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

December, 2013

Abstract

Economic variables are often measured with an error and may be endogenous. This paper gives new identification results for the linear index model with measurement error and endogeneity. It also shows how 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 negative while the elasticity for married women with no or one child is positive.

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

JEL codes: C20, J20, C27

*Correspondence address: Yingyao Hu, Department of Economics, 3400 N. Charles Street, Baltimore, MD 21218; email: yhu@jhu.edu.

†We thank Richard Blundell and Ronald Oaxaca for helpful comments and discussions and Stephanie Barana for research assistance.

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 is measured with some error. 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. In particular, we allow for measurement error and endogeneity and consider the following semiparametric model,

$$y = m(\theta x^* + w, \eta), \tag{1}$$

$$x = x^* + \varepsilon, \tag{2}$$

$$x^* = g(z, w) + u, \tag{3}$$

where we observe $\{y, x, z, w\}$ and $\theta \neq 0$. The measurement error ε may be correlated with η 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 θ which measures the relative effects for continuous variables x^* and w.[1] In the next section, we show that the parameter θ 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. 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 do not work. Thus, we allow labor force participation to change as the wage changes. In general, we allow equation (2) to be replaced by the requirement that the function $g(z, w)$ in equation (3) is identified.

If we ignore measurement error in equation (2), i.e., $x = x^*$, the proposed model becomes a single index control function model which is closely related to Imbens and Newey (2009).[2]

[1] The ratio of partial effects corresponding to x^* and w on y can be given by: $\frac{\partial m(\theta x^* + w, \eta)/\partial x^*}{\partial m(\theta x^* + w, \eta)/\partial w} = \theta$ if η is independent of x^* and w.

[2] There may exist a correlation between x^* and η in equation (1) and the nonseparable term $g(z, w)$ can be regarded as a control function to because $g(z, w)$ is independence of η and a correlation between x^* and η should come from a correlation between u and η.

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 equation (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 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 negative while the coefficient for married women with no or one child is positive. Thus, our findings for men are in agreement with Pencavel (2002) who used cohorts to estimate the labor supply elasticity for men but we also estimate the labor supply elasticity for women with children, for which Pencavel's (2002) methodology is not suitable.

This paper is organized as follows. We discuss identification and estimation of the model in section 2. We then show how these tools can be used to determine whether the labor supply elasticity is positive or negative in section 3. We present a simulation study in section 4 and our empirical results in section 5. We discuss possible extensions in section 6 and conclude in section 7.

2. Single Index Models

In this section, we consider the identification of θ. We assume that the error terms (u, η, ε) in equations (1)-(3) are independent of z, w. We normalize $E(u)$ in equation (3) to be zero. 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

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

Taking the derivative of the above equation with respective to z, w, we have

$$\begin{aligned} \frac{\partial E[y|z,w]}{\partial w} &= \int m'(\theta g(z,w) + w + \theta u, \eta)\{\theta \frac{\partial}{\partial w} g(z,w) + 1\} f_{u,\eta}(u,\eta) du d\eta \\ &= [\theta \frac{\partial}{\partial w} g(z,w) + 1] \int m'(\theta g(z,w) + w + \theta u, \eta) f_{u,\eta}(u,\eta) du d\eta, \end{aligned} \tag{4}$$

and

$$\begin{aligned} \frac{\partial E[y|z,w]}{\partial z} &= \int m'(\theta g(z,w) + w + \theta u, \eta) \frac{\partial}{\partial z} \theta g(z,w) f_{u,\eta}(u,\eta) du d\eta \\ &= \theta \frac{\partial}{\partial z} g(z,w) \int m'(\theta g(z,w) + w + \theta u, \eta) f_{u,\eta}(u,\eta) du d\eta. \end{aligned} \tag{5}$$

Dividing equation (4) by equation (5) to eliminate $\int m'(\theta g(z,w) + w + \theta u, \eta) f_{u,\eta}(u,\eta) du d\eta$, we obtain

