The Japanese Economic Review
The Journal of the Japanese Economic Association

The Japanese Economic Review
Vol. 68, No. 1, March 2017

doi: 10.1111/jere.12091

MOMENT ESTIMATION OF THE PROBIT MODEL WITH AN ENDOGENOUS CONTINUOUS REGRESSOR

By DAIJI KAWAGUCHI†, YUKITOSHI MATSUSHITA‡ and HISAHIRO NAITO§

†University of Tokyo ‡Tokyo Institute of Technology
§University of Tsukuba

We propose a generalized method of moments (GMM) estimator with optimal instruments for a probit model that includes a continuous endogenous regressor. This GMM estimator incorporates the probit error and the heteroscedasticity of the error term in the first-stage equation in order to construct the optimal instruments. The estimator estimates the structural equation and the first-stage equation jointly and, based on this joint moment condition, is efficient within the class of GMM estimators. To estimate the heteroscedasticity of the error term of the first-stage equation, we use the *k*-nearest neighbour (*k*-nn) non-parametric estimation procedure. Our Monte Carlo simulation shows that in the presence of heteroscedasticity and endogeneity, our GMM estimator outperforms the two-stage conditional maximum likelihood estimator. Our results suggest that in the presence of heteroscedasticity in the first-stage equation, the proposed GMM estimator with optimal instruments is a useful option for researchers.

1. Introduction

Consider the following probit model, with a continuous endogenous regressor:

$$y_{1i}^* = \alpha y_{2i} + x_i\beta + u_i, \tag{1}$$

$$y_{1i} = 1 \ \text{if} \quad y_{1i}^* > 0, \tag{2}$$

$$y_{1i} = 0 \ \text{if} \quad y_{1i}^* \le 0, \tag{3}$$

$$y_{2i} = z_i\gamma + v_i, \tag{4}$$

$$E(v_i|z_i) = 0, \tag{5}$$

$$\begin{gathered} u_i = \rho v_i + e_i, \\ e_i|z_i, v_i \sim N(0,1), \\ \text{for } \ i = 1, 2, \ldots, n, \end{gathered} \tag{6}$$

where x_i is a $(1 \times k)$-row vector of exogenous explanatory variables, y_{2i} is a continuous explanatory variable, and z_i is a $(1 \times l)$-row vector of instrumental variables, which includes x_i as a subset. We assume that the first k elements of z_i are x_i. Then, we assume there are h excluded variables, which implies that $i = k + h$. The system of equations is assumed to be just identified or overidentified (i.e. $h \geq 1$ and the part of γ that corresponds to the excluded variables includes at least one non-zero element). Following convention, we refer to Equation (1) as the structural equation and Equation (4) as the first-stage equation.

A probit model with a continuous endogenous regressor is typically estimated using the maximum likelihood method, assuming the multivariate normality of u_i and v_i. However, the normality assumption on v_i is a strong assumption because it excludes the case where v_i exhibits heteroscedasticity. It also excludes cases where y_{2i} has limited support and, thus, where v_i has limited support.

Rivers and Vuong (1988) propose a two-step conditional maximum likelihood (2SCML) estimation in which they assume the joint normality of u_i and v_i.[1] The 2SCML estimation estimates γ using the ordinary least squares (OLS) method in the first stage, and introduces $y_{2i} - z_i\widehat{\gamma}$ as an additional regressor in the second-stage probit.

Now, consider the following equation:

$$y_{1i}^* = \alpha y_{2i} + x_i\beta + \rho v_i + e_i.$$

Because $e_i|z_i, v_i \sim N(0,1)$, the conditional expectation of y_{1i} on y_{2i} x_i, and v_i is given as:

$$E\left(y_{1i}|y_{2i}, x_{i,}, v_i\right) = \Phi(\alpha y_{2i} + x_i\beta + \rho v_i), \tag{7}$$

where Φ is the standard normal distribution function. Thus, we have the conditional moment condition:

$$E[y_{1i} - \Phi(\alpha y_{2i} + x_i\beta + \rho v_i)|y_{2i}, x_i, v_i] = 0. \tag{8}$$

Let Z_{1i} be a $1\times m_1$-row vector, where each element is a function of y_{2i}, x_i and v_i, and where $m_1 \geq k+2$. From the law of iterated expectations, we have

$$E\left[Z_{1i}'(y_{1i} - \Phi(\alpha y_{2i} + x_i\beta + \rho v_i))\right] = \mathbf{0}_{m_1\times 1}. \tag{9}$$

With regard to y_{2i}, the conditional expectation of y_{2i} on z_i is $E(y_{2i}|z_i) = z_i\gamma$. Thus, we have

$$E[y_{2i} - z_i\gamma|z_i] = 0. \tag{10}$$

Let Z_{2i} be a $1\times m_2$-row vector, where each element is a function of z_i, and where $m_2 \geq 1$ Then, we have

$$E\left[Z_{2i}'(y_{2i} - z_i\gamma)\right] = \mathbf{0}_{m_2\times 1}. \tag{11}$$

Note that in our above formulation, we do not assume the error term of the first-stage equation is homoscedastic. In microdata, heteroscedasticity is quite common, and often happens naturally when y_{2i} has limited support. Therefore, we believe that incorporating heteroscedasticity is important and, thus, allow for heteroscedasticity in the error term, v_i.

The above two moment conditions (9) and (11) naturally suggest using a generalized method of moments (GMM) estimator. In this study, we use these two orthogonal conditions to estimate the parameters of both the first-stage equation and the structural equation. In our GMM estimator, the optimal instruments use information on the conditional variance of the probit error and the conditional variance of the error of the first-stage equation. In contrast, the 2SCML estimator of Rivers and Vuong (1988) only uses information on the probit error. With regard to the first-stage equation, the 2SCML estimator uses the OLS method. Here, we use the GMM estimator with optimal instruments, utilizing the heteroscedasticity information

[1] Rivers and Vuong (1988) adopt the joint normality of u and v throughout their analysis. However, they mention in footnote 1 that the joint normality assumption is stronger than necessary for the validity of their estimator, and that their result is valid without the assumption of joint normality.

in the first-stage error term. Thus, the main difference between our GMM estimator and the 2SCML estimator is the treatment of the first-stage equation.

