Tsukuba Economics Working Papers
No. 2015-003

Moment Estimation of the Probit Model with an Endogenous Continuous Regressor

By

Daiji Kawaguchi
Hitotsubashi University

Yukitoshi Matsushita
Tokyo Institute of Technology

Hisahiro Naito
University of Tsukuba

December 2015
(Forthcoming from *Japanese Economic Review*)

UNIVERSITY OF TSUKUBA
Graduate School of Humanities and Social Sciences
1-1-1 Tennodai
Tsukuba, Ibaraki 305-8571
JAPAN

Moment Estimation of the Probit Model with an Endogenous Continuous Regressor*

Daiji Kawaguchi†
Graduate School of Economics
Hitotsubashi University

Yukitoshi Matsushita‡
Department of Mechanical and Environmental Informatics
Tokyo Institute of Technology

Hisahiro Naito§
Graduate School of Humanities and Social Sciences
University of Tsukuba

December 21, 2015

Forthcoming
Japanese Economic Review

*This is the longer version of the paper with the same tittle that is forthcoming from *the Japanese Economic Review*. This version contains the derivation of several equations and a mathematical appendix that are omitted in the paper from the JER due to the limitation of the space.

†Address: Naka 2-1, Kunitachi, Tokyo 186-8601; e-mail: kawaguch@econ.hit-u.ac.jp
‡Address: 2-12-1 Ookayama, Meguro-ku, Tokyo, 152-8552; e-mail: matsushita.y.ab@m.titech.ac.jp
§Address: Tennodai 1-1-1, Tsukuba, Ibaraki, 305-8573; e-mail: naito@dpipe.tsukuba.ac.jp

Abstract

We propose a 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 neighbor (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 (2SCML) 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.

JEL Classification: C25

Keywords: Probit, Continuous endogenous regressor, Moment estimation.

1 Introduction

Consider the following probit model with a continuous endogenous regressor:

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

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

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

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

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

$$u_i = \rho v_i + e_i, \tag{6}$$

$$e_i|z_i, v_i \sim N(0,1),$$

$$\text{for } i = 1, 2, ..., n,$$

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 that 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 $l = k + h$. The system of equations is assumed to be just or over-identified (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 (1) as the structural equation and (4) as the first-stage equation.

The 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) proposed a two-step conditional maximum likelihood (2SCML) estimation in which they assume the joint normality of u_i and v_i.[1] The

[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.

2SCML estimation estimates γ using the ordinary least squares (OLS) method in the first stage, and introduces $y_{2i} - z_i\hat{\gamma}$ 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:

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

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 iterated of expectations, we have

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

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 $m_2 \geq l$. Then, we have

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

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 the generalized method of moments (GMM). 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, our optimal instruments use the 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 the information on the probit error. With regard to the first-stage equation, the 2SCML estimator uses the OLS method. Here, we use the GMM with the optimal instruments, utilizing the heteroscedasticity information 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 for estimating the parameters of the structural equation? This is a natural question since, 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. In the standard 2SLS, if the predicted variable made by the instrumental variable is imprecise, the precision of the 2SLS becomes low. However, by modeling the heteroscedasticity of the first-stage equation, we can make the prediction of the endogenous variable more precise, which means we can estimate the structural equation more precisely. Our GMM with the optimal instruments adopts the same logic, increasing the precision of the estimated parameters of the structural equation.

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

In this study, we propose using the k-th nearest neighbor (k-nn) non-parametric estimation method to estimate the heteroscedasticity of the error term of the first-stage equation. In a very important paper, Robinson (1987) studied the feasible generalized least squares estimator when the functional form of the error term is unknown. He showed 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) analyzed the instrumental variable estimation using the k-nn non-parametric estimation of the conditional expectation and the heteroscedasticity. In these studies, he proved 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, 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 section 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 section 2.2, we analyze 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 the 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 with optimal instruments and the 2SCML methods. Lastly, 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 section 2.2, where we provide the main result, we characterize the optimal instrument in a more general class.

From (9), consider the following sample moment:

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

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

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

For consistency, we need a stronger condition than (12). Let $\theta_1 = [\alpha \ \beta \ \rho]$. Let Θ_1 be the possible parameter space. We assume that the following uniform convergency holds:

$$\max_{\theta_1 \in \Theta_1} \left| \frac{1}{n}\sum_{i=1}^{n} Z_{1i}^{'}\{y_{1i} - \Phi(\alpha y_{2i} + x_i\beta + \rho v_i)\} - E[Z_{1i}^{'}(y_{1i} - \Phi(\alpha y_{2i} + x_i\beta + \rho v_i))] \right| \xrightarrow{p} 0. \tag{13}$$

Similarly, for the first-stage equation, we assume that

$$\max_{\gamma \in \theta_\gamma} \left| \frac{1}{n}\sum_{i=1}^{n} Z_{2i}^{'}\{y_{2i} - z_i\gamma\} - E[Z_{1i}^{'}(y_{1i} - z_i\gamma)] \right| \xrightarrow{p} 0,$$

where Θ_γ is the possible space of γ. From the definition of $v_i = y_{1i} - z_i\gamma$. Let $\theta_3 = [\alpha\ \beta\ \rho\ \gamma]$. Let Θ_3 be the possible space of θ_3. We also assume that the following uniform convergency holds:

$$\max_{\theta_3 \in \Theta_3} \left| \frac{1}{n} \sum_{i=1}^{n} Z_{1i}^{'} \{ y_{1i} - \Phi(\alpha y_{2i} + x_i\beta + \rho(y_2 - z_i\gamma)) \} \right.$$
$$\left. - E[Z_{1i}^{'} \{ y_{1i} - \Phi(\alpha y_{2i} + x_i\beta + \rho(y_2 - z_i\gamma)) \} \right| \xrightarrow{p} 0. \tag{14}$$