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

It follows that

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

2.1. Multivariate case

Note that the method developed in this section can be applied to a more general case containing two mismeasured independent variables. For example, the parametric model has the form where the variables $\{y, x, w, z, \tilde{z}\}$ are observed. We assume $E[\varepsilon|z, \tilde{z}]$, and $E[u|z, \tilde{z}]$ are all equal to zeros and $(\eta, u_1, u_2, \varepsilon_1, \varepsilon_2)$ is independent of $(z, \tilde{z})$. We assume

Assumption 2.1. *An i.i.d. random sample $\{y, x, w, z, \tilde{z}\}$ satisfies*

$$\begin{aligned} y &= m(\theta x^* + w^*, \eta), \\ x &= x^* + \varepsilon_1, \\ x^* &= g_1(z, \tilde{z}) + u_1, \\ w &= w^* + \varepsilon_2, \\ w^* &= g_2(z, \tilde{z}) + u_2, \end{aligned}$$

where (x, w) are the measurement of the latent explanatory variables (x^*, w^*), $(z, \tilde{z})$ are instruments, the functions $m(\cdot,\cdot)$, $g_1(\cdot,\cdot)$, $g_2(\cdot,\cdot)$ 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 the last assumption, we can replace $x = x^* + \varepsilon_1$ by the assumption that $g_1(z, \tilde{z})$ is identified and we can 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 2.2. *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

(6)
$$\begin{aligned} \frac{\partial E[y|z,\tilde{z}]}{\partial \tilde{z}} &= \int m'(\theta g_1(z,\tilde{z}) + g_2(z,\tilde{z}) + \theta u_1 + u_2, \eta)[\theta \frac{\partial}{\partial \tilde{z}} g_1(z,\tilde{z}) + \frac{\partial}{\partial \tilde{z}} g_2(z,\tilde{z})] f_{u_1,u_2,\eta}(u_1,u_2,\eta) du_1 du_2 d\eta \\ &= [\theta \frac{\partial}{\partial \tilde{z}} g_1(z,\tilde{z}) + \frac{\partial}{\partial \tilde{z}} g_2(z,\tilde{z})] \int m'(\theta g_1(z,\tilde{z}) + g_2(z,\tilde{z}) + \theta u_1 + u_2, \eta) f_{u_1,u_2,\eta}(u_1,u_2,\eta) du_1 du_2 d\eta, \end{aligned}$$

and

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

Assumption 2.3. *There exists a set $\mathcal{Z}$ in the support of $(z, \tilde{z})$ such that $\Pr(\mathcal{Z}) > 0$ and for any $(z, \tilde{z}) \in \mathcal{Z}$, $\frac{\partial}{\partial z} E[y|z,\tilde{z}] \neq 0$ 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 were $E[y|z,\tilde{z}] = aE[x|z,\tilde{z}] + b$ for some a, b, which rarely happens. Dividing equation (4) by equation (5) 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

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

Under Assumption 2.3, we may then solve for θ. We summarize the results as follows:

Theorem 2.1. *Suppose that assumptions 2.1, 2.2, and 2.3 hold. Then the parameter of interest θ is identified. In particular,*

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

Notice that the result degenerates to the previous case when there is no measurement error in w, i.e., $w = \tilde{z} = w^*$. Also, suppose that 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 θ. Härdle (1990) and Härdle and Linton (1994) provide an overview of nonparametric estimators for such derivatives and we state an estimator in the next section.

2.2. Estimation

The kernel estimator of the joint densities $f(z, w)$ is

$$\widehat{f}(z,w) = \frac{1}{n\sigma^2} \sum_{i=1}^{n} K(\frac{z_i - z}{\sigma}) K(\frac{w_i - w}{\sigma}).$$

Also, let

$$\widehat{g}(z,w) = \frac{1}{n\sigma^2} \sum_{i=1}^{n} y_i K(\frac{z_i - z}{\sigma}) K(\frac{w_i - w}{\sigma}).$$

