

Identification and estimation of single-index models with measurement error and endogeneity

YINGYAO HU[†], JI-LIANG SHIU[‡] AND TIEMEN WOUTERSEN[§]

[†]*Department of Economics, Johns Hopkins University, 440 Mergenthaler Hall,
3400 N. Charles Street, Baltimore, MD 21218, USA.*

E-mail: yhu@jhu.edu

[‡]*Hanqing Advanced Institute of Economics and Finance, Renmin University of China,
Beijing 100872, P. R. China.*

E-mail: jishiu@ruc.edu.cn

[§]*Department of Economics, University of Arizona, McClelland Hall 401, P.O. Box 210108,
Tucson, AZ 85721-0108, USA.*

E-mail: woutersen@email.arizona.edu

First version received: January 2014; final version accepted: June 2015

Summary Economic variables are often measured with an error and may be endogenous. In this paper, we give 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. We apply these tools to estimate the labour-supply elasticity and find that the labour-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: *Endogeneity, Labour-supply elasticity, Measurement error, Nonclassical measurement error.*

1. INTRODUCTION

In many applied problems in economics, some variables are measured with error and/or are endogenous. For example, when estimating the labour-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 labour-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,

$$y = m(\theta x^* + w, \eta), \quad (1.1)$$

$$x = x^* + \varepsilon, \quad (1.2)$$

$$x^* = g(z, w) + u, \quad (1.3)$$

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 η , ε and u . The error terms (η, ε, u) in (1.1)–(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 η 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 η are generally correlated. The measurement error ε may be correlated with η so that the procedures by 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 θ , which measures the relative effects for continuous variables x^* and w . If x^* and w are independent of η , then θ can be regarded as the ratio of partial effects corresponding to x^* and w on y because

$$\theta = \frac{\partial m(\theta x^* + w, \eta) / \partial x^*}{\partial m(\theta x^* + w, \eta) / \partial w}.$$

Chesher (2003) and Schennach et al. (2012) also use ratios of derivatives to identify quantities on nonseparable models.¹

In the next section, we show that the parameter θ is identified. One of the many models that fit this framework is a new model for labour supply that allows for measurement error and endogeneity. Let y be log hours and let x^* be log wage. Then, the parameter θ represents the ratio of labour-supply elasticity over the partial effect of w on y . We present and estimate this model to illustrate our methodology but we 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 labour force participation to change as the wage changes. In general, we allow (1.2) to be replaced by the requirement that the function $g(z, w)$ in (1.3) is identified.

If we ignore measurement error in (1.2), i.e., $x = x^*$, then the proposed model becomes a single-index control function model, which is closely related to Imbens and Newey (2009).² 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 (1.1)–(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 η and the other is that the measurement error ε may also be correlated with η . These changes allow the method to be used for economic models with endogeneity.

The estimation of the labour-supply elasticity has been of considerable interest in labour economics. For example, Ashenfelter and Heckman (1974), Killingsworth and Heckman (1986), Pencavel (1986, 2002) and Blundell et al. (1998) 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

¹ The appendix contains an example that illustrates the fact that using a linear approximation may yield the wrong sign for the causal effect.

² There may exist a correlation between x^* and η in (1.1) and the nonseparable term $g(z, w)$ can also be regarded as a control function because $g(z, w)$ is independent of η and a correlation between x^* and η should come from a correlation between u and η .

estimate the labour-supply elasticity, and also discusses the problems caused by measurement error. Keane (2011) reviews the literature on the effects of taxes on labour supply. Our paper introduces a new methodology that allows for measurement error. We find that the labour-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.

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. Then, in Section 4, we show how these tools can be used to determine whether the labour-supply elasticity is positive or negative. We conclude in Section 5.

