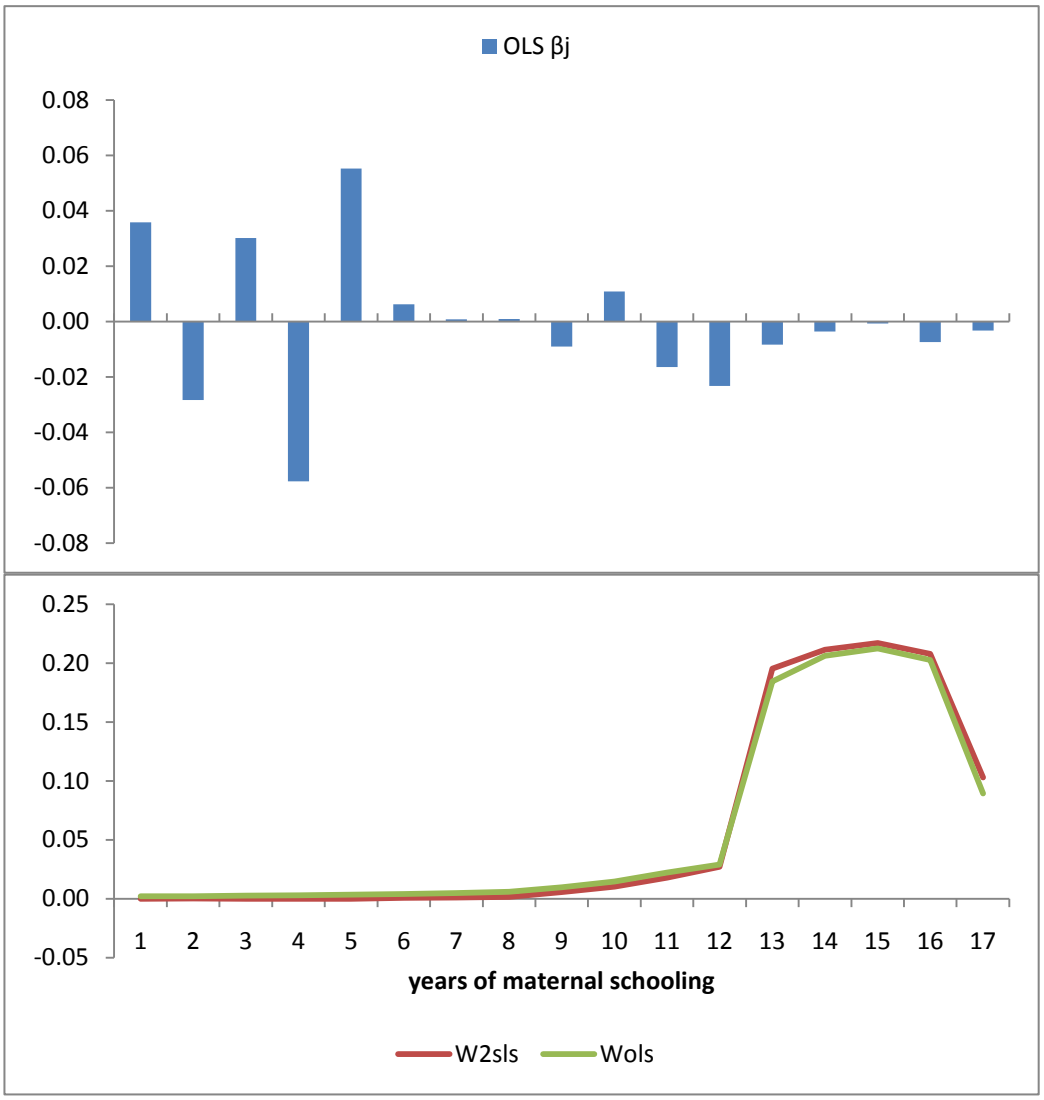

Figure 5: Effects of Maternal Schooling on Probability of Pre-Term Birth
(Estimated OLS Effects and Weights)