Then the Nadarya-Watson kernel regression estimator of the conditional mean $E[y|z,w]$ is

$$\widehat{m}(z,w) = \frac{\sum_{i=1}^{n} y_i K(\frac{z_i - z}{\sigma}) K(\frac{w_i - w}{\sigma})}{\sum_{i=1}^{n} K(\frac{z_i - z}{\sigma}) K(\frac{w_i - w}{\sigma})} = \frac{\widehat{g}(z,w)}{\widehat{f}(z,w)}.$$

This suggests that a consistent estimator for the derivative of $E[y|z,w]$ with respect to z can be obtained by

$$\frac{\partial \widehat{E[y|z,w]}}{\partial z} = \frac{1}{2\delta}[\widehat{m}(z+\delta,w) - \widehat{m}(z-\delta,w)]$$

using a small value for δ. This suggests the following estimator for θ,

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

Define $\overline{\delta}$ as the vector of all δs in the construction of $\widehat{\theta}$. Write $y = E[y|z,w] + \xi_1$ and $x = E[x|z,w] + \xi_2$ such that $E[\xi_i|z,w] = 0$ for $i = 1,2$. Assume

Assumption 2.4. *For each i, the error term ξ_i is i.i.d. with zero mean and the variance of ξ_i conditional on w, z is uniformly finite for each i, i.e., $E(\xi_i^2|z,w) < c < \infty$ for some positive c.*

Assumption 2.5. *The conditional expectations $E[y|z,w]$ and $E[x|z,w]$ and the joint density function $f(z,w)$ are twice continuously differential in a neighbor of each point (z,w).*

Assumption 2.6. *The bandwidth $\sigma \to 0$, $\overline{\delta} \to 0$ and $n\sigma^2 \to \infty$ as $n \to \infty$.*

Assumption 2.7. *Let the kernel function $K(\psi)$ be the class of all Borel measurable, bounded, real-valued function satisfying (i)$\int K(\psi)d\psi = 1$, (ii)$\int |K(\psi)|d\psi < \infty$, (iii)$|\psi||K(\psi)| \to 0$ as $|\psi| \to \infty$, (iv)$\sup|K(\psi)| < \infty$, and (v)$\int K(\psi)^2 d\psi < \infty$,*

Theorem 2.2. *When assumptions 2.1-2.7 hold, $\underset{n\to\infty}{plim}\widehat{\theta} = \theta$.*

Proof. See the appendix.

3. Application: the Labor Supply Elasticity

A major difficulty for estimating the labor supply elasticity is that the wages and the number of hours worked is 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 section 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,

$$\ln(wage_i) = schooling_i \beta + W_i \gamma + \varepsilon_i \tag{7}$$

where the error term ε_i is assumed to be independent of the exogenous regressors in the vector W_i and also independent of the instrument Z_i. One can use 2SLS or IV quantile regression to estimate the parameters in the wage equation. After estimating the model, one can then calculate the predicted log wage,

$$\widehat{\ln wage}_i = schooling_i \hat{\beta} + W_i \hat{\gamma}.$$

Let the potential wage, the wage that somebody could earn if that person would work, is 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^*,

$$labor\ supply_i = m(\theta \cdot x_i^* + S_i \delta, \eta_i), \tag{8}$$

where S_i is a vector of exogenous regressors. Note that using $\widehat{\ln wage}$ and 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,

$$hours_i = h(\theta \cdot x_i^* + H_i\kappa, \tau_i).$$

where H_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. We found that the Nadarya-Watson kernel estimator performs well, and discuss this in more detail in the next section.

