The Stata Journal (2016)
16, Number 3, pp. 761–777

A test for exogeneity in the presence of nonlinearities

Michael P. Babington
Department of Economics
Florida State University
Tallahassee, FL
mb13m@my.fsu.edu

Javier Cano-Urbina
Department of Economics
Florida State University
Tallahassee, FL
jcanourbina@fsu.edu

Abstract. We provide a command, `locmtest`, that implements a test for exogeneity that is robust when the true relationship between the outcome variable and a discrete potentially endogenous variable is nonlinear. This test was developed in Lochner and Moretti (2015, *Review of Economics and Statistics* 97: 387–397), and it can be implemented even when only a single valid instrument is available. We present the motivation and general idea of the test. We also describe `locmtest`, which calculates the test, and provide empirical applications of the test and the command.

Keywords: st0454, locmtest, exogeneity, nonlinear

1 Introduction

Recent work by Lochner and Moretti (2015) develops a new test for exogeneity that is robust to nonlinearities in the relationship between the outcome variable and a discrete potentially endogenous variable. Nonlinear relationships arise naturally in many economic applications, as in the case of schooling and earnings where there is the possibility of finding sheepskin effects of education on earnings. In such cases, even though the potentially endogenous variable enters the relationship nonlinearly, the estimation approach often assumes it enters linearly and applies a Hausman test. However, in the presence of these nonlinearities, the Hausman test is uninformative about exogeneity of the potentially endogenous variable.

To better understand Lochner and Moretti's (2015) test, let's suppose we are interested in estimating the effect of variable s_i on outcome y_i, where $s_i \in \{0, 1, 2, 3, \ldots, S\}$, and that in the population, the conditional mean of y_i is given by

$$y_i = \sum_{j=1}^{S} D_{ij}\beta_j + \mathbf{x}_i'\gamma + \varepsilon_i \qquad (1)$$

where $D_{ij} = \mathbf{1}(s_i \geq j)$ is an indicator that equals 1 if $s_i \geq j$ and 0 otherwise,[1] $E(\varepsilon_i) = 0$, and $E(\varepsilon_i|\mathbf{x}_i) = 0$, so that $\mathbf{x}_i$ is a $k \times 1$ vector of exogenous covariates (including an intercept).[2] If s_i is potentially endogenous in (1), instrumental-variable (IV) estimation requires several valid instruments, yet we are typically limited with the number of valid instruments at hand.[3] As a result, researchers typically estimate specifications that assume a linear relation between y_i and s_i such as

$$y_i = s_i\beta^L + \mathbf{x}_i'\gamma^L + \nu_i \tag{2}$$

where $\mathbf{x}_i$ is the same vector of covariates (including an intercept) that appears in (1).[4] Equation (2) is commonly used in empirical studies, and it implicitly assumes that the effect of s_i on the outcome is uniform across all levels of s_i, while (1) allows s_i to affect the outcome variable differently for different levels of s_i.

Lochner and Moretti (2015) show that estimating (2) when the true relationship is described by (1) can lead to different ordinary least-squares (OLS) and IV or two-stage least-squares (2SLS) estimates even in the absence of endogeneity. The problem is that conclusions about the exogeneity of s_i in (2) are typically based on the comparison of the OLS and IV or 2SLS estimates (as in the standard Hausman test)—hence the potential of incorrectly rejecting exogeneity of s_i when the problem is a misspecification of the relationship between s_i and the outcome variable. The test proposed by Lochner and Moretti (2015) is robust to this misspecification and has the advantage of requiring only a single instrument that can be binary.

Note that the test developed in Lochner and Moretti (2015) does not apply to all nonlinear models, only the specific case described in (1) and (2). It also follows from these two equations that three conditions are crucial: i) a single finite-valued discrete potentially endogenous regressor s_i must be present; ii) exogenous regressors $\mathbf{x}_i$ are additively separable and enter the equation linearly; and iii) all coefficients are homogeneous in the population. These assumptions are usually made in many empirical studies.

Section 2 shows that estimates of $\widehat{\beta}^L$ in the linear specification (2) when the nonlinear specification (1) is the correct one lead to different weighted averages for OLS and IV or 2SLS, which can in turn lead to the wrong conclusion if the standard Hausman test is applied. Section 3 describes the test for exogeneity. Section 4 describes the `locmtest` command to compute the Lochner–Moretti (LM) test for exogenity. Finally, section 5 presents three examples in which the LM test and the command can be used to test exogeneity of years of education with a discrete potentially endogenous variable in many empirical studies.

1. Note that to avoid perfect multicollinearity, we provide no dummy variable for the first category of s_i.
2. Because (1) is a specification for the conditional mean of y_i, exogeneity of $\mathbf{x}_i$ requires $E(\varepsilon_i|\mathbf{x}_i) = 0$. However, all results presented below hold under the weaker assumption $E(\mathbf{x}_i\varepsilon_i) = 0$.
3. For example, if s_i represents years of education, S could be 18, and so we would need to estimate 18 different grade-specific β_j parameters associated with all levels of schooling.
4. Recall that this vector of covariates is exogenous in (1).