For identification, we assume that α, β, and ρ that satisfy $E[Z_1^{'}\{y_{1i} - \Phi(\alpha y_{2i} + x_i\beta + \rho v_i)\}] = 0$ 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 $y_{2i} - z_i\gamma = v_i$. Thus, we henceforth assume that γ is known in the instrument Z_{1i} and that Z_{1i} is a function of y_{2i}, x_i, and z_i. Since z_i includes x_i, we can state that Z_{1i} is a function of y_{2i} and z_i. Later, we analyze the effect of replacing γ with $\widehat{\gamma}$, namely the consistent estimate of γ, when constructing Z_{1i}.

Next, we define an $m_1 \times 1$ vector $q_{1i}(\theta, \widehat{\gamma})$ as follows:

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

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

$$q_{2i}(\gamma) = Z_{2i}^{'}(y_{2i} - z_i\gamma). \tag{15}$$

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 minimize the following objective function:

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

Let $\widehat{\theta}_1$ be the solution of the above GMM problem. Let θ_1^o be the true parameter value. Let γ^o be the true parameter value of γ. Then, the standard GMM argument implies that

$$\sqrt{n}(\widehat{\theta}_1 - \theta_1^o) = -[G_1^{'}W_1G_1]^{-1}G_1^{'}W_1\frac{1}{\sqrt{n}}\sum_{i=1}^{n} q_{1i}(\theta_1^o, \widehat{\gamma}) + o_p(1)$$

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

Note that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} q_{1i}(\theta_1^o, \widehat{\gamma}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(q_{1i}(\theta_1^o, \gamma^o) + \frac{\partial q_{1i}(\theta_1^o, \gamma^o)}{\partial \gamma}\frac{1}{\sqrt{n}}\sqrt{n}(\widehat{\gamma} - \gamma^o)\right) + o_p(1), \quad (16)$$

$$\text{where } \frac{\partial q_{1i}(\theta_1^o, \gamma^o)}{\partial \gamma} \text{ is an } m_1 \times l \text{ matrix} \quad (17)$$

$$\text{and } \sqrt{n}(\widehat{\gamma} - \gamma^o) \text{ is an } l \times 1 \text{ random vector} \quad (18)$$

We assume that γ is estimated by solving the following GMM problem:

$$\min_{\gamma \in \Theta_\gamma} \qquad (\frac{1}{n}\sum_{i=1}^{n} q_{2i}(\gamma))'\widehat{W}_2(\frac{1}{n}\sum_{i=1}^{n} q_{2i}(\gamma)),$$

where $q_{2i}(\gamma)$ is defined in (15). Then, $\sqrt{n}(\widehat{\gamma} - \gamma^o)$ becomes

$$\begin{aligned}
\sqrt{n}(\widehat{\gamma} - \gamma^o) &= -[G_2^{'}W_2G_2]^{-1}G_2^{'}W_2\frac{1}{\sqrt{n}}\sum_{i=1}^{n} q_{2i}(\gamma^o) + o_p(1) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} s_i + o_p(1), \\
\text{where } G_2 &= E[\frac{\partial q_{2i}(\gamma^o)}{\partial \gamma}] \text{ and is an } m_2 \times l \text{ matrix;} \\
s_i &= -[G_2^{'}W_2G_2]^{-1}G_2^{'}W_2q_{2i} \\
&\equiv -[G_2^{'}W_2G_2]^{-1}G_2^{'}W_2Z_{2i}^{'}(y_{2i} - z_i\gamma^o) \text{ and is an } l \times 1 \text{ matrix.}
\end{aligned}$$

Thus, (16) can be written as follows:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} q_{1i}(\theta_1^o, \widehat{\gamma}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} q_{1i}(\theta_1^o, \gamma^o) + \frac{1}{n}\sum_{i=1}^{n}\frac{\partial q_{1i}(\theta_1^o, \gamma^o)}{\partial \gamma}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} s_i + o_p(1). \quad (19)$$

Note that $\frac{1}{n}\sum_{i=1}^{n}\frac{\partial q_{1i}(\theta_1^o, \gamma^o)}{\partial \gamma} \xrightarrow{p} E[\frac{\partial q_{1i}(\theta_1^o, \gamma^o)}{\partial \gamma}]$. Let $A = E[\frac{\partial q_{1i}(\theta_1^o, \gamma^o)}{\partial \gamma}]$. The above equation now becomes

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} q_{1i}(\theta_1^o, \widehat{\gamma}) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} q_{1i}(\theta_1^o, \gamma^o) + A\frac{1}{\sqrt{n}}\sum_{i=1}^{n} s_i + o_p(1) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(q_{1i}(\theta_1^o, \gamma^o) + As_i\right) + o_p(1), \\ \text{where } A &= E\left[\frac{\partial q_{1i}(\theta_1^o, \gamma^o)}{\partial \gamma}\right].\end{aligned}$$

