ISSN 0971-1554

JOURNAL OF QUANTITATIVE ECONOMICS

New Series Volume 10, No.1	***January, 2012***
Old Series Volume 26, No. 1	*January, 2010*

JOURNAL OF QUANTITATIVE ECONOMICS

(Journal of the Indian Econometric Society)

New Series Volume 10, No. 1 — ***January, 2012***

Old Series Volume 26, No. 1 — *January, 2010*

Managing Editor: **D.M. Nachane** and *Joint Managing Editors:* **G. Mythili and Rupayan Pal**

CONTENTS **Page**

Journal of Quantitative Economics, Vol. 10, No. (1), January 2012

EFFICIENT SEMIPARAMETRIC INSTRUMENTAL VARIABLE ESTIMATION UNDER CONDITIONAL HETEROSKEDASTICITY

FENG YAO[1]

Abstract

We consider the estimation of a semiparametric regression model where the data is independently and identically distributed. Our primary interest is the estimation of the parametric vector, where the associated regressors are correlated with the errors, contain both continuous and discrete variables, and the error term is conditionally heteroskedastic. Under general conditional heteroskedasticity that depends on both excluded and included exogenous variables, we propose a new estimator based on Yao and Zhang's (2011, Efficient semiparametric instrumental variable estimation, working paper, Economics Department, West Virginia University) framework and establish its asymptotic properties. It is consistent and asymptotically normally distributed. It allows the reduced form to be nonparametric and is efficient as it reaches the semiparametric efficiency bound. Furthermore, it is asymptotically equivalent to a GMM estimator that optimally select the instrumental variables with conditional moment restriction, and thus is also efficient among a class of semiparametric IV estimators. We perform a Monte Carlo study, which illustrates its finite sample properties and confirms our theoretical result.

Keywords: Instrumental variables, semiparametric regression, efficient estimation.

JEL Classifications: C14, C21

[1] Department of Economics, West Virginia University, Morgantown, WV 26505 USA
email: feng.yao@mail.wvu.edu

I am indebted to the associate editor, AmanUllah, whose detailed comments greatly improves this paper.

1. Introduction

There is now a vast and increasing literature on nonparametric and semiparametric estimation of structural models with endogenous regressors (Blundell and Powell 2003 and references). We specifically consider the literature on nonparametric and semiparametric extensions of the classical simultaneous equations models. Here, endogenous and predetermined variables are related through a system of equations, so endogeneity arises from the feedback of the dependent variable to the explanatory variable. Many recent nonparametric and semiparametric approaches account for the endogeneity in the spirit of classical two-stage least squares (2SLS) (see Matzkin 1994, 2008; Imbens and Newey 2009). The estimation and inference relax the tight parametric assumptions on the functional forms of structural equations or the error term distribution, thus are robust to potential models misspecification.

In this paper, we consider a semiparametric additive regression model

$$Y_t = g(Z_{1t}, X_t, \varepsilon_t) = m(Z_{1t}) + X_t\beta + \varepsilon_t, \qquad t = 1,\cdots,n. \qquad \text{... (1)}$$

We denote the endogenous explanatory variables explicitly by $X'_t \in \Re^K$, the included exogenous variable by $Z_{1t}' \in \Re^{l_1}$, and the excluded exogenous variable by $Z_{2t}' \in \Re^{l_2}$, with $l_1 + l_2 = l$. We explicitly consider continuous and discrete variables in $X_t = (X_t^c, X_t^d)$, where $X_t^c\ ' \in \Re^{K_c}$ are the continuous variables, $X_t^d\ ' \in \Re^{K_d}$ discrete variables, and $K_c + K_d = K$. Similarly, $Z_{1t} = (Z_{1t}^c, Z_{1t}^d)$, $Z_{2t} = (Z_{2t}^c, Z_{2t}^d)$, where $Z_{1t}^c\ ' \in \Re^{l_{1c}}$, $Z_{2t}^c\ ' \in \Re^{l_{2c}}$ are the continuous variables, and $Z_{1t}^d\ ' \in \Re^{l_{1d}}$, $Z_{2t}^d\ ' \in \Re^{l_{2d}}$ discrete variables.

The semiparametric regression model in (1) when X_t is considered exogenous is familiar in the literature (see Robinson 1988 and Speckman 1988 under iid assumption with continuous regressors; Delgado and Mora 1995 for discrete regressor and weakerconditions than those in Robinson 1988). Theoretical results on semiparametric estimation without endogenous variables are available for dependent data (see Aneiros and Quintela 2001;Aneiros-Perez and Quintela-del-Rio 2002; and the monograph on partially linear models by Härdle, Liang and Gao 2000). Li (2000), Fan et al. (1998) and Chen and Shen (1998) consider estimation in additive partially linear models, where the nonparametric component is further assumed to be additive functions.

The partially linear model in (1) provides much needed flexibility through the nonparametric control function $m(Z_{1t})$, while the endogenous variables X_t enter the model parametrically, allowing for convenient interpretation, faster convergence rateand easy implementation. First, consider the problem of estimating returns to schooling in Card (1995), where Y_t is the log of wage, X_t is the years of schooling and Z_{1t} includes potential experience (treated as exogenous), and demographic variables, i.e., race, southern residence, and residence in a standard metropolitan statistical area. Parameter β can be conveniently interpreted as the marginal return to schooling, provided there is an exogenous source of variation in the education choices for estimation. Second, since we model the parametric component in a linear fashion, relative to the fully nonparametric model $g(Z_{1t}, X_t, \varepsilon_t)$ (see Matzkin 1994; Imben and Newey 2009) or $g(Z_{1t}, X_t, \varepsilon_t) = \tilde{g}(Z_{1t}, X_t) + \varepsilon_t$ (see e.g., Das 2005; Newey and Powell 2003; Darolles et al. 2011; Newey et al. 1999; Ng and Pinkse 1995;Pinkse 2000; Florens et al. 2008), we have less variables in the nonparametric control function $m(Z_{1t})$, thus the convergence rate could be faster. For example, it can be convenient when X_t is a vector of discrete variables that take a large number of different values, or when X_t consists of categorical variables. Third, it is not a trivial task to incorporate both continuous and discrete variables as in Das (2005),Newey and Powell (2003), and Darolles et al. (2011). In the control function approach considered by Newey et al. (1999), Ng and Pinkse (1995), Pinkse (2000), and Florens et al. (2008), an important assumption is on the control variable $V_t = X_t - E(X_t \mid \mathbf{Z_t})$, where $\mathbf{Z_t} = (Z_{1t}, Z_{2t})$. The fact that X_t and V_t are functionally related generally requires that endogenous components X_t and V_t be continuously distributed with unbounded support conditional on $\mathbf{Z_t}$. The requirement that X_t is continuously distributed limits the application of the control function approach. In model (1), we can easily incorporate both continuous and discrete variables in X_t, Z_{1t} and Z_{2t}.

With endogenous variables, the precision of the estimator for β is a concern because the restrictions imposed on $m(Z_{1t})$ and on the reduced form $E(X_t \mid \mathbf{Z_t})$ are so weak. Newey (1990, 1993) considers efficient estimation of a parametrically specified structural model with conditional moment restrictions using nearest neighbor and series estimators. Ai and Chen (2003) consider a more general semiparametric model with conditional moment restrictions, where sieve estimation is employed andthe parametric component estimator reaches the semiparametric efficiency bound. Chen and Pouzo (2009) consider semiparametric efficient estimation of conditional moment models with possibly nonsmooth residuals using penalized sieve minimum distance estimator, where the *square-root-n* normality and the optimal convergence rate are achieved for the parametric and nonparametric estimators. Otsu (2007) proposes a penalized empirical likelihood estimation for semiparametric models, with the parametric component being efficiently estimated. Florens et al. (2009) consider an instrumental regression in partially linear models with endogenous variables in $m(\cdot)$, where the estimation of the parametric components could be ill-posed. A locally efficient estimators for semiparametric measurement error model is also considered by Ma and Carroll (2006), using an estimating equation approach. Efficient instrumental variable estimation of a semiparametric dynamic panel data model is considered in Li and Stengos (1996), Li andUllah (1998), Berg et al. (2001), and Baltagi and Li (2002). Yao and Zhang (2011) consider efficient semiparametric instrumental variable (IV) estimation in model (1) and propose three estimators for the parametric vector. However, whenconditional heteroskedasticity depends on both the included and excluded exogenous variables, their estimators are not efficient.

We cite Li and Racine (2007, p237) to illustrate the potential difficulty in the efficient estimation of β, “We point out that an efficient semiparametric estimator of β is quite complex. It requires estimation of a nonparametric model with dimension $p+q$ (the dimension of (X_t, Z_{1t})), while the estimation of β and $m(\cdot)$ involves only nonparametric estimation with dimension q.

