Identification and Estimation of Single Index Models with Measurement Error and Endogeneity*

Yingyao Hu, Ji-Liang Shiu, and Tiemen Woutersen †
Johns Hopkins University, Renmin University of China, and University of Arizona

May 2015

Abstract

Economic variables are often measured with an error and may be endogenous. This paper gives new identification results for the ratio of partial effects in linear index models with measurement error and endogeneity. The identification restrictions include independence of covariates and error terms, and the derivative of some conditional mean functions being nonzero. We propose a local polynomial regression estimator to estimate the single index parameters. The paper applies these tools to estimate the labor supply elasticity and finds that the labor supply elasticity for married men is positive, while the coefficients for married women are negative for the full sample and positive for the working sample.

Keywords: Nonclassical measurement error, measurement error and endogeneity, labor supply elasticity, labor force participation.

JEL codes: C20, J20, C27

*We thank Richard Blundell and Ronald Oaxaca for helpful comments and discussions and Maddie Pickens and Stephanie Barana for research assistance. We also thank a co-editor and three anonymous referees. This article is forthcoming in the Econometrics Journal.

†Correspondence address: Tiemen Woutersen, Department of Economics, Eller College of Management, University of Arizona, P.O. Box 210108, Tucson, AZ 8572; email: woutersen@email.arizona.edu.
1. Introduction

In many applied problems in economics some variables are measured with error and/or are endogenous. For example, when estimating the labor supply elasticity, it is likely that the wage and the number of hours worked are measured with errors. In this paper, we use the estimation of the labor supply elasticity as a leading example, but the estimation procedure is more general than that. We consider the models with the combination of measurement error and endogeneity. The model takes the following semiparametric form,

\[ y = m(\theta x^* + w, \eta), \]
\[ x = x^* + \varepsilon, \]
\[ x^* = g(z, w) + u, \]

where we observe \( \{y, x, z, w\} \). The model involves a dependent variable \( y \), a true regressor \( x^* \), a mismeasured regressor \( x \), a correctly measured regressor \( w \) and an instrument \( z \) while the unobservables are \( \eta, \varepsilon, \) and \( u \). The error terms \( (\eta, \varepsilon, u) \) in equations (1)-(3) are independent of \( z, w \) with the following conditional mean assumptions: \( E(u|z, w) = 0 \) and \( E(\varepsilon|z, w) = 0 \). In particular, there is no restriction about the dimension of the disturbance \( \eta \) in the regression function. The true regressor \( x^* \) is endogenous in the sense that it is determined by \( g(z, w) + u \) and \( u \) and \( \eta \) are generally correlated. The measurement error \( \varepsilon \) may be correlated with \( \eta \) so that Hu (2008) and Hu and Schennach (2008) do not apply. Horowitz (1998) gives a nice review of the single index model with exogenous regressors and the techniques reviewed by him do not apply either.

The parameter of interest is \( \theta \), which measures the relative effects for continuous variables \( x^* \) and \( w \). If \( x^* \) and \( w \) are independent of \( \eta \), then \( \theta \) can be regarded as the ratio of partial effects corresponding to \( x^* \) and \( w \) on \( y \) since \( \theta = \left( \frac{\partial m(\theta x^* + w, \eta)}{\partial x^*} \right) / \left( \frac{\partial m(\theta x^* + w, \eta)}{\partial w} \right) \). Chesher (2003), Schennach, White, and Chalak (2012) also use ratios of derivatives to identify quantities on nonseparable models.\footnote{Appendix 1 contains an example that illustrates that using a linear approximation may yield the wrong sign for the causal effect.}

In the next section, we show that the parameter \( \theta \) is identified. One of the many models that fits this framework is a new model for labor supply that allows for measurement error and
endogeneity. Let $y$ be log hours and let $x^*$ be log wage. Then, the parameter $\theta$ represents the ratio of labor supply elasticity over the partial effect of $w$ on $y$. We present and estimate this model to illustrate our methodology but also view this elasticity as an important parameter for policy analysis. Also, we allow the wage to be measured with error for those who work and to be unobserved for those who do not work. Thus, we allow labor force participation to change as the wage changes. In general, we allow Eq. (2) to be replaced by the requirement that the function $g(z, w)$ in Eq. (3) is identified.

If we ignore measurement error in Eq. (2), i.e., $x = x^*$, then the proposed model becomes a single index control function model which is closely related to Imbens and Newey (2009). They use the conditional cumulative distribution function of the endogenous variable given the instruments as a control variable to identify and estimate nonseparable models. Another related paper is Schennach (2007) which provides a closed form solution for the identification of nonparametric regression models in the presence of measurement error. Her approach for nonlinear error-in-variables models can handle the case where the true regressor is continuously distributed. The proposed model in Eq. (1)-(3) differs in two respects from the model considered by Schennach (2007). One is that the true regressor $x^*$ may be correlated with the disturbance $\eta$ and the other is that the measurement error $\varepsilon$ may also be correlated with $\eta$. These changes allow the method to be used for economic models with endogeneity.

The estimation of the labor supply elasticity has been of considerable interest in labor economics. For example, Ashenfelter and Heckman (1974), Killingsworth and Heckman (1986), Pencavel (1986), Blundell, Duncan, Meghir (1998), and Pencavel (2002) provide estimates, and Blundell and MaCurdy (1999) give a general review. Card and Hyslop (2005) estimate a related quantity, the effect of a subsidy for people leaving welfare. Borjas (2009) gives an overview of empirical studies that estimate the labor supply elasticity and also discusses the problems caused by measurement error. Keane (2011) reviews the literature on the effects of taxes on labor supply. Our paper introduces a new methodology that allows for measurement error. We find that the labor supply elasticity for married men is positive, while the coefficients for married women are negative for the full sample and positive for the working sample.

---

3There may exist a correlation between $x^*$ and $\eta$ in Eq. (1) and the nonseparable term $g(z, w)$ can be regarded as a control function too because $g(z, w)$ is independence of $\eta$ and a correlation between $x^*$ and $\eta$ should come from a correlation between $u$ and $\eta$. 

3
This paper is organized as follows. We discuss identification and estimation of the model in Section 2. We present a simulation study in Section 3. We then show how these tools can be used to determine whether the labor supply elasticity is positive or negative in Section 4. We conclude in Section 5.