Does using the first-stage heteroscedasticity generate an efficiency gain over the 2SCML estimator when estimating the parameters of the structural equation? This is a natural question because, in most cases, researchers are interested in the parameters of the structural equation rather than the parameters of the first-stage equation. We argue that it is possible to make the estimation of the structural equation more efficient by estimating the first-stage equation more efficiently. To see why, consider the standard linear two-stage least squares (2SLS) estimator. Here, if the variable predicted by the instrumental variable is imprecise, the precision of the 2SLS decreases. However, by modelling the heteroscedasticity of the first-stage equation, we can increase the precision of the prediction of the endogenous variable, which means we can estimate the structural equation more precisely. Our GMM with optimal instruments adopts the same logic, increasing the precision of the estimated parameters of the structural equation.

Note that to increase the efficiency by modelling the heteroscedasticity of the error term of the first-stage equation, we need to estimate this heteroscedasticity. However, because we do not know its functional form, it may seem that this efficiency gain is difficult to achieve in practice.

We propose using the *k*th nearest neighbour (*k*-nn) non-parametric estimation method to estimate the heteroscedasticity of the error term of the first-stage equation. In a very important study, Robinson (1987) examines the feasible generalized least squares (GLS) estimator when the functional form of the error term is unknown. He shows that by using the *k*-nn non-parametric method, the estimator of the parameter of the main equation performs as well as the GLS estimator.

Extending the result of Robinson (1987), Newey (1990) and Newey (1993) analyse the instrumental variable estimation using the *k*-nn non-parametric estimation of the conditional expectation and the heteroscedasticity. In these studies, Newey proves that the estimated coefficients of the structural equation behave as if the conditional variance is known, rather than estimated. This implies that when we apply the *k*-nn non-parametric method to estimate the conditional variance in our GMM estimator, we can conduct statistical inferences as if the conditional variance is known.

A natural question to the above argument is to what extent the estimate of the structural estimate is improved by considering the heteroscedasticity in the first-stage equation. We conduct a Monte Carlo simulation to investigate this issue, and show that there is a substantial efficiency gain when estimating the parameters of the structural equation. The order of improvement in terms of the root-mean-square error (RMSE) from the true parameter value varies from 0 to 50%, depending on the heteroscedasticity in the first-stage equation and on the endogeneity of one of the explanatory variables.

In the literature, a GMM estimator of the probit model with continuous endogenous regressors was originally suggested by Grogger (1990). However, Dagenais (1999) and Lucchetti (2002) have shown the inconsistency of the proposed GMM estimator. Here, we first propose a consistent estimator that is efficient within the class of GMM estimators.

The remainder of this paper is organized as follows. In Subsection 2.1, we consider the optimal instrument in the case of the two-step estimation procedure. In this setting, we illustrate the importance of considering the heteroscedasticity in the first-stage equation when constructing the optimal instrument. In Subsection 2.2, we analyse the optimal instrument in a more general setting and characterize the optimal instrument. In Section 3, we discuss our Monte Carlo simulation. Here, we show there is a substantial efficiency gain from using the GMM estimator with optimal instruments when endogeneity and heteroscedasticity are present in the first-stage equation. In Section 4, we present an empirical analysis using the proposed GMM estimator with optimal instruments and the 2SCML methods. Section 5 concludes the paper.

2. Analysis

2.1 Optimal instrument in the two-step estimation

In this section, we illustrate the role of considering the heteroscedasticity in the first-stage equation when constructing the optimal instrument. Here, we consider the efficient estimator within the class of two-step estimators and characterize the optimal instrument. In this setting, we show that the optimal instrument must utilize the information on the heteroscedasticity of the first-stage equation. Then, in Subsection 2.2, where we provide the main result, we characterize the optimal instrument in a more general class.

From Equation (9), consider the following sample moment:

$$\frac{1}{n}\sum_{i=1}^{n} Z_{1i}^{'}\{y_{1i}-\Phi(\alpha y_{2i}+x_i\beta+\rho v_i)\}.$$

Because $y_{1i}-\Phi(\alpha y_{2i}+x_i\beta+\rho v_i)=e_i$ and e_i are i.i.d., it is natural to assume that

$$\frac{1}{n}\sum_{i=1}^{n} Z_{1i}^{'}\{y_{1i}-\Phi(\alpha y_{2i}+x_i\beta+\rho v_i)\}\xrightarrow{p} E\left[Z_{1i}^{'}\{y_{1i}-\Phi(\alpha y_{2i}+x_i\beta+\rho v_i)\}\right]. \quad (12)$$

For consistency, we assume that the condition of uniform convergence holds. For identification, we assume that α, β and ρ, which satisfy $E\left[Z_{1i}^{'}\{y_{1i}-\Phi(\alpha y_{2i}+x_i\beta+\rho v_i)\}\right]=0$, are the true parameters and are unique. With regard to the instrument Z_{1i}, until now, we have assumed that Z_{1i} is a function of y_{2i}, x_i and v_i. Note that when γ is known, knowing y_{2i} and z_i is equivalent to knowing v_i, from $y_{2i}-z_i\gamma=v_i$. Because z_i includes x_i, Z_{1i} is a function of y_{2i} and z_i, given γ. Later, we analyse the effect of replacing γ with the consistent estimate of γ. Let $\theta_1=[\alpha\ \beta\ \rho]$. Let Θ_1 be the possible parameter space. Let $\widehat{\gamma}$ be the consistent estimate of γ. Let $\theta_3=[\alpha\ \beta\ \rho\ \gamma]$. Let Θ_3 be the possible space of θ_3. Then, we define an $m_1\times 1$ vector $q_{1i}(\theta_1,\widehat{\gamma})$ as follows:

$$q_{1i}(\theta_1,\widehat{\gamma}) \equiv Z_{1i}^{'}\{y_{1i}-\Phi(\alpha y_{2i}+x_i\beta+\rho(y_{2i}-z_i\widehat{\gamma}))\}. \quad (13)$$

Similarly, define an $m_2\times 1$ vector $q_{2i}(\gamma)$ as follows:

$$q_{2i}(\gamma)=Z_{2i}^{'}(y_{2i}-z_i\gamma). \quad (14)$$

Let W_1 and W_2 be $m_1\times m_1$ and $m_2\times m_2$ symmetric positive definite weighting matrices, and let $\widehat{W}_1$ and $\widehat{W}_2$ be estimated matrices of W_1 and W_2, respectively, where plim $(\widehat{W}_1)=W_1$ and plim