Therefore, the “curse of dimensionality” may prevent researchers from applying efficient estimation procedures to a partially linear model when the error is conditionally heteroskedastic.”

In this paper, we propose a new kernel based estimator in the simple model (1) under general conditional heteroskedasticity structure. The partially linear model has been considered by many literatures, however, we notice the efficient estimation of the parametric component β associated with endogenous variables using the popular kernel estimator has not been formally considered. We show that the kernel based estimator with both continuous and discrete endogenous variables is easy to construct, and furthermore it offers the following advantages. Both β and the nonparametric component $m(\cdot)$ are easily identified or well-posed with fairly reasonable assumptions. We demonstrate the β's estimator is *square-root-n* normal and its asymptotic distribution is not impacted by $m(\cdot)$'s estimation, taking advantage of the kernel based estimator by using a higher order kernel and assuming the higher order differentiability on $m(\cdot)$. We only assume $m(\cdot)$ is smooth, but does not have to be in a compact function space. The efficient estimation requires the information of the heteroskedasticity function and we propose a simple estimator for it, whose properties are well studied in the asymptotic results. The estimator for β is feasible, can be written in explicit form and is a GMM based estimator that optimally selects the instrument variables, enabling easy construction and interpretation. We show it to be consistent, and asymptotically normally distributed. Its limiting distribution, i.e., the asymptotic variance used for inference purpose, can be easily estimated without resorting to the simulation based bootstrap procedure. Last, the new estimator is efficient as it reaches the semiparametric efficiency bounds. We notethat above mentioned papers could be applied in more general situations, however, the implementation could be involved. For example, implementing the efficient sieve minimum distance estimator proposed by Ai and Chen (2003) entails the numerical optimization under the partially linear model. Our proposed estimator can be expressed in closed form, can be easily implemented, and we demonstrate that our estimator outperforms theirs in the simulation study.

Our semiparametric model (1) could be considered as a special case of the functional coefficient instrumental variable (IV) model (2.1) in Su et al. (2011) in the sense that they allow the vector $(m(Z_{1t}),\beta')$ to be fully nonparametric. Although their model is quite general, our model (1) and the proposed estimator provide a useful alternative. First, our semiparametric specification in $(m(Z_{1t}),\beta')$ allows estimation of β at the $\sqrt{n}$ parametric rate (see $\tilde{\beta}^E$ and Theorem 2 in section 3), while the estimator for the nonparametric local parameters in Su et al. suffers from the curse of dimensionality. Second, the construction of local linear GMM estimator in Su et al. calls for the choice of IV. The chosen IV in their (2.12) is computationally simple but it may not be optimal in the sense of minimizing the asymptotic variance for the class of local linear GMM estimators under the conditional moment condition. It is not clear how to construct the optimal instruments in their framework (see Remark 1 in Su et al.). On the other hand, utilizing the semiparametric structure in $(m(Z_{1t}),\beta')$, we construct feasible efficient estimator $\tilde{\beta}^E$ for β that fully explores the moment conditions $E(\varepsilon_t \mid \mathbf{Z_t})=0$ and $E(\varepsilon_t^2 \mid \mathbf{Z_t})=\sigma^2(\mathbf{Z_t})$. We show in section 3 that $\tilde{\beta}^E$ is asymptotically equivalent to a GMM estimator that optimally selects the IV, which is given by $-\frac{\tilde{W}_t}{\sigma^2(\mathbf{Z}_t)}$, where $\tilde{W}_t$ is defined in assumption A6(2) (see the discussion following Theorem 3). We show it is semiparametrically efficient as it reaches the semiparametric efficiency bound. Third, we notice our model (1) further allows both continuous and discrete endogenous variables X_t.

In what follows, we introduce the semiparametric model and propose the new estimator in section 2, provide the asymptotic properties in section 3, perform a Monte Carlo study to demonstrate its finite sample properties and to compare with previously considered estimators in section 4, and conclude in section 5. The tables and graphs are relegated to appendix 1, and all proofs are collected in Yao (2011).

2. Semiparametric Model

Consider the model in Equation (1) and assume the existence of instrumental variables $\mathbf{Z_t} = (Z_{1t}, Z_{2t})$ with $E(\varepsilon_t \mid \mathbf{Z_t}) = 0$ for all t. Suppose the true conditional expectation $E(Y_t \mid Z_{1t}) = m(Z_{1t}) + E(X_t \mid Z_{1t})\beta$ is known. Hence, we could subtract $E(Y_t \mid Z_{1t})$ from (1) to obtain

$$Y_t - E(Y_t \mid Z_{1t}) = (X_t - E(X_t \mid Z_{1t}))\beta + \varepsilon_t. \qquad \text{... (2)}$$

The conditional expectations are generally unknown, but we can replace them with nonparametric conditional mean estimators $\hat{E}(Y_t \mid Z_{1t})$ and $\hat{E}(X_t \mid Z_{1t})$. However, due to the correlation between ε_t and X_t, we can not apply Robinson's (1988) estimator by regressing $Y_t - \hat{E}(Y_t \mid Z_{1t})$ on $X_t - \hat{E}(X_t \mid Z_{1t})$. Since $E(X_t - E(X_t \mid Z_{1t}) \mid \mathbf{Z_t})$ is a function of $\mathbf{Z_t}$ and is uncorrelated with ε_t, Yao and Zhang (2011) propose to use $\hat{E}(X_t \mid \mathbf{Z_t}) - \hat{E}(X_t \mid Z_{1t})$ as the instrumental variables, where $\hat{E}(X_t \mid \mathbf{Z_t})$ is a nonparametric estimator for $E(X_t \mid \mathbf{Z_t})$, and estimate β by

$$\tilde{\beta} = (\hat{W}'\hat{W})^{-1}\hat{W}'(\hat{E}(Y \mid \vec{Z}) - \hat{E}(Y \mid \vec{Z}_1)). \qquad \text{... (3)}$$

where $\hat{W} = \begin{bmatrix} \hat{W}_1 \\ \hat{W}_2 \\ \vdots \\ \hat{W}_n \end{bmatrix} = \begin{bmatrix} \hat{W}_{1,1} & \hat{W}_{1,2} & \vdots & \hat{W}_{1,K} \\ \hat{W}_{2,1} & \hat{W}_{2,2} & \cdots & \hat{W}_{2,K} \\ \cdots & \vdots & \vdots & \vdots \\ \hat{W}_{n,1} & \hat{W}_{n,2} & \cdots & \hat{W}_{n,K} \end{bmatrix}, \hat{W}_{t,k} = \hat{E}(X_{t,k} \mid \mathbf{Z_t}) - \hat{E}(X_{t,k} \mid Z_{1t})$,

and $X_{t,k}$ is the kth element of random vector X_t. Furthermore, $\hat{E}(Y \mid \vec{Z}) = (\hat{E}(Y_1 \mid \mathbf{Z}_1), \hat{E}(Y_2 \mid \mathbf{Z}_2), \cdots, \hat{E}(Y_n \mid \mathbf{Z_n}))'$, and $\hat{E}(Y \mid \vec{Z}_1) = (\hat{E}(Y_1 \mid Z_{11}), \hat{E}(Y_2 \mid Z_{12}), \cdots, \hat{E}(Y_n \mid Z_{1n}))'$. When conditional heteroskedasticity depends only on the included exogenous variables, that is, $E(\varepsilon_t^2 \mid \mathbf{Z_t}) = \sigma^2(Z_{1t})$, they propose to account for the heteroskedasticity by

$$\tilde{\beta}^H = (\hat{W}'\hat{\Omega}^{-1}(\vec{Z}_1)\hat{W})^{-1}\hat{W}'\hat{\Omega}^{-1}(\vec{Z}_1)(\hat{E}(Y \mid \vec{Z}) - \hat{E}(Y \mid \vec{Z}_1)), \qquad \text{... (4)}$$

where the conditional variance is estimated with $\hat{\sigma}^2(Z_{1t}) = \hat{E}(\tilde{\varepsilon}_t^2 \mid Z_{1t})$, with $\tilde{\varepsilon}_t = Y_t - \hat{E}(Y_t \mid Z_{1t}) - (X_t - \hat{E}(Y_t \mid Z_{1t}))\tilde{\beta}$. The conditional covariance matrix $\Omega(\vec{Z}_1)$ is estimated by $\hat{\Omega}(\vec{Z}_1)$, which is a diagonal matrix with the $t-\text{th}$ element as $\hat{\sigma}^2(Z_{1t})$. These estimators are consistent and asymptotically normally distributed. Under conditional homoskedasticity, $\tilde{\beta}$ is efficient relative to previously considered estimators, i.e., Li and Stengos (1996), and reach the semiparametric efficiency bound in Chamberlain (1992). Under special conditional heteroskedasticity that depends only on the included exogenous variables, they show $\tilde{\beta}^H$ also reaches the semiparametric efficiency bound.