2. SINGLE-INDEX MODELS

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

ASSUMPTION 2.1. *The observables z and w are independent of the error terms (u, η, ε) in (1.1)–(1.3). The error terms u and ε satisfy the conditional mean restrictions $E[u|z, w] = 0$ and $E[\varepsilon|z, w] = 0$, respectively.*

By (1.3), taking the expectation of x conditional on z and w yields

$$\begin{aligned} E[x|z, w] &= E[x^* + \varepsilon|z, w] \\ &= E[g(z, w) + u + \varepsilon|z, w] \\ &= g(z, w). \end{aligned}$$

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

$$\begin{aligned} E[y|z, w] &= E[m(\theta g(z, w) + \theta u + w, \eta)|z, w] \\ &= \int m(\theta g(z, w) + \theta u + w, \eta) f_{u,\eta}(u, \eta) du d\eta. \end{aligned}$$

We need the following regularity condition for the identification result.

ASSUMPTION 2.2. *The instruments (z, w) each contain a continuous element and the functions $m(\cdot, \cdot)$ and $g(\cdot, \cdot)$ are differentiable.*

Taking the derivative of the conditional mean function $E[y|z, w]$ with respect to z, w , we have

$$\begin{aligned} \frac{\partial E[y|z, w]}{\partial w} &= \int m'(\theta g(z, w) + w + \theta u, \eta) \left(\theta \frac{\partial}{\partial w} g(z, w) + 1 \right) f_{u,\eta}(u, \eta) du d\eta \\ &= \left(\theta \frac{\partial}{\partial w} g(z, w) + 1 \right) \int m'(\theta g(z, w) + w + \theta u, \eta) f_{u,\eta}(u, \eta) du d\eta, \end{aligned} \tag{2.1}$$

and

$$\begin{aligned} \frac{\partial E[y|z, w]}{\partial z} &= \int m'(\theta g(z, w) + w + \theta u, \eta) \frac{\partial}{\partial z} \theta g(z, w) f_{u,\eta}(u, \eta) du d\eta \\ &= \theta \frac{\partial}{\partial z} g(z, w) \int m'(\theta g(z, w) + w + \theta u, \eta) f_{u,\eta}(u, \eta) du d\eta. \end{aligned} \tag{2.2}$$

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 \mathcal{Z} in support of (z, w) such that $\Pr(\mathcal{Z}) > 0$ and for any $(z, w) \in \mathcal{Z}$, $(\partial/\partial z)E[y|z, w] \neq 0$ 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 $(\partial/\partial z)E[y|z, w] \neq 0$ in Assumption 2.3, we can divide (2.1) by (2.2) to eliminate $\int m'(\theta g(z, w) + w + \theta u, \eta) f_{u,\eta}(u, \eta) du d\eta$. This yields

$$\begin{aligned} \frac{(\partial/\partial w)E[y|z, w]}{(\partial/\partial z)E[y|z, w]} &= \frac{\theta(\partial/\partial w)g(z, w) + 1}{\theta(\partial/\partial z)g(z, w)} \\ &= \frac{\theta(\partial/\partial w)E[x|z, w] + 1}{\theta(\partial/\partial z)E[x|z, w]}. \end{aligned}$$

It follows that

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

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 $W = (w_1, \tilde{W})$:

$$y = m(\theta x^* + w_1 + h(\tilde{W}), \eta), \tag{2.3}$$

$$x = x^* + \varepsilon, \tag{2.4}$$

$$x^* = g(z, W) + u. \tag{2.5}$$

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

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

The last equality holds for every $(z, w) \in \mathbb{R}^2$. In order to estimate θ , 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*

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

$$x = x^* + \varepsilon_1,$$

$$\begin{aligned} x^* &= g_1(z, \tilde{z}) + u_1, \\ w &= w^* + \varepsilon_2, \\ w^* &= g_2(z, \tilde{z}) + u_2, \end{aligned}$$

where (x, w) are the measurement of the latent explanatory variables (x^*, w^*) , (z, \tilde{z}) are instruments, the functions $m(\cdot, \cdot)$, $g_1(\cdot, \cdot)$ and $g_2(\cdot, \cdot)$ are unknown, θ 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 (η, 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}

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

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} ,

$$\begin{aligned} E[y|z, \tilde{z}] &= E[m(\theta g_1(z, \tilde{z}) + g_2(z, \tilde{z}) + \theta u_1 + u_2, \eta)|z, \tilde{z}] \\ &= \int m(\theta g_1(z, \tilde{z}) + g_2(z, \tilde{z}) + \theta u_1 + u_2, \eta) f_{u_1, u_2, \eta}(u_1, u_2, \eta) du_1 du_2 d\eta. \end{aligned}$$

We further assume the following.