$(\widehat{W}_2) = W_2$. Then, we choose θ_1 to solve the following GMM problem regarding Equation (13) with the weighting matrix $\widehat{W}_1$:

$$\min_{\theta_1 \in \Theta_1} \left(\frac{1}{n} \sum_{i=1}^{n} q_{1i}(\theta_1, \widehat{\gamma}) \right)' \widehat{W}_1 \left(\frac{1}{n} \sum_{i=1}^{n} q_{1i}(\theta_1, \widehat{\gamma}) \right). \tag{15}$$

Let $\widehat{\theta}_1$ be the solution of the above GMM problem Equation (15), and let θ_1^o be the true parameter value of θ_1. We assume that $\widehat{\gamma}$ is obtained by solving another GMM problem regarding Equation (14) with the weighting matrix $\widehat{W}_2$. Let γ^o be the true parameter value of γ. Then, the standard GMM argument implies that:[2]

$$Avar\left(\sqrt{n}\left(\widehat{\theta}_1 - \theta_1^o\right)\right)$$

$$= \underbrace{\left[G_1' W_1 G_1\right]^{-1} G_1' W_1 E\left[q_{1i}(\theta_1^o, \gamma^o) q_{1i}'(\theta_1^o, \gamma^o)\right] W_1 G_1 \left[G_1' W_1 G_1\right]^{-1}}_{\text{standard covariance of GMM}} \tag{16}$$

$$+ \underbrace{\left[G_1' W_1 G_1\right]^{-1} G_1' W_1 A E\left[s_i s_i'\right] A' W_1 G_1 \left[G_1' W_1 G_1\right]^{-1}}_{\text{the effect of sampling error of } \widehat{\gamma}}, \tag{17}$$

where $A = E\left[\frac{\partial q_{1i}(\theta_1^o, \gamma^o)}{\partial \gamma}\right]$; $s_i = -\left[G_2' W_2 G_2\right]^{-1} G_2' W_2 Z_{2i}'(y_{2i} - z_i \gamma^o)$;

$G_1 = E\left[\frac{\partial q_1(\theta_1^o, \gamma^o)}{\partial \theta_1}\right]$ and it is $m_1 \times (k+2)$ matrix;

$G_2 = E\left[\frac{\partial q_{2i}(\gamma^o)}{\partial \gamma}\right]$ and it is an $m_2 \times l$ matrix.

Equation (16) is the covariance matrix of the GMM estimator when γ^o is known. Equation (17) is the covariance term measuring the sampling error from using the estimated value $\widehat{\gamma}$ instead of the true value γ^o.

Now, consider choosing the instruments Z_{1i} and Z_{2i} optimally. Note that Z_{1i} affects G_1 and Z_{2i} affects $E\left[s_i s_i'\right]$. Thus, Z_{1i} affects both Equations (16) and (17). In contrast, Equation (17) is positive semi-definite. Thus, irrespective of Z_{1i}, it is optimal to minimize $E\left[s_i s_i'\right]$ in the sense of the matrix. To do so, we apply the standard argument of the optimal instrument because $E\left[s_i s_i'\right]$ is the asymptotic covariance of $\widehat{\gamma}$. The optimal instrument Z_{2i}^* is a $1 \times l$ vector, defined as

[2] For the derivation, see our discussion paper (Kawaguchi *et al.*, 2015).

$$Z_{2i}^{*} = \frac{1}{\sigma_2^2(z_i)} J_2, \tag{18}$$

$$\text{where } J_2 = E\left[\frac{\partial r_{2i}}{\partial \gamma}\middle| z_i\right] \text{ and it is } 1\times l \text{ matrix}$$

$$\sigma_2^2(z_i) = E\left[r_2^2 \middle| z_i\right], \quad r_{2i} = y_{2i} - z_i\gamma^o.$$

Thus, Z_{2i}^{*} becomes

$$Z_{2i}^{*} = \frac{1}{\sigma_2^2(z_i)} z_i. \tag{19}$$

When Z_{2i}^{*} is chosen according to Equation (19), $E\left[s_i s_i^{'}\right]$ becomes

$$E\left[s_i s_i^{'}\right] = E\left[J_2^{'} \sigma_2^2(z_i)^{-1} J_2\right]^{-1}.$$

With regard to the choice of Z_{1i}, it is not possible to characterize the optimal instrument of Z_{1i} in the class of two-step estimators unless it is just identified. This is because Z_{1i} affects both Equations (16) and (17). Therefore, an alternative solution is to find Z_{1i} that minimizes Equation (16). Thus, let Z_{1i}^{*} be Z_{1i} that minimizes Equation (16). Then, Z_{1i}^{*} becomes

$$Z_{1i}^{*} = \frac{1}{\Phi_i(1-\Phi_i)} [\phi_i y_2, \phi_i z_{1i}, \phi_i z_{2i}, .., \phi_i z_{ki}, \phi_i v_i] \tag{20}$$

where z_{ji} is the jth element of z_i;
$\phi_i = \phi(\alpha^o y_{2i} + x_i\beta^o + \rho^o(y_{2i} - z_i\gamma^o))$;
ϕ is the standard normal density function;
α^o, β^o and ρ^o are true parameter value of α, β and ρ.

When Z_{1i}^{*} and Z_{2i}^{*} are chosen according to Equations (19) and (20), we solve the following moment conditions:

$$\sum_{i=1}^{n} Z_{1i}^{*'} \{y_{1i} - \Phi(\alpha y_{2i} + x_i\beta + \rho(y_{2i} - z_i\widehat{\gamma}))\} = \mathbf{0}_{(k+2)\times 1} \tag{21}$$

$$\sum_{i=1}^{n} Z_{2i}^{*'} \{y_{2i} - z_i\gamma\} = \mathbf{0}_{l\times 1}, \tag{22}$$

where $\widehat{\gamma}$ is the solution of Equation (22).

The above result shows one of the key differences between the GMM estimator and the 2SCML method proposed by Rivers and Vuong (1988). In the case of the 2SCML method, the first stage is estimated using the OLS. In the GMM estimator, the first stage is estimated using the (F)GLS in order to utilize the heteroscedasticity information in the first-stage equation. As Equation (17) shows, utilizing the heteroscedasticity of the error term of the first-stage equation increases the efficiency of the estimated parameters of both the first-stage equation and the structural equation.