However, conditional heteroskedasticity generally depends on the included and excluded exogenous variables, i.e., $E(\varepsilon_t^2 \mid \mathbf{Z_t}) = \sigma^2(\mathbf{Z_t})$. Let's consider the returns to schooling example in the introduction. The heteroskedasticity may depend not only on the included exogenous variables, i.e., the potential experience, and indicators for race, southern residence, and residence in a standard metropolitan statistical area, but also on the excluded exogenous variable, i.e., the proximity to a four year college. Thus, there is a need for the efficient estimator to account for the information properly. First, let's examine the semiparametric efficiency bound derived in Chamberlain (1992). Under assumption A1(2) in section 3 that $E(\varepsilon_t \mid \mathbf{Z_t}) = 0$ for all t, our Equation (1) considers the estimation of β in the model of $\varepsilon_t = Y_t - m(Z_{1t}) - X_t\beta$ with the conditional moment restriction $E(\varepsilon_t \mid \mathbf{Z_t}) = 0$. Since $\frac{\partial \varepsilon_t}{\partial \beta'} = -X_t$, define $D_0(\mathbf{Z_t}) \equiv E(\frac{\partial \varepsilon_t}{\partial \beta'} \mid \mathbf{Z_t}) = -E(X_t \mid \mathbf{Z_t})$, $\Sigma_0(\mathbf{Z_t}) \equiv E(\varepsilon_t \varepsilon'_t \mid \mathbf{Z_t}) = \sigma^2(\mathbf{Z_t})$ with the general heteroskedasticity structure in assumption A4(3), $H_0(\mathbf{Z_t}) \equiv E(\frac{\partial \varepsilon_t}{\partial r} \mid \mathbf{Z_t}) = -1$ for $r = m(Z_{1t})$. The inverse of the semiparametric efficiency bound for β is

$$\begin{aligned}
J_0 = \; & E\{E(D_0(\mathbf{Z_t})'\Sigma_0(\mathbf{Z_t})^{-1}D_0(\mathbf{Z_t}) \mid Z_{1t}) - E(D_0(\mathbf{Z_t})'\Sigma_0(\mathbf{Z_t})^{-1}H_0(\mathbf{Z_t}) \mid Z_{1t}) \\
& \times [E(H_0(\mathbf{Z_t})'\Sigma_0(\mathbf{Z_t})^{-1}H_0(\mathbf{Z_t}) \mid Z_{1t})]^{-1}E(H_0(\mathbf{Z_t})'\Sigma_0(\mathbf{Z_t})^{-1}D_0(\mathbf{Z_t}) \mid Z_{1t})\} \\
= \; & E\{[E(X_t \mid \mathbf{Z_t}) - (E(\tfrac{1}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t}))^{-1}E(\tfrac{E(X_t \mid \mathbf{Z_t})}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t})]' \tfrac{1}{\sigma^2(\mathbf{Z_t})} \\
& \times [E(X_t \mid \mathbf{Z_t}) - (E(\tfrac{1}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t}))^{-1}E(\tfrac{E(X_t \mid \mathbf{Z_t})}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t})]\} \\
= \; & E(\tilde{W}_t' \tfrac{1}{\sigma^2(\mathbf{Z_t})} \tilde{W}_t),
\end{aligned}$$

and the semiparametric efficiency bound is J_0^{-1}. Define $W_t = E(X_t \mid \mathbf{Z_t}) - E(X_t \mid Z_{1t})$. Under homoskedasticity, $E(\varepsilon_t \varepsilon'_t \mid \mathbf{Z_t}) = \sigma^2$, then $J_0^{-1} = (\frac{1}{\sigma^2}E(W_t' W_t))^{-1}$. When theheteroskedasticity function depends only on the included exogenous variables, $E(\varepsilon_t \varepsilon_t' \mid \mathbf{Z_t}) = \sigma^2(Z_{1t})$, $J_0^{-1} = (E(\frac{1}{\sigma^2(Z_{1t})} W'_t W_t))^{-1}$. They are the asymptotic variance of estimators $\tilde{\beta}$ and $\tilde{\beta}^H$, derived in Theorem 1 and 4 respectively in Yao and Zhang (2011). But the semiparametric efficiency bound is different under general heteroskedasticity structure. Second, inspection of the construction of

$\tilde{\beta}^H$ indicates the semiparametric efficiency bound is not achieved. So we propose a new estimator that accomplishes this goal.

To motivate this estimator, let's consider Equation (1), which together with assumption A1(2) that $E(\varepsilon_t \mid \mathbf{Z_t}) = 0$ implies $E(\frac{Y_t}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t}) = E(\frac{1}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t})m(Z_{1t}) + E(\frac{X_t}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t})\beta$. Here we use the fact that $E(\frac{\varepsilon_t}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t}) = E(E(\frac{\varepsilon_t}{\sigma^2(\mathbf{Z_t})} \mid \mathbf{Z_t}) \mid Z_{1t}) = E(\frac{1}{\sigma^2(\mathbf{Z_t})} E(\varepsilon_t \mid \mathbf{Z_t}) \mid Z_{1t}) = 0$. Thus, given $\sigma^2(\mathbf{Z_t}) > C > 0$ in assumption A6(1), we multiply above by $(E(\frac{1}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t}))^{-1}$, then subtract it from Equation (1) to obtain

$$Y_t - (E(\frac{1}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t}))^{-1} E(\frac{Y_t}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t}) = (X_t - (E(\frac{1}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t}))^{-1} E(\frac{X_t}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t}))\beta + \varepsilon_t. \quad \text{... (5)}$$

Suppose the conditional expectations are known. However, due to the correlation between X_t and ε_t , we cannot regress $Y_t - (E(\frac{1}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t}))^{-1} E(\frac{Y_t}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t})$ on $X_t - (E(\frac{1}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t}))^{-1} E(\frac{X_t}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t})$. Different from Yao and Zhang (2011), we use estimates of $E(X_t \mid \mathbf{Z_t}) - (E(\frac{1}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t}))^{-1} E(\frac{X_t}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t})$ as instrumental variables. Since

$$E(Y_t \mid \mathbf{Z_t}) - (E(\frac{1}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t}))^{-1} E(\frac{Y_t}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t}) = (E(X_t \mid \mathbf{Z_t}) - (E(\frac{1}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t}))^{-1} E(\frac{X_t}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t}))\beta, \quad \text{... (6)}$$

we consider

$$\tilde{\beta}^E = (\tilde{W}^{F\prime} \hat{\Omega}^{-1}(\vec{Z}) \tilde{W}^F)^{-1} \tilde{W}^{F\prime} \hat{\Omega}^{-1}(\vec{Z}) [\hat{E}(Y \mid \vec{Z}) - \hat{E}^*(Y \mid \vec{Z}_1)]. \quad \text{... (7)}$$

Here, $\tilde{W}^F$ is a $n \times K$ matrix with the (t,k)th element $\tilde{W}^F_{t,k} = \hat{E}(X_{t,k} \mid \mathbf{Z_t}) - (\hat{E}(\frac{1}{\hat{\sigma}^2(\mathbf{Z_t})} \mid Z_{1t}))^{-1} \hat{E}(\frac{X_{t,k}}{\hat{\sigma}^2(\mathbf{Z_t})} \mid Z_{1t})$. $\hat{\sigma}^2(\mathbf{Z_t}) = \hat{E}(\tilde{\varepsilon}_t^2 \mid \mathbf{Z_t})$, where $\tilde{\varepsilon}_t = Y_t - \hat{E}(Y_t \mid Z_{1t}) - (X_t - \hat{E}(X_t \mid Z_{1t}))\tilde{\beta}$, where $\tilde{\beta}$ estimator is consistent under our assumptions. The diagonal matrix $\hat{\Omega}(\vec{Z})$ has the $t-\text{th}$ element as $\hat{\sigma}^2(\mathbf{Z_t})$. Furthermore, $\hat{E}^*(Y \mid \vec{Z}_1)$ is a $n \times 1$ vector with the $t-\text{th}$ element as $\hat{E}^*(Y_t \mid Z_{1t}) \equiv (\hat{E}(\frac{1}{\hat{\sigma}^2(\mathbf{Z_t})} \mid Z_{1t}))^{-1} \hat{E}(\frac{Y_t}{\hat{\sigma}^2(\mathbf{Z_t})} \mid Z_{1t})$.