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

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

$$\begin{aligned} \frac{\partial E[y|z, \tilde{z}]}{\partial z} &= (\theta \frac{\partial}{\partial z} g_1(z, \tilde{z}) + \frac{\partial}{\partial z} g_2(z, \tilde{z})) \int m'(\theta g_1(z, \tilde{z}) + g_2(z, \tilde{z}) + \theta u_1 + u_2, \eta) \\ &\quad \times f_{u_1, u_2, \eta}(u_1, u_2, \eta) du_1 du_2 d\eta. \end{aligned} \tag{2.6}$$

and

$$\begin{aligned} \frac{\partial E[y|z, \tilde{z}]}{\partial \tilde{z}} &= (\theta \frac{\partial}{\partial \tilde{z}} g_1(z, \tilde{z}) + \frac{\partial}{\partial \tilde{z}} g_2(z, \tilde{z})) \int m'(\theta g_1(z, \tilde{z}) + g_2(z, \tilde{z}) + \theta u_1 + u_2, \eta) \\ &\quad \times f_{u_1, u_2, \eta}(u_1, u_2, \eta) du_1 du_2 d\eta, \end{aligned} \tag{2.7}$$

ASSUMPTION 2.7. *There exists a set \mathcal{Z} in support of (z, \tilde{z}) such that $\Pr(\mathcal{Z}) > 0$ and for any $(z, \tilde{z}) \in \mathcal{Z}$, $(\partial/\partial z)E[y|z, \tilde{z}] \neq 0$ 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 (2.6) by (2.7) to eliminate $\int m'(\theta g_1(z, \tilde{z}) + g_2(z, \tilde{z}) + \theta u_1 + u_2, \eta) f_{u_1, u_2, \eta}(u_1, u_2, \eta) du_1 du_2 d\eta$, we obtain

$$\begin{aligned} \frac{(\partial E[y|z, \tilde{z}]/\partial z)}{(\partial E[y|z, \tilde{z}]/\partial \tilde{z})} &= \frac{\theta(\partial/\partial z)g_1(z, \tilde{z}) + (\partial/\partial z)g_2(z, \tilde{z})}{\theta(\partial/\partial \tilde{z})g_1(z, \tilde{z}) + (\partial/\partial \tilde{z})g_2(z, \tilde{z})} \\ &= \frac{\theta(\partial/\partial z)E[x|z, \tilde{z}] + (\partial/\partial z)E[w|z, \tilde{z}]}{\theta(\partial/\partial \tilde{z})E[x|z, \tilde{z}] + (\partial/\partial \tilde{z})E[w|z, \tilde{z}]} \end{aligned}$$

Under Assumption 2.7, we can then solve for θ . We summarize the results as follows.

THEOREM 2.1. *Suppose that Assumptions 2.4–2.7 hold. Then, the parameter of interest θ is identified. In particular,*

$$\theta = \frac{(\partial E[y|z, \tilde{z}]/\partial z)(\partial E[w|z, \tilde{z}]/\partial \tilde{z}) - (\partial E[y|z, \tilde{z}]/\partial \tilde{z})(\partial E[w|z, \tilde{z}]/\partial z)}{(\partial E[y|z, \tilde{z}]/\partial \tilde{z})(\partial E[x|z, \tilde{z}]/\partial z) - (\partial E[y|z, \tilde{z}]/\partial z)(\partial E[x|z, \tilde{z}]/\partial \tilde{z})}$$