Then, $E[q_{1i}] = 0$ and $E[s_i] = 0$ from the law of the iterated expectations, and $q_{1i}(\theta_1^o, \gamma^o) + As_i$ is independent across i. Thus, $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(q_{1i}(\theta_1^o, \gamma^o) + As_i\right)$ is distributed with mean 0 and variance $E[(q_{1i} + As_i)(q_{1i} + As_i)']$. In addition, note that $E[q_{1i}s_i] = E[E[q_{1i}s_i|y_{2i}, z_i]] = E[s_i E[q_{1i}|y_{2i}, z_i]] = 0$. Thus,

$$\begin{aligned}& Avar(\sqrt{n}(\widehat{\theta} - \theta_1^o)) \\ &= [G_1'W_1G_1]^{-1}G_1'W_1E[(q_{1i}(\theta_1^o, \gamma^o) + As_i)(q_{1i}(\theta_1^o, \gamma^o) + As_i)']W_1G_1[G_1'W_1G_1]^{-1} \qquad (20)\\ &= \underbrace{[G_1'W_1G_1]^{-1}G_1'W_1E[q_{1i}(\theta_1^o, \gamma^o)q_{1i}'(\theta_1^o, \gamma^o)]W_1G_1[G_1'W_1G_1]^{-1}}_{\text{standard covariance of GMM}} \qquad (21)\\ &+ \underbrace{[G_1'W_1G_1]^{-1}G_1'W_1AE[s_is_i']A'W_1G_1[G_1'W_1G_1]^{-1}}_{\text{the effect of sampling error of } \widehat{\gamma}}, \qquad (22)\end{aligned}$$

$$\text{where } A = E\left[\frac{\partial q_{1i}(\theta_1^o, \gamma^o)}{\partial \gamma}\right] \quad \text{and } s_i = -[G_2'W_2G_2]^{-1}G_2'W_2Z_{2i}'(y_{2i} - z_i\gamma^o).$$

In the above equation, (21) is the covariance matrix of the GMM estimator when γ is known. The term (22) is the covariance term measuring the sampling error due to the estimated value $\widehat{\gamma}$ being used 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 $E[s_is_i']$. Thus, Z_{1i} affects both (21) and (22). On the other hand, (22) is positive semi-definite. Thus, irrespective of Z_{1i}, it is optimal to minimize $E[s_is_i']$ in the sense of the matrix. When $E[s_is_i']$ becomes smaller in the sense of the matrix, then (22) becomes smaller. By the definition of s_i, $E[s_is_i']$ becomes

$$E[s_is_i'] = [G_2'W_2G_2]^{-1}G_2'W_2E[q_{2i}(\theta_1^o, \gamma^o)q_{2i}(\theta_1^o, \gamma^o)']W_2G_2[G_2'W_2G_2]^{-1}.$$

This equation implies that $E[s_i s_i^{'}]$ is the asymptotic covariance of $\widehat{\gamma}$. Thus, choosing Z_{2i} to make $\widehat{\gamma}$ efficient will also make $\widehat{\theta}_1$ more efficient when there is endogeneity on y_{2i} and there is heteroscedasticity in the error term of the first-stage equation. When y_{2i} is not endogenous, $\rho = 0$ and $A = 0$. Thus, making $E[s_i s_i^{'}]$ small does not affect $Avar(\sqrt{n}(\widehat{\theta} - \theta_1^o))$. To make $E[s_i s_i^{'}]$ smallest in the sense of the matrix, we apply the standard argument of the optimal instrument because $E[s_i s_i^{'}]$ is the asymptotic covariance of $\widehat{\gamma}$. The optimal instrument Z_{2i}^* is a $1 \times l$ vector, defined as

$$\begin{aligned} Z_{2i}^* &= \frac{1}{\sigma_2^2(z_i)} J_2 \qquad (23) \\ \text{where } J_2 &= E[\frac{\partial r_{2i}}{\partial \gamma}|_{\gamma=\gamma^o} z_i] \text{ and it is } 1 \times l \text{ matrix} \\ \sigma_2^2(z_i) &= E[r_2^2|_{\gamma=\gamma^o} z_i]; \\ r_{2i} &= y_{2i} - z_i \gamma. \end{aligned}$$

Thus, Z_{2i}^* becomes

$$Z_{2i}^* = \frac{1}{\sigma_2^2(z_i)} z_i. \qquad (24)$$

When $Z_{2i}^*$is chosen, $E[s_i s_i^{'}]$ becomes

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

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. In other words, it is not possible to find Z_{1i} that minimizes (20). This is because Z_{1i} affects both (21) and (22). Therefore, one alternative solution would be to find Z_{1i} that minimizes (21).

Let Z_{1i}^* be Z_{1i} that minimizes (21). Then, Z_{1i}^* becomes

$$Z_{1i}^* = \frac{1}{\sigma_1^2(y_{2i}, z_i)} J_1 \tag{25}$$

$$\text{where } J_1 = E[\frac{\partial r_{1i}(\gamma)}{\partial \theta_1}\Big|_{\theta=\theta^o,\gamma=\gamma^o} y_{2i}, z_i] \text{ and is a } 1 \times (k+2) \text{ matrix}$$

$$\sigma_1^2(y_{2i}, z_i) = E[r_{1i}(\gamma)^2\big|_{\theta=\theta^o,\gamma=\gamma^o} y_{2i}, z_i] = \Phi_i(1-\Phi_i)$$

$$r_{1i} = y_{1i} - \Phi(\alpha y_{2i} + x_i\beta + \rho(y_{2i} - z_i\gamma)) \tag{26}$$

$$\Phi_i = \Phi(\alpha^o y_{2i} + x_i\beta^o + \rho^o(y_{2i} - z_i\gamma^o)) \tag{27}$$

$$a^o, \beta^o, \rho^o \text{ and } \gamma^o \text{ are true value of } \alpha, \beta, \rho \text{ and } \gamma \tag{28}$$

This implies that Z_{1i}^* can be written as follows:

$$Z_{1i}^* = \frac{1}{\Phi_i(1-\Phi_i)}[\phi_i y_2, \phi_i z_{1i}, \phi_i z_{2i}, .., \phi_i z_{ki}, \phi_i v_i] \tag{29}$$

$$\text{where } z_{ji} \text{ is the j-th element of } z_i. \tag{30}$$

$$\phi_i = \phi(\alpha^o y_{2i} + x_i\beta^o + \rho^o(y_{2i} - z_i\gamma^o)) \tag{31}$$

$$\phi \text{ is the standard normal density function} \tag{32}$$

The above equation has a quite intuitive interpretation. In each case, the observation with the higher variance should have the lower weight for the instrument. Furthermore, this variance depends on $\Phi_i(1-\Phi_i)$. Here, $\Phi_i(1-\Phi_i)$ is the variance term that comes from the probit error.