4. Simulation

This section studies the finite sample performance of the proposed estimator for the parameter of interest θ. The simulation models in this study are based on nonparametric estimation of derivatives by kernel methods given in Section 2.2. 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

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

The wage is observed with measurement error ε,

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

There are two DGPs in the experiment:

$$\begin{aligned}
\text{DGP I: } & g(z, w) = 2 + 0.09z, \\
& m(\theta x^* + w, \eta) = exp(\theta x^* + w + \eta), \\
\text{DGP II: } & g(z, w) = 2 + 0.09z^2 + w, \\
& m(\theta x^* + w, \eta) = (\theta x^* + w)^3 + \eta,
\end{aligned}$$

where $z \sim N(0,1)$, $w \sim N(0,1)$, $u \sim N(0,1)$, $\varepsilon \sim N(0,1/10)$, $\eta \sim N(0,1)$ 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: $\theta = -0.5, -1$, and -1.5. In addition, three different sample sizes N are considered: 500, 1000, 2000. We used 100 simulation replications for for each sample size in order to estimate the standard errors.

Tables 1, 2 and 3 present the simulation results of the proposed kernel estimator. The simulation results of DGP I (the exponential case) show a small downward bias. In DGP I, the mean and median of coefficients are almost the same. In addition, the standard errors of θ are larger if the absolute value of θ is larger. As for DGP II, when $\theta = -1$, the bias and standard error are relative small compared to the other two cases. For $\theta = -1.5$, the mean and median of the estimator are somewhat different, implying that the estimator has a somewhat skewed distribution. Nevertheless, the mean and median estimation values are close to the true values and standard errors do not vary much. Thus, the Monte Carlo simulation in this study shows that the proposed estimator works very well.

5. Empirical Analysis

In 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 variables are measured with error and/or are 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 number of children, 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. Using the tools that we introduce in the earlier sections to estimate the labor supply elasticity, we find that the coefficient for married men is negative while the coefficient for married women with no or one child is positive.[3] Thus, our findings for men are in agreement with Pencavel (2002) who used cohorts to estimate the labor supply elasticity for men. However, our methodology also recovers the sign of the labor supply elasticity for women with children, for which Pencavel's (2002) methodology is not suitable. 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. Thus, for men, the number of hours worked is decreasing in the wage, i.e. the income effect

[3] We use the normal multivariate kernel with the bandwidth as the Silverman's rule of thumb, i.e.$h = c \cdot n^{1/5}$. The constant $c = 1$ in most cases. As a robustness check, we also try $c = 2$ and $c = 3$ and refer to those cases as double and triple bandwidth respectively.

('buy more leisure') is larger than the substitution effect ('leisure is more expensive if the wage increases'). For women with children, the income effect is smaller than the substitution effect, so that the labor supply elasticity is positive. We use the normal multivariate kernel with the bandwidth as the Silverman's rule of thumb. If we use double or triple the bandwidth of this rule of thumb then we describe it as double or triple bandwidth respectively.

6. Extensions

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.

7. Conclusion

This paper gives new identification results for the single index model with measurement error and endogeneity and also shows how to estimate the single index parameter. The paper applies these tools to estimate the labor supply elasticity and finds that the coefficient for married men is negative while the coefficient for married women with no or one child is positive. The new estimator allows for endogeneity and measurement error, a situation one often has to deal with in empirical work.

A. Appendix: Proof of Theorem 2.2

:

Assumptions 2.3 and 2.6 imply that theorem 2.2 holds if we can show that $\underset{n\to\infty}{plim}\widehat{m}(z,w) = E[y|z,w]$. Using the standard multivariate kernel density techniques, we know that $\underset{n\to\infty}{plim}\widehat{f}(z,w) = f(z,w)$. This reduces the problem into having to prove that $\underset{n\to\infty}{plim}\widehat{g}(z,w) = E[y|z,w]f(z,w)$ since $\underset{n\to\infty}{plim}\widehat{m}(z,w) = \frac{\underset{n\to\infty}{plim}\,\widehat{g}(z,w)}{\underset{n\to\infty}{plim}\,\widehat{f}(z,w)}$. Thus, we start with deriving an expression for $\widehat{g}(z,w)$. Set $\overline{z} = (z_1,...,z_n)$ and $\overline{w} = (w_1,...,w_n)$.