We use different instrument variables in $\tilde{W}^F$ and transformed regressand $\hat{E}^*(Y \mid \vec{Z}_1)$ in constructing $\tilde{\beta}^E$, compared to $\tilde{\beta}^H$ in Yao and Zhang. The estimator is constructed by weighting the data in a specialfashion. However, we note first that the weighting scheme enables us to remove the $m(\cdot)$ in equation (5). Second, the weighting scheme is what is called for by the semiparametric efficiency bound. Third, we demonstrate in next section that $\tilde{\beta}^E$ is a GMM based estimator that optimally selects the instrumental variables. We will show that $\tilde{\beta}^E$ is consistent, asymptotically normally distributed with variance J_0^{-1}, thus it reaches the semiparametric efficiency bound.

Here, we define the density of Z_{1t} at z_{10} as $f_1(z_{10})$, and the density of $\mathbf{Z_t}$ at $\mathbf{z}_0 = (z_{10}, z_{20})$ as $f(\mathbf{z}_0)$. We estimate them with the Rosenblatt density estimators with both continuous and discrete variables, use the Nadaraya-Watson estimators for $E(A_0 \mid z_{10})$, and $E(A_0 \mid \mathbf{z}_0)$ for its simplicity. Specifically,

$$\hat{f}_1(z_{10}) = \frac{1}{nh_1^{l_{1c}}} \sum_{t=1}^{n} K_1\left(\frac{Z_{1t}^c - z_{10}^c}{h_1}\right) I(Z_{1t}^d = z_{10}^d),$$

$$\hat{f}(\mathbf{z}_0) = \frac{1}{nh_2^{l_{1c}+l_{2c}}} \sum_{t=1}^{n} K_2\left(\frac{Z_{1t}^c - z_{10}^c}{h_2}, \frac{Z_{2t}^c - z_{20}^c}{h_2}\right) I(Z_{1t}^d = z_{10}^d, Z_{2t}^d = z_{20}^d),$$

$$\hat{E}(A_0 \mid z_{10}) = \frac{\frac{1}{nh_1^{l_{1c}}} \sum_{t=1}^{n} K_1\left(\frac{Z_{1t}^c - z_{10}^c}{h_1}\right) I(Z_{1t}^d = z_{10}^d) A_t}{\hat{f}_1(z_{10})},$$

$$\hat{E}(A_0 \mid \mathbf{z}_0) = \frac{\frac{1}{nh_2^{l_{1c}+l_{2c}}} \sum_{t=1}^{n} K_2\left(\frac{Z_{1t}^c - z_{10}^c}{h_2}, \frac{Z_{2t}^c - z_{20}^c}{h_2}\right) I(Z_{1t}^d = z_{10}^d, Z_{2t}^d = z_{20}^d) A_t}{\hat{f}(\mathbf{z}_0)},$$

where h_1 and h_2 are bandwidths which go to zero as $n \to \infty$. $K_1(\cdot), K_2(\cdot)$, and $I(\cdot)$ are the kernel and indicator functions. We expect the use of the local linear estimator will work and produce the same asymptotic theory as well due to its well known properties. Furthermore, for ease of notation, we suppress the use of a trimming function that trim out small $\hat{f}_1(\cdot)$ and $\hat{f}(\cdot)$, or replace them with a small but positive constant. It avoids the technical difficulty due to the random denominators $\hat{f}_1(\cdot)$ and $\hat{f}(\cdot)$, which could be small. This will not change the asymptotic results so we will not introduce it in the definition explicitly.

3. Asymptotic Properties

Similar to the parametric instrumental variable estimation, our semiparametric instrumental variable estimator is likely to be biased. We investigate its asymptotic properties with the following assumptions.

We let C denote a generic constant below, which can vary from one place to another. Let $G = G_1^c \times G_1^d \times G_2^c \times G_2^d \subset \Re^l$, G is compact. $G_1 = G_1^c \times G_1^d \subset \Re^{l_1}$ and $G_2 = G_2^c \times G_2^d \subset \Re^{l_2}$.

Let's denote a generic function $g(Z_{1t}) \in C_1^s$ if $g(Z_{1t})$ is s times continuously differentiable w.r.t. Z_{1t}^c, with its sth order derivative uniformly continuous on G_1^c, and for $|j| = 1, 2, \cdots, s$, $\sup_{Z_{1t} \in G_1} \left| \frac{\partial^j}{\partial (Z_{1t}^c)^j} g(Z_{1t}) \right| < \infty$. Here, we denote the $|j|$th order derivative as $\frac{\partial^j}{\partial (Z_{1t}^c)^j} g(Z_{1t}) \equiv \frac{\partial^{|j|} g(Z_{1t}^c, Z_{1t}^d)}{\partial (Z_{1t,1}^c)^{j_1} \partial (Z_{1t,2}^c)^{j_2} \cdots \partial (Z_{1t,l_{1c}}^c)^{j_{l_{1c}}}}$. We adopt the notation that $j = (j_1, j_2, \cdots, j_{l_{1c}})'$, $|j| = \sum_{i=1}^{l_{1c}} j_i$. For future purposes, we denote $\sum_{0\le|j|\le s} = \sum_{|j|=0} + \sum_{|j|=1} + \cdots + \sum_{|j|=s}$, $j! = j_1! \times j_2! \times \cdots \times j_{l_{1c}}!$, $(Z_{1t}^c)^j = (Z_{1t,1}^c)^{j_1} \times (Z_{1t,2}^c)^{j_2} \times \cdots \times (Z_{1t,l_{1c}}^c)^{j_{l_{1c}}}$, where $Z_{1t,i}^c$ refers to the $i-\text{th}$ element in Z_{1t}^c.

Denote a generic function $g(Z_{1t}) \in C_1^{s_1}$ if $g(Z_{1t})$ is s_1 times continuously differentiable w.r.t. Z_{1t}^c, with its sth and s_1th order derivative uniformly continuous on G_1^c, and for $|j| = 1, 2, \cdots, s_1$, $\sup_{Z_{1t}\in G_1} |\frac{\partial^j}{\partial (Z_{1t}^c)^j} g(Z_{1t})| < \infty$. Denote a generic function $g(\mathbf{Z_t}) \in C_{1,2}^{s_1}$ if $g(\mathbf{Z_t})$ is s_1 times continuously differentiable w.r.t. $\mathbf{Z_t^c} = (Z_{1t}^c, Z_{2t}^c)$, with its sth and s_1th order derivative uniformly continuous on $G^c = G_1^c \times G_2^c$, and for $|j| = 1, 2, \cdots, s_1$, $\sup_{\mathbf{Z_t}\in G} |\frac{\partial^j}{\partial (\mathbf{Z_t^c})^j} g(\mathbf{Z_t})| < \infty$. Here, we denote the $|j|$th order derivative as $\frac{\partial^j}{\partial (\mathbf{Z_t^c})^j} g(\mathbf{Z_t}) \equiv \frac{\partial^{|j|} g(Z_{1t}^c, Z_{1t}^d, Z_{2t}^c, Z_{2t}^d)}{\partial (Z_{1t,1}^c)^{j_1} \partial (Z_{1t,2}^c)^{j_2} \cdots \partial (Z_{1t,l_{1c}}^c)^{j_{l_{1c}}} \partial (Z_{2t,1}^c)^{j_{l_{1c}+1}} \partial (Z_{2t,2}^c)^{j_{l_{1c}+2}} \cdots \partial (Z_{2t,l_{2c}}^c)^{j_{l_{1c}+l_{2c}}}}$.

A1: (1) $\{Y_t, X_t, \mathbf{Z_t}\}_{t=1}^n$ is an independent and identically distributed (iid) sequence of random vectors related as in Equation (1). (2) $E(\varepsilon_t \mid \mathbf{Z_t}) = 0$ for all t.

(3) Let $W_t = E(X_t \mid \mathbf{Z_t}) - E(X_t \mid Z_{1t})$, $E(W_t' W_t)$ is a symmetric and positive definite matrix.

A2: (1) Denote the density of Z_{1t} at z_{10} by $f_1(z_{10}^c, z_{10}^d)$. Assume $f_1(z_{10}^c, z_{10}^d) \in C_1^s$.

(2) $0 < C < f_1(z_{10}^c, z_{10}^d) < \infty$, for all $z_{10} \in G_1$.

(3) $X_{t,k} = E(X_{t,k} \mid Z_{1t}) + X_{t,k} - E(X_{t,k} \mid Z_{1t}) = g_{1,k}(Z_{1t}) + e_{1,kt}$, and $\forall z_{10} \in G_1$, assume $g_{1,k}(z_{10}) \in C_1^{s_1}$. The conditional density of Z_{1t} given $e_{1,kt}$ is bounded, and the conditional density of $X_{t,k}$ given Z_{1t} is continuous at Z_{1t}^c.

(4) Denote the density of $\mathbf{Z_t}$ at $\mathbf{z}_0$ by $f_z(z_{10}^c, z_{10}^d, z_{20}^c, z_{20}^d) = f_z(\mathbf{z}_0)$, and assume $f_z(\mathbf{z}_0) \in C_{1,2}^{s_1}$.