2. Single Index Models

In this section, we first introduce conditions for the identification of $\theta$ and establish the result. Then, we will extend the result of the identification in a variety of useful directions.

Assumption 2.1. The observables $z, w$ are independent of the error terms $(u, \eta, \varepsilon)$ in Eq. (1)-(3). The error terms $u$ and $\varepsilon$ satisfy the conditional mean restrictions $E(u|z, w) = 0$ and $E(\varepsilon|z, w) = 0$, respectively.

By Eq. (3), taking the expectation of $x$ conditional on $z$ and $w$ yields

$$E[x|z, w] = E[x^* + \varepsilon|z, w]$$

$$= E[g(z, w) + u + \varepsilon|z, w]$$

$$= g(z, w).$$

Also, by independence between $u$ and $z, w$ we have

$$E[y|z, w] = E[m(\theta g(z, w) + \theta u + w, \eta)|z, w]$$

$$= \int m(\theta g(z, w) + \theta u + w, \eta)f_{u, \eta}(u, \eta)dud\eta.$$
we have

\[ \frac{\partial E[y|z,w]}{\partial w} = \int m'(\theta g(z, w) + w + \theta u, \eta) \{ \theta \frac{\partial}{\partial w} g(z, w) + 1 \} f_{u,\eta}(u, \eta) dud\eta \]

\[ = \left[ \theta \frac{\partial}{\partial w} g(z, w) + 1 \right] \int m'(\theta g(z, w) + w + \theta u, \eta) f_{u,\eta}(u, \eta) dud\eta, \]

and

\[ \frac{\partial E[y|z,w]}{\partial z} = \int m'(\theta g(z, w) + w + \theta u, \eta) \frac{\partial}{\partial z} g(z, w) f_{u,\eta}(u, \eta) dud\eta \]

\[ = \theta \frac{\partial}{\partial z} g(z, w) \int m'(\theta g(z, w) + w + \theta u, \eta) f_{u,\eta}(u, \eta) dud\eta. \]

Here we use the notation \( m'(\cdot, \cdot) \) to denote the partial derivative of the regression function in the first argument.

**Assumption 2.3.** There exists a set \( Z \) in the support of \((z, w)\) such that \( \Pr(Z) > 0 \) and for any \((z, w) \in Z, \frac{\partial}{\partial z} E[y|z,w] \neq 0 \) and

\[ \frac{\partial E[y|z,w]}{\partial w} \frac{\partial E[x|z,w]}{\partial z} - \frac{\partial E[y|z,w]}{\partial z} \frac{\partial E[x|z,w]}{\partial w} \neq 0. \]

Because \( \frac{\partial}{\partial z} E[y|z,w] \neq 0 \) in Assumption 2.3 we can divide Eq. (4) by Eq. (5) to eliminate \( \int m'(\theta g(z, w) + w + \theta u, \eta) f_{u,\eta}(u, \eta) dud\eta \). This yields

\[ \frac{\frac{\partial}{\partial w} E[y|z,w]}{\frac{\partial}{\partial z} E[y|z,w]} = \frac{\theta \frac{\partial}{\partial z} g(z, w) + 1}{\theta \frac{\partial}{\partial z} g(z, w)} \]

\[ = \frac{\theta \frac{\partial}{\partial w} E[x|z,w] + 1}{\theta \frac{\partial}{\partial z} E[x|z,w]} . \]

It follows that

\[ \theta = \frac{\frac{\partial}{\partial w} E[y|z,w] \frac{\partial}{\partial z} E[x|z,w] - \frac{\partial}{\partial z} E[y|z,w] \frac{\partial}{\partial w} E[x|z,w]}{\frac{\partial}{\partial w} E[y|z,w]} \frac{\partial}{\partial z} E[y|z,w] \frac{\partial}{\partial w} E[x|z,w], \]

where the denominator is nonzero by Assumption 2.3.

We can easily adjust our method to the following model with a vector of multiple regressors
\[ W = (w_1, \tilde{W}) : \]

\( y = m(\theta x^* + w_1 + h(\tilde{W}), \eta), \)  

\( x = x^* + \varepsilon, \)

\( x^* = g(z, W) + u. \)

Following the similar manner in the previous derivation, we can obtain

\[ \theta = \frac{\frac{\partial}{\partial w_1} E[y|z, W]}{\frac{\partial}{\partial w_1} E[y|z, W] \frac{\partial}{\partial x} E[x|z, W] - \frac{\partial}{\partial x} E[y|z, W] \frac{\partial}{\partial w_1} E[x|z, W]}. \]

The last equality holds for every \((z, w) \in \mathbb{R}^2\). In order to estimate \(\theta\) one can use a weighted average over \((z, w)\) using for example the density of \((z, w)\) as the weighting function.

### 2.1. Multivariate case

Note that the method developed in this section can be applied to a more general case containing more than two mismeasured independent variables.

**Assumption 2.4.** Assume \(E[\varepsilon|z, \tilde{z}], \) and \(E[u|z, \tilde{z}]\) are all equal to zero and the vector of unobservables \((\eta, u_1, u_2, \varepsilon_1, \varepsilon_2)\) is independent of \((z, \tilde{z})\).

**Assumption 2.5.** An i.i.d. random sample \(\{y, x, w, z, \tilde{z}\}\) satisfies

\[ y = m(\theta x^* + w^*, \eta), \]

\[ x = x^* + \varepsilon_1, \]

\[ x^* = g_1(z, \tilde{z}) + u_1, \]

\[ w = w^* + \varepsilon_2, \]

\[ w^* = g_2(z, \tilde{z}) + u_2, \]

where \((x, w)\) are the measurement of the latent explanatory variables \((x^*, w^*)\), \((z, \tilde{z})\) are instruments, the functions \(m(\cdot, \cdot), g_1(\cdot, \cdot), g_2(\cdot, \cdot)\) are unknown, \(\theta\) is the unknown parameter of interest, and the unobservables \((\eta, \varepsilon_1, \varepsilon_2, u_1, u_2)\) are jointly independent of \((z, \tilde{z})\). The errors \((\eta, u_1, u_2)\) are normalized to have zero expectation.
In Assumption 2.5, we can replace \( x = x^* + \varepsilon_1 \) by the assumption that \( g_1(z, \tilde{z}) \) is identified and replace \( w = w^* + \varepsilon_2 \) by the assumption that \( g_2(z, \tilde{z}) \) is identified. This allows us to deal with the fact that for some the wage is measured with error while for others (who do not work) the wage is not observed at all. We take an expectation of \( x \) conditional on \( z, \tilde{z} \)

\[
E[x|z, \tilde{z}] = E[x^* + \varepsilon_1 | z, \tilde{z}]
= E[g_1(z, \tilde{z}) + u_1 + \varepsilon_1 | z, \tilde{z}]
= g_1(z, \tilde{z}).
\]