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

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 $X = (x_1, \dots, x_k)$ for a true regressor $X^* = (x_1^*, \dots, x_k^*)$:

$$y = m(\theta_1 x_1^* + \dots + \theta_k x_k^* + w^*, \eta); \tag{2.8}$$

$$x_i = x_i^* + \varepsilon_i, \quad i = 1, \dots, k; \tag{2.9}$$

$$x_i^* = g_i(z, \tilde{z}) + u_i; \tag{2.10}$$

$$w = w^* + \tilde{\varepsilon}; \tag{2.11}$$

$$w^* = \tilde{g}(z, \tilde{z}) + \tilde{u}. \tag{2.12}$$

Here, $z = (z_1, \dots, z_k)$ and \tilde{z} are instruments, the functions $m(\cdot, \cdot)$, $g_i(\cdot, \cdot)$ and $\tilde{g}(\cdot, \cdot)$ are unknown, $\theta_1, \dots, \theta_k$ are the unknown parameters of interest and $(\eta, \varepsilon_1, \dots, \varepsilon_k, u_1, \dots, u_k, \tilde{\varepsilon}, \tilde{u})$ are the unobservables 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 θ_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 $u = (u_1, \dots, u_k)$, and so we have

$$E[x_i|z, \tilde{z}] = g_i(z, \tilde{z}), \quad \text{for } i = 1, \dots, k,$$

$$E[w|z, \tilde{z}] = \tilde{g}(z, \tilde{z}),$$

$$\begin{aligned} E[y|z, \tilde{z}] &= E[m(\theta_1 x_1^* + \dots + \theta_k x_k^* + w^*, \eta)|z, \tilde{z}] \\ &= \int m(\theta_1 g_1(z, \tilde{z}) + \dots + \theta_k g_k(z, \tilde{z}) + \tilde{g}(z, \tilde{z}) + \theta_1 u_1 + \dots + \theta_k u_k + \tilde{u}, \eta) \\ &\quad \times f_{u, \tilde{u}, \eta}(u, \tilde{u}, \eta) du d\tilde{u} d\eta. \end{aligned}$$

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

$$\begin{aligned} \frac{\partial E[y|z, \tilde{z}]}{\partial z_i} &= \left(\theta_1 \frac{\partial g_1(z, \tilde{z})}{\partial z_i} + \dots + \theta_k \frac{\partial g_k(z, \tilde{z})}{\partial z_i} + \frac{\partial \tilde{g}(z, \tilde{z})}{\partial z_i} \right) \\ &\quad \times \int m'(\theta_1 g_1(z, \tilde{z}) + \dots + \theta_k g_k(z, \tilde{z}) + \tilde{g}(z, \tilde{z}) + \theta_1 u_1 + \dots + \theta_k u_k + \tilde{u}, \eta) \\ &\quad \times f_{u, \tilde{u}, \eta}(u, \tilde{u}, \eta) du d\tilde{u} d\eta, \end{aligned}$$

and

$$\begin{aligned} \frac{\partial E[y|z, \tilde{z}]}{\partial \tilde{z}} &= \left(\theta_1 \frac{\partial g_1(z, \tilde{z})}{\partial \tilde{z}} + \dots + \theta_k \frac{\partial g_k(z, \tilde{z})}{\partial \tilde{z}} + \frac{\partial \tilde{g}(z, \tilde{z})}{\partial \tilde{z}} \right) \\ &\quad \times \int m'(\theta_1 g_1(z, \tilde{z}) + \dots + \theta_k g_k(z, \tilde{z}) + \tilde{g}(z, \tilde{z}) + \theta_1 u_1 + \dots + \theta_k u_k + \tilde{u}, \eta) \\ &\quad \times f_{u, \tilde{u}, \eta}(u, \tilde{u}, \eta) du d\tilde{u} d\eta. \end{aligned}$$