Note that in the two-step GMM estimator, the instrument for the structural equation is not optimal unless it is just identified.[3] To find a fully efficient estimator, we need to consider a more general class of estimator. In the next subsection, we consider the class of GMM estimators in which both the first-stage equation and the structural equation are estimated jointly. Because the two-step estimator can always be written as a joint estimator, considering a one-step estimator implies that we are looking for an efficient estimator within a broader class of estimators.

2.2 Optimal instruments in the class of joint estimation

In the previous subsection, we characterized the optimal instruments in the class of two-step estimators. We were able to characterize the optimal instrument for the first-stage equation, but were not able to find the optimal instrument for the structural equation. In this subsection, we consider the optimal instrument in the class of joint estimations. As we show, in this case, we can find the optimal instrument for both the first-stage equation and the structural equation. We characterize the form of the optimal instrument in this setting and show that this instrument achieves the minimum bound in the class of GMM estimators.

It may seem that finding the optimal instrument is straightforward, based on the result of Chamberlain (1987). However, in our model, the information sets used for the conditional expectation are different for the two equations, owing to the endogeneity. This implies that the optimal instruments in our setting are not obvious.

Now, define the residuals r_{1i} and r_{2i} as follows:

$$\begin{aligned} r_{1i} &= y_{1i} - \Phi(\alpha y_{2i} + x_i\beta + \rho(y_{2i} - z_i\gamma)) \\ r_{2i} &= y_{2i} - z_i\gamma. \end{aligned}$$

Let Z_{1i}' be the $m\times 1$ instrument vector that corresponds to r_{1i}, where $m\geq k+l+2$. We assume that Z_{1i}' is a function of y_{2i} and z_i. Because z_i includes x_i, this is not restrictive. Let Z_{2i}' be an $m\times 1$ instrument matrix that corresponds to r_{2i}. Here, we assume that Z_{2i}' is a function of z_i. Using the conditional expectation, we have

$$E\left[Z_{1i}'r_{1i}\right] = 0 \text{ and } E\left[Z_{2i}'r_{2i}\right] = 0.$$

Note that $\theta_3=[\alpha\ \ \beta\ \ \rho\ \ \gamma]'$, and Θ_3 is the possible parameter space of θ_3.
Then, we define Z_{3i} as

$$Z_{3i}' = \left[Z_{1i}'\ Z_{2i}'\right] \text{ which is } m\times 2 \text{ matrix.} \tag{23}$$

Let r_{3i} be an 2×1 matrix, where

$$r_{3i} = \begin{pmatrix} r_{1i} \\ r_{2i} \end{pmatrix}. \tag{24}$$

[3] We show this result in the next subsection.

Now, consider $Z_{3i}'r_{3i}$. Because $E\left[Z_{3i}'r_{3i}\right] = 0$, we can consider the standard GMM and the optimal instrument. Let Z_{3i}^* be the optimal instrument, and let $\delta \equiv [\alpha, \beta, \rho]$. Now, we have the following proposition:

Proposition: *The* $2\times(2+k+l)$ *matrix*, Z_{3i}^*, *is the optimal instrument*:

$$\text{where } Z_{3i}^* \equiv \Omega(z_i, y_{2_i})^{-1} \times R(z_i, y_{2i}); \tag{25}$$

$$\Omega(z_i, y_{2_i})^{-1} = \begin{pmatrix} \sigma_1^2(y_{2i}, z_i)^{-1} & 0 \\ 0 & \sigma_2^2(z_i)^{-1} \end{pmatrix} \text{ and is a } 2\times2 \text{ matrix;}$$

$$\sigma_1^2(z_i) = E\left[r_1^2 | z_i, y_{2i}\right] \text{ and } \sigma_2^2(z_i) = E\left[r_2^2 | z_i\right];$$

$$R(y_{2i}, z_i) = \begin{pmatrix} E[\nabla_\delta r_1 | z_i, y_{2i}] & E[\nabla_\gamma r_1 | z_i, y_{2i}] \\ \mathbf{0}_{1\times(2+k)} & E[\nabla_\gamma r_2 | z_i] \end{pmatrix} \text{ and is a } 2\times(2+k+l) \text{ matrix;}$$

$\nabla_\delta r_1, \nabla_\gamma r_1, \nabla_\gamma r_2$ are derivative matrices, and $\nabla_\delta r_1 = \dfrac{\partial r_1}{\partial \delta}$, $\nabla_\gamma r_1 = \dfrac{\partial r_1}{\partial \gamma}, \nabla_\gamma r_2 = \dfrac{\partial r_2}{\partial \gamma}$.

The asymptotic variance of $\sqrt{n}\left(\widehat{\theta_3} - \theta_3^o\right)$ *is given by*:

$$Avar\left(\sqrt{n}(\theta_3 - \theta_3^o)\right) = E\left[R(y_{2i}, z_i)' \Omega(z_i, y_{2_i})^{-1} R(y_{2i}, z_i)\right]^{-1}. \tag{26}$$

The proof is provided in our discussion paper, Kawaguchi *et al.* (2015).

Note that when Z_{3i}^* is defined in this way, the moment condition becomes

$$\frac{1}{n}\sum_{i=1}^n Z_{3i}^{*'} r_{3i} = \mathbf{0}_{(k+2+l)\times1}. \tag{27}$$

Thus, we have the following moment conditions:

$$\frac{1}{n}\sum_{i=1}^n \frac{1}{\Phi_i(1-\Phi_i)} \phi_i y_{2i} r_{1i} = 0, \tag{28}$$

$$\frac{1}{n}\sum_{i=1}^n \frac{1}{\Phi_i(1-\Phi_i)} \phi_i x_i' r_{1i} = \mathbf{0}_{k\times1}, \tag{29}$$

$$\frac{1}{n}\sum_{i=1}^n \frac{1}{\Phi_i(1-\Phi_i)} \phi_i (y_{2i} - z_i\gamma) r_{1i} = 0, \tag{30}$$

$$\frac{1}{n}\sum_{i=1}^n \left\{\frac{1}{\Phi_i(1-\Phi_i)} \phi_i z_i' r_{1i} - \sigma_2^2(z_i)^{-1} z_i' r_{2i}\right\} = \mathbf{0}_{l\times1}. \tag{31}$$

We can make several observations on the above moment conditions. First, the conditions again clearly show the importance of considering the heteroscedasticity in the first-stage equation, which is captured by $\sigma_2^2(z_i)^{-1}$. Second, they show that for excluded variables, the moment conditions of the first-stage equation and the structural equation need to be stacked together, which is absent in the case of the 2SCML estimator when it is overidentified. Third, in the just-identified case, the above joint estimation can be implemented in two steps.[4] Thus, in the just-identified case, the moment conditions considered in the previous section, Equations (21) and (22) are also fully efficient.