When Z_{1i}^* and Z_{2i}^* are chosen according to (24) and (29), we solve the following moment conditions:

$$\sum_{i=1}^{n} Z_{1i}^{*'}\{y_{1i} - \Phi(\alpha y_{2i} + x_i\beta + \rho(y_{2i} - z_i\widehat{\gamma}\ ))\} = \mathbf{0}_{(k+2)\times 1} \tag{33}$$

$$\sum_{i=1}^{n} Z_{2i}^{*'}\{y_{2i} - z_i\gamma\} = \mathbf{0}_{l\times 1} \tag{34}$$

$$\text{where } \widehat{\gamma} \text{ is the solution of (34)}$$

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 an OLS approach. In the GMM estimator, the first stage is estimated using (F)GLS in order to utilize the heteroscedasticity information in the first-stage equation. As equation (22) 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.[2] 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. Since 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 estimator.

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 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 from 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.

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

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

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

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. Since z_i includes x_i, this is not restrictive. Let Z'_{2i} be an $m \times 1$ instrument matrix that corresponds to r_{2i} . We assume that Z'_{2i} is a function of z_i. Using the conditional expectation, we have

$$E[Z'_{1i}r_{1i}] = 0 \text{ and } E[Z'_{2i}r_{2i}] = 0.$$

Let $\theta_3 = [\alpha\ \beta\ \rho\ \gamma]'$, and let Θ_3 be the possible parameter space of θ_3.

Then, we define Z_{3i} as

$$Z'_{3i} = [Z'_{1i}\ Z'_{2i}] \text{ which is } m \times 2 \text{ matrix.} \tag{35}$$

Let r_{3i} be an 2×1 matrix, where

$$r_{3i} = \begin{pmatrix} r_{1i} \\ r_{2i}. \end{pmatrix} \tag{36}$$

Now, consider $Z'_{3i}r_{3i}$. This is an $m \times 1$ matrix and

$$\begin{aligned} Z'_{3i}r_{3i} &= [Z'_{1i}\ Z'_{2i}] \begin{pmatrix} r_{1i} \\ r_{2i} \end{pmatrix} \\ &= Z'_{1i}r_{1i} + Z'_{2i}r_{2i}. \end{aligned} \tag{37}$$

Then, we have

$$E[Z'_{3i}r_{3i}] = E[Z'_{1i}r_{1i}] + E[Z'_{2i}r_{2i}]. \tag{38}$$

Because we have $E[Z'_{1i}r_{1i}] = 0$ and $E[Z'_{2i}r_{2i}] = 0$, we have

$$E[Z'_{3i}r_{3i}] = 0. \tag{39}$$

Since $E[Z_{3i}' r_{3i}] = 0$, it is natural to consider the following sample moment:

$$\frac{1}{n}\sum_{i=1}^{n} Z_{3i}' r_{3i}.$$

Define q_{3i} as $q_{3i} \equiv Z_{3i}' r_{3i}$. Let $\widehat{W}_3$ be an $m \times m$ positive definite matrix, where $\text{plim}\widehat{W}_3 = W_3$. Let θ_3 be $\theta_3 = [\alpha, \beta, \rho, \gamma]$ and let Θ_3 be the possible space of θ_3. Then, the GMM problem becomes

$$\min_{\theta_3 \in \Theta_3} \qquad (\frac{1}{n}\sum_{i=1}^{n} q_{3i})' \widehat{W}_3 (\frac{1}{n}\sum_{i=1}^{n} q_{3i}).$$

Let θ_3^o be the true value. Using the standard GMM argument, we have

$$\sqrt{n}(\widehat{\theta}_3 - \theta_3^o) = -[G_3' W_3 G_3]^{-1} G_3' W_3 \frac{1}{\sqrt{n}} \sum_{i=1}^{n} q_{3i}(\theta_3^o) + o_p(1) \ , \qquad (40)$$

where $G_3 = E\left[\dfrac{\partial q_3}{\partial \theta_3}\right]$ is an $m \times (k + l + 2)$ matrix.