2 The motivation of the LM test

To develop their test for exogeneity, Lochner and Moretti (2015) start by noticing that under standard IV assumptions, IV estimates of $\widehat{\beta}^L$ in the linear specification (2) converge in probability to a weighted average of the level-specific β_j from the nonlinear specification (1).[5] To see this, consider the case in which one valid instrument z_i is available, and it is used to estimate β^L in (2), which is the coefficient in the potentially endogenous variable s_i. Define the annihilator matrix $\mathbf{M}_x = I - x(x'x)^{-1}x'$, and for any variable w, let $\widetilde{w} = \mathbf{M}_x w$. Then, the IV estimator of β^L is given by

$$\begin{aligned}
\widehat{\beta}^L_{\text{IV}} &= (z'\mathbf{M}_x s)^{-1} z'\mathbf{M}_x y \\
&= (z'\mathbf{M}_x s)^{-1} z'\mathbf{M}_x \left(\sum_{j=1}^{S} \mathbf{D}_j \beta_j + x\gamma + \boldsymbol{\varepsilon} \right) \\
&= \sum_{j=1}^{S} \widehat{\omega}_j^{\text{IV}} \beta_j + (\widetilde{z}'\widetilde{s})^{-1} \widetilde{z}\boldsymbol{\varepsilon}
\end{aligned}$$

where $\mathbf{D}_j$ is the $N \times 1$ vector of indicator variables corresponding to level j, $\widehat{\omega}_j^{\text{IV}} = (\widetilde{z}'\widetilde{s})^{-1}\widetilde{z}'\mathbf{D}_j$, and $\boldsymbol{\varepsilon}$ is the $N \times 1$ vector of error terms in (1).

Next, to derive the probability limit of $\widehat{\beta}^L_{\text{IV}}$ and the necessary assumptions for convergence to that limit, we consider two linear projections: i) the linear projection of s_i on $\mathbf{x}_i$ given by $s_i = \mathbf{x}_i'\delta_s + \eta_i$, where $\delta_s = \{E(\mathbf{x}_i\mathbf{x}_i')\}^{-1}E(\mathbf{x}_i s_i)$ by construction and $E(\mathbf{x}_i\eta_i) = 0$; and ii) the linear projection of z_i on $\mathbf{x}_i$ given by $z_i = \mathbf{x}_i'\delta_z + \zeta_i$, where $\delta_z = \{E(\mathbf{x}_i\mathbf{x}_i')\}^{-1}E(\mathbf{x}_i z_i)$ by construction and $E(\mathbf{x}_i\zeta_i) = 0$. Finally, consider the following assumption:

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

Then, Lochner and Moretti (2015) show that if assumption 1 holds,

$$\widehat{\beta}^L_{\text{IV}} \xrightarrow{p} \sum_{j=1}^{S} \omega_j^{\text{IV}} \beta_j \tag{3}$$

and the weights ω_j^{IV} sum to 1 over all $j = 1, 2, \ldots, S$ and can be written as

$$\omega_j^{\text{IV}} = \frac{\Pr(s_i \geq j|\mathbf{x}_i)E(\zeta_i|s_i \geq j)}{\sum_{q=1}^{S} \Pr(s_i \geq q|\mathbf{x}_i)E(\zeta_i|s_i \geq q)} \geq 0$$

5. Other studies have also noticed that estimates from a misspecified linear model yield weighted averages of the level-specific effects, for example, Yitzhaki (1996), Angrist and Imbens (1995), and Heckman, Urzua, and Vytlacil (2006).

The sample analogs of the IV weights ω_j^{IV} are given by $\widehat{\omega}_j^{\mathrm{IV}} = (\widetilde{z}'\widetilde{s})^{-1}\widetilde{z}'\mathbf{D}_j$, which are simply the IV estimates of the coefficient on s_i using z_i as an instrument for s_i in the regressions given by

$$D_{ij} = \omega_j s_i + \mathbf{x}_i'\alpha_j + \psi_{ij}, \quad j = 1, 2, \ldots, S \tag{4}$$

Given the assumptions above, it can be shown that $\widehat{\omega}_j^{\mathrm{IV}} \xrightarrow{p} \omega_j^{\mathrm{IV}}$.

A difficulty with the IV weights is that they can be negative, which complicates their interpretation. If a condition that Lochner and Moretti (2015) call "monotonicity" holds, then the weights are all nonnegative. This condition is very similar to the monotonicity assumption in Imbens and Angrist (1994) and requires that increases in z_i causes everyone to weakly increase s_i or causes everyone to weakly decrease s_i.[6]

Note that monotonicity is often a point of contention in applied work (see Barua and Lang [2009], de Chaisemartin [2014], Aliprantis [2012], and Klein [2010]).[7] However, the monotonicity assumption is necessary neither to derive the probability limit of $\widehat{\beta}_{\mathrm{IV}}^L$ nor to derive the Wald test presented below. Nonetheless, if satisfied, the monotonicity assumption facilitates interpretation of the IV weights.

The case of estimation with multiple instruments $z_i = (z_{i1}, \ldots, z_{iI})$ that leads to the 2SLS estimator of β^L in (2) also converges in probability to a weighted average, and the weights can also be consistently estimated from the data. For this result to hold, it is necessary to have sufficient variation in the instruments conditional on $\mathbf{x}_i$. That is, consider the linear projection of each instrument z_{il} on $\mathbf{x}_i$ given by $z_{il} = \mathbf{x}_i'\delta_{zl} + \zeta_{il}$, where $\delta_{zl} = \{E(\mathbf{x}_i\mathbf{x}_i')\}^{-1}E(\mathbf{x}_i z_{il})$ by construction and $E(\mathbf{x}_i\zeta_{il}) = 0$ for $l = 1, 2, \ldots, I$. Let $\zeta_i = (\zeta_{i1}, \ldots, \zeta_{iI})'$. Then, for the result to hold, it is necessary that the following assumption holds:

Assumption 2. *The covariance matrix for z_i after controlling for $\mathbf{x}_i$, $E(\zeta_i\zeta_i')$, is full rank.*

Thus Lochner and Moretti (2015) show that if assumptions 1 and 2 hold,

$$\widehat{\beta}_{\mathrm{2SLS}}^L \xrightarrow{p} \sum_{j=1}^{S} \omega_j^{\mathrm{2SLS}}\beta_j \tag{5}$$

where the weights ω_j^{2SLS} sum to 1 over all $j = 1, 2, \ldots, S$. The overall intuition is similar to the case of a single instrument, but the weights formulation is more cumbersome, so we leave it to the reader to verify these in the article (see proposition 2 of Lochner and Moretti [2015, 391]). Nonetheless, the 2SLS weights ω_j^{2SLS} can also be consistently estimated, and they are the 2SLS estimates of the coefficient on s_i using $z_i = (z_{i1}, \ldots, z_{iI})$ as instruments for s_i in the regressions in (4).