2.3 Estimating $\sigma_2^2(z_i)$

To use the instrument Z_{3i}^*, we need to estimate $\sigma_2^2(z_i)$. To estimate $\sigma_2^2(z_i)$, we need to know the functional form of the conditional variance. When researchers are not confident of the functional form of the conditional variance of the error term of the first-stage equation, a reasonable approach is to use a non-parametric method. In the literature on non-parametric methods, many methods have been proposed for non-parametric estimations of heteroscedasticity (e.g. Carroll, 1982 Robinson, 1987) (For a survey of the literature, see Li and Racine (2007) and Wolfgangs and Linton (1994).) Among these methods, that of Robinson (1987) is prominent. Robinson (1987) studies the performance of the estimator in the feasible GLS model when the functional form of the conditional variance of the error term is unknown. He shows that by using the k-nearest neighbour non-parametric method, the estimator of the main equation behaves at the same speed as when estimated using the GLS. Newey (1990) and Newey (1993) use non-parametric estimations of the conditional expectation and the heteroscedasticity for the instrumental variable estimation. Here, we follow the procedure suggested by Newey (1990). Let z_{ji} be the jth element of z_i, and let σ_{z_j} be the sample standard deviation of the jth element of z_i. For $i \neq t$, calculate the standardized distance of observation i and $t(i \neq t)$, $d(i, t)$, as follows:

$$d(i,t) = \left\{ \sum_{j=1}^{l} \left(z_{ji} - z_{jt}\right)^2 / \sigma_{z_j}^2 \right\}^{1/2}. \tag{32}$$

For each observation z_i, rank the observation according to $d(i, t)$, where $t \neq i$. Then, choose the first k observations according to $d(i, t)$ from those that satisfy $d(i, t) > 0$. Define $N(z_i)$ as follows:

$$N(z_i) = \{t : z_t \text{ is one of the first } k \text{ observations according to } d(i,t), \text{ where } d(i,t) > 0.\}.$$

Let $\widehat{\sigma}_{2i}^2$ be the calculated squared residual of the error of the first-stage equation at z_i, where the residual is calculated using the OLS approach. Next, the conditional variance of kth nearest neighbour at z_i is calculated as follows:

$$\widehat{m}_k(z_i) = \frac{1}{k} \sum_{t \in N(z_i)} \widehat{\sigma}_{2t}^2, \tag{33}$$

where $\widehat{\sigma}_{2i}$ is the calculated residual of observation of i.

---

4 We can prove the just-identified case as follows. Note that the first k elements of z_i are x_i. Thus, from Equations (29) and (31), we have $\frac{1}{n}\sum_{i=1}^n \sigma_2^2(z_i)^{-1} x_i' r_{2i} = \mathbf{0}_{k\times 1}$. For the excluded variable, from Equations (28), (29) and (30), we have $\frac{1}{n}\sum_{i=1}^n \frac{1}{\Phi_i(1-\Phi_i)} \phi_i z_{i,(k+1)} r_{1i} = 0$, where $z_{i,(k+1)}$ is the $(k+1)$th element of z_i. From Equation (31), we have that $\frac{1}{n}\sum_{i=1}^n \sigma_2^2(z_i)^{-1} z_{i,(k+1)} r_{2i} = 0$. Thus, the moment conditions coincide with the moment conditions in the two-step case.

To choose an optimal value of k, we use the cross-validation (CV) method. Because $\widehat{m}_k(z_i)$ is already calculated by excluding observation i, the CV function is

$$CV(k) = \sum_{i=1}^{n} \left\{\widehat{\sigma}_{2i}^2 - \widehat{m}_k(z_i)\right\}^2. \tag{34}$$

The CV method chooses k to minimize $CV(k)$.

The theoretical results of Robinson (1987), Newey (1990) and Newey (1993) show that, asymptotically, the estimates of the parameters of the main equation behave as if the conditional variance is known, rather than estimated, when the above k-nn non-parametric method is used. We can apply their results directly to our GMM estimator. A natural question is how the GMM estimator behaves in practice when the k-nn non-parametric estimation method is used. In Section 3, we conduct a small-scale Monte Carlo simulation using this method to estimate the conditional variance of the error term of the first-stage equation. The simulation shows that when there is heteroscedasticity and endogeneity, applying both the k-nn non-parametric estimation of the heteroscedasticity of the first-stage error and the GMM estimator results in a substantial efficiency gain when estimating the parameters of the structural equation.

2.4 The effect of the preliminary estimates

In order to construct Z_{3i}^*, we need to calculate Φ_i. To calculate Φ_i, we need preliminary estimates of α, β, ρ and γ, the probability limits of which are the true values. In addition, we need to use the estimated value of $\sigma_2^2(z_i)$ to construct Z_{3i}^*. Here, we determine the asymptotic behaviour of $\sqrt{n}(\widehat{\theta}_3 - \theta_3^o)$ when the estimated values of $\alpha, \beta, \rho, \gamma$ and $\sigma_2^2(z_i)$ are used. Let $\widehat{Z}_{3i}^*$ be the estimated value of Z_{3i}^*, based on the preliminary consistent estimate of $\alpha, \beta, \rho, \gamma$ and $\sigma_2^2(z_i)$. As in the previous subsection, we assume that $\sigma_2^2(z_i)$ is estimated using the k-nn non-parametric method. We further assume that the CV method is used to find the optimal k. Then, it can be shown that the mean value expansion of $n^{-1/2}\sum \widehat{Z}_{3i}^* r_{3i}$ is the same as the mean value expansion of $n^{-1/2}\sum Z_{3i}^* r_{3i}$, based on the arguments of Newey (1990) and Newey (1993), and, in particular, on theorem 1 of Newey (1993). Thus, using the preliminary consistent estimates to construct the instrument does not affect the asymptotic distribution of $\sqrt{n}\left(\widehat{\theta}_3 - \theta_3^o\right)$.