In addition, the asymptotic variance of $\sqrt{n}(\widehat{\theta}_3 - \theta_3^o)$ is

$$Avar(\sqrt{n}(\widehat{\theta}_3 - \theta_3^o)) = [G_3' W_3 G_3]^{-1} G_3' W_3 E[q_{3i} q_{3i}'] W_3 G_3 [G_3' W_3 G_3]^{-1}.$$

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:*

$$Z_{3i}^{*} \equiv \Omega(z_i, y_{2_i})^{-1} \times R(z_i, y_{2i}) \tag{41}$$

$$\text{, where } Z_{3i}^{*} \text{ is a } 2 \times (2+l+k) \text{ matrix,} \tag{42}$$

$$\Omega(z_i, y_{2_i})^{-1} = \begin{pmatrix} \sigma_1^2(y_{2i}, z_i)^{-1} & 0 \\ 0 & \sigma_2^2(z_i)^{-1} \end{pmatrix} \text{ and is a } 2 \times 2 \text{ matrix;} \tag{43}$$

$$\sigma_1^2(z_i) = E[r_1^2 | z_i, y_{2i}] = \text{ and } \sigma_2^2(z_i) = E[r_2^2 | z_i]; \tag{44}$$

$$R(y_{2i}, z_i) = \begin{pmatrix} E[\nabla_\delta r_1 | z_i, y_{2i}] & E[\nabla_\gamma r_1 | z_i, y_{2i}] \\ \mathbf{0}_{1\times(2+k)} & E[\nabla_\gamma r_2 | z_i] \end{pmatrix} \text{ and is a } 2 \times (2+k+l) \text{ matrix} \tag{45}$$

$$\nabla_\delta r_1, \nabla_\gamma r_1, \nabla_\gamma r_2 \text{ are derivative matrices, and } \nabla_\delta r_1 = \frac{\partial r_1}{\partial \delta}, \nabla_\gamma r_1 = \frac{\partial r_1}{\partial \gamma}, \nabla_\gamma r_2 = \frac{\partial r_2}{\partial \gamma} \tag{46}$$

The asymptotic variance of $\sqrt{n}(\widehat{\theta}_3 - \theta_3^o)$ *is given by:*

$$Avar(\sqrt{n}(\widehat{\theta}_3 - \theta_3^o)) = E[R(y_{2i}, z_i)' \Omega(z_i, y_{2_i})^{-1} R(y_{2i}, z_i)]^{-1} \tag{47}$$

The proof consists of several steps. First, we show that $G_3 = G_3'$. This implies that $Avar = G_3^{-1} E[q_{3i} q_{3i}'] G_3^{-1}$. Second, we show that $G_3 = E[q_{3i} q_{3i}']$. This implies that $G_3^{-1} E[q_{3i} q_{3i}'] G_3^{-1} = G_3^{-1}$. Third, we show that $E[R(y_{2i}, z_i)' \Omega(z_i, y_{2_i})^{-1} R(y_{2i}, z_i)] = G_3$. Then, we have $Avar = E[R(y_{2i}, z_i)' \Omega(z_i, y_{2_i})^{-1} R(y_{2i}, z_i)]^{-1}$. Fourth, we show that if the asymptotic covariance has this form, it achieves the minimum bound. The proof is provided in the appendix.

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

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

This is equivalent to

$$\frac{1}{n}\sum_{i=1}^{n}(Z_{1i}^{*'}Z_{2i}^{*'})\begin{pmatrix} r_{1i} \\ r_{2i} \end{pmatrix} = \mathbf{0}_{(k+2+l)\times 1}$$

$$\text{where } Z_{1i}^{*'} = \begin{pmatrix} \sigma_1^2(y_{2i}, z_i)^{-1}E[\nabla_\delta r_1|z_i, y_{2i}]' \\ \sigma_1^2(y_{2i}, z_i)^{-1}E[\nabla_\gamma r_1|z_i, y_{2i}]' \end{pmatrix}$$

$$Z_{2i}^{*'} = \begin{pmatrix} \mathbf{0}_{(2+k)\times 1} \\ \sigma_2^2(z_i)^{-1}E[\nabla_\gamma r_2|z_i]' \end{pmatrix}.$$

Thus, we have the following moment conditions:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\Phi_i(1-\Phi_i)}\phi_i y_{2i} r_{1i} = 0 \tag{48}$$

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\Phi_i(1-\Phi_i)}\phi_i x_i^{'} r_{1i} = \mathbf{0}_{k\times 1} \tag{49}$$

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\Phi_i(1-\Phi_i)}\phi_i (y_{2i} - z_i\gamma) r_{1i} = 0 \tag{50}$$

$$\frac{1}{n}\sum_{i=1}^{n}\{\frac{1}{\Phi_i(1-\Phi_i)}\phi_i z_i^{'} r_{1i} - \sigma_2^2(z_i)^{-1} z_i' r_{2i}\} = \mathbf{0}_{l\times 1}. \tag{51}$$

We can make several observations on the above moment conditions. First, the above conditions again clearly show the importance of considering the heteroscedasticity in the first-stage equation, which is captured by $\sigma_1^2(z_i)^{-1}$. Second, the moment conditions 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 over-identified. Third, in the just-identified case, the above joint-estimation can be implemented in two steps[3]. Thus, in the just-identified case, the moment conditions considered in the previous section, (33) and (34), are also fully efficient.

[3]We can prove the just-identified case as follows. Note that the first k elements of z_i are x_i. Thus, from (49) and (51), 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 (48), (49), and (50), 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 (51), 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.

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 (Carroll (1982) and Robinson (1987))(For 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) studied 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 showed that by using the k-nearest neighbor 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) used 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 j-th element of z_i. Let σ_{z_j} be the sample standard deviation of the j-th 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:

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

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 that satisfy $d(i,t) > 0$. Define $\mathcal{N}(z_i)$ as follows:

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

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 k-th nearest neighbor at z_i is calculated as follows:

$$\widehat{m}_k(z_i) = \frac{1}{k} \sum_{t \in \mathcal{N}(z_i)} \widehat{\sigma}^2_{2t} \ , \tag{52}$$

$$\text{where } \widehat{\sigma}_{2i} \text{ is the calculated residual of observation of } i. \tag{53}$$

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

$$CV(k) = \sum_{i=1}^{n} \{\widehat{\sigma}^2_{2i} - \widehat{m}_k(z_i)\}^2. \tag{54}$$

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 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, whose probability limits are the true values. In addition, we need to use the estimate value of $\sigma^2_2(z_i)$ to construct Z^*_{3i}. Here, we determine the asymptotic behavior of $\sqrt{n}(\widehat{\theta}_3 - \theta^o_3)$ when the estimated values of α, β, ρ, γ, 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 α, β, ρ, γ, and $\sigma^2_2(z_i)$. As in 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 $\sqrt{n}(\widehat{\theta}_3 - \theta_3^o)$.