6. That is, if two levels of the IV z_i, $z_i = z_0$ and $z_i = z_1$, where $z_1 > z_0$, then $\Pr\{s_i(z_1) < s_i(z_0)\} = 0$ if s_i is weakly increasing in z_i or $\Pr\{s_i(z_1) > s_i(z_0)\} = 0$ if s_i is weakly decreasing in z_i.

7. We thank an anonymous referee for pointing out this to us.

Finally, to see why the standard Hausman test can lead to the incorrect conclusion in this framework, we consider the OLS estimation of (2). The OLS estimator of β^L is given by

$$\begin{aligned}\widehat{\beta}^L_{\text{OLS}} &= (s'\mathbf{M}_x s)^{-1} s'\mathbf{M}_x y \\ &= \sum_{j=1}^{S} \widehat{\omega}_j^{\text{OLS}} \beta_j + (\widetilde{s}'\widetilde{s})^{-1}\widetilde{s}\boldsymbol{\varepsilon}\end{aligned}$$

where $\widehat{\omega}_j^{\text{OLS}} = (\widetilde{s}'\widetilde{s})^{-1}\widetilde{s}'\mathbf{D}_j$. Then, in the absence of endogeneity—that is, if the condition $E(\boldsymbol{\varepsilon}|s_i) = 0$ holds—it can be shown that

$$\widehat{\beta}^L_{\text{OLS}} \xrightarrow{p} \sum_{j=1}^{S} \omega_j^{\text{OLS}} \beta_j \tag{6}$$

where the weights ω_j^{OLS} are nonnegative, sum to 1 over all $j = 1, 2, \ldots, S$, and can be written as

$$\omega_j^{\text{OLS}} = \frac{\Pr(s_i \geq j|\mathbf{x}_i) E(\eta_i|s_i \geq j)}{\sum_{q=1}^{S} \Pr(s_i \geq q|\mathbf{x}_i) E(\eta_i|s_i \geq q)}$$

The sample analog of the OLS weights ω_j^{OLS} is simply the OLS estimates of the coefficient on s_i in the regressions in (4), and given the exogeneity assumptions above, it can be shown that $\widehat{\omega}_j^{\text{OLS}} \xrightarrow{p} \omega_j^{\text{OLS}}$.

The standard Hausman test is based on the difference between the OLS and IV or 2SLS estimates. However, the analysis presented above illustrates that even in the absence of endogeneity, the OLS and IV or 2SLS estimators converge to different weighted averages; see (3), (5), and (6). If the OLS and IV or 2SLS weights are substantially different, this can produce different OLS and IV or 2SLS estimates, which in turn can lead to rejecting exogeneity using the standard Hausman test. The next section presents the test for exogeneity developed in Lochner and Moretti (2015) that is robust to this type of misspecification.

3 Wald test for consistency of OLS of all β_j's

Lochner and Moretti (2015) notice if s_i is exogenous, so that $E(\varepsilon_i|s_i) = 0$, then OLS estimates of $B = (\beta_1, \ldots, \beta_S)$ in the nonlinear specification (1) are consistent. Thus this result combined with the results in the probability limit of $\widehat{\beta}^L_{\text{2SLS}}$, given in (5), suggests that if one uses i) consistent estimators of the 2SLS weights, $\omega'_{\text{2SLS}} = (\omega_1^{\text{2SLS}}, \ldots, \omega_S^{\text{2SLS}})$, and ii) consistent estimators for the level-specific effects of s_i on outcome y_i, that is $B = (\beta_1, \ldots, \beta_S)$, the weighted average $\widehat{\omega}^{\text{2SLS}}\widehat{B}$ would have the same probability limit as $\widehat{\beta}^L_{\text{2SLS}}$. In other words,

$$\widehat{\beta}^L_{\text{2SLS}} - \widehat{\omega}'_{\text{2SLS}}\widehat{B} \xrightarrow{p} 0 \tag{7}$$

which will not be true in general if $E(\varepsilon_i D_{ij}) \neq 0$ for any j.

The test is derived in two steps: i) frame estimation of all parameters as a stacked generalized method of moments (GMM) problem to derive their asymptotic distribution and ii) use of the delta method to derive the asymptotic distribution of the transformation suggested by (7). To write the system GMM, we define the following vectors and matrices, $\mathbf{X}_{1i} = (D_i' \; \mathbf{x}_i')$, $\mathbf{X}_{2i} = (s_i \; \mathbf{x}_i')$, $\mathbf{Z}_{2i} = (z_i' \; \mathbf{x}_i')$, and

$$\mathbf{X}_i = \begin{pmatrix} \mathbf{X}_{1i} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_2 \otimes \mathbf{X}_{2i} \end{pmatrix}, \quad \mathbf{Z}_i = \begin{pmatrix} \mathbf{X}_{1i} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_2 \otimes \mathbf{Z}_{2i} \end{pmatrix}, \quad \mathbf{Y}_i = \begin{pmatrix} y_i \\ y_i \\ D_i \end{pmatrix}, \quad \text{and}$$

$$\mathbf{U}_i = \begin{pmatrix} \varepsilon_i \\ \nu_i \\ \psi_i \end{pmatrix}$$

where $\mathbf{I}_2$ is an $(S+1)$ identity matrix and the $\mathbf{0}$s reflect conformable matrices of 0s. We define the vector of parameters as

$$\boldsymbol{\theta} = (B', \gamma', \beta^L, \gamma^{L'}, \omega_1, \alpha_1', \ldots, \omega_S, \alpha_S')$$

which is of dimension $\{2S + 1 + (S+2)k\} \times 1$, so the system of equations is given by