Similarly,

\[
E[w|z, \tilde{z}] = g_2(z, \tilde{z}).
\]

Note that, by independence between \( u = (u_1, u_2) \) and \( z, \tilde{z} \),

\[
E[y|z, \tilde{z}] = E[m(\theta g_1(z, \tilde{z}) + g_2(z, \tilde{z}) + \theta u_1 + u_2, \eta)]|z, \tilde{z}]
= \int m(\theta g_1(z, \tilde{z}) + g_2(z, \tilde{z}) + \theta u_1 + u_2, \eta)f_{u_1,u_2,\eta}(u_1, u_2, \eta)du_1du_2d\eta.
\]

We further assume the following.

**Assumption 2.6.** The instruments \((z, \tilde{z})\) each contain a continuous element and the functions \( m(\cdot, \cdot), g_1(\cdot, \cdot), g_2(\cdot, \cdot) \) are differentiable.

Taking the derivative of the above equation with respective to \( z, \tilde{z} \), we have

\[
\frac{\partial E[y|z, \tilde{z}]}{\partial z} = \left[ \theta \frac{\partial}{\partial z} g_1(z, \tilde{z}) + \frac{\partial}{\partial z} g_2(z, \tilde{z}) \right] \int m'(\theta g_1(z, \tilde{z}) + g_2(z, \tilde{z}) + \theta u_1 + u_2, \eta) f_{u_1,u_2,\eta}(u_1, u_2, \eta)du_1du_2d\eta,
\]

and

\[
\frac{\partial E[y|z, \tilde{z}]}{\partial \tilde{z}} = \left[ \theta \frac{\partial}{\partial \tilde{z}} g_1(z, \tilde{z}) + \frac{\partial}{\partial \tilde{z}} g_2(z, \tilde{z}) \right] \int m'(\theta g_1(z, \tilde{z}) + g_2(z, \tilde{z}) + \theta u_1 + u_2, \eta) f_{u_1,u_2,\eta}(u_1, u_2, \eta)du_1du_2d\eta,
\]

**Assumption 2.7.** There exists a set \( Z \) in the support of \((z, \tilde{z})\) such that \( \Pr(Z) > 0 \) and for
any \((z, \tilde{z}) \in Z, \frac{\partial}{\partial z} E[y|z, \tilde{z}] \neq 0\) and

\[
\frac{\partial E[y|z, \tilde{z}]}{\partial \tilde{z}} \frac{\partial E[x|z, \tilde{z}]}{\partial z} - \frac{\partial E[y|z, \tilde{z}]}{\partial z} \frac{\partial E[x|z, \tilde{z}]}{\partial \tilde{z}} \neq 0.
\]

Notice that this assumption is imposed on observables and is directly testable from the sample. A condition that we impose to ensure that a denominator is nonzero is excluding the case where \(E[y|z, \tilde{z}] = aE[x|z, \tilde{z}] + b\) for some \(a, b\), which rarely happens. Dividing Eq. (9) by Eq. (10) to eliminate \(R_{m0}(g_1(z, \tilde{z}) + g_2(z, \tilde{z}) + u_1 + u_2, \eta)f_{u_1, u_2, \eta}(u_1, u_2, \eta)du_1du_2d\eta\), we obtain

\[
\frac{\partial E[y|z, \tilde{z}]}{\partial \tilde{z}} = \frac{\theta}{\theta} \frac{\partial E[y|z, \tilde{z}]}{\partial z} + \frac{\partial}{\partial z} g_1(z, \tilde{z}) + \frac{\partial}{\partial z} g_2(z, \tilde{z})
\]

Under Assumption 2.7, we may then solve for \(\theta\). We summarize the results as follows:

**Theorem 2.1.** Suppose that Assumptions 2.4, 2.5, 2.6, and 2.7 hold. Then the parameter of interest \(\theta\) is identified. In particular,

\[
\theta = \frac{\partial E[y|z, \tilde{z}]}{\partial z} \frac{\partial E[x|z, \tilde{z}]}{\partial z} - \frac{\partial E[y|z, \tilde{z}]}{\partial z} \frac{\partial E[w|z, \tilde{z}]}{\partial \tilde{z}} - \frac{\partial E[w|z, \tilde{z}]}{\partial \tilde{z}} \frac{\partial E[y|z, \tilde{z}]}{\partial z}.
\]

Notice that the result degenerates to the previous case when there is no measurement error in \(w\), i.e., \(w = \tilde{z} = w^\ast\). If the denominator in the last expression is nonzero, we can then replace the derivatives of the expectations by consistent estimators, which yields a consistent estimator for \(\theta\).

The results presented so far can be extended in a more general framework for handling more than two variables. Consider the following model with a vector of mismeasured regressors
\( X = (x_1, ..., x_k) \) for a true regressor \( X^* = (x_1^*, ..., x_k^*) \):

\[
y = m(\theta_1 x_1^* + \ldots + \theta_k x_k^* + w^*, \eta),
\]
\[
x_i = x_i^* + \varepsilon_i, \quad i = 1, ..., k
\]
\[
x_i^* = g_i(z, \tilde{z}) + u_i,
\]
\[
w = w^* + \tilde{\varepsilon},
\]
\[
w^* = \tilde{g}(z, \tilde{z}) + \tilde{u},
\]

where \( z = (z_1, ..., z_k) \) and \( \tilde{z} \) are instruments, the functions \( m(\cdot, \cdot), g_i(\cdot, \cdot), \tilde{g}(\cdot, \cdot) \) are unknown, \( \theta_1, ..., \theta_k \) are the unknown parameters of interest, and the unobservables \((\eta, \varepsilon_i, ..., \varepsilon_k, u_i, ..., u_k, \tilde{z}, \tilde{u})\) are jointly independent of \((z, \tilde{z})\). The unobservables are normalized to have zero expectation.