2.5 Estimation procedure

The actual procedure for the estimation is as follows:

1. Run the OLS regression y_{2i} on z_i, keeping the residual as $\widehat{v}_i$. Calculate $\widehat{v}_i^2$.
2. Run the probit regression y_{1i} on $y_{2i}, x_i, \widehat{v}_i$, and estimate α, β and ρ. Denote the estimates of these parameters as $\widehat{\alpha}, \widehat{\beta}$ and $\widehat{\rho}$, respectively. Using $\widehat{\alpha}, \widehat{\beta}$ and $\widehat{\rho}$, calculate $\widehat{\Phi}_i$ as $\widehat{\Phi}_i = \Phi\left(\widehat{\alpha} y_{2i} + \widehat{\beta} x_i + \widehat{\rho}\widehat{v}_i\right)$ and $\widehat{\phi}_i = \phi\left(\widehat{\alpha} y_{2i} + \widehat{\beta} x_i + \widehat{\rho}\widehat{v}_i\right)$.
3. Apply the k-nn non-parametric method to estimate $\sigma_2^2(z_i)$ from $\widehat{v}_i^2$. Because we do not know which element of z_i should be included in order to calculate $\sigma_2^2(z_i)$, we include all elements of z_i when applying the k-nn non-parametric estimation of $\sigma_2^2(z_i)$. Use the CV method to

choose the optimal k for the k-nn non-parametric estimation. Once k is determined, apply the k-nn estimator to estimate $\sigma_2^2(z_i)$. Let $\widehat{\sigma}_2^2(z_i)$ be the estimate of $\sigma_2^2(z_i)$.

4. Conduct the GMM estimation using the following moment conditions, assuming that the weighting matrix is the $(k+2+l)\times(k+2+l)$ identity matrix:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\widehat{\Phi}_i\left(1-\widehat{\Phi}_i\right)}\widehat{\phi}_i y_{2i} r_{1i}=0 \tag{35}$$

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\widehat{\Phi}_i\left(1-\widehat{\Phi}_i\right)}\widehat{\phi}_i x_i' r_{1i}=\mathbf{0}_{k\times 1} \tag{36}$$

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\widehat{\Phi}_i\left(1-\widehat{\Phi}_i\right)}\widehat{\phi}_i'\widehat{v}_i r_{1i}=0 \tag{37}$$

$$\frac{1}{n}\sum z_i'\left\{\frac{1}{\widehat{\Phi}_i\left(1-\widehat{\Phi}_i\right)}\widehat{\phi}_i' r_{1i}-\widehat{\sigma}_2^2(z_i)^{-1}r_{2i}\right\}=\mathbf{0}_{l\times 1}. \tag{38}$$

5. In order to calculate the asymptotic variance of $\sqrt{n}\left(\widehat{\theta}_3-\theta_3^o\right)$, we use Equation (26). Let $\widetilde{R}(y_{2i}z_i)$ be the calculated residual after following the above GMM procedure. Let $\widetilde{\Omega}\left(z_i y_{2_i}\right)$ be the estimated conditional variance. Then, the asymptotic variance is calculated as

$$Avar\left(\sqrt{n}\left(\widehat{\theta}_3-\theta_3^o\right)\right)=\left[n^{-1}\sum_{i=1}^{n}\widetilde{R}(y_{2i},z_i)'\widetilde{\Omega}\left(z_i,y_{2_i}\right)^{-1}\widetilde{R}(y_{2i},z_i)\right]^{-1}. \tag{39}$$

3. Monte Carlo simulation

We conduct Monte Carlo simulations to assess the performance of our proposed GMM estimator. Here, we consider the following model:

$$y_1^*=\alpha y_2+\beta_0+x_1\beta_1+\rho v+e \tag{40}$$

$$y_1=1 \text{ if } y_1^*>0 \tag{41}$$

$$y_1=0 \text{ if } y_1^*\leq 0 \tag{42}$$

$$y_2=\gamma_0+x_1\gamma_1+z_1\gamma_2+z_2\gamma_3+v. \tag{43}$$

We set the parameter values as $\alpha=1$, $\beta_0=1$, $\beta_1=-1$, $\gamma_0=1$, $\gamma_1=1$, $\gamma_2=-1$ and $\gamma_3=-1$. The error term e is drawn randomly from the standard normal distribution. The error term v follows a normal distribution with heteroscedasticity. We assume that $v\sim N\left(0,e^{\lambda z_2}\right)$, where λ is a parameter showing the degree of heteroscedasticity. When $\lambda=0$, the variance of v is constant. We set $\lambda=0,1,1.5$, and compare the 2SCML estimator and the GMM estimator with optimal instruments, where the heteroscedasticity is estimated using the k-nn non-

[5] To save space, we include the results for these parameter values only. Comprehensive results of the Monte Carlo simulations are available on the authors' website.

parametric method. Here, ρ shows the degree of endogeneity of y_2 and y_1. For the value of ρ, we consider the cases where $\rho=0,1,2,-1,-2$.[5] The random variables $\{x_1, z_1, z_2\}$ are drawn from the multivariate normal distribution, with $Var(x_1)=Var(z_1)=Var(z_2)=1$ and $Cov(z_1,z_2)=Cov(x_1,z_1)=Cov(x_1,z_2)=0.5$. These settings are similar to the assumptions in Rivers and Vuong (1988). The number of observations in each replication is set to 500 and we repeat the estimation 500 times. In order to conduct the k-nn non-parametric estimation of the conditional variance of the error of the first-stage equation, we need to set the parameter value of k. This is a smoothing parameter in the k-nn non-parametric estimation. Calculating k is time consuming because it is necessary to conduct a cross-validation for all possible values from 1 to the number of observations. It is not realistic to calculate k in each replication separately. Thus, in this simulation, we first conduct 20 replications for $\rho=1$ and $\lambda=1$, then find the optimal k in each replication. Next, we calculate the average value for k, which we find to be 14 in these 20 replications. Thus, we fix k at 14 for all replications, and for different values of ρ and λ. Note that the optimal k can be different for different values of ρ and λ. In addition, even if the values of ρ and λ are the same, the optimal k may differ across replications. Thus, our Monte Carlo simulations underestimate the performance of the GMM with optimal instruments. However, we demonstrate that, even here, our estimator outperforms the 2SCML method.

Table 1 shows the bias and RMSE of the estimated α, β_0, β_1 and ρ of the two estimators. With regard to the bias, there is little difference between the 2SCML and GMM estimators. However, with regard to the RMSE, when there is endogeneity ($\rho \neq 0$) and a large degree of heteroscedasticity ($\lambda=1$), there is a substantial efficiency gain when using the GMM with optimal instruments. For example, when $\rho=2$ and $\lambda=1$, the RMSE of α when using the 2SCML is 50% greater than when using the GMM with optimal instruments. When there is no endogeneity or no heteroscedasticity, there is no efficiency gain when using the GMM with optimal instruments.