The quotient of these two conditional expectations gives

$$\begin{aligned} \frac{(\partial E[y|z, \tilde{z}]/\partial z_i)}{(\partial E[y|z, \tilde{z}]/\partial \tilde{z})} &= \frac{\theta_1(\partial g_1(z, \tilde{z})/\partial z_i) + \dots + \theta_k(\partial g_k(z, \tilde{z})/\partial z_i) + (\partial \tilde{g}(z, \tilde{z})/\partial z_i)}{\theta_1(\partial g_1(z, \tilde{z})/\partial \tilde{z}) + \dots + \theta_k(\partial g_k(z, \tilde{z})/\partial \tilde{z}) + (\partial \tilde{g}(z, \tilde{z})/\partial \tilde{z})} \\ &= \frac{\theta_1(\partial E[x_1|z, \tilde{z}]/\partial z_i) + \dots + \theta_k(\partial E[x_k|z, \tilde{z}]/\partial z_i) + (\partial E[w|z, \tilde{z}]/\partial z_i)}{\theta_1(\partial E[x_1|z, \tilde{z}]/\partial \tilde{z}) + \dots + \theta_k(\partial E[x_k|z, \tilde{z}]/\partial \tilde{z}) + (\partial E[w|z, \tilde{z}]/\partial \tilde{z})} \\ &\quad \text{for } i = 1, \dots, k. \end{aligned}$$

Cross-multiplying the above quotient terms yields

$$\begin{aligned} &\theta_1 \left(\frac{\partial E[x_1|z, \tilde{z}]}{\partial \tilde{z}} \frac{\partial E[y|z, \tilde{z}]}{\partial z_i} - \frac{\partial E[x_1|z, \tilde{z}]}{\partial z_i} \frac{\partial E[y|z, \tilde{z}]}{\partial \tilde{z}} \right) \\ &\quad + \dots + \theta_k \left(\frac{\partial E[x_k|z, \tilde{z}]}{\partial \tilde{z}} \frac{\partial E[y|z, \tilde{z}]}{\partial z_i} - \frac{\partial E[x_k|z, \tilde{z}]}{\partial z_i} \frac{\partial E[y|z, \tilde{z}]}{\partial \tilde{z}} \right) \\ &= - \left(\frac{\partial E[w|z, \tilde{z}]}{\partial \tilde{z}} \frac{\partial E[y|z, \tilde{z}]}{\partial z_i} - \frac{\partial E[w|z, \tilde{z}]}{\partial z_i} \frac{\partial E[y|z, \tilde{z}]}{\partial \tilde{z}} \right) \quad \text{for } i = 1, \dots, k. \end{aligned}$$

Let A denote a $k \times k$ matrix whose (i, j) th element is

$$\frac{\partial E[x_j|z, \tilde{z}]}{\partial \tilde{z}} \frac{\partial E[y|z, \tilde{z}]}{\partial z_i} - \frac{\partial E[x_j|z, \tilde{z}]}{\partial z_i} \frac{\partial E[y|z, \tilde{z}]}{\partial \tilde{z}}$$

and let 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 $A\vec{\theta} = B$, where $\vec{\theta} = (\theta_1, \dots, \theta_k)'$. If the $k \times k$ matrix of partial derivatives A is invertible, then the vector of parameters of interest is identified by $\vec{\theta} = A^{-1}B$.

2.2. Estimation

The main result in Theorem 2.1 is expressed in terms of the first derivatives of the conditional expectations and we propose the use of 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 Theorem 2.1,

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

works for the domain of (z, w) . As such, we can use

$$\theta = \int \int \frac{(\partial/\partial z)E[y|z, w]}{(\partial/\partial w)E[y|z, w](\partial/\partial z)E[x|z, w] - (\partial/\partial z)E[y|z, w](\partial/\partial w)E[x|z, w]} \times \Omega(z, w) dz dw, \tag{2.13}$$

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, $\{y_i, x_i, z_i, w_i\}_i^n$, we can apply local polynomial estimators for the derivatives of the conditional expectations, $(\partial/\partial z)E[y|z, w]$, $(\partial/\partial w)E[y|z, w]$, $(\partial/\partial z)E[x|z, w]$ and $(\partial/\partial w)E[x|z, w]$. Let

$$\frac{\partial \widehat{E[y|z, w]}}{\partial z}, \quad \frac{\partial \widehat{E[y|z, w]}}{\partial w}, \quad \frac{\partial \widehat{E[x|z, w]}}{\partial z} \quad \text{and} \quad \frac{\partial \widehat{E[x|z, w]}}{\partial w}$$

be local polynomial estimators for these derivatives. This suggests a sample counterpart estimator of the parameter of interest in (2.13),

$$\widehat{\theta} = \sum_i \frac{(\partial \widehat{E[y|z_i, w_i]}/\partial z)}{(\partial \widehat{E[y|z_i, w_i]}/\partial w)(\partial \widehat{E[x|z_i, w_i]}/\partial z) - (\partial \widehat{E[y|z_i, w_i]}/\partial z)(\partial \widehat{E[x|z_i, w_i]}/\partial w)} \times \eta(z_i, w_i). \tag{2.14}$$