Under this setting, the variable \( w^* \) is treated as a base variable and the parameter \( \theta_i \) represents the ratio of partial effects corresponding to \( x_i^* \) and \( w^* \) on \( y \).

Similar derivations can be applied to this general model using the inverse of some \( k \times k \) matrix of partial derivatives of conditional expectations. First, we set \( u = (u_1, ..., u_k) \), and so we have

\[
E[x_i|z, \tilde{z}] = g_i(z, \tilde{z}), \quad \text{for } i = 1, ..., k,
\]
\[
E[w|z, \tilde{z}] = \tilde{g}(z, \tilde{z}),
\]
\[
E[y|z, \tilde{z}] = E[m(\theta_1 x_1^* + \ldots + \theta_k x_k^* + w^*, \eta)|z, \tilde{z}]
\]
\[
= \int m(\theta_1 g_1(z, \tilde{z}) + \ldots + \theta_k g_k(z, \tilde{z}) + \tilde{g}(z, \tilde{z}) + \theta_1 u_1 + \ldots + \theta_k u_k + \tilde{u}, \eta)
\]
\[
\times f_{u,\tilde{u},\eta}(u, \tilde{u}, \eta)dud\tilde{u}d\eta.
\]

Partially differentiating with respect to \( z_i \) or \( \tilde{z} \) on the conditional expectation \( E[y|z, \tilde{z}] \) yields

\[
\frac{\partial E[y|z, \tilde{z}]}{\partial z_i} = \left[ \theta_1 \frac{\partial g_1(z, \tilde{z})}{\partial z_i} + \ldots + \theta_k \frac{\partial g_k(z, \tilde{z})}{\partial z_i} + \frac{\partial \tilde{g}(z, \tilde{z})}{\partial z_i} \right]
\]
\[
\times \int m'(\theta_1 g_1(z, \tilde{z}) + \ldots + \theta_k g_k(z, \tilde{z}) + \tilde{g}(z, \tilde{z}) + \theta_1 u_1 + \ldots + \theta_k u_k + \tilde{u}, \eta)
\]
\[
\times f_{u,\tilde{u},\eta}(u, \tilde{u}, \eta)dud\tilde{u}d\eta,
\]

9
and
\[
\frac{\partial E[y, z, \tilde{z}]}{\partial \tilde{z}} = \left[ \theta_1 \frac{\partial g_1(z, \tilde{z})}{\partial \tilde{z}} + ... + \theta_k \frac{\partial g_k(z, \tilde{z})}{\partial \tilde{z}} + \frac{\partial \tilde{g}(z, \tilde{z})}{\partial \tilde{z}} \right] \times \int m'(\theta_1g_1(z, \tilde{z}) + ... + \theta_kg_k(z, \tilde{z}) + \tilde{g}(z, \tilde{z}) + \theta_1u_1 + ... + \theta_ku_k + \bar{u}, \eta) \\
\times f_u, \bar{u}, \eta(u, \bar{u}, \eta)du d\eta.
\]

The quotient of these two conditional expectations gives
\[
\frac{\partial E[y, z, \tilde{z}]}{\partial z_i} = \frac{\theta_1 \frac{\partial E[x_1|z, \tilde{z}]}{\partial z_i} + ... + \theta_k \frac{\partial E[x_k|z, \tilde{z}]}{\partial z_i} + \frac{\partial E[w|z, \tilde{z}]}{\partial z_i}}{\theta_1 \frac{\partial E[x_1|z, \tilde{z}]}{\partial z_i} + ... + \theta_k \frac{\partial E[x_k|z, \tilde{z}]}{\partial z_i} + \frac{\partial E[w|z, \tilde{z}]}{\partial z_i}}
\]
for \(i = 1, ..., k\).

Cross multiplying the above quotient terms yields
\[
\theta_1 \left( \frac{\partial E[x_1|z, \tilde{z}]}{\partial \tilde{z}} \frac{\partial E[y|z, \tilde{z}]}{\partial z_i} - \frac{\partial E[x_1|z, \tilde{z}]}{\partial z_i} \frac{\partial E[y|z, \tilde{z}]}{\partial \tilde{z}} \right) + ... \\
+ \theta_k \left( \frac{\partial E[x_k|z, \tilde{z}]}{\partial \tilde{z}} \frac{\partial E[y|z, \tilde{z}]}{\partial z_i} - \frac{\partial E[x_k|z, \tilde{z}]}{\partial z_i} \frac{\partial E[y|z, \tilde{z}]}{\partial \tilde{z}} \right) \\
= - \left( \frac{\partial E[w|z, \tilde{z}]}{\partial \tilde{z}} \frac{\partial E[y|z, \tilde{z}]}{\partial z_i} - \frac{\partial E[w|z, \tilde{z}]}{\partial z_i} \frac{\partial E[y|z, \tilde{z}]}{\partial \tilde{z}} \right) \text{ for } i = 1, ..., k.
\]

Let \(A\) denote a \(k \times k\) matrix whose \((i, j)\)th element is \(\frac{\partial E[x_i|z, \tilde{z}]}{\partial z_i} \frac{\partial E[y|z, \tilde{z}]}{\partial z_j} - \frac{\partial E[x_i|z, \tilde{z}]}{\partial z_j} \frac{\partial E[y|z, \tilde{z}]}{\partial z_i}\) and \(B\) denote a \(k \times 1\) vector whose \(i\)th element is \(-\left( \frac{\partial E[w|z, \tilde{z}]}{\partial \tilde{z}} \frac{\partial E[y|z, \tilde{z}]}{\partial z_i} - \frac{\partial E[w|z, \tilde{z}]}{\partial z_i} \frac{\partial E[y|z, \tilde{z}]}{\partial \tilde{z}} \right)\). Using these notations, we can express the \(k\) equations as \(A\tilde{\theta} = B\), where \(\tilde{\theta} = (\theta_1, ..., \theta_k)'\). If the \(k \times k\) matrix of partial derivatives \(A\) is invertible, then the vector of parameters of interest is identified by \(\tilde{\theta} = A^{-1}B\).