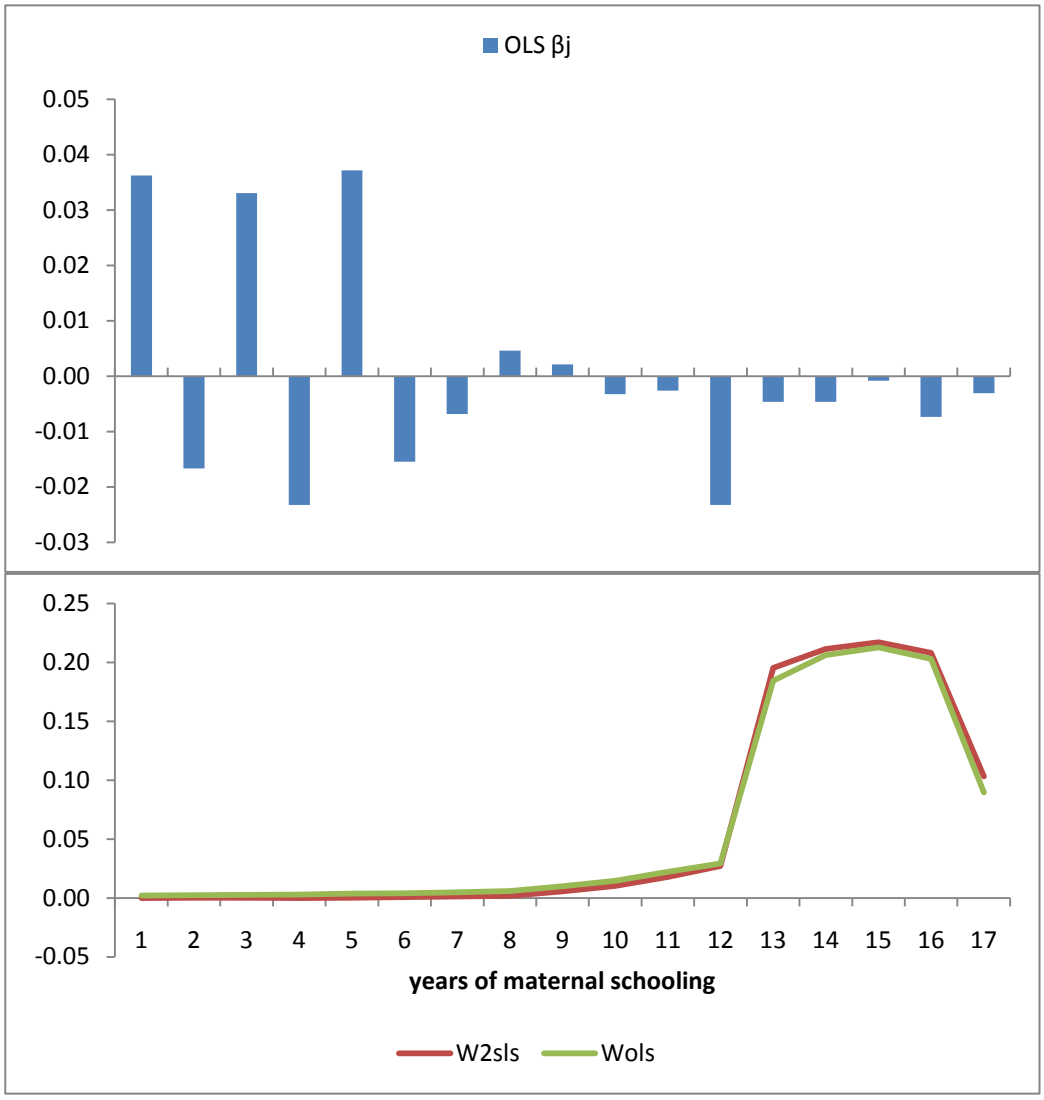

Figure 6: Effects of Schooling on Log Annual Earnings for Men
(Estimated OLS Effects and Weights)