$$
\begin{aligned}
\mathrm{E}[\widehat{g}(z,w)] &= \mathrm{E}[\frac{1}{n\sigma^2}\sum_{i=1}^{n} y_i K(\frac{z_i - z}{\sigma})K(\frac{w_i - w}{\sigma})] \\
&= \mathrm{E}\left[\mathrm{E}[\frac{1}{n\sigma^2}\sum_{i=1}^{n} y_i K(\frac{z_i - z}{\sigma})K(\frac{w_i - w}{\sigma})|\overline{z},\overline{w}]\right] \\
&= \mathrm{E}\left[\frac{1}{n\sigma^2}\sum_{i=1}^{n} K(\frac{z_i - z}{\sigma})K(\frac{w_i - w}{\sigma})\mathrm{E}[y_i|\overline{z},\overline{w}]\right] \\
&= \frac{1}{\sigma^2}\int K(\frac{z_1 - z}{\sigma})K(\frac{w_1 - w}{\sigma})\mathrm{E}[y|z_1,w_1]]f(z_1,w_1)dz_1dw_1 \\
&= \int K(\psi_1)K(\psi_2)\mathrm{E}[y|z+\sigma\psi_1, w+\sigma\psi_2]]f(z+\sigma\psi_1, w+\sigma\psi_2)d\psi_1 d\psi_2.
\end{aligned}
$$

By Assumptions 2.5 and 2.7, we obtain that $\mathrm{E}[\widehat{g}(z,w)]$ converges to $E[y|z,w]f(z,w)$. Also, we can write the variance of $\widehat{g}(z,w)$ as

$$
\begin{aligned}
&\mathrm{var}[\widehat{g}(z,w)] \\
&= \mathrm{E}[\mathrm{var}(\widehat{g}(z,w)|\overline{z},\overline{w})]] + \mathrm{var}[\mathrm{E}(\widehat{g}(z,w)|\overline{z},\overline{w})]] \\
&= \frac{1}{n^2\sigma^4}\mathrm{E}(\sum_{i=1}^{n} K^2(\frac{z_i - z}{\sigma})K^2(\frac{w_i - w}{\sigma})\mathrm{E}(\xi_{1i}^2|z,w)) + \mathrm{var}[\frac{1}{n\sigma^2}\sum_{i=1}^{n} K(\frac{z_i - z}{\sigma})K(\frac{w_i - w}{\sigma})\mathrm{E}[y|z_i,w_i]],
\end{aligned}
$$

where we have used (i) $\widehat{g}(z,w) - \mathrm{E}(\widehat{g}(z,w)|\overline{z},\overline{w}) = \frac{1}{n\sigma^2}\sum_{i=1}^{n} K(\frac{z_i-z}{\sigma})K(\frac{w_i-w}{\sigma})\xi_{1i}$, and the error term ξ_{1i} is i.i.d.. The first term of $\mathrm{var}[\widehat{g}(z,w)]$ can be further written as

$$
\begin{aligned}
&\mathrm{E}[\mathrm{var}(\widehat{g}(z,w)|\overline{z},\overline{w})]] \\
&= \frac{1}{n\sigma^2}\int K^2(\psi_1)K^2(\psi_2)\mathrm{E}[\xi_{1i}^2|z+\sigma\psi_1, w+\sigma\psi_2]]f(z+\sigma\psi_1, w+\sigma\psi_2)d\psi_1 d\psi_2,
\end{aligned}
$$