2.5 Estimation Procedure

The actual procedure for the estimation is as follows:

1. Run the OLS regression y_{2i} on z_i, and keep the residual as $\hat{v}_i$. Calculate $\hat{v}_i^2$.

2. Run the probit regression y_{1i} on $y_{2i}, x_i, \hat{v}_i$, 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 $\widehat{\Phi}_i = \Phi(\widehat{\alpha}y_{2i} + \widehat{\beta}x_i + \widehat{\rho}\hat{v}_i)$ and $\widehat{\phi}_i = \phi(\widehat{\alpha}y_{2i} + \widehat{\beta}x_i + \widehat{\rho}\hat{v}_i)$.

3. Apply the k-nn non-parametric method to estimate $\sigma_2^2(z_i)$ from $\hat{v}_i^2$. Since 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 $\sigma_2^2(z_i)$. For the choice of k, 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:

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

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

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\widehat{\Phi}_i(1-\widehat{\Phi}_i)}\widehat{\phi}_i \hat{v}_i r_{1i} = 0$$

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

5. In order to calculate the asymptotic variance of $\sqrt{n}(\widehat{\theta}_3 - \theta_3^o)$, we use equation (47). Let $\widetilde{R}(y_{2i}, z_i)$ be the calculated residual after following the above GMM procedure. Let $\widetilde{\Omega}(z_i, y_{2_i})$ be the estimated conditional variance. Then, the asymptotic variance is calculated as

$$Avar(\sqrt{n}(\widehat{\theta}_3 - \theta_3^o)) = \left[n^{-1} \sum_{i=1}^{n} \widetilde{R}(y_{2i}, z_i)' \widetilde{\Omega}(z_i, y_{2_i})^{-1} \widetilde{R}(y_{2i}, z_i) \right]^{-1}. \quad (55)$$

3 Monte Carlo Simulation

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

$$y_1^* = \alpha y_2 + \beta_0 + x_1\beta_1 + \rho v + e \quad (56)$$

$$y_1 = 1 \text{ if } y_1^* > 0 \quad (57)$$

$$y_1 = 0 \text{ if } y_1^* \leq 0 \quad (58)$$

$$y_2 = \gamma_0 + x_1\gamma_1 + z_1\gamma_2 + z_2\gamma_3 + v. \quad (59)$$

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 $v \sim N(0, e^{\lambda z_2})$, 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-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 $\rho = 0, 1, 2, -1, -2$. [4] 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

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

is set to 500 and we replicate 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. To calculate k is time consuming because it needs 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 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 simulation underestimates the performance of the GMM with optimal instruments. We will demonstrate that, even in such a situation, our estimator outperforms the 2SCML method.

Table 1 shows the bias and RMSE of the estimated α, β_0, β_1, and ρ for 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 percent 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 The Effect of Non-labor Income on Married Women's Labor-force Participation

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

In addition to a married woman's own labor 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 labor-force participation. Although we often observe a negative impact of husband's earnings on wife's labor-force participation, note that the husband's earnings could be endogenous. If a wife has a high ability in market work, she is more likely to work and 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 efforts between market and household production makes the other household income (the household income less the married woman's labor earnings) endogenous. To deal with this potential endogeneity in the other household income, the variable is instrumented using dummy variables indicating the husband's occupation.

This empirical exercise is implemented by 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 labor-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, once endogeneity is considered. The estimated ρ is negative and statistically significant. As expected, the unobserved determinants of the wife's labor-force participation and other income are negatively correlated, perhaps resulting from the husband's endogenous effort allocation in market work.

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 in (22) 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 simulation. Thus, this empirical example confirms the Monte Carlo simulation result that the benefit of using 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 neighbor non-parametric estimator to characterize the heteroscedasticity of the error term of the first-stage equation. A Monte Carlo simulation reveals that the GMM estimator based on optimal instruments performs better than the 2SCML estimator 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 tool for empirical research.

Acknowledgments

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.

References

Carroll, Raymond J., "Adapting for heteroscedasticity in linear models," *Annals of Statistics*, 1982, *10*, 1224–1233.

Chamberlain, Gary, "Asymptotic efficiency in estimation with conditional moment restrictions," *Journal of Econometrics*, March 1987, *34* (3), 305–334.

Dagenais, Marcel G., "Inconsistency of a proposed nonlinear instrumental variables estimator for probit and logit models with endogenous regressors," *Economics Letters*, April 1999, *63* (1), 19–21.

Grogger, Jeffrey, "A simple test for exogeneity in probit, logit, and Poisson regression models," *Economics Letters*, August 1990, *33* (4), 329–332.

Lee, Myoung-Jae, "Semi-parametric estimation of simultaneous qquations with limited dependent variables: A case study of female labour supply," *Journal of Applied Econometrics*, April-Jun 1995, *10* (2), 187–200.

Li, Qi and Jeffrey S. Racine, *Nonparametric Econometrics: Theory and Practice*, Princeton University Press, 2007.