(5) $0 < C < f_z(\mathbf{z}_0) < \infty$, for all $\mathbf{z}_0 \in G$.

(6) $X_{t,k} = E(X_{t,k} \mid \mathbf{Z_t}) + X_{t,k} - E(X_{t,k} \mid \mathbf{Z_t}) = g_k(\mathbf{Z_t}) + e_{kt}$, and $\forall \mathbf{z}_0 \in G$ assume $g_k(\mathbf{z}_0) \in C_{1,2}^{s_1}$. The conditional density of $\mathbf{Z_t}$ given e_{kt} is bounded, and the conditional density of $X_{t,k}$ given $\mathbf{Z_t}$ is continuous at $\mathbf{Z_t^c}$. (7) $m(z_{10}) \in C_1^{s_1}$.

A3: (1) For $x \in \Re^d$, $d = l_{1c}$ or l_{2c}, the kernel function $K(x)(K_1(x)$ or $K_2(x))$ is bounded with bounded support, and it is of order $3s_1$.

(2) $|u^i K(u) - v^i K(v)| \le C_K \| u - v \|$, $|i| = 0, 1, 2, \cdots, s_1$.

A4: (1) For some $\delta > 0$, $E(X_{t,k}^{2+\delta} \mid \mathbf{Z_t}), E(X_{t,k}^{2+\delta} \mid Z_{1t}) < \infty$, $|g_k(\mathbf{Z_t})|, |g_{1,k}(Z_{1t})| < \infty$ almost everywhere.

(2) $E(\varepsilon_t^{2+\delta} \mid \mathbf{Z_t}), E(\varepsilon_t^{2+\delta} \mid Z_{1t}) < \infty$. (3) $E(\varepsilon_t^2 \mid \mathbf{Z_t}) = \sigma^2(\mathbf{Z_t})$.

(4) The conditional density of Z_{1t} given ε_t is bounded, and the conditional density of ε_t given Z_{1t} is continuous at Z_{1t}^c. The conditional density of $\mathbf{Z_t}$ given ε_t is bounded, and the conditional density of ε_t given $\mathbf{Z_t}$ is continuous at $\mathbf{Z_t^c}$.

A5: (1) $nh_1^{2l_{1c}} \to \infty$. (2) $nh_2^{2(l_{1c}+l_{2c})} \to \infty$. (3) $nh_1^{2(s+1)}, nh_2^{2(s_1+1)} \to 0$.

A6: (1) (i) $0 < C < \sigma^2(\mathbf{Z_t}) < \infty$. (ii) $\sigma^2(\mathbf{Z_t}) \in C_{1,2}^{s_1}$.

(2) Define $\tilde{W}_t = E(X_t \mid \mathbf{Z_t}) - (E(\frac{1}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t}))^{-1} E(\frac{E(X_t \mid \mathbf{Z_t})}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t})$. Let $E(\tilde{W}_t{}' \frac{1}{\sigma^2(\mathbf{Z_t})} \tilde{W}_t)$ bea symmetric and positive definite matrix.

(3) For some $\delta < 0$, $E(|X_{t,k}|^{4+\delta} | \mathbf{Z_t}) < \infty$, and $E(|\varepsilon_t|^{4+\delta} | \mathbf{Z_t}) < \infty$. The conditional density of $\mathbf{Z_t}$ given $|X_{t,k}\varepsilon_t|$ is bounded, and the conditional density of $|X_{t,k}\varepsilon_t|$ given $\mathbf{Z_t}$ is continuous at $\mathbf{Z_t^c}$. The conditional density of $\mathbf{Z_t}$ given $e_{1,kt}e_{1,k't}$ is bounded, and the conditional density of $|X_{t,k}X_{t,k'}|$ given $\mathbf{Z_t}$ is continuous at $\mathbf{Z_t^c}$ for all $k, k' \in \{1, \cdots, K\}$.

(4) We assume $E(\frac{1}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t}) \in C_1^{s_1}$ and $E(\frac{E(X_{t,k} \mid \mathbf{Z_t})}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t}) \in C_1^{s_1}$ for $k = \{1, \cdots, K\}$.

Assumptions A1-A5 are used in Yao and Zhang to identify parameter β and to obtain asymptotic properties of estimators $\tilde{\beta}$. Since our new estimator $\tilde{\beta}^E$ is constructed with $\tilde{\beta}$, we adopt them here. In A1(2), we require the conditional expectation of the error term ε_t given $\mathbf{Z_t}$ to be zero, but we allow ε_t's and X_t's to be possibly correlated, and thus $\mathbf{Z_t}$ plays the role of instrumental variables. A1(3) guarantees that β in equation (1) is identified. First, X_t cannot contain a constant. Second, it implies that $E(X_t \mid \mathbf{Z_t}) \neq E(X_t \mid Z_{1t})$. Since $\mathbf{Z_t} = (Z_{1t}, Z_{2t})$, any element of X_t cannot be perfectly a.s. predictable by Z_{1t}, i.e., X_t cannot be some function of Z_{1t} only. Obviously, Z_{2t} cannot simply be a linear combination of Z_{1t}, so Z_{2t} needs to contain variables that are linearly independent of Z_{1t}. A1(3) forbids more general forms of dependence. Third, because W_t cannot be a.s. zero, no elements of $E(X_t \mid \mathbf{Z_t}) - E(X_t \mid Z_{1t})$ are multicollinear. This fails if X_t is collinear. Alternative but similar condition A6(3) also guarantees that β is identified.

Assumptions A2(1), (2), (4), (5), and (7) require the densities $f_1(z_{10})$, $f_z(\mathbf{z}_0)$, and $m(z_{10})$ to be continuously differentiable w.r.t. its continuous components and bounded. These assumptions are commonly used in nonparametric kernelregression, enabling the use of Taylor expansion (Martins-Filho and Yao 2007). They are similar in spirit to the smoothness and boundedness condition in Definition 2 of Robinson (1988), or the assumption A1 of Li and Stengos (1996). A2(3) and (6) explicitly assume the relationship between X_t and Z_{1t} and between X_t and $\mathbf{Z_t}$. Similar assumptions have been maintained in Aneiros and Quintela (2001) and Speckman (1988) in the fixed design case. Assumption A3 requires the kernel functionto be smooth and bounded (Martins-Filho and Yao 2007). Since the kernel function depends on n, the asymptotic distribution of $\tilde{\beta}$ is established using Liapunov's central limit theorem, with conditional moments assumption of ε_t and X_t given $\mathbf{Z_t}$ or Z_{1t} in A4. The bandwidth assumptions A5(1) and (2) are in line with those used in the literature (Martins-Filho and Yao 2007). A3, together with A5(3), specifies the kernel properties and the rate of decay for the bandwidths. They are used to control the bias introduced in the nonparametric regression, which is similar to assumptions in, for example, Robinson (1988) and Li and Stengos (1996). However, A5(3) is stronger than that maintained in Li and Stengos, Li andUllah, or Robinson. As $\tilde{\beta}$ involves estimation of W_t, the bias arises not only from the estimation of $E(X_t \mid Z_{1t})$, but also from the estimation of $E(X_t \mid \mathbf{Z_t})$. A5 requires choosing a higher order kernel to eliminate the biasasymptotically, so the asymptotic distribution is not impacted by the estimation of $m(\cdot)$. However, $m(\cdot)$ may not be estimated at the optimal rate. But given the important work on twicing kernels by Newey et al. (2004), we expect using the twicing kernels will enable us to estimate $m(\cdot)$ at its optimal rate and maintain $\sqrt{n}$ normality of estimators for β. Asymptotic properties of $\tilde{\beta}$ are obtained for general heteroskedasticity structure in A4(3). We do not need theparameter space for β to be compact, as the model is partially linear. In the GMM setting, for example, it is well known that the compactness can be relaxed if the objective function is concave.