### 2.2. Estimation

The main results in Theorems 2.1 and 2.2 are all expressed in terms of the first derivatives of the conditional expectations and we propose to use local polynomial regression methods to estimate these derivatives. Fan and Gijbels (1996) provide an overview of local polynomial estimators for such derivatives and we state an estimator in this section.
The main result in Theorems 2.1,
\[ \theta = \frac{\partial}{\partial z}E[y|z, w] \frac{\partial}{\partial w}E[x|z, w] \]
works for the domain of of \((z, w)\). As such, we can use
\[ \theta = \int \int \frac{\partial}{\partial z}E[y|z, w] \frac{\partial}{\partial w}E[x|z, w] \Omega(z, w) dz dw \]
where \(\Omega(z, w)\) is a weighting function and one can choose \(\Omega(z, w) = f_{ZW}(z, w)\eta(z, w)\) including the density \(f_{ZW}\) and some other weighting \(\eta(z, w)\).

Given a random sample of data, \(\{y_i, x_i, z_i, w_i\}_{i=1}^n\), we can apply local polynomial estimators for the derivatives of the conditional expectations, \(\frac{\partial}{\partial z}E[y|z, w]\), \(\frac{\partial}{\partial w}E[x|z, w]\), and \(\frac{\partial}{\partial w}E[x|z, w]\). Let \(\frac{\partial E[y|z, w]}{\partial z}, \frac{\partial E[y|z, w]}{\partial w}, \frac{\partial E[x|z, w]}{\partial z}, \text{and} \frac{\partial E[x|z, w]}{\partial w}\) be local polynomial estimators for these derivatives. This suggests a sample counterpart estimator of the parameter of interest in Eq. (??),
\[ \hat{\theta} = \sum \frac{\partial E[y|z, i, w_i]}{\partial z} \frac{\partial E[x|z, i, w_i]}{\partial w} \eta(z_i, w_i). \]
Set \((\bar{z}, \bar{w})\) is the sample mean of \(\{z_i, w_i\}_{i=1}^n\). Our choice of the weight function with the bandwidth \(h_1\) is the following:
\[ \frac{1}{N_{h_1}} T \left( \frac{||(z_i, w_i) - (\bar{z}, \bar{w})||}{h_1} \right), \]
where \(T\) is the tri-cube weight function such that
\[ T(u) = \begin{cases} \begin{array}{ll} (1 - |u|^3)^3 & \text{if } |u| \leq 1, \\ 0 & \text{else.} \end{array} \end{cases} \]
The distance \(|| \cdot ||\) is defined by
\[ ||(z_i, w_i) - (\bar{z}, \bar{w})|| = \left( \frac{z_i - \bar{z}}{\text{var}(z_i)} \right)^2 + \left( \frac{w_i - \bar{w}}{\text{var}(w_i)} \right)^2, \]
and \(N_{h_1}\) is the number of points in \(\{z_i, w_i\}_{i=1}^n\) close to the sample mean \((\bar{z}, \bar{w})\) within the distance \(h_1\) defined by \(T\) and \(|| \cdot ||\). The weight function \(\eta(z_i, w_i)\) is supposed to pick up points close to the sample mean \((\bar{z}, \bar{w})\). If we assume the denominator term in Eq. ?? to be bounded
away from zero at the sample mean \((\bar{z}, \bar{w})\), by continuity the weight function removes the points which make the denominator term close to zero. Hence, the weight function can be regarded as a trimming function and prevents so-called ratio bias, as discussed in Hoderlein and Sherman (2015).

The proposed local polynomial estimator should be consistent under general conditions and we will analyze it through Monte Carlo study in the next section. The rest of the section will illustrate how we apply local polynomial regression to two-dimensional functions. We start with a local quadratic approximation for a function \(E[y|z, w]\) around a point \((z_0, w_0)\),

\[
E[y|z, w] \approx \beta_0 + \beta_1(z - z_0) + \beta_2(w - w_0) + \beta_3(z - z_0)(w - w_0) + \beta_4(z - z_0)^2 + \beta_5(w - w_0)^2.
\]

Let \(h_2\) be a positive value for the bandwidth. To obtain the local polynomial regression estimate we try to find the vector of parameters \(\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5)'\) that minimizes the following objective function:

\[
\sum_{i} \left\{ y_i - \beta_0 - \beta_1(z_i - z_0) - \beta_2(w_i - w_0) - ... - \beta_5(w_i - w_0)^2 \right\}^2 \cdot \frac{T\left(\frac{\|z_i, w_i\|}{h_2}\right)}{h_2^2}.
\]

This is a weighted least squares problem where the weights are given by the functions \(T(\cdot)\). The estimation gives \(\frac{\partial}{\partial z} E[y|z_0, w_0] = \beta_1, \) and \(\frac{\partial}{\partial w} E[y|z_0, w_0] = \beta_2.\)

3. Simulation

This section studies the finite sample performance of the proposed local polynomial regression estimator for the parameter of interest \(\theta\). The simulation models in this study are based on nonparametric estimation of derivatives by the method presented in Section 2. The data generating process (DGP) for single index models with measurement error and endogeneity in the Monte Carlo experiments are generated according to the wage equation and the hours worked equation. Consider the wage equation

\[
x^* = g(z, w) + u,
\]
and the hours worked equation
\[ y = m(\theta x^* + w, \eta). \]

The wage is observed with measurement error \( \varepsilon \),
\[ x = x^* + \varepsilon. \]

There are two DGPs in the experiment:

DGP I: \( g(z, w) = z, \)
\[ m(\theta x^* + w, \eta) = \exp(\theta x^* + w + \eta), \]
DGP II: \( g(z, w) = 2 + 0.1z^2 + w, \)
\[ m(\theta x^* + w, \eta) = (\theta x^* + w)^3 + \eta, \]

where \( z \sim N(0, 1), w \sim N(0, 1), u \sim N(0, 0.2), \varepsilon \sim N(0, 0.2), \eta \sim N(0, 1) \) and these normal distributions are jointly independent. These two DGPs have nonlinear functional forms for the hours worked equation; DGP I is an exponential function and DGP II is a polynomial. We consider three different values of \( \theta \) in the experiments: \( \theta = -0.8, -1, \) and \(-1.2 \). In addition, three different sample sizes \( N \) are considered: 500, 1000, 2000. We used 200 simulation replications for each sample size in order to estimate the standard errors.