Lucchetti, Riccardo, "Inconsistency of naive GMM estimation for QR models with endogenous regressors," *Economics Letters*, April 2002, *75* (2), 179–185.

Newey, Whitney K., "Efficient instrumental variables estimation of nonlinear models," *Econometrica*, July 1990, *58* (4), 809–837.

_ , "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, Elsevier Sience, 1993.

Rivers, Douglas and Quang H. Vuong, "Limited information estimators and exogeneity tests for simultaneous probit models," *Journal of Econometrics*, November 1988, *39* (3), 347–366.

Robinson, Peter. M., "Efficient estimation in the presence of heteroskedasticity of unknown form," *Econometrica*, 1987, *55* (4), 875–891.

Wolfgangs, Härdle and Oliver Linton, "Chapter 18: Applied nonparametric methods," *Handbook of Econometrics*, 1994, *IV.*

Table 1: The bias and RMSE of α, β0, β1and ρ

parameter		method	α		β0		β1		ρ	
ρ	λ		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: RV stands for the 2SCML by Rivers and Voung. The sample size is 500. The number of replication is 500 for each case.

Table 2: Descriptive statistics of the analysis sample of married women

Variable name	Description	Mean	S.D.
emp	Wife reports positive hour	0.735	0.441
ohhinc	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
children0-5	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 cfraftsman	0.302	0.459
hus farm	1 if the husband is farm-related worker	0.025	0.157

Note: The number of observations is 3382. These data were extracted from 1987 wave of Michigan Panel Study of Income Dynamics (PSID). Details are available in Lee (1995).

Table 3: The determination of labor force participation of married women

	(1)	(2)	(3)	(5)	(6)
Method	Probit	Rivers-Voung		GMM	
Dep. Var.	Emp	Other income	Emp	Other income	Emp
Other income	-0.006	-	0.002	-	-0.00013
	(0.001)		(0.004)		(0.0015)
Age	-0.032	0.641	-0.038	0.717	-0.034
	(0.003)	(0.044)	(0.005)	(0.108)	(0.002)
Educ	0.134	2.196	0.108	1.782	0.117
	(0.011)	(0.205)	(0.023)	(0.198)	(0.012)
Children 0-5	-0.506	0.809	-0.510	1.126	-0.520
	(0.036)	(0.648)	(0.04)	(0.529)	(0.038)
Nonwhite	0.134	-6.522	0.209	-6.149	0.195
	(0.057)	(1.021)	(0.069)	(0.950)	(0.059)
Husband manager	-	16.647	-	8.477	-
		(1.221)		(3.046)	
Husband sales	-	6.296	-	13.590	-
		(1.108)		(13.00)	
Husband farm	-	-3.802	-	-20.181	-
		(2.903)		(24.119)	
V hat / Rho	-	-	-0.009	-	-0.006
			(0.004)		(0.0012)
Constant	0.620	-26.490	0.876	-24.568	0.907
	(0.185)	(3.396)	(0.369)	(8.940)	(0.227)
R2 / Log likelihood	-1711.416	0.196	-1709.156	-	-

Note: Number of observations is 3382. Standard errors are in parethesis.

Appendix : Proof of the optimality of the proposed instrument

In the first step, we show that $G_3 = G_3^{'}$. When Z_{3i}^* is used, then $q_{3i} = Z_{3i}^{*'} r_{3i}$. By the definition of G_3, we have

$$
\begin{aligned}
G_3 = & \ E[\frac{\partial q_3}{\partial \theta_3}] \\
& = E[Z_{3i}^{*'} \frac{\partial r_{3i}}{\partial \theta_3}] \\
& = E[R(y_{2i}, z_i)^{'} \Omega(z_i, y_{2_i})^{-1} \frac{\partial r_{3i}}{\partial \theta_3}]
\end{aligned}
$$

Now we calculate $R(y_{2i}, z_i)^{'} \Omega(z_i, y_{2_i})^{-1}$. This can be calculated as follows:

$$
\begin{aligned}
& R(y_{2i}, z_i)^{'} \Omega(z_i, y_{2_i})^{-1} \\
& = \begin{pmatrix} E[\nabla_\delta r_1 | z_i, y_{2i}], & E[\nabla_\gamma r_1 | z_i, y_{2i}] \\ \mathbf{0}_{1\times(2+k)}, & E[\nabla_\gamma r_2 | z_i] \end{pmatrix}^{'} \Omega(z_i, y_{2_i})^{-1} \\
& = \begin{pmatrix} E[\nabla_\delta r_1 | z_i, y_{2i}]', & \mathbf{0}_{(2+k)\times 1} \\ E[\nabla_\gamma r_1 | z_i, y_{2i}]', & E[\nabla_\gamma r_2 | z_i]' \end{pmatrix} \Omega(z_i, y_{2_i})^{-1} \\
& = \begin{pmatrix} \frac{1}{\sigma_1^2(y_{2i}, z_i)} E[\nabla_\delta r_1 | z_i, y_{2i}]', & \mathbf{0}_{(2+k)\times 1} \\ \frac{1}{\sigma_1^2(y_{2i}, z_i)} E[\nabla_\gamma r_1 | z_i, y_{2i}]', & \frac{1}{\sigma_2^2(z_i)} E[\nabla_\gamma r_2 | z_i]' \end{pmatrix}
\end{aligned}
$$

Next, we calculate $R(y_{2i}, z_i)^{'} \Omega(z_i, y_{2_i})^{-1} \frac{\partial r_{3i}}{\partial \theta_3}$.