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\theta} + \mathbf{U}_i$$

and the GMM estimator of $\boldsymbol{\theta}$ is defined by

$$\widehat{\theta} = \arg \min_{\theta \in \Theta} \left\{ \frac{1}{n} \sum_i \mathbf{Z}_i'(\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\theta}) \right\}' \widehat{\boldsymbol{\Omega}} \left\{ \frac{1}{n} \sum_i \mathbf{Z}_i'(\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\theta}) \right\}$$

where the weighting matrix $\widehat{\boldsymbol{\Omega}} = [1/n \sum_i \mathbf{Z}_i' \mathbf{Z}_i] \xrightarrow{p} E(\mathbf{Z}_i' \mathbf{Z}_i)^{-1} = \Omega$, which is a symmetric and positive-definite matrix. Then, if assumptions 1 and 2 hold,

$$\sqrt{n} \left(\widehat{\theta} - \boldsymbol{\theta} \right) \xrightarrow{d} N(0, V)$$

and it is straightforward to show that

$$\begin{aligned} \mathbf{V} = & \{E(\mathbf{X}_i'\mathbf{Z}_i)\Omega E(\mathbf{Z}_i'\mathbf{X}_i)\}^{-1} \; E(\mathbf{X}_i'\mathbf{Z}_i) \, \Omega \, E(\mathbf{Z}_i'\mathbf{U}_i\mathbf{U}_i'\mathbf{Z}_i) \, \Omega \, E(\mathbf{Z}_i'\mathbf{X}_i) \\ & \{E(\mathbf{X}_i'\mathbf{Z}_i)\Omega E(\mathbf{Z}_i'\mathbf{X}_i)\}^{-1} \end{aligned}$$

or, more compactly,

$$\mathbf{V} = \mathbf{A} \boldsymbol{\Lambda} \mathbf{A}'$$

where

$$\begin{aligned} \mathbf{A} &= \{E(\mathbf{X}_i'\mathbf{Z}_i)\Omega E(\mathbf{Z}_i'\mathbf{X}_i)\}^{-1} \; E(\mathbf{X}_i'\mathbf{Z}_i) \, \Omega \\ \boldsymbol{\Lambda} &= E(\mathbf{Z}_i'\mathbf{U}_i\mathbf{U}_i'\mathbf{Z}_i) \end{aligned}$$

and the sample analogs of A and Λ are given by

$$\widehat{\mathbf{A}} = \begin{pmatrix} [\mathbf{X}_1'\mathbf{X}_1]^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_2 \otimes \{\mathbf{X}_2'\mathbf{Z}_2[\mathbf{Z}_2'\mathbf{Z}_2]^{-1}\mathbf{Z}_2'\mathbf{X}_2\}^{-1}\mathbf{X}_2'\mathbf{Z}_2[\mathbf{Z}_2'\mathbf{Z}_2]^{-1} \end{pmatrix} \tag{8}$$

and

$$\widehat{\mathbf{\Lambda}} = \frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix} \widehat{\varepsilon}_i^2(\mathbf{X}_{1i}'\mathbf{X}_{1i}) & \widehat{\varepsilon}_i\widehat{\nu}_i(\mathbf{X}_{1i}'\mathbf{Z}_{2i}) & \widehat{\varepsilon}_i\widehat{\Psi}_i'\otimes(\mathbf{X}_{1i}'\mathbf{Z}_{2i}) \\ \widehat{\varepsilon}_i\widehat{\nu}_i(\mathbf{Z}_{2i}'\mathbf{X}_{1i}) & \widehat{\nu}_i^2(\mathbf{Z}_{2i}'\mathbf{Z}_{2i}) & \widehat{\nu}_i\widehat{\Psi}_i'\otimes(\mathbf{Z}_{2i}'\mathbf{Z}_{2i}) \\ \widehat{\varepsilon}_i\widehat{\Psi}_i\otimes(\mathbf{Z}_{2i}'\mathbf{X}_{1i}) & \widehat{\nu}_i\widehat{\Psi}_i\otimes(\mathbf{Z}_{2i}'\mathbf{Z}_{2i}) & \widehat{\Psi}_i\widehat{\Psi}_i'\otimes(\mathbf{Z}_{2i}'\mathbf{Z}_{2i}) \end{pmatrix} \tag{9}$$

where $\widehat{\varepsilon}_i = y_i - D_i'\widehat{B} - \mathbf{x}_i'\widehat{\gamma}$, $\widehat{\nu}_i = y_i - s_i\widehat{\beta}_{2\text{SLS}}^L - \mathbf{x}_i'\widehat{\gamma}^L$, and $\widehat{\Psi}_i = D_i - s_i\widehat{\omega}^{2\text{SLS}} - \widehat{\alpha}'\mathbf{x}_i$.

Next, we consider the transformation $T(\cdot):\mathbb{R}^{(2S+1+(S+2)k)}\to\mathbb{R}^1$, defined by $T(\boldsymbol{\theta}) = (\beta^L - \omega_{2\text{SLS}}'B)$, which has continuous first derivatives given by

$$\mathbf{G} = \{-\omega', \mathbf{0}_x', 1, \mathbf{0}_x', (-\beta_1, \mathbf{0}_x'), \ldots (-\beta_s, \mathbf{0}_x')\}$$

where $\mathbf{0}_x$ is a $k\times 1$ vector of 0s, so that the dimension of the Jacobian vector $\mathbf{G}$ is equal to $1\times\{2S+1+(S+2)k\}$. Then, using (7) and the delta method, we see that

$$\sqrt{n}\left\{T\left(\widehat{\theta}\right) - T(\boldsymbol{\theta})\right\} \xrightarrow{d} N(0, \mathbf{GVG}')$$

A consistent estimator of the asymptotic variance of $T(\widehat{\theta})$ is given by $\widehat{\mathbf{G}}\widehat{\mathbf{V}}\widehat{\mathbf{G}}'$, where

$$\widehat{\mathbf{G}} = \left\{-\widehat{\omega}', \mathbf{0}_x', 1, \mathbf{0}_x', \left(-\widehat{\beta}_1, \mathbf{0}_x'\right), \ldots \left(-\widehat{\beta}_s, \mathbf{0}_x'\right)\right\}$$

and $\widehat{\mathbf{V}} = \widehat{\mathbf{A}}\widehat{\mathbf{\Lambda}}\widehat{\mathbf{A}}'$ is given by (8) and (9).