Set (\bar{z}, \bar{w}) is the sample mean of $\{z_i, w_i\}_i^n$. 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

$$T(u) = \begin{cases} (1 - |u|^3)^3 & \text{if } |u| \leq 1, \\ 0 & \text{else.} \end{cases}$$

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 $\{z_i, w_i\}_i^n$ 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 that the denominator term in (2.14) is bounded away from zero at the sample mean (\bar{z}, \bar{w}) , by continuity the weight function removes the points that 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 by Hoderlein and Sherman (2015).

The proposed local polynomial estimator should be consistent under general conditions and we analyse it using Monte Carlo study in the next section. In the rest of this section, we 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:

$$\sum_i^n (y_i - \beta_0 - \beta_1(z_i - z_0) - \beta_2(w_i - w_0) - \dots - \beta_5(w_i - w_0)^2)^2 T\left(\frac{\|(z_i, w_i) - (z_0, w_0)\|}{h_2}\right).$$

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

3. SIMULATION

In this section, we study the finite sample performance of the proposed local polynomial regression estimator for the parameter of interest θ . 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 ε ,

$$x = x^* + \varepsilon.$$

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: $g(z, w) = 2 + 0.1z^2 + w,$

$$m(\theta x^* + w, \eta) = (\theta x^* + w)^3 + \eta,$$

Table 1. Finite sample performance of the local polynomial regression estimator.

		$\theta = -0.8$	$\theta = -1$	$\theta = -1.2$
<i>N</i> = 500				
DGP I	Mean	-0.965	-1.116	-1.334
	Median	-0.879	-1.110	-1.304
	Std dev.	0.460	0.815	0.502
DGP II	Mean	-0.783	-0.987	-1.160
	Median	-0.817	-1.013	-1.172
	Std dev.	0.802	0.570	1.050
<i>N</i> = 1000				
DGP I	Mean	-0.850	-1.065	-1.292
	Median	-0.839	-1.075	-1.299
	Std dev.	0.108	0.163	0.324
DGP II	Mean	-0.813	-0.994	-1.137
	Median	-0.805	-1.005	-1.180
	Std dev.	0.738	0.114	0.850
<i>N</i> = 2000				
DGP I	Mean	-0.827	-1.048	-1.267
	Median	-0.820	-1.042	-1.256
	Std dev.	0.068	0.086	0.104
DGP II	Mean	-0.811	-1.017	-1.163
	Median	-0.809	-1.005	-1.202
	Std dev.	0.316	0.185	0.464

Note: Standard deviations of the estimators are computed by using the sample standard deviation of 200 replications.

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 θ in the experiments: $\theta = -0.8$, -1 and -1.2 . In addition, three different sample sizes N are considered: 500, 1000 and 2000. We used 200 simulation replications for each sample size in order to estimate the standard errors.

Table 1 presents 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 θ are larger if the absolute value of θ 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 θ . 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.

4. EMPIRICAL APPLICATION: THE LABOUR-SUPPLY ELASTICITY

A major difficulty for estimating the labour-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, e.g., Chapter 2 of Borjas (2009). The tools of the previous sections are exactly designed to deal with a measurement error problem like this. Consider the following model, which has a wage equation, a participation equation and an hours equation.