Tables 1, 2 and 3 present the simulation results of the proposed local polynomial estimator. The simulation results of DGP I (the exponential case) show a small downward bias. In DGP I, the mean and median of the coefficients are almost the same. In addition, the standard errors of \( \theta \) are larger if the absolute value of \( \theta \) is larger except for the case \( N = 500 \). As for DGP II, when \( \theta = -1 \), the bias and standard error are relatively small compared to the other two values of \( \theta \). Nevertheless, the mean and median estimation values are closer to the true values and standard errors decrease as sample sizes grow. Therefore, the Monte Carlo simulation in this study shows that the proposed local polynomial estimator works very well.

A major difficulty for estimating the labor supply elasticity is that both the wages and the number of hours worked may be measured with error. For example, people may not recall the exact amount that they earned per hour in the last year, see for example Borjas (2009, chapter 2). The tools of the previous sections are exactly designed to deal with a measurement error problem like this. Consider the following model that has a wage equation, participation equation, and an hours equation.

The wage equation is as follows,

\[(16) \quad \ln(wage_i) = schooling_i \beta + H_i \gamma + \varepsilon_i \]

where the error term \(\varepsilon_i\) is assumed to be independent of the exogenous regressors in the vector \(H_i\) and also independent of the instrument \(Z_i\). Let the potential wage, the wage that somebody could earn if that person would work, be denoted by \(x^*\). Note that for some, but not all, individuals we observe both predicted log wage and the potential wage measured with error. Next, consider the participation equation. The probability that somebody participates in the labor force is a function of the potential wage \(x^*\),

\[(17) \quad labor\ supply_i = l(\theta_p \cdot x_i^* + S_i \delta, \tau_i), \]

where \(S_i\) is a vector of exogenous regressors. Note that using \(x_i^*\) solves the problem that (a) we do not observe the wage for those who do not work and (b) we have a measurement error problem, even if we do observe the wage.

The hours worked equation is also a function of the potential wage,

\[ \ln(hours_i) = m(\theta \cdot x_i^* + W_i \kappa, \eta_i). \]

where \(W_i\) is a vector of exogenous regressors. Traditional labor models use a wage equation, participation equation and an hours equation. The hours equation is then written as a function of the observed wage rather than the potential wage. Our methodology allows us to write the participation equation and the hours equation in terms of potential wage, i.e. the wage
that somebody could earn if that person would work. Therefore, we can express the labor supply elasticity directly as a function of this potential wage that is unobserved for some and measured with error for others. In particular, an increase in this wage has a direct effect on the hours (e.g. from zero hours to a positive number of hours). Thus, using $x^*$ yields a very natural approach to deal with the intensive and extensive margin. Alternatively, one could only consider the wage equation and the participation equation but we do not pursue that strategy here. Also, note that measurement error plays a role in both the labor supply equation and in the hours worked equation.

In the rest of this section, we report empirical estimates of the model presented using data in US Panel Study of Income Dynamics (PSID). The advantage of our methodology is that it allows that the wage variable is measured with error and/or is endogenous. Many survey data have these features.

We use Wave 35 of the Michigan Panel Study of Income Dynamics 2007 as the source of data for the empirical work. Our sample consists of 880 married couples, age 25-55 in 2007. Other sample restrictions include eliminating observations where the husband or the wife reported that he or she was disabled, removing observations where the couple reports self-employment or farm income, and elimination of observations with missing data. Hours of work is measured in terms of hours on all jobs held in 2007. This variable was the product of weeks worked times average hours worked per week over all jobs. Wage rate is calculated by dividing labor income by hours of work. Nonlabor income is calculated by summing income from rent, dividends, interest, trust funds, and royalties. Also, we use the education of the father as an instrument.

Descriptive statistics for the variables used in the labor supply estimation are presented in Table 4. There are two samples including full sample and working women subsample. Since the data is fairly recent (2007), the labor force participation rate for both men and women is high, 0.966 for men and 0.896 for women. It follows that the subsample of families in which the wife works does not seem to vary significantly from the full sample in terms of hours of work or wage rate of either spouse. Worker characteristics, including age, education, and race, were included to represent human capital factors. Husbands and wives both had 13 years of education on average, but the wives were about one and half years younger than their husbands in the full sample. For simplicity, we consider the variable $w$ to be a scalar.
Therefore, the empirical model becomes

\[ y_i = m(\theta x^* + w_i, \eta_i), \]
\[ x_i = x_i^* + \varepsilon_i, \]
\[ x_i^* = g(z_i, w_i) + u_i. \]

In this empirical application, the dependent variable \( y \) is the natural logarithm of hours of work plus one, the covariate \( x \) is the natural logarithm of the wage rate plus one, \( z \) is the education level of the father, and \( w \) is nonlabor income. Using the tools that we introduce in the earlier sections to estimate the labor supply elasticity, we can only estimate the ratio of partial effects corresponding to \( x^* \) and \( w \) on \( y \) and we need to know the direction of the effect of \( w \) to know the direction of the effect of \( x^* \). In the present empirical example, the choice of \( w \) is nonlabor income of each sample member. An increase in nonlabor income reduces the number of hours that one works so that the sign of the effect of \( w \) is negative. This implies that the sign of the effect of \( x^* \) is the opposite sign of the estimated \( \theta \). In other words an increase in the wage has an income and a substitution effect but an increase in nonlabor income only has an income effect so that we can use the direction of this effect.