The main result of Lochner and Moretti (2015) is summarized in the following theorem:

Theorem 3. *Under assumptions 1 and 2, if* $E(\varepsilon_i|s_i) = 0$, *then*

$$W_n = n\left\{\frac{\left(\widehat{\beta}_{2\text{SLS}}^L - \widehat{\omega}_{2\text{SLS}}'\widehat{B}\right)^2}{\widehat{\mathbf{G}}\widehat{\mathbf{V}}\widehat{\mathbf{G}}'}\right\} \xrightarrow{d} \chi^2(1)$$

As Lochner and Moretti (2015) note, this Wald test represents only a partial solution to the problem of estimating multiple per-unit treatment effects with limited instruments. This is because if the test fails to reject exogeneity, then researchers can gain some confidence in the OLS estimates of $B = (\beta_1, \ldots, \beta_S)$ in (1); but if the test rejects exogeneity, it does not provide any help in fitting the true model.

Lochner and Moretti (2015) also note that failure to reject exogeneity, which suggests that $T(\widehat{\theta}) \xrightarrow{p} 0$, may not imply that the OLS estimates of $B = (\beta_1, \ldots, \beta_S)$ are consistent for two reasons. First, this test cannot tell anything about whether $\widehat{\beta}_j \xrightarrow{p} \beta_j$ for some j if $\omega_j = 0$; however, the rest of the OLS estimates would be consistent. Second, $\widehat{\beta}_j$ may be upward biased for some j and downward biased for others; however, if $\omega_j > 0$ for all j and $E(\varepsilon_j D_{ij})$ are either all nonnegative or all nonpositive, then all $\widehat{\beta}_j$ are biased in the same direction, so failure to reject exogeneity implies that $\widehat{B}$ are consistent.

Monte Carlo simulations performed by Lochner and Moretti (2015) show that for 1,000 observations, the Wald test proposed by Lochner and Moretti (2015) performs similarly to the Durbin–Wu–Hausman (DWH) test when s_i is exogenous and the true specification is given by (2), both rejecting 5% of the time at a significance level of 0.05. If the true relation between y_i and s_i is nonlinear, the LM test continues to reject 5% of the time, while the DWH rejects at an increasing rate as the degree of nonlinearity increases (see online appendix B of Lochner and Moretti [2015]). The three numerical examples provided in section 5 have sample sizes larger than 1,000 observations, so the asymptotic results should work well.

4 The locmtest command

4.1 Syntax

`locmtest` *depvar* `(`*varlist1* `=` *varlist_iv*`)` [*indepvars*] [*if*] [`,` **graph** **coefficients**]

where *depvar* is the dependent variable to be used, *varlist1* is the discrete endogenous variable, *varlist_iv* is the set of instruments, and *indepvars* is a list of exogenous variables. While this command permits factor variables in *indepvars* (see [U] **11.4.3 Factor variables**), it does not permit factor variables in *varlist_iv*.

4.2 Options

graph displays a graph of the estimated level-specific OLS coefficients from (1), $B = (\beta_1, \dots, \beta_S)$, the estimated OLS weights, $\widehat{\omega}'_{\text{OLS}} = (\widehat{\omega}_1^{\text{OLS}}, \dots, \widehat{\omega}_S^{\text{OLS}})$, and the estimated 2SLS weights, $\widehat{\omega}'_{\text{2SLS}} = (\widehat{\omega}_1^{\text{2SLS}}, \dots, \widehat{\omega}_S^{\text{2SLS}})$.

coefficients displays a matrix of the estimated level-specific OLS coefficients from (1), $B = (\beta_1, \dots, \beta_S)$, the estimated OLS weights, $\widehat{\omega}'_{\text{OLS}} = (\widehat{\omega}_1^{\text{OLS}}, \dots, \widehat{\omega}_S^{\text{OLS}})$, the estimated 2SLS weights, $\widehat{\omega}'_{\text{2SLS}} = (\widehat{\omega}_1^{\text{2SLS}}, \dots, \widehat{\omega}_S^{\text{2SLS}})$, and their standard errors.

4.3 Stored results

`locmtest` stores the following in `e()`:

Scalars

`e(BLols)`	OLS coefficient on the endogenous regressor from the linear equation
`e(SDBLols)`	standard error of OLS coefficient on the endogenous regressor from the linear equation
`e(BLiv)`	2SLS coefficient on the endogenous regressor from the linear equation
`e(SDBLiv)`	standard error of 2SLS coefficient on the endogenous regressor from the linear equation
`e(DIVOLS)`	difference between `e(BLiv)` and `e(BLols)`
`e(SDDIVOLS)`	standard error of difference between `e(BLiv)` and `e(BLols)`
`e(WBols)`	reweighted OLS using 2SLS weights
`e(SDWBols)`	standard error of reweighted OLS
`e(T)`	difference between `e(BLiv)` and `e(WBols)`
`e(SDT)`	standard error of difference between `e(BLiv)` and `e(WBols)`
`e(wm)`	LM test statistic
`e(pwm)`	p-value of LM test statistic
`e(nw)`	naïve Wald test statistic
`e(pnw)`	p-value of naïve Wald test statistic
`e(dwh)`	DWH test statistic
`e(pdwh)`	p-value of DWH test statistic

Matrices

`e(B)`	OLS coefficients on dummies in the nonlinear equation
`e(VB)`	standard errors of OLS coefficients on dummies in the nonlinear equation
`e(Wols)`	OLS weights
`e(VWols)`	standard errors of OLS weights
`e(W)`	2SLS weights
`e(VW)`	standard errors of 2SLS weights

5 Implementing the LM test

5.1 Example 1

To demonstrate this test, we first use data from Card (1995a), who estimates the effect of education on earnings using the following specification,

$$\ln w_i = s_i \beta^L + \mathbf{x}_i' \gamma^L + \nu_i \tag{10}$$

where $\ln w_i$ is the logarithm of the hourly earnings, $s_i \in \{1, 2, \ldots, 18\}$ is years of education, and $\mathbf{x}_i$ is a vector of covariates including the constant. The concern is that in this case, s_i is endogenous because there are unobserved individual characteristics that determine both earnings and the years of education, for example, an individual's unobserved ability. Card (1995a) uses an IV approach to estimate β^L, and the instrument used, z_i, is an indicator variable equal to one if the individual grew up in a local labor market with a four-year college and zero otherwise.