4. Effect of non-labour income on married women's labour-force participation

As an empirical application of the proposed method, we consider married women's labour-force participation decisions. The standard labour supply model predicts that married women participate in the labour force when their offered wages exceed their reservation wages. Because the offered wage of non-participants is not observed, it is typically modelled as a linear function of demographic characteristics, such as age, education and race. In the following example, the reservation wage is modelled as a linear function of the number of children between ages 0 and 5, household income, and the married woman's own labour earnings.

In addition to a married woman's own labour earnings, a typical source of household income is spousal earnings. An increase in spousal earnings generally increases the reservation wage of a married woman, and reduces her probability of labour-force participation. Although we often observe a negative impact of a husband's earnings on wife's labour-force participation, the husband's earnings could be endogenous. If a wife has a high ability level in a market, she is more likely to work and to spend less time on household production, such as child rearing. The husband, accordingly, may reduce his market work to share the burden of household production. The husband's endogenous allocation of effort between market and household production makes the other household income (the household income less the married woman's labour earnings) endogenous. To deal with this potential endogeneity in

TABLE 1
The bias and RMSE of α, β_0, β_1 and ρ

Parameter			α		β_0		β_1		ρ	
ρ	λ	Method	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
0	0	RV	0.0249	0.1258	0.0318	0.133	−0.0298	0.1427	0.003	0.1239
0	0	GMM	0.025	0.1259	0.0317	0.1331	−0.0298	0.1426	0.003	0.1241
0	0.5	RV	0.0327	0.1344	0.0327	0.1404	−0.0376	0.1605	0.0059	0.1213
0	0.5	GMM	0.0326	0.1346	0.0329	0.1398	−0.0375	0.1604	0.0061	0.1215
0	1	RV	0.0355	0.1578	0.039	0.1582	−0.0374	0.1871	0.0077	0.1413
0	1	GMM	0.0344	0.1565	0.0408	0.1552	−0.0377	0.1845	0.0098	0.1435
1	0	RV	0.0296	0.1284	0.0345	0.163	−0.0306	0.1561	0.0226	0.1759
1	0	GMM	0.0291	0.1288	0.0354	0.1655	−0.0321	0.1589	0.0235	0.1763
1	0.5	RV	0.0342	0.1504	0.048	0.2015	−0.0358	0.1793	0.0361	0.1957
1	0.5	GMM	0.0344	0.148	0.0487	0.1963	−0.0379	0.1737	0.0399	0.1955
1	1	RV	0.0445	0.2123	0.0416	0.2916	−0.0321	0.2535	0.0095	0.2529
1	1	GMM	0.0482	0.1729	0.0434	0.2113	−0.0427	0.1948	0.045	0.2268
2	0	RV	0.0346	0.1604	0.0412	0.2098	−0.0334	0.1858	0.0647	0.2862
2	0	GMM	0.0339	0.1613	0.0437	0.2181	−0.0369	0.1913	0.0684	0.2885
2	0.5	RV	0.0451	0.1972	0.0552	0.2744	−0.0348	0.2287	0.0825	0.3512
2	0.5	GMM	0.0485	0.1774	0.0592	0.2618	−0.043	0.2069	0.0994	0.3548
2	1	RV	0.0412	0.327	0.0334	0.4912	−0.0207	0.3969	−0.0184	0.4679
2	1	GMM	0.066	0.2086	0.0557	0.2892	−0.0586	0.2428	0.1121	0.4121
−1	0	RV	0.026	0.1239	0.0261	0.1329	−0.038	0.1452	−0.0224	0.1625
−1	0	GMM	0.0275	0.1252	0.026	0.1349	−0.0373	0.1468	−0.024	0.1641
−1	0.5	RV	0.0254	0.1311	0.0185	0.151	−0.0382	0.1528	−0.0258	0.1485
−1	0.5	GMM	0.0262	0.124	0.022	0.1405	−0.0369	0.1461	−0.0258	0.1416
−1	1	RV	0.0189	0.1913	−0.0164	0.2878	−0.0423	0.2236	−0.0199	0.1961
−1	1	GMM	0.0255	0.1304	0.0159	0.1566	−0.0369	0.1457	−0.0255	0.1353
−2	0	RV	0.0236	0.1418	0.0182	0.1715	−0.0327	0.1771	−0.0425	0.2401
−2	0	GMM	0.031	0.1487	0.0237	0.1827	−0.0359	0.1833	−0.0564	0.2498
−2	0.5	RV	0.0276	0.1621	0.0127	0.2281	−0.0317	0.2069	−0.0535	0.2523
−2	0.5	GMM	0.0329	0.1469	0.0209	0.2113	−0.0323	0.1805	−0.0642	0.246
−2	1	RV	0.0177	0.3173	−0.0302	0.4814	−0.0455	0.3925	−0.0144	0.4157
−2	1	GMM	0.0414	0.1851	0.029	0.2316	−0.0479	0.2114	−0.0748	0.3565

Notes: The sample size is 500. The number of replication is 500 for each case. RMSE, root-mean-square error; RV, the a two-step conditional maximum likelihood by Rivers and Voung.

other household income, the variable is instrumented using dummy variables that indicate the husband's occupation.

This empirical exercise is implemented using the data set provided by Lee (1995), which was originally extracted from the 1987 wave of the Michigan Panel Study of Income Dynamics. The sample includes 3,382 married women, whose descriptive statistics are shown in Table 2.

Table 3 reports the regression results. Column (1) reports the probit regression, without considering the endogeneity of other household income. The increase in other household income decreases the labour-force participation rate in a statistically significant way. An increase in a husband's income decreases the wife's employment via the income effect. The estimated coefficients for the other covariates are standard.