which converges to zero as $n \to \infty$. Similarly, the second term becomes

$$
\begin{aligned}
&\mathrm{var}[\frac{1}{n\sigma^2}\sum_{i=1}^{n} K(\frac{z_i - z}{\sigma})K(\frac{w_i - w}{\sigma})\mathrm{E}[y|z_i, w_i]] \\
&= \frac{1}{n^2\sigma^4}\mathrm{E}[\left(\sum_{i=1}^{n} K(\frac{z_i - z}{\sigma})K(\frac{w_i - w}{\sigma})\mathrm{E}[y|z_i, w_i]\right)^2] \\
&\quad - \left(\mathrm{E}[\frac{1}{n\sigma^2}\sum_{i=1}^{n} K(\frac{z_i - z}{\sigma})K(\frac{w_i - w}{\sigma})\mathrm{E}[y|z_i, w_i]\right)^2 \\
&= \frac{1}{n\sigma^2}\int K^2(\psi_1)K^2(\psi_2)\mathrm{E}[y|z_i, w_i]f(z + \sigma\psi_1, w + \sigma\psi_2)d\psi_1 d\psi_2 \\
&\quad + \frac{n(n-1)}{n^2}\left(\int K(\psi_1)K(\psi_2)\mathrm{E}[y|z_i, w_i]f(z + \sigma\psi_1, w + \sigma\psi_2)d\psi_1 d\psi_2\right)^2 \\
&\quad - \left(\int K(\psi_1)(\psi_2)\mathrm{E}[y|z_i, w_i]f(z + \sigma\psi_1, w + \sigma\psi_2)d\psi_1 d\psi_2\right)^2,
\end{aligned}
$$

which also converges to zero as $n \to \infty$. Applying the Chebychev's inequality yields the desired result $\underset{n\to\infty}{plim}\widehat{g}(z, w) = E[y|z, w]f(z, w)$. QED.

Table 1: Finite Sample Performance of the Kernel Estimator (N=500)

DGP	MEAN	MEDIAN	STD
DGP I			
$\theta = -0.5$	-0.529	-0.534	0.191
$\theta = -1$	-1.123	-1.092	0.313
$\theta = -1.5$	-1.553	-1.596	0.645
DGP II			
$\theta = -0.5$	-0.596	-0.588	0.547
$\theta = -1$	-1.094	-1.088	0.046
$\theta = -1.5$	-1.374	-1.385	0.983

Standard errors of the parameters are computed by using sample standard deviation of 100 replications.

Table 2: Finite Sample Performance of the Kernel Estimator (N=1000)

DGP	MEAN	MEDIAN	STD
DGP I			
$\theta = -0.5$	-0.541	-0.508	0.181
$\theta = -1$	-1.022	-0.999	0.298
$\theta = -1.5$	-1.624	-1.562	0.499
DGP II			
$\theta = -0.5$	-0.568	-0.519	0.170
$\theta = -1$	-1.090	-1.063	0.167
$\theta = -1.5$	-1.245	-1.371	0.907

Standard errors of the parameters are computed by using sample standard deviation of 100 replications.

Table 3: Finite Sample Performance of the Kernel Estimator (N=2000)

DGP	MEAN	MEDIAN	STD
DGP I			
$\theta = -0.5$	-0.535	-0.553	0.158
$\theta = -1$	-1.014	-1.050	0.277
$\theta = -1.5$	-1.549	-1.567	0.481
DGP II			
$\theta = -0.5$	-0.337	-0.505	0.832
$\theta = -1$	-1.052	-1.048	0.045
$\theta = -1.5$	-1.494	-1.370	0.766

Standard errors of the parameters are computed by using sample standard deviation of 100 replications.

Table 4: Sample Statistics

Variables	Full Sample		Working Women Subsample	
	Mean	Std. Dev.	Mean	Std. Dev.
No children	0.389	0.488	0.408	0.492
One child	0.242	0.429	0.237	0.426
More children	0.369	0.483	0.355	0.479
Husband's				
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
Nonlabor 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
Mother's Education	4.661	1.498	4.622	1.491
Wife's				
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
Nonlabor 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	1.8119
Mother's Education	4.814	1.566	4.809	1.564
Sample size	880		789	

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 are 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

	No Children	One Child	Children
Men	-0.229	-0.055	-0.306
	(0.006)	(0.005)	(0.005)
Women	0.665	0.700	-2.32
	(5.487)	(0.393)	(109.59)

Note: a. Standard errors are in parentheses.
b. The exogenous regressors are income, age, and an indicator for race.
c. The instrumental variable in these specifications is father's education.