In this case, the monotonicity condition would require that an individual who grew up in a local labor market with a four-year college would get at least as much education as that individual would have gotten if he or she had grown up in a local labor market without a four-year college. However, remember that this condition is not necessary for the validity of the results that we present below.

Now, suppose that in the population, the relationship between $\ln w_i$ and s_i is nonlinear because of sheepskin effects and can be written as

$$\ln w_i = \sum_{j=2}^{18} D_{ij}\beta_j + \mathbf{x}_i'\gamma + \varepsilon_i$$

Then, even if s_i is exogenous so that $E(\varepsilon_i|s_i) = 0$, we could obtain different estimates from OLS and IV using (10), which can lead to the incorrect conclusion if we use only the standard Hausman test—hence, the relevance of the LM test.

To implement the LM test, we have $\ln w =$ `lwage`, $s =$ `educ`, and $z =$ `near4`. The covariates include only the constant, labor market experience, and its square. Applying the test yields the following results:

```
. use http://www.stata.com/data/jwooldridge/eacsap/card
. locmtest lwage (educ = nearc4) exper expersq

 ===================================================
  Output for the Lochner & Moretti (2015) Wald Test
 ===================================================

 Output Variable y: lwage
 Endogenous Variable s: educ
 Instruments z: nearc4

 Number of observations = 3010
 Number of Categories of Endogenous variable is: 18
 Number of Dummies is: 17

 The number of Excluded Instruments is: 1

 Estimated Coefficients

             |      Coef.    Std. Err.
-------------+--------------------------
         OLS |   .09317071   .00357785
          IV |   .25871555   .03373941
       RWOLS |   .09072257   .00573885
-------------+--------------------------

 Estimated Test Statistics
             |      Test      p-value
-------------+--------------------------
     LM-Wald |   24.196549   8.699e-07
  Naive Wald |   30.124769   4.051e-08
    DWH Test |   41.823869   1.162e-10
-------------+--------------------------

 NOTES:

 RWOLS = Reweighted OLS using TSLS Weights

 LM-Wald = Lochner-Moretti Wald Test

 Naive Wald = [ (IV - OLS) / SD(IV-OLS) ]^2

 DWH Test: Durbin-Wu-Hausman Test (Augmented Regression)
```

Note that the OLS and IV estimates of β^L from (10) are different. In particular, the IV estimate is larger than the OLS estimate, contrary to the common idea that the endogeneity of schooling would overestimate the effect of education in OLS. While there are several explanations to this result, such as measurement error and individual heterogeneity in the effects of schooling (see Card [1995b]), nonlinearity in the earnings–schooling relationship may play a role. This seems particularly relevant given the possible sheepskin effects. In this framework, both the naïve Wald and DWH tests reject exogeneity. The LM test also rejects exogeneity, which lowers the concerns that the conclusions from the former tests are due to misspecification.

5.2 Example 2

Next, we use one of the examples of Lochner and Moretti (2015). This example is based on Lochner and Moretti (2004), who study the effect of education on crime. The empirical specification they estimate is given by

$$p_i = s_i\beta^L + \mathbf{x}_i'\gamma^L + \nu_i \tag{11}$$

where p_i is an indicator equal to 1 if the respondent is in prison and 0 otherwise, $s_i \in \{0, 1, 2, \ldots, 18\}$ is years of education, and $\mathbf{x}_i$ is a vector of covariates including the constant. The concern here is that years of education, s_i, may be endogenous because there are unobserved factors (for example, patience) that determine an individual's educational attainment and that may also determine that individual's propensity to commit crime.

Lochner and Moretti (2004) propose using as instruments for s_i the compulsory attendance laws in the state of birth of individuals when they were age 14. This instrument was previously used in Acemoglu and Angrist (2001).[8] The compulsory attendance laws are then summarized in four indicator variables: `ca8`$_i$, `ca9`$_i$, `ca10`$_i$, and `ca11`$_i$.[9] Because the four dummies represent collectively exhaustive events, `ca8` is an omitted category for the instrument.

For this example, the monotonicity condition would require that a given individual born in a state with tough compulsory attendance laws when he or she was age 14 would have at least as much education as he or she would if that individual had been born in a state with weak compulsory attendance laws when he or she was age 14.[10]

8. See Acemoglu and Angrist (2001) for a detailed explanation of the instruments.
9. These indicator variables are the following: i) `ca8` equals 1 if the compulsory attendance law is 8 or less years of schooling and 0 otherwise; ii) `ca9` equals 1 if the compulsory attendance law is 9 years of schooling and 0 otherwise; iii) `ca10` equals 1 if the compulsory attendance law is 10 years of schooling and 0 otherwise; and iv) `ca11` equals 1 if the compulsory attendance law is 11 or more years of schooling and 0 otherwise.
10. Note that in this case, the instrument has no effect on individuals already intending to stay in school longer than the compulsory schooling age; it has an effect only on those who intended to leave before the compulsory schooling age (see Oreopoulos [2006]). These individuals are commonly referred to as the "compliers" (see Angrist, Imbens, and Rubin [1996]). Compulsory schooling age is associated with finishing high school; thus the monotonicity assumption implies that the effect of this instrument not only increases the fraction of high school graduates but also increases the fraction of college graduates. We thank an anonymous referee for providing us with this example.