Columns (2) and (3) report the estimation results for the 2SCML procedure (Rivers and Vuong, 1988). The first-stage regression reported in column (2) indicates that the husband's occupation strongly predicts other household income. The second-stage probit regression result in column (3) indicates that other income does not explain the wife's employment, after considering the endogeneity. The estimated ρ is negative and statistically significant. As

TABLE 2
Descriptive statistics of the analysis sample of married women

Variable name	Description	Mean	Standard deviation
Employed	Wife reports positive hour	0.735	0.441
Other income	The other household income in \$1000	29.69	28.8
Age	Age of the wife	36.813	11.353
Educ	Education years of the wife	12.553	2.416
Children 0–5	Number of childrent for ages 0 to 5	0.507	0.763
Nonwhite	1 if non-white	0.296	0.456
Hus manager	1 if the husband is manager or professional	0.284	0.451
Hus sales	1 if the husband is sales worker or clerical or craftsman	0.302	0.459
Hus farm	1 if the husband is farm-related worker	0.025	0.157

Note: The number of observations is 3,382. These data were extracted from the 1987 wave of Michigan Panel Study of Income Dynamics (PSID). Details are available in Lee (1995).

TABLE 3
The determination of labour force participation of married women

Method	(1) Probit	(2) Rivers–Voung	(3)	(5) GMM	(6)
Dependent variable	*Employed*	*Other income*	*Employed*	*Other income*	*Employed*
Other income	−0.006 (0.001)	—	0.002 (0.004)	—	−0.00013 (0.0015)
Age	−0.032 (0.003)	0.641 (0.044)	−0.038 (0.005)	0.717 (0.108)	−0.034 (0.002)
Educ	0.134 (0.011)	2.196 (0.205)	0.108 (0.023)	1.782 (0.198)	0.117 (0.012)
Children 0–5	−0.506 (0.036)	0.809 (0.648)	−0.510 (0.04)	1.126 (0.529)	−0.520 (0.038)
Non-white	0.134 (0.057)	−6.522 (1.021)	0.209 (0.069)	−6.149 (0.950)	0.195 (0.059)
Husband manager	—	16.647 (1.221)	—	8.477 (3.046)	—
Husband sales	—	6.296 (1.108)	—	13.590 (13.00)	—
Husband farm	—	−3.802 (2.903)	—	−20.181 (24.119)	—
V hat/Rho	—	—	−0.009 (0.004)	—	−0.006 (0.0012)
Constant	0.620 (0.185)	−26.490 (3.396)	0.876 (0.369)	−24.568 (8.940)	0.907 (0.227)
R^2/log-likelihood	−1, 711.416	0.196	−1, 709.156	—	—

expected, the unobserved determinants of the wife's labour-force participation and other income are negatively correlated, perhaps resulting from the husband's endogenous effort allocation in the labour market.

The GMM estimations reported in columns (4) and (5) are almost identical to the results using the 2SCML procedure. Thus, in this empirical example, the GMM with optimal instruments does not increase the precision of the estimator. The reason for this can be seen in column (1), which shows that ρ, which estimates the degree of endogeneity, is quite low compared with the other parameters. When ρ is low, A in Equation (17) is also low. In such a case, there is little gain from estimating the first-stage equation more efficiently, as we show in our Monte Carlo simulations. Thus, this empirical example confirms the simulation results that the benefit of using the GMM with optimal instruments depends on the heteroscedasticity in the first-stage equation and the endogeneity in the structural equation.

5. Conclusion

This study proposed an efficient GMM estimator for a probit model with a continuous regressor. We derived the optimal instruments that attain asymptotic efficiency within the class of

GMM estimators in the presence of heteroscedasticity in the first-stage equation. Here, we use the *k*th nearest neighbour non-parametric estimator to characterize the heteroscedasticity of the error term of the first-stage equation. Monte Carlo simulations reveal that the GMM estimator based on optimal instruments performs better than the 2SCML estimator does in the presence of heteroscedasticity. These results suggest that the GMM estimator, with a non-parametric estimation of the heteroscedasticity of the first-stage equation, is a useful empirical research tool.

Acknowledgements

We thank two anonymous referees and the editor, Yoshihiko Nishiyama, for their helpful comments on the previous version of the paper. One of the referees suggested the literature on non-parametric estimations. Using the non-parametric method improved the estimation procedure of the study substantially. We also thank Hidehiko Ichimura and Yu-Chin Hsu for their comments, which improved the paper significantly.

Final version accepted 21 December 2015.

REFERENCES

Carroll, R. J. (1982) "Adapting for Heteroscedasticity in Linear Models", *Annals of Statistics*, Vol. 10, pp. 1224–1233.

Chamberlain, G. (1987) "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions", *Journal of Econometrics*, Vol. 34, No. 3, pp. 305–334.

Dagenais, M. G. (1999) "Inconsistency of a Proposed Nonlinear Instrumental Variables Estimator for Probit and Logit Models with Endogenous Regressors", *Economics Letters*, Vol. 63, No. 1, pp. 19–21.

Grogger, J. (1990) "A Simple Test for Exogeneity in Probit, Logit, and Poisson Regression Models", *Economics Letters*, Vol. 33, No. 4, pp. 329–332.

Kawaguchi, D., Y. Matsushita and H. Naito (2015) "Moment Estimation of the Probit Model with an Endogenous Continuous Regressor", *Tsukuba Economics Working Papers Series*, Vol. 3, pp. 1–22.

Lee, M-J (1995) "Semi-parametric Estimation of Simultaneous Equations with Limited Dependent Variables: A Case Study of Female Labour Supply", *Journal of Applied Econometrics*, Vol. 10, No. 2, pp. 187–200.

Li, Q. and J. S. Racine (2007) Nonparametric Econometrics: Theory and Practice, Princeton, NJ: Princeton University Press.

Lucchetti, R. (2002) "Inconsistency of Naive GMM Estimation for QR Models with Endogenous Regressors", *Economics Letters*, Vol. 75, No. 2, pp. 179–185.

Newey, W. K. (1990) "Efficient Instrumental Variables Estimation of Nonlinear Models", *Econometrica*, Vol. 58, No. 4, pp. 809–837.

—— (1993) "Efficient Estimation of Models of Conditional Moment Restrictions", in C. R. Rao, G. S. Maddala and H. D. Vinnot, eds, Handbook of Statistics, Vol. 11, Amsterdam: Elsevier Scence.

Rivers, D. and Q. H. Vuong (1988) "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models", *Journal of Econometrics*, Vol. 39, No. 3, pp. 347–366.

Robinson, P. M. (1987) "Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form", *Econometrica*, Vol. 55, No. 4, pp. 875–891.

Wolfgangs, H. and O. Linton (1994) "Chapter 18: Applied Nonparametric Methods", Handbook of Econometrics, Vol. IV, Amsterdam: Elsevier.