The wage equation is

$$\ln(\text{wage}_i) = \text{schooling}_i \beta + H_i \gamma + \varepsilon_i, \quad (4.1)$$

where the error term ε_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 (i.e., 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 labour force is a function of the potential wage x^* ,

$$\text{labour supply}_i = l(\theta_p \cdot x_i^* + S_i \delta, \tau_i), \quad (4.2)$$

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,

$$\ln(\text{hours}_i) = m(\theta \cdot x_i^* + W_i \kappa, \eta_i),$$

where W_i is a vector of exogenous regressors. Traditional labour models use a wage equation, a 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 labour-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 labour-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 the US Panel Study of Income Dynamics (PSID). The advantage of our methodology is that it allows the wage variable to be measured with error and/or to be endogenous. Many survey data have these features.

We use Wave 35 of the Michigan PSID 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

Table 2. Sample statistics.

Variables	Full sample		Working women subsample	
	Mean	Std dev.	Mean	Std dev.
Husband				
Age	40.316	8.849	40.760	8.827
Participation	0.966	0.182	0.965	0.185
Hours of work	2217.176	528.480	2200.583	512.658
Wage rate	26.804	20.570	26.848	19.300
Nonlabour income	488.285	1879.806	511.137	1945.202
Black	0.168	0.374	0.166	0.372
Education	5.526	1.513	5.535	1.520
Father's education	4.713	1.738	4.699	1.751
Wife				
Age	38.881	8.876	39.310	8.830
Participation	0.896	0.305	1.000	0.000
Hours of work	1698.382	662.855	1756.705	622.529
Wage rate	21.255	14.943	21.717	15.212
Nonlabour income	396.992	1605.627	410.303	1649.953
Black	0.161	0.368	0.160	0.366
Education	5.752	1.478	5.780	1.478
Father's education	4.902	1.790	4.877	1.812
Sample size	880		789	

Notes: Standard errors are in parentheses. Wage rate is calculated by dividing labour income by hours of work. Nonlabour income is calculated by summing income from rent, dividends, interest, trust funds and royalties. 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).

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 labour income by hours of work. Nonlabour 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 labour-supply estimation are presented in Table 2. There are two samples including the full sample and the subsample of working women. Because the data are fairly recent (2007), the labour 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 the 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

Table 3. Estimation results of relative effects.

		Full sample	Working subsample
Men	Mean	-3.595	-2.345
	Median	-3.575	-3.090
	Std dev.	(7.199)	(17.285)
Women	Mean	0.084	-1.146
	Median	1.037	-1.146
	Std dev.	(4.377)	(0.948)

Note: The estimates reported here are the relative effect of the potential wage and the nonlabour income on hours worked. 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 this table. As a robustness check, we also use $c = 4$ in Table 4 and refer to it as double bandwidth. The instrumental variable in these specifications is the father’s education.

Table 4. Estimation results of relative effects (double bandwidth).

		Full sample	Working subsample
Men	Mean	-15.231	-22.170
	Median	-8.515	-11.200
	Std dev.	(36.997)	(74.714)
Women	Mean	42.923	-7.942
	Median	23.258	-7.942
	Std dev.	(54.092)	(2.288)

Note: The estimates reported here are the relative effect of the potential wage and the nonlabour income on hours worked. 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 3. As a robustness check, we use $c = 4$ in this table. The instrumental variable in these specifications is the father’s education.

$$y_i = m(\theta x^* + w_i, \eta_i),$$

$$x_i = x_i^* + \varepsilon_i,$$

$$x_i^* = g(z_i, w_i) + u_i.$$

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 nonlabour income. Using the tools that we introduced in the earlier sections to estimate the labour-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 nonlabour income of each sample member. An increase in nonlabour 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 θ . In other words, an increase in the wage has an income and a substitution effect but an increase in nonlabour income only has an income effect so that we can use the direction of this effect.

Tables 3 and 4 show the results of the estimation of the parameter θ . 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

Table 5. Estimated signs of the labour supply elasticities.

		Full sample	Working subsample
Men	Mean	+	+
	Median	+	+
Women	Mean	-	+
	Median	-	+

Note: Because the direction of the effect of the nonlabour 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 θ .

the estimated coefficients. Therefore, we focus on the sign of the labour-supply elasticity. Table 5 presents the estimated sign of the labour-supply elasticities by assuming that the effect of the nonlabour income on hours of work 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; that is, 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. We apply these tools to estimate the labour-supply elasticity, and find that the labour-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 that often has to be dealt 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.