Once again, suppose that in the population, the incarceration–schooling relationship is given by

$$p_i = \sum_{j=1}^{18} D_{ij}\beta_j + \mathbf{x}_i'\gamma + \varepsilon_i \tag{12}$$

For this example, we focus on the analysis that Lochner and Moretti (2004) do on black males; the next example focuses on white males. The data were obtained from Moretti's webpage, and we use the same covariates as described in their article. To implement the LM test, we have p = `prison`, s = `educ`, and z = `ca9`, `ca10`, `ca11`. The covariates include a constant and dummies for age groups, year, state of residence, and state of birth. For black males, the covariates also include state-of-birth dummies interacted with a dummy for black men born in the South who turned age 14 in 1958 or later to account for the impact of *Brown v. Board of Education*. Applying the test yields the following results:

(*output omitted*)

```
. locmtest prison (educ = ca9 ca10 ca11) i.rage i.year i.state i.birthpl
> i.birthpl#i.BBeduc

 ====================================================
  Output for the Lochner & Moretti (2015) Wald Test
 ====================================================

 Output Variable y: prison
 Endogenous Variable s: educ
 Instruments z: ca9 ca10 ca11

 Number of observations = 401529
 Number of Categories of Endogenous variable is: 19
 Number of Dummies is: 18

 The number of Excluded Instruments is: 3

 Estimated Coefficients

             |      Coef.    Std. Err.
-------------+--------------------------
         OLS |  -.00369034   .00008333
          IV |   -.0047513   .00115743
       RWOLS |  -.00073792   .00017873
-------------+--------------------------

 Estimated Test Statistics
             |       Test     p-value
-------------+--------------------------
     LM-Wald |   11.944147   .00054819
  Naive Wald |   .97566386   .32327168
    DWH Test |   .51540357   .47280942
-------------+--------------------------
```

```
NOTES:

RWOLS = Reweighted OLS using TSLS Weights

LM-Wald = Lochner-Moretti Wald Test

Naive Wald = [ (IV - OLS) / SD(IV-OLS) ]^2

DWH Test: Durbin-Wu-Hausman Test (Augmented Regression)
```

Note that the OLS and IV estimates of $\widehat{\beta}^L$ from (11) are different. In particular, the IV estimate is larger in magnitude than the OLS estimate. Both the naïve Wald and DWH tests fail to reject exogeneity, but the LM test rejects it. The reweighted OLS reveals that on average, the OLS estimates of β_j from (12) are significantly biased toward zero because the reweighted OLS is substantially smaller in magnitude than the 2SLS estimate of β^L from (11). This is an example in which differences in level-specific effects may lead the standard Hausman test to fail to reject exogeneity when it should be rejected.

5.3 Example 3

Now, we implement the LM test for the analysis of Lochner and Moretti (2004) described in example 2 but focus on white males. All the variables have the same definition; however, for the sample of white males, the authors do not account for the effect of *Brown v. Board of Education*. The linear and nonlinear specifications are once again described by (11) and (12), respectively, and the monotonicity condition is the same as for black males. The implementation of the test using the `graph` and `coefficients` options is as follows:

(output omitted)

```
. locmtest prison (educ = ca9 ca10 ca11) i.rage i.year i.state i.birthpl,
> graph coefficients

 ===================================================
  Output for the Lochner & Moretti (2015) Wald Test
 ===================================================

 Output Variable y: prison
 Endogenous Variable s: educ
 Instruments z: ca9 ca10 ca11

 Number of observations = 3209138
 Number of Categories of Endogenous variable is: 19
 Number of Dummies is: 18

 The number of Excluded Instruments is: 3
```

```
Estimated Coefficients

             |      Coef.    Std. Err.
-------------+------------------------
         OLS |  -.00099111   .00001191
          IV |  -.00114869   .00036243
       RWOLS |  -.00120313     .000034
-------------+------------------------

Estimated Test Statistics
             |      Test     p-value
-------------+------------------------
     LM-Wald |   .02247255   .88083682
  Naive Wald |   .20211212   .65302138
    DWH Test |   .16365057   .68581755
-------------+------------------------

NOTES:

RWOLS = Reweighted OLS using TSLS Weights

LM-Wald = Lochner-Moretti Wald Test

Naive Wald = [ (IV - OLS) / SD(IV-OLS) ]^2

DWH Test: Durbin-Wu-Hausman Test (Augmented Regression)

Estimated Coefficients:
             |         B        seB      W2SLS    seW2sls       Wols     seWols
-------------+------------------------------------------------------------------
           1 |  -.0004088   .0011741   .0037614   .0004353   .0072537   .0000132
           2 |   .0045219   .0012997   .0061937     .00047   .0085221   .0000143
           3 |   -.001657   .0009236     .01063   .0005368   .0109205   .0000164
           4 |  -.0010993   .0007393   .0224401   .0006503   .0145451   .0000194
           5 |   .0012797   .0006386   .0393125   .0008261   .0187105   .0000225
           6 |  -.0001663   .0005139   .0590797   .0010652   .0236794   .0000259
           7 |   .0000373   .0003851   .0833742   .0013781   .0316599   .0000307
           8 |  -.0020164    .000276   .1191036   .0018339   .0426187   .0000363
           9 |   .0013537    .000223   .1539329   .0021541   .0668838   .0000444
          10 |  -.0023834   .0002292   .1452373   .0019494   .0809498   .0000476
          11 |   -.001019   .0002225   .1335972   .0018293   .0928632   .0000511
          12 |  -.0046185   .0001729    .151474   .0020407   .0986295    .000055
          13 |  -.0014042   .0001508   .0016496   .0028537   .1147939   .0000602
          14 |  -.0004354   .0001853   .0166923    .002551   .1120058    .000057
          15 |  -.0014717   .0002178   .0173397   .0023118   .0993689   .0000538
          16 |   .0007271   .0002121   .0225779   .0020952   .0889882   .0000522
          17 |  -.0003403     .00023   .0088103    .001667   .0512865   .0000451
          18 |   .0009268   .0002471   .0047934   .0014149   .0363205   .0000395
```