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

	No Children	One Child	Children
Men (Double Bandwith)	-0.185	-0.073	-0.024
	(0.003)	(0.004)	(0.008)
Women (Double Bandwith)	-0.013	0.356	-0.692
	(0.034)	(0.178)	(0.045)
Women (Triple Bandwith)	0.083	0.093	-0.412
	(0.065)	(0.021)	(0.017)

Note: a. Standard errors are in parentheses.
b. The exogenous regressors income, age, and an indicator for race.
c. The instrumental variable in these specifications is father's education.

Table 7: Estimation Results of Relative Effects

	No Children	One Child	Children
Men	0.180	0.055	-0.401
	(0.024)	(0.005)	(0.02)
Women	0.006	-0.458	0.109
	(0.008)	(0.021)	(0.005)

Note: a. Standard errors are in parentheses.
b. The exogenous regressors are income, age, and an indicator for race.
c. The instrumental variable in these specifications is the regressor for itself, i.e. wage is assumed to be exogenous.

Table 8: Estimation Results of Relative Effects (Double Bandwith)

	No Children	One Child	Children
Men	-0.133	0.055	1.956
	(0.020)	(0.006)	(0.529)
Women	-0.008	-0.247	0.018
	(0.009)	(0.009)	(0.004)

Note: a. Standard errors are in parentheses.
b. The exogenous regressors are income, age, and an indicator for race.
c. The instrumental variable in these specifications is the regressor for itself, i.e. wage is assumed to be exogenous

References

[16] 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.

[16] Blundell, R., A. Duncan, and C. Meghir (1998): "Estimating Labor Supply Responses Using Tax Reforms," *Econometrica*, 66: 827-861.

[16] Blundell, R. and T. MaCurdy (1999): "Labor Supply: A Review of Alternative Approaches,", in *Handbook of Labor Economics*, O.C. Ashenfelter and D. Card, editors, Volume 3A, Amsterdam: Elsevier, 1559-1695.

[16] Borjas, G.J. (2009): *Labor Economics,* 5th edition, McGraw-Hill/Irwin.

[16] Card, D., and D. R. Hyslop (2005): "Estimating the Effects of a Time-Limited Earnings Subsidy for Welfare-Leavers," *Econometrica*, 73: 1723-177.

[16] Härdle, W. (1990): *Applied Nonparametric Regression.* Cambridge: Cambridge University Press.

[16] Härdle, W., and O. Linton (1994): "Applied Nonparametric Methods," in *Handbook of Econometrics,* Volume 4, edited by R. F. Engle and D. McFadden. Amsterdam: North-Holland.

[16] Horowitz, J. L. (1998): *Semiparametric Methods in Econometrics Semiparametric Methods in Econometrics,* Springer-Verlag, New York.

[16] Hu, Y. (2008): "Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution," *Journal of Econometrics*, 144: 27-61.

[16] Hu, Y. and S. M. Schennach (2008): "Instrumental variable treatment of nonclassical measurement error models," *Econometrica*, 76(1):195-216.

[16] Imbens, G. W., and Newey, W. K. (2009): "Identification and Estimation of Triangular Simultaneous Equations Models without Additivity," *Econometrica*, 77: 1481-1512.

[16] Keane, M. P. (2011): "Labor Supply and Taxes: A Survey," *Journal of Economic Literature*, 49: 961–1075.

[16] Killingsworth, M. and J. J. Heckman (1986): "Female Labor Supply: A Survey," in *Handbook of Labor Economics,* O. Ashenfelter and R. Layard, editors, Volume 1, Amsterdam and New York: North-Holland, 1986.

[16] Pencavel, J. (1986): "Male Labor Supply: A Survey," in *Handbook of Labor Economics,* O. Ashenfelter and R. Layard, editors, Volume 1, Amsterdam and New York: North-Holland, 1986.

[16] Pencavel, J. (2002): "A Cohort Analysis of the Association between Work Hours and Wages among Men," *Journal of Human Resources,* 37: 251-274.

[16] Schennach, S. M. (2007): "Instrumental Variable Estimation of Nonlinear Errors-in-Variables Models," *Econometrica*, 75: 201-239.