For efficient estimation using $\tilde{\beta}^E$, we place additional assumptions in A6. Different from $\tilde{\beta}^H$ in Yao and Zhang, we allow the heteroskedasticity to depend on both included and excluded exogenous variables. A6(1)(i) require $\sigma^2(\mathbf{Z_t})$ to be bounded away from zero. (ii) requires the heteroskedasticity function to be smooth in the sense that $\sigma^2(\mathbf{Z_t}) \in C^{s_1}_{1,2}$. Because estimated $(\sigma^2(\mathbf{Z_t}))^{-1}$ is used in constructing $\tilde{\beta}^E$, A6(1) enables us to perform Taylor expansion and to guarantee that the estimator will behave properly. A6(2) requires the semiparametric information bound $J_0 = E(\tilde{W}_t{}'\frac{1}{\sigma^2(\mathbf{Z_t})}\tilde{W}_t)$ be a symmetric and positive definite matrix. From Equation (6), we obtain

$$\beta = (E(\tilde{W}_t{}'\tfrac{1}{\sigma^2(\mathbf{Z_t})}\tilde{W}_t))^{-1}E(\tilde{W}_t{}'\tfrac{1}{\sigma^2(\mathbf{Z_t})}[E(Y_t \mid \mathbf{Z_t}) - (E(\tfrac{1}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t}))^{-1}E(\tfrac{E(Y_t \mid \mathbf{Z_t})}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t})]).$$

Since conditional expectation is identified, β is identified in Equation (1) with A6(2). Assumption A6(3) provides higher moments, additional boundedness and smoothness conditions on the conditional distribution, enabling us to obtain the asymptotic results for $\tilde{\beta}^E$, which involves estimation of the conditional covariance matrix of ε_t. Assumption A6(4) further requires

$E(\frac{1}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t})$ and $E(\frac{E(X_t|\mathbf{Z_t})}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t})$ to be smooth, so that we could use Taylor expansion. Lemma 1 in the Appendix 2 of Yao (2011) establishes the order in probability of certain linear combinations of kernel functions that appear repeatedly in the component expressions of our estimators. Lemma 2 collects useful results obtained in Yao and Zhang (2011). We use them in the proofs of the Theorems below.

First, suppose $\sigma^2(\mathbf{Z_t})$ is known and we consider the infeasible estimator as

$$\tilde{\beta}^I = (\tilde{W}^{I\prime}\Omega^{-1}(\vec{Z})\tilde{W}^I)^{-1}\tilde{W}^{I\prime}\Omega^{-1}(\vec{Z})[\hat{E}(Y \mid \vec{Z}) - \hat{E}^{*I}(Y \mid \vec{Z}_1)], \qquad \text{... (8)}$$

where $\tilde{W}^I$ is a $n \times K$ matrix with the (t,k)th element $\tilde{W}^I_{t,k} = \hat{E}(X_{t,k} \mid \mathbf{Z_t}) - (\hat{E}(\frac{1}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t}))^{-1}\hat{E}(\frac{X_{t,k}}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t})$. The diagonal matrix $\Omega(\vec{Z})$ has its $t-\text{th}$ element as $\sigma^2(\mathbf{Z_t})$. Furthermore, $\hat{E}^{*I}(Y \mid \vec{Z}_1)$ is a $n \times 1$ vector with the $t-\text{th}$ element as $\hat{E}^{*I}(Y_t \mid Z_{1t}) \equiv (\hat{E}(\frac{1}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t}))^{-1}\hat{E}(\frac{Y_t}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t})$. Thus, the true heteroskedasticity information is utilized in $\tilde{W}^I$, $\Omega(\vec{Z})$ and $\hat{E}^{*I}(Y \mid \vec{Z}_1)$. The asymptotic property of $\tilde{\beta}^I$ given in Theorem 1represents the "oracle" property.

Theorem 1 Given assumptions A1-A5, A6(1) and A6(2), we have

$$\sqrt{n}(\tilde{\beta}^I - \beta) \xrightarrow{d} N(0, ((E(\tilde{W}_t{}'\frac{1}{\sigma^2(\mathbf{Z_t})}\tilde{W}_t))^{-1}).$$

Theorem 1 shows the infeasible estimator $\tilde{\beta}^I$ reaches the semiparametric efficiency bound. Since the structure of $\sigma^2(\mathbf{Z_t})$ is generally unknown, we estimate it in $\tilde{\beta}^E$ with $\hat{\sigma}^2(\mathbf{Z_t}) = \hat{E}(\tilde{\varepsilon}_t^2 \mid \mathbf{Z_t})$, where $\tilde{\varepsilon}_t = Y_t - \hat{E}(Y_t \mid Z_{1t}) - (X_t - \hat{E}(X_t \mid Z_{1t}))\tilde{\beta}$. In Lemma 3 in Appendix 2 of Yao (2011), we establish the uniform convergence result of $\hat{\sigma}^2(\mathbf{Z_t})$. We use it in Theorem 2 to show $\tilde{\beta}^E$ and $\tilde{\beta}^I$ are asymptotically equivalent at rate $\sqrt{n}$, thus $\tilde{\beta}^E$ inheritsthe "oracle" property of $\tilde{\beta}^I$ and we conclude $\tilde{\beta}^E$ reaches the semiparametric efficiency bound as well.

Theorem 2 Given assumptions A1-A6, we have

$$(1)\ \sqrt{n}(\tilde{\beta}^I - \tilde{\beta}^E) = o_p(1).$$

$$(2)\ \text{Given Theorem 1, we have } \sqrt{n}(\tilde{\beta}^E - \beta) \xrightarrow{d} N(0, ((E(\tilde{W}_t{}'\frac{1}{\sigma^2(\mathbf{Z_t})}\tilde{W}_t))^{-1}).$$

To conduct statistical inference on the parametric vector β, we need a consistent estimator of the covariance matrix. We provide one such estimator, which consistently estimates each component of $(E(\tilde{W}_t{}'\frac{1}{\sigma^2(\mathbf{Z_t})}\tilde{W}_t))^{-1}$ given in Theorem 2. Note from Lemma 3 in the Appendix

2 of Yao (2011), a consistent estimator of $\sigma^2(\mathbf{Z_t})$ already exists. We propose to estimate $(E(\tilde{W}_t'\frac{1}{\sigma^2(\mathbf{Z_t})}\tilde{W}_t))^{-1}$ with $(\frac{1}{n}\tilde{W}^{F}{}'\hat{\Omega}^{-1}\tilde{W}_t^F)^{-1}$. Theorem 3 below shows it is a consistent estimator.

Theorem3 Given assumption A1-A6, we have $(\frac{1}{n}\tilde{W}^{F}{}'\hat{\Omega}^{-1}\tilde{W}_t^F)^{-1} - (E(\tilde{W}_t'\frac{1}{\sigma^2(\mathbf{Z_t})}\tilde{W}_t))^{-1} = o_p(1)$.

To shed light on the theoretical results above, let us consider a class of semiparametric IV estimators based on the model in Equation (1) that satisfies the conditional moment restriction $E(\varepsilon_t \mid \mathbf{Z_t}) = 0$, where

$\varepsilon_t = \varepsilon_t(\beta) = Y_t - \left(E\left(\frac{1}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t}\right)\right)^{-1} E\left(\frac{Y_t}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t}\right) - \left(X_t - \left(E\left(\frac{1}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t}\right)\right)^{-1} E\left(\frac{X_t}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t}\right)\right)\beta$ as in Equation (5). Then, $\tilde{\beta}^E$ is asymptotically equivalent to the generalized method of moments (GMM) estimator that optimally selects instrument variables. Thus, $\tilde{\beta}^E$ is efficient among this class of semiparametric IV estimators in the sense that its asymptotic variance is smallest. To establish this, suppose $E\left(\frac{Y_t}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t}\right)$, $E\left(\frac{X_t}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t}\right)$ and $E\left(\frac{1}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t}\right)$ are known, or can be consistently estimated at a certain rate, then let $H(\mathbf{Z_t})$ denote an $h \times 1$ vector of functions of $\mathbf{Z_t}$, $h \geq K$. By law of iterated expectation we have $E(H(\mathbf{Z_t})\varepsilon_t(\beta)) = 0$. Following Yao and Zhang (2011), we could construct the IV estimators using the GMM estimator. It is defined as

$$\beta^{GMM} = \text{argmin}_\beta \hat{g}_n(\beta)'\hat{P}\hat{g}_n(\beta), \quad \hat{g}_n(\beta) = \frac{1}{n}\sum_{t=1}^{n} H(\mathbf{Z_t})\varepsilon_t(\beta),$$

for $h \times h$ positive semidefinite matrix $\hat{P}$, which may be random. Since $\varepsilon_t(\beta)$ is linear in β, by solving the minimization problem we easily obtain,

$$\begin{aligned}
&\sqrt{n}(\beta^{GMM} - \beta) \\
= &\left[\frac{1}{n}\sum_t \left(X_t - \left(E\left(\frac{1}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t}\right)\right)^{-1} E\left(\frac{X_t}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t}\right)\right)' H(\mathbf{Z_t})'\hat{P}\right. \\
&\left.\times \frac{1}{n}\sum_t H(\mathbf{Z_t})\left(X_t - \left(E\left(\frac{1}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t}\right)\right)^{-1} E\left(\frac{X_t}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t}\right)\right)\right]^{-1} \\
&\times \frac{1}{n}\sum_t \left(X_t - \left(E\left(\frac{1}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t}\right)\right)^{-1} E\left(\frac{X_t}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t}\right)\right)' H(\mathbf{Z_t})'\hat{P}\sqrt{n}\frac{1}{n}\sum_t H(\mathbf{Z_t})\varepsilon_t,
\end{aligned}$$