ACKNOWLEDGEMENTS

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.

REFERENCES

Ashenfelter, O. and J. J. Heckman (1974). The estimation of income and substitution effects in a model of family labor supply. *Econometrica* 42, 73–86.

- Blundell, R., A. Duncan and C. Meghir (1998). Estimating labor supply responses using tax reforms. *Econometrica* 66, 827–61.
- Blundell, R. and T. MaCurdy (1999). Labor supply: a review of alternative approaches. In O. C. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics, Volume 3A*, 1559–695. Amsterdam: North-Holland.
- Borjas, G. J. (2009). *Labor Economics* (5th ed.). New York, NY: McGraw-Hill.
- Card, D. and D. R. Hyslop (2005). Estimating the effects of a time-limited earnings subsidy for welfare-leavers. *Econometrica* 73, 1723–70.
- Chesher, A. (2003). Identification in nonseparable models. *Econometrica* 71, 1405–41.
- Chernozhukov, V., G. W. Imbens and W. K. Newey (2007). Instrumental variable estimation of nonseparable models. *Journal of Econometrics* 139, 4–14.
- Fan, J. and I. Gijbels (1996). *Local Polynomial Modelling and Its Applications: Monographs on Statistics and Applied Probability, Volume 66*. London: Chapman and Hall.
- Hoderlein, S. and R. Sherman (2015). Identification and estimation in a correlated random coefficients binary response model. *Journal of Econometrics* 188, 135–149.
- Horowitz, J. L. (1998). *Semiparametric Methods in Econometrics*. New York, NY: Springer.
- Hu, Y. (2008). Identification and estimation of nonlinear models with misclassification error using instrumental variables: a general solution. *Journal of Econometrics* 144, 27–61.
- Hu, Y. and S. M. Schennach (2008). Instrumental variable treatment of nonclassical measurement error models. *Econometrica* 76, 195–216.
- Imbens, G. W. and W. K. Newey (2009). Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica* 77, 1481–512.
- Keane, M. P. (2011). Labor supply and taxes: a survey. *Journal of Economic Literature* 49, 961–1075.
- Killingsworth, M. and J. J. Heckman (1986). Female labor supply: a survey. In O. C. Ashenfelter and R. Layard (Eds.), *Handbook of Labor Economics, Volume 1*, 103–204. Amsterdam: North-Holland.
- Pencavel, J. (1986). Labor supply of men: a survey. In O. C. Ashenfelter and R. Layard (Eds.), *Handbook of Labor Economics, Volume 1*, 3–102. Amsterdam: North-Holland.
- Pencavel, J. (2002). A cohort analysis of the association between work hours and wages among men. *Journal of Human Resources* 37, 251–74.
- Schennach, S. M. (2007). Instrumental variable estimation of nonlinear errors-in-variables models. *Econometrica* 75, 201–39.
- Schennach, S. M., H. White and K. Chalak (2012). Local indirect least squares and average marginal effects in nonseparable structural systems. *Journal of Econometrics* 166, 282–302.

APPENDIX A

The following example shows that a naïve 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 that we include an example to illustrate this. Consider the following linear version of the proposed model

$$\begin{aligned}
 y &= \theta x^* + w + \eta, \\
 x &= x^* + \varepsilon, \\
 x^* &= g(z, w) + u.
 \end{aligned}$$

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 naïve regression of y on x yields the following estimator for θ :

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

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 (η, ε, 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 naïve regression depends on several factors including the covariance of x and w , the covariances of the measurement error ε between the disturbance η and the endogenous error u , and the endogenous term $\text{cov}(u, \eta)$. It is easy to see that the sign of θ is not identified by the naïve regression. If the measurement error ε is uncorrelated to the disturbance η and the endogenous error u with small variation, and if the covariance between x and w is 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 η) is strong enough to reverse the sign of the effect under the following conditions:

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

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher’s website:

Replication files