Tables 5 and 6 show the results of the estimation of the parameter \( \theta \). The mean and median estimates present similar results in terms of directions of the relative effects and estimated coefficients. Given the flexible nature of the estimation approach, it is hard to compare the size of the estimated coefficients. Therefore, we focus on the sign of the labor supply elasticity. Table 7 presents the estimated sign of the labor supply elasticities by assuming that the effect of the nonlabor income on hours of working is negative. The coefficients for married men are positive, while the coefficients for married women are negative for the full sample and positive for the working sample. Thus, for men, the number of hours worked is increasing in the wage, i.e. the income effect (‘buy more leisure’) is smaller than the substitution effect (‘leisure is more expensive if the wage increases’). For women, the income effect is smaller than the substitution effect in the working sample while the income effect is larger than the substitution effect in the full sample.
5. Conclusion

This paper gives new identification results for the ratio of partial effects in the single index model with measurement error and endogeneity and also shows how to estimate the single index parameter by a local polynomial regression estimator. The paper applies these tools to estimate the labor supply elasticity, and finds that the labor supply elasticity for married men is positive, while the coefficients for married women are negative for the full sample and positive for the working sample. The new estimator allows for endogeneity and measurement error, a situation one often has to deal with in empirical work. An aspect that was not explored in this paper is the use of additional restrictions. For example, if, in a subset of the data, a substantial fraction of the women do not work, then it may still be possible to estimate the median income for the whole subset. This gives the applied researcher additional restrictions and more precise estimates.
Table 1: Finite Sample Performance of the Local Polynomial Regression Estimator (N=500)

\[ \theta = -0.8 \quad \theta = -1 \quad \theta = -1.2 \]

**DGP I**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>( \theta = -0.8 )</th>
<th>( \theta = -1 )</th>
<th>( \theta = -1.2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>-0.965</td>
<td>-1.116</td>
<td>-1.334</td>
</tr>
<tr>
<td>Median</td>
<td>-0.879</td>
<td>-1.110</td>
<td>-1.304</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.460</td>
<td>0.815</td>
<td>0.502</td>
</tr>
</tbody>
</table>

**DGP II**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>( \theta = -0.8 )</th>
<th>( \theta = -1 )</th>
<th>( \theta = -1.2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>-0.783</td>
<td>-0.987</td>
<td>-1.160</td>
</tr>
<tr>
<td>Median</td>
<td>-0.817</td>
<td>-1.013</td>
<td>-1.172</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.802</td>
<td>0.570</td>
<td>1.050</td>
</tr>
</tbody>
</table>

Standard deviations of the parameters are computed by using sample standard deviation of 200 replications.

Table 2: Finite Sample Performance of the Local Polynomial Regression Estimator (N=1000)

\[ \theta = -0.8 \quad \theta = -1 \quad \theta = -1.2 \]

**DGP I**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>( \theta = -0.8 )</th>
<th>( \theta = -1 )</th>
<th>( \theta = -1.2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>-0.850</td>
<td>-1.065</td>
<td>-1.292</td>
</tr>
<tr>
<td>Median</td>
<td>-0.839</td>
<td>-1.075</td>
<td>-1.299</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.108</td>
<td>0.163</td>
<td>0.324</td>
</tr>
</tbody>
</table>

**DGP II**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>( \theta = -0.8 )</th>
<th>( \theta = -1 )</th>
<th>( \theta = -1.2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>-0.813</td>
<td>-0.994</td>
<td>-1.137</td>
</tr>
<tr>
<td>Median</td>
<td>-0.805</td>
<td>-1.005</td>
<td>-1.180</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.738</td>
<td>0.114</td>
<td>0.850</td>
</tr>
</tbody>
</table>

Standard deviations of the estimators are computed by using sample standard deviation of 200 replications.
Table 3: Finite Sample Performance of the Local Polynomial Regression Estimator (N=2000)

\[
\begin{align*}
\theta = -0.8 & \quad \theta = -1 & \quad \theta = -1.2 \\
\hline
\text{DGP I} & \\
\text{mean} & -0.827 & -1.048 & -1.267 \\
\text{median} & -0.820 & -1.042 & -1.256 \\
\text{std. dev.} & 0.068 & 0.086 & 0.104 \\
\hline
\text{DGP II} & \\
\text{mean} & -0.811 & -1.017 & -1.163 \\
\text{median} & -0.809 & -1.005 & -1.202 \\
\text{std. dev.} & 0.316 & 0.185 & 0.464 \\
\end{align*}
\]

Standard deviations of the estimators are computed by using sample standard deviation of 200 replications.
Table 4: Sample Statistics

<table>
<thead>
<tr>
<th>Variables</th>
<th>Full Sample</th>
<th>Working Women Subsample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Dev.</td>
</tr>
<tr>
<td>Husband’s</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>40.316</td>
<td>8.849</td>
</tr>
<tr>
<td>Participation</td>
<td>0.966</td>
<td>0.182</td>
</tr>
<tr>
<td>Hours of work</td>
<td>2217.176</td>
<td>528.480</td>
</tr>
<tr>
<td>Wage rate</td>
<td>26.804</td>
<td>20.570</td>
</tr>
<tr>
<td>Nonlabor income</td>
<td>488.285</td>
<td>1879.806</td>
</tr>
<tr>
<td>Black</td>
<td>0.168</td>
<td>0.374</td>
</tr>
<tr>
<td>Education</td>
<td>5.526</td>
<td>1.513</td>
</tr>
<tr>
<td>Father’s Education</td>
<td>4.713</td>
<td>1.738</td>
</tr>
<tr>
<td>Wife’s</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>38.881</td>
<td>8.876</td>
</tr>
<tr>
<td>Participation</td>
<td>0.896</td>
<td>0.305</td>
</tr>
<tr>
<td>Hours of work</td>
<td>1698.382</td>
<td>662.855</td>
</tr>
<tr>
<td>Nonlabor income</td>
<td>396.992</td>
<td>1605.627</td>
</tr>
<tr>
<td>Black</td>
<td>0.161</td>
<td>0.368</td>
</tr>
<tr>
<td>Education</td>
<td>5.752</td>
<td>1.478</td>
</tr>
<tr>
<td>Father’s Education</td>
<td>4.902</td>
<td>1.790</td>
</tr>
<tr>
<td>Sample size</td>
<td>880</td>
<td></td>
</tr>
</tbody>
</table>

Note: a. Standard errors are in parentheses. b. No kid, One kid, and More kid are indicators for having no child, having one child, and having two or more children, respectively. c. Wage rate is calculated by dividing labor income by hours of work. d. Nonlabor income is calculated by summing income from rent, dividends, interest, trust funds, and royalties. e. Education is imputed from the following categorical scheme: 1 = '0-5 grades' (2.5 years); 2 = '6-8' (7 years); 3 = '9-11' (10 years); 4 = '12' (12 years); 5 = '12 plus non-academic training' (13 years); 6 = 'some college' (14-15 years); 7 = 'college degree, not advanced' (16 years); 8 = 'college advanced degree' (17 years).
Table 5: Estimation Results of Relative Effects