Assume $\hat{P} \xrightarrow{p} P$, a positive semi-definite matrix,

$\frac{1}{n}\sum_t H(\mathbf{Z_t})\left(X_t - \left(E\left(\frac{1}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t}\right)\right)^{-1} E\left(\frac{X_t}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t}\right)\right) \xrightarrow{p} -E(H(\mathbf{Z_t})\frac{\partial\varepsilon_t(\beta)}{\partial\beta}) = -G$, and define $V = EH(\mathbf{Z_t})\varepsilon_t\varepsilon_t{}'H(\mathbf{Z_t})' = E\sigma^2(\mathbf{Z_t})H(\mathbf{Z_t})H(\mathbf{Z_t})'$, then

$$\sqrt{n}(\beta^{GMM} - \beta) \xrightarrow{d} N(0,(G'PG)^{-1}G'PVP'G(G'PG)^{-1}).$$

Let $A = G'PH(\mathbf{Z_t})\varepsilon_t(\beta)$, $B = -\left[E\left(X_t - \left(E\left(\frac{1}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t}\right)\right)^{-1} E\left(\frac{X_t}{\sigma^2(\mathbf{Z_t})} \mid Z_{1t}\right) \mid \mathbf{Z_t}\right)\right]' \frac{\varepsilon_t(\beta)}{\sigma^2(\mathbf{Z_t})}$, then

$$(G'PG)^{-1}G'PVP'G(G'PG)^{-1}) - (E(\tilde{W}_t{}'\frac{1}{\sigma^2(\mathbf{Z_t})}\tilde{W}_t))^{-1}$$
$$= E\{(EAB')^{-1}[A - (EAB')(EBB')^{-1}B][A' - B'(EBB')^{-1}(EBA')](EBA')^{-1}\}$$

which is a quadratic form, so the difference is positive semidefinite. We note the asymptotic variances will be the same if we let the optimal instrumental variable to be $H(\mathbf{Z_t}) = -\frac{\tilde{W}_t}{\sigma^2(\mathbf{Z_t})}$ and let $P = V^{-1}$. Thus, asymptotically, $\tilde{\beta}^E$ behaves like a GMM estimator that optimally selects the instrumental variables. Comparing $\tilde{\beta}^E$ and β^{GMM} from a technical point of view, we notice there are extra steps taken in $\tilde{\beta}^E$, i.e., X_t and Y_t are further projected with the conditional expectation operator $E(\cdot \mid \mathbf{Z_t})$. This step is well justified. By performing an expansion of β^{GMM}, we notice it contains an extra term relative to $\tilde{\beta}^E$. The magnitude of this extra term can be controlled asymptotically, but the existence of it impacts β^{GMM}'s finite sample performance in separate experiments we run, thus it is not considered further in the Monte Carlo study. Chen et al.(2003) provides a general framework and sufficient conditions for asymptotic properties of a class of semiparametric estimators. Our result further illustrates careful choices of instrumental variables can lead to efficient estimators.

4. Monte Carlo Study

To implement our efficient semiparametric instrumental variable estimator, and to evaluate its finite sample performance, we perform a Monte Carlo study. The simulation is conducted on a data-generating process in Baltagi and Li(2002) and Yao and Zhang (2011), adapted to the iid set-up as

$$Y_t = \beta X_t + \alpha_1 Z_{1t} + \alpha_2 Z_{1t}^2 + \varepsilon_t, \text{ and } X_t = g_i(Z_{1t}, Z_{1t}) + U_t.$$

Here, we fix $\beta = \alpha_1 = \alpha_2 = 1$, so the nonlinear function $m(Z_{1t})$ is $Z_{1t} + Z_{1t}^2$. We generate the exogenous variables Z_{1t} and Z_{2t} independently from a standard normal distribution, truncated to $[-1,1]$. We select two functions for g_i, where $g_1(Z_{1t}, Z_{2t}) = Z_{1t} + Z_{2t}$ and $g_2(Z_{1t}, Z_{2t}) = Z_{1t}^2 + Z_{2t}^2$. Hence, the endogenous variable X_t depends on $\mathbf{Z_t}$ linearly with $g_1(Z_{1t}, Z_{2t})$, while X_t related to $\mathbf{Z_t}$ in a nonlinear fashion in $g_2(Z_{1t}, Z_{2t})$. Conditional on $\mathbf{Z_t}$, we generate ε_t and U_t from a bivariate normal distribution with zero mean, variance $\sigma_i^2(\mathbf{Z_t})$, and correlation θ. We truncate ε_t and U_t to $[-1,1]\times[-1,1]$. We choose $\theta = 0.2, 0.5$ and 0.8. As θ increases, the correlation between X_t and ε_t increases, thus endogeneity is magnified. We consider $\sigma_1^2(\mathbf{Z_t}) = (Z_{1t} + Z_{2t})^2$, and $\sigma_2^2(Zt) = \exp(Z_{1t} + 2Z_{2t})$ for the heteroskedasticity case, and $\sigma_3^2(\mathbf{Z_t}) = 1$ for the homoskedasticity case. It is easy to verify that assumptions maintained in A1, A2, A4 and A6 are satisfied. We consider two sample sizes, $n = 200$ and 400, and perform 1000 repetitions for each experimental design.

To implement out estimator $\tilde{\beta}^E$, we need to select the bandwidth sequences h_1 and h_2. We select the bandwidth $\hat{h}_1$ using the*rule-of-thumb*data driven plug-in method of Ruppert et al. (1995). We select $\hat{h}_2$ with $1.25SD(\mathbf{Z_t})n^{-\frac{1}{6}}$, where $SD(\mathbf{Z_t})$ is the standard deviation of $\mathbf{Z_t}$. We choose a second order Epanechnikov kernel, which satisfies our assumption A3(2) and part of A3(1). Assumption A3(1) further requires higher order kernel to make the bias disappear as in A5(3). With the choice of bandwidth and kernel function, our assumption A5(3) is violated. Thus, we investigate the robustness of our estimators against a popular kernel function of lower order.

Aside from the estimator we propose, we also include the semiparametric estimator $\breve{\beta}^{(1)}$ without considering the endogenous variable as in Robinson, two estimators $\tilde{\beta}$ and $\tilde{\beta}^H$ proposed in Yao and Zhang (2011). $\breve{\beta}^{(1)}$ serves as the benchmark because it ignores the endogeneity problem. Following the 2SLS literatures, we expect larger bias but smaller variance for $\breve{\beta}^{(1)}$ relative to the other estimators considered. The performance of $\tilde{\beta}$ and $\tilde{\beta}^H$ relative to other popular estimators, i.e., Li and Stengos (1996), has been investigated in Yao and Zhang (2011). It indicates $\tilde{\beta}$ and $\tilde{\beta}^H$ perform better, especially when $g_i(Z_{1t}, Z_{2t})$ is nonlinear and heteroskedasticity is present. So we do not consider them here. We evaluate the performance of each estimator using bias (B), standard deviation (S), and root mean squared error (R) as criteria. In Appendix 1, we summarize the experiment results with $\sigma_1^2(\mathbf{Z_t}) = (Z_{1t} + Z_{2t})^2$ in Table 1 for g_1 and Table 2 for g_2. The results with $\sigma_2^2(\mathbf{Z_t}) = \exp(Z_{1t} + 2Z_{2t})$ are provided in Tables 3 and 4. The homoskedastic results with $\sigma_3^2(\mathbf{Z_t}) = 1$ are listed in Tables 5 and 6.

We notice all estimators carry positive bias. As the sample size n increases, estimators $\tilde{\beta}^E$, $\tilde{\beta}$ and $\tilde{\beta}^H$ perform better in terms of smaller bias, standard deviation and root mean squared error. This observation confirms our asymptotic results in Theorems 1, 2 as well as the asymptotic results obtained in Yao and Zhang (2011) that the three estimators are consistent. $\breve{\beta}^{(1)}$'s performance does not improve because its bias does not go down as sample size increases. It is expected since $\breve{\beta}^{(1)}$ ignores the endogeneity problem. As θ increases, the endogeneity problem is magnified and correspondingly, it is more difficult to estimate β. We observe all estimators' bias increases. Though their standard deviation drops slightly, the drop is dominated by the increase in bias. Thus the root mean square error increases with θ. It is harder to estimate β when $g_2(Z_{1t}, Z_{2t})$ is in the data-generating process, as indicated by the larger magnitudes on bias, standard deviation and root mean squared error for all estimators in $g_2(Z_{1t}, Z_{2t})$ (Tables 2, 4, and 6) relative to those in $g_1(Z_{1t}, Z_{2t})$ (Tables 1, 3, and 5).