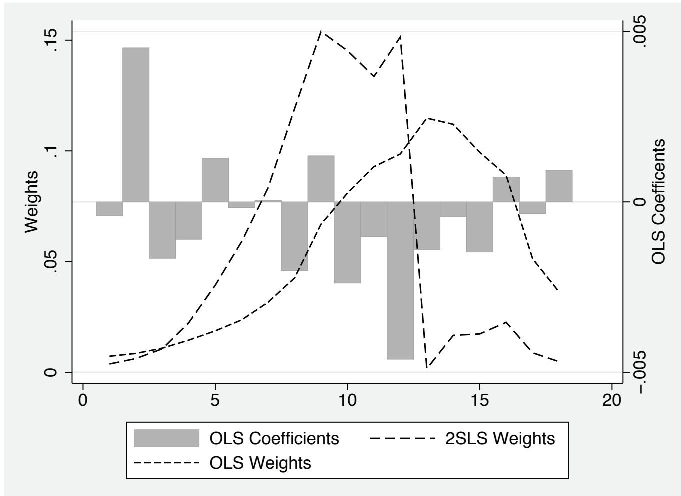

Figure 1. Effect of schooling on the probability of incarceration for white males

The results indicate that the 2SLS estimate is 10% larger in magnitude than the OLS estimate. In this framework, both the naïve Wald and DWH tests fail to reject exogeneity of s_i. That the LM test also fails to reject exogeneity of s_i lowers the concern that the conclusion of the former tests are due to nonlinearities in the incarceration–schooling relationship, which in the previous example proved to be important in the case of black males.

Because both DWH and the LM tests fail to reject exogeneity of s_i, this suggests that the OLS estimates of the β_j from (12) are consistent. These estimates are depicted in figure 1 and in the matrix of estimated coefficients from the Stata output. Both the matrix and figure 1 also present the estimates for the OLS and 2SLS weights. Figure 1 shows that the OLS weights are high for years of schooling between 12 and 16, while the 2SLS are high between 9 and 12. This implies that the effect of the transitions between 9 and 12 years of schooling have a substantial effect on the 2SLS estimate, which could partly explain the higher magnitude of the 2SLS compared with the OLS estimate.

6 Conclusions

We provided a command, `locmtest`, that implements a test for exogeneity recently proposed by Lochner and Moretti (2015). This test is robust to nonlinearities in the relationship between the potentially endogenous variable and the outcome variable. The assumptions necessary for the main results are very standard in the literature on IV estimation. We provided detailed description of the test and presented three examples to illustrate the implementation of `locmtest` and how the results can be used to test exogeneity. Additionally, if exogeneity is not rejected, we can use the results from `locmtest` to give a partial explanation of the differences between the OLS and 2SLS estimates.

7 Acknowledgments

We would like to thank Joe Newton and an anonymous referee for the valuable comments and suggestions which greatly improved this article.

8 References

Acemoglu, D., and J. Angrist. 2001. How large are human-capital externalities? Evidence from compulsory-schooling laws. In *NBER Macroeconomics Annual 2000*, ed. B. S. Bernanke and K. Rogoff, 9–74. Cambridge, MA: National Bureau of Economic Research.

Aliprantis, D. 2012. Redshirting, compulsory schooling laws, and educational attainment. *Journal of Educational and Behavioral Statistics* 37: 316–338.

Angrist, J. D., and G. W. Imbens. 1995. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association* 90: 431–442.

Angrist, J. D., G. W. Imbens, and D. B. Rubin. 1996. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 91: 444–455.

Barua, R., and K. Lang. 2009. School entry, educational attainment and quarter of birth: A cautionary tale of LATE. NBER Working Paper No. 15236, The National Bureau of Economic Research. http://www.nber.org/papers/w15236.

Card, D. E. 1995a. Using geographic variation in college proximity to estimate the return to schooling. In *Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp*, ed. L. N. Christofides, E. K. Grant, and R. Swidinsky, 201–222. Toronto, Canada: University of Toronto Press.

———. 1995b. Earnings, schooling, and ability revisited. *Research in Labor Economics* 14: 23–48.

de Chaisemartin, C. 2014. Tolerating defiance? Local average treatment effects without monotonicity. Working Paper 197, University of Warwick. http://www2.warwick.ac.uk/fac/soc/economics/research/centres/cage/manage/publications/197-2014_chaisemartin.pdf.

Heckman, J. J., S. Urzua, and E. J. Vytlacil. 2006. Understanding instrumental variables in models with essential heterogeneity. *Review of Economics and Statistics* 88: 389–432.

Imbens, G. W., and J. D. Angrist. 1994. Identification and estimation of local average treatment effects. *Econometrica* 62: 467–475.

Klein, T. J. 2010. Heterogeneous treatment effects: Instrumental variables without monotonicity? *Journal of Econometrics* 155: 99–116.

Lochner, L., and E. Moretti. 2004. The effect of education on crime: Evidence from prison inmates, arrests, and self-reports. *American Economic Review* 94: 155–189.

———. 2015. Estimating and testing models with many treatment levels and limited instruments. *Review of Economics and Statistics* 97: 387–397.

Oreopoulos, P. 2006. Estimating average and local average treatment effects of education when compulsory schooling laws really matter. *American Economic Review* 96: 152–175.

Yitzhaki, S. 1996. On using linear regressions in welfare economics. *Journal of Business and Economic Statistics* 14: 478–486.

About the authors

Michael P. Babington is a PhD student in the Department of Economics at Florida State University.

Javier Cano-Urbina is an assistant professor in the Department of Economics at Florida State University.