<table>
<thead>
<tr>
<th></th>
<th>Full Sample</th>
<th>Working Subsample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td>mean</td>
<td>-3.595</td>
</tr>
<tr>
<td></td>
<td>median</td>
<td>-3.575</td>
</tr>
<tr>
<td></td>
<td>std. dev.</td>
<td>(7.199)</td>
</tr>
<tr>
<td>Women</td>
<td>mean</td>
<td>0.084</td>
</tr>
<tr>
<td></td>
<td>median</td>
<td>1.037</td>
</tr>
<tr>
<td></td>
<td>std. dev.</td>
<td>(4.377)</td>
</tr>
</tbody>
</table>

Note: a. The estimates reported here are the relative effect of the potential wage and the nonlabor income on hours worked.
b. We use Silverman’s rule of thumb, i.e. \( h_i = c \cdot n^{1/5} \) for \( i = 1, 2 \). The constant \( c = 2 \) in table 5. As a robustness check, we use \( c = 4 \) in table 6 and refer to it as double bandwidth.
c. The instrumental variable in these specifications is father’s education.

Table 6: Estimation Results of Relative Effects (Double Bandwith)

<table>
<thead>
<tr>
<th></th>
<th>Full Sample</th>
<th>Working Subsample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td>mean</td>
<td>-15.231</td>
</tr>
<tr>
<td></td>
<td>median</td>
<td>-8.515</td>
</tr>
<tr>
<td></td>
<td>std. dev.</td>
<td>(36.997)</td>
</tr>
<tr>
<td>Women</td>
<td>mean</td>
<td>42.923</td>
</tr>
<tr>
<td></td>
<td>median</td>
<td>23.258</td>
</tr>
<tr>
<td></td>
<td>std. dev.</td>
<td>(54.092)</td>
</tr>
</tbody>
</table>

Note: a. The estimates reported here are the relative effect of the potential wage and the nonlabor income on hours worked.
b. We use the bandwidths as the Silverman’s rule of thumb, i.e. \( h_i = c \cdot n^{1/5} \) for \( i = 1, 2 \). The constant \( c = 2 \) in table 5. As a robustness check, we use \( c = 4 \) in this table.
c. The instrumental variable in these specifications is father’s education.
Table 7: Estimated Signs of the Labor Supply Elasticities

<table>
<thead>
<tr>
<th></th>
<th>Full Sample</th>
<th>Working Subsample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men: mean</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Men: median</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Women: mean</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>Women: median</td>
<td>–</td>
<td>+</td>
</tr>
</tbody>
</table>

a. Because the direction of the effect of the nonlabor income \( w \) on hours of working is negative, the sign of the effect of potential wage \( x^* \) on hours of working is the opposite sign of the estimated \( \theta \).

References


Appendix 1

The following example shows that a naive regression of $y$ on $x$ may yield the wrong sign on the causal effect. In other words, the measurement error and/or endogeneity may reverse the sign of the effect. We thank an anonymous referee for suggesting to include an example to illustrate this. Consider the following linear version of the proposed model

$$y = \theta x^* + w + \eta,$$
$$x = x^* + \varepsilon,$$
$$x^* = g(z, w) + u.$$

This implies that $y = \theta x^* + w + \eta = \theta x + (w + \eta - \theta \varepsilon)$. Set $v = w + \eta - \theta \varepsilon$ and then $y = \theta x + v$. A naive regression of $y$ on $x$ yields the following estimator for $\theta$,

$$\tilde{\theta} = \frac{\sum_i (y_i - \bar{y})(x_i - \bar{x})}{\sum_i (x_i - \bar{x})^2} = \theta + \frac{\sum_i (x_i - \bar{x})(v_i - \bar{v})}{\sum_i (x_i - \bar{x})^2}.$$

By the definition of $v$, we have $v_i - \bar{v} = (w_i - \bar{w}) + (\eta_i - \bar{\eta}) - \theta(\varepsilon_i - \bar{\varepsilon})$. Plugging the relationship back into the above equation yields

$$\tilde{\theta} = \theta + \frac{\sum_i (x_i - \bar{x})[(w_i - \bar{w}) + (\eta_i - \bar{\eta}) - \theta(\varepsilon_i - \bar{\varepsilon})]}{\sum_i (x_i - \bar{x})^2}.$$

In large samples $\tilde{\theta}$ converges in probability to

$$\theta + \frac{\text{cov}(x, w) + \text{cov}(x, \eta) - \theta \text{cov}(x, \varepsilon)}{\text{var}(x)}.$$

Because $x = g(z, w) + \varepsilon + u$ and $(\eta, \varepsilon, u)$ are independent of $z, w$, the probability limit $\tilde{\theta}$ is

$$\theta + \frac{\text{cov}(x, w) + \text{cov}(\varepsilon + u, \eta) - \theta \text{cov}(\varepsilon + u, \varepsilon)}{\text{var}(x)}.$$

This shows that the bias of the naive regression depends on several factors including the covariance of $x$ and $w$, the covariances of the measurement error $\varepsilon$ between the disturbance $\eta$ and the endogenous error $u$, and the endogenous term $\text{cov}(u, \eta)$. It is easy to see that the sign
of $\theta$ is not identified by the naive regression. If (i) the measurement error $\varepsilon$ is uncorrelated to the disturbance $\eta$ and the endogenous error $u$ with small variation and (ii) the covariance between $x$ and $w$ are small, then the probability limit is

$$\theta + \frac{\text{cov}(u, \eta)}{\text{var}(x)}.$$ 

In this case the endogeneity (nonzero covariance between $u$ and $\eta$) is strong enough to reverse the sign of the effect under the following conditions,

$$\begin{cases} 
\frac{\text{cov}(u, \eta)}{\text{var}(x)} < -\theta & \text{if } \theta > 0, \\
\frac{\text{cov}(u, \eta)}{\text{var}(x)} > |\theta| & \text{otherwise}.
\end{cases}$$