Across all experiment designs, $\breve{\beta}^{(1)}$ carries the largest bias, generally smallest standard deviation, but largest root mean squared error, relative to the other estimators considered. So we focus on the relative performance of $\tilde{\beta}^{E}$, $\tilde{\beta}$ and $\tilde{\beta}^{H}$. When heteroskedasticities are present in the data-generating process, i.e., Tables 1-4 with $\sigma_1^2(\mathbf{Z_t})$ and $\sigma_2^2(\mathbf{Z_t})$, $\tilde{\beta}^{E}$ performs best in terms of its smallest bias, standard deviation and root mean squared error, followed by $\tilde{\beta}^{H}$, then by $\tilde{\beta}$. It confirms our asymptotic results in Theorems 1 and 2 that $\tilde{\beta}^{E}$ reaches the semiparametric efficiency bound. It makes sense to take the conditional heteroskedasticity information into account when constructing estimators, like $\tilde{\beta}^{E}$ and $\tilde{\beta}^{H}$. Note even though $\sigma_1^2(\mathbf{Z_t})$ and $\sigma_2^2(\mathbf{Z_t})$ are functions of both included and excluded exogenous variables Z_{1t} and Z_{2t}, which is not the maintained heteroskedasticity structure for $\tilde{\beta}^{H}$ in assumption A6 of Yao and Zhang (2011), $\tilde{\beta}^{H}$ still improves on $\tilde{\beta}$, which does not utilize any heteroskedasticity information. However, it pays to account for the information properly as in $\tilde{\beta}^{E}$. With $\tilde{\beta}^{E}$, we notice the further reduction of standard deviation and root mean squared error relative to $\tilde{\beta}^{H}$ is over 15%. On the other hand, when homoskedasticity is present in Tables 5 and 6, $\tilde{\beta}$ performs best, followed by $\tilde{\beta}^{H}$ and then by $\tilde{\beta}^{E}$. The outcome is expected as well. Since $\tilde{\beta}^{H}$ and $\tilde{\beta}^{E}$ involve additional estimation of the heteroskedasticity function which is unnecessary in the case of homoskedasticity, they generate additional noise in the estimation process. We notice the performances of the three estimators are fairly close in this case.

To further illustrate the performances in finite sample, we provide Rosenblatt density estimates of the four estimates centered around the true value when $\sigma_1^2(\mathbf{Z_t})$, $g_2(Z_{1t}, Z_{2t})$ and $\theta = 0.8$ are used in the data-generating process in Figure 1 for $n = 200$ and Figure 2 when $n = 400$. The experiment designs correspond to the case of heteroskedasticity, $E(X_t \mid \mathbf{Z_t})$ is nonlinear, and severe endogeneity presented in Table 2. Similar graphs can be generated for other heteroskedastic experiments. We notice the density for $\breve{\beta}^{(1)}$ is generally taller, but centered farther away from zero, indicating smaller variance, but much larger bias that we observe above. As sample size increases, we note the density estimates for $\tilde{\beta}^{E}$, $\tilde{\beta}^{H}$ and $\tilde{\beta}$ get spikier, and centered closer around zero, confirming the asymptotic results. The density estimate for $\tilde{\beta}^{E}$ is significantly more tightly centered around zero, relative to those of the others, which indicates the finite sample improvement of $\tilde{\beta}^{E}$ over the other estimators considered.

We further compare our efficient estimator $\tilde{\beta}^E$ with the efficient sieve minimum distance estimator $\tilde{\beta}^S$ proposed by Ai and Chen (2003) in the simulation, whose implementation calls for numerical optimizations. Since the two estimators are constructed with different nonparametric estimators, we follow their simulation data generating process carefully for meaningful comparisons. The only changes we made here are that the endogenous variable X_t shows up only in the parametric part, and we investigate their performances with sample sizes 200, 400, and perform 200 repetitions. Larger value for the parameter ρ denotes higher degree of endogeneity(they use R, but we use ρ since R is reserved here for the root mean squared error)and we set $\rho = 0.1$ and 0.9 according to their prescription. The results aresummarized in Table A. We observe that as the sample size increases, both estimators' performance improves, and it is harder to estimate for both when ρ gets larger. The performance of $\tilde{\beta}^E$ is better than that of $\tilde{\beta}^S$, illustrating the finite sample gains obtained by using our kernel based $\tilde{\beta}^E$, which is explicitly constructed and easy to implement.

5. Conclusion

We propose an estimator for the parametric vector in a semparametric regression model, in which the associated regressors are correlated with the errors. We assume the data is iid, but can contain both continuous and discrete variables. Under general conditional heteroskedasticity structure, we show that our new estimator is consistent and asymptotically normally distributed. It allows the reduced form to be nonparametric and incorporates the heteorskedasticity information. We show it is efficient in the sense that it reaches the semiparametric efficiency bound. It is also asymptotically equivalent to semiparametric instrumental variable estimators that optimally select the instrumental variables, and thus are efficient among a class of semiparametric IV estimators with conditional moment restrictions. We perform a Monte Carlo study that implements the estimator, and illustrates its finite sample performance, which confirms our asymptotic result.

Florens, J-P., Johannes, J. and Van Bellegem, S. (2009*) Instrumental regression in partially linear models*, IDEI working paper series, Num. 613

Härdle, W., Liang, H. and J. Gao(2000) *Partially linear models*, Physica-Verlag, Heidelberg

m en G W and W Ne e (200) Identification and estimation of triangular simultaneous equations models wit t additi it *Econometrica*, 77(5),1481-1512

i (2000) Efficient estimation of additive partially linear models *International Economic Review*, 41(4), 1073-1092

Li, Q. and J. S. Racine(2007)*Nonparametric Econometrics*, Princeton University Press, Princeton

i and teng (1) Semiparametric estimation of partially linear panel data mode *Journal of Econometrics*, 71(1-2), 389-397

i and U a (1 8) Estimating partially linear panel data models with one-way error mp nent *Econometric Reviews*, 17(2), 145-166

Ma Y and Carr (200) a e i ient e timat r r emiparametri m de it mea rement err r *Journal of the American Statistical Association*, 101(476), 1465-1474

Martins- i C and Ya (200) Nonparametric frontier estimation via a inear regre i n *Journal of Econometrics*, 141(1), 283-319

Mat in (1) Restrictions of economic theory in nonparametric methods n Eng e and D. L. McFadden (Eds.) *Handbook of Econometrics*: Vol. IV, Elsevier Science, North Holland

Mat in (2008) Identification in nonparametric simultaneous equations , *Econometrica*, 76(5), 945-978

Newey, W. K. (1 0) Efficient instrumental variable estimation of nonlinear models *Econometrica*, 58(4), 809-837

Ne e W (1) Efficient estimation of models with conditional moment restrictions inMaddala, G. S., C.R. Rao, and H.D. Vinod (Eds.) *Handbook of Statistics*, Vol. 11, 419-454, Amsterdam: North-Holland

Newey, W. K., F. Hsieh and J. M. Robins(200) Twicing kernels and a small bias property of emiparametri e timat r *Econometrica*, 72(3), 947-962

Newey, W.K. and J.L. Powell(200) Nonparametric instrumental variables estimation of nonparametric models *r* , 71(5), 1565-1578

Newey, W. P e and Ve a (1) Nonparametric estimation of triangular simultaneous equations models *Econometrica*, 67(3), 565 603

Ng, S. and J. Pinkse(1995)*Nonparametric two step estimation of unknown regression function when the regressors and the regressor error are not independent*,working paper, University of Montreal

Pin e (2000) Nonparametric two-step regression functions when regressors and error are dependent *Canadian Journal of Statistics*, 28(1), 289-300

Otsu, T.(2007) Penalized empirical likelihood estimation of semiparametric models *Journal of Multivariate Analysis*, 98(10), 1923-1954

Robinson, P.(1988) Root-n-consistent semiparametric regression *Econometrica*, 56(4), 931-954

Speckman, P.(1988) Kernel smoothing in partial linear models *Journal of the Royal Statistical Society, Series B*, 50(3), 413-436

Ruppert, D., S. J. Sheather, M. P. Wand(1995) An effective bandwidth selector for local least squares regression *Journal of the American Statistical Association*, 90(432), 1257-1270

Su, L., I. Murtazashvili, and A. Ullah(2011*) Local linear GMM estimation of functional coefficient IV models with an application to estimating the rate of return to schooling*, working paper, Singapore Management University

Yao, F.(2011)*Appendix to: Efficient semiparametric instrumental variable estimation under conditional heteroskedasticity*,Economics Department, West Virginia University.http://community.wvu.edu/~fy006/hp/yao2011a.pdf

Yao, F. and Zhang, J.(2011)*Efficient semiparametric instrumental variable estimation*, working paper, Economics Department, West Virginia University. http://community.wvu.edu/~fy006/hp/yao-zhang2009.pdf