$$
\begin{aligned}
& R(y_{2i}, z_i)^{'} \Omega(z_i, y_{2_i})^{-1} \frac{\partial r_{3i}}{\partial \theta_3}. \\
& = \begin{pmatrix} \frac{1}{\sigma_1^2(y_{2i}, z_i)} E[\nabla_\delta r_1 | z_i, y_{2i}]', & \mathbf{0}_{(2+k)\times 1} \\ \frac{1}{\sigma_1^2(y_{2i}, z_i)} E[\nabla_\gamma r_1 | z_i, y_{2i}]', & \frac{1}{\sigma_2^2(z_i)} E[\nabla_\gamma r_2 | z_i]' \end{pmatrix} \frac{\partial r_{3i}}{\partial \theta_3} \\
& = \begin{pmatrix} \frac{1}{\sigma_1^2(y_{2i}, z_i)} E[\nabla_\delta r_1 | z_i, y_{2i}]' & \mathbf{0}_{(2+k)\times 1} \\ \frac{1}{\sigma_1^2(y_{2i}, z_i)} E[\nabla_\gamma r_1 | z_i, y_{2i}]' & \frac{1}{\sigma_2^2(z_i)} E[\nabla_\gamma r_2 | z_i]' \end{pmatrix} \begin{pmatrix} \nabla_\delta r_1 & \nabla_\gamma r_1 \\ \mathbf{0}_{1\times(2+k)} & \nabla_\gamma r_2 \end{pmatrix} \\
& = \left(\begin{matrix} \frac{1}{\sigma_1^2(y_{2i}, z_i)} E[\nabla_\delta r_1 | z_i, y_{2i}]' \nabla_\delta r_1, \\ \frac{1}{\sigma_1^2(y_{2i}, z_i)} E[\nabla_\gamma r_1 | z_i, y_{2i}]' \nabla_\delta r_1, \end{matrix} \right. \\
& \qquad \left. \begin{matrix} \frac{1}{\sigma_1^2(y_{2i}, z_i)} E[\nabla_\delta r_1 | z_i, y_{2i}]' \nabla_\gamma r_1 \\ \frac{1}{\sigma_1^2(y_{2i}, z_i)} E[\nabla_\gamma r_1 | z_i, y_{2i}]' \nabla_\gamma r_1 + \frac{1}{\sigma_2^2(z_i)} E[\nabla_\gamma r_2 | z_i]' \nabla_\gamma r_2 \end{matrix} \right)
\end{aligned}
$$

Thus, G_3 becomes

$$E\begin{pmatrix} \frac{1}{\sigma_1^2(y_{2i},z_i)}E[\nabla_\delta r_1|z_i,y_{2i}]'\nabla_\delta r_1, & \frac{1}{\sigma_1^2(y_{2i},z_i)}E[\nabla_\delta r_1|z_i,y_{2i}]'\nabla_\gamma r_1 \\ \frac{1}{\sigma_1^2(y_{2i},z_i)}E[\nabla_\gamma r_1|z_i,y_{2i}]'\nabla_\delta r_1, & \frac{1}{\sigma_1^2(y_{2i},z_i)}E[\nabla_\gamma r_1|z_i,y_{2i}]'\nabla_\gamma r_1 + \frac{1}{\sigma_2^2(z_i)}E[\nabla_\gamma r_2|z_i]'\nabla_\gamma r_2 \end{pmatrix}$$

$$= \begin{pmatrix} E[\frac{1}{\sigma_1^2(y_{2i},z_i)}E[\nabla_\delta r_1|z_i,y_{2i}]'\nabla_\delta r_1], & E[\frac{1}{\sigma_1^2(y_{2i},z_i)}E[\nabla_\delta r_1|z_i,y_{2i}]'\nabla_\gamma r_1] \\ E[\frac{1}{\sigma_1^2(y_{2i},z_i)}E[\nabla_\gamma r_1|z_i,y_{2i}]'\nabla_\delta r_1], & E[\frac{1}{\sigma_1^2(y_{2i},z_i)}E[\nabla_\gamma r_1|z_i,y_{2i}]'\nabla_\gamma r_1 + \frac{1}{\sigma_2^2(z_i)}E[\nabla_\gamma r_2|z_i]'\nabla_\gamma r_2] \end{pmatrix}$$

For the top right, top left, lower left parts, we take the conditional expectation on y_{2i} and z_i and apply the law of iterated expectation.

$$\begin{pmatrix} E[E[\frac{1}{\sigma_1^2(y_{2i},z_i)}E[\nabla_\delta r_1|z_i,y_{2i}]'\nabla_\delta r_1|y_{2i},z_i]], & E[E[\frac{1}{\sigma_1^2(y_{2i},z_i)}E[\nabla_\delta r_1|z_i,y_{2i}]'\nabla_\gamma r_1|y_{2i},z_i]] \\ E[E[\frac{1}{\sigma_1^2(y_{2i},z_i)}E[\nabla_\gamma r_1|z_i,y_{2i}]'\nabla_\delta r_1|y_{2i},z_i]], & E[\frac{1}{\sigma_1^2(y_{2i},z_i)}E[\nabla_\gamma r_1|z_i,y_{2i}]'\nabla_\gamma r_1 + \frac{1}{\sigma_2^2(z_i)}E[\nabla_\gamma r_2|z_i]'\nabla_\gamma r_2] \end{pmatrix}$$

$$= \begin{pmatrix} E[\frac{1}{\sigma_1^2(y_{2i},z_i)}E[\nabla_\delta r_1|z_i,y_{2i}]'E[\nabla_\delta r_1|y_{2i},z_i]], & E[\frac{1}{\sigma_1^2(y_{2i},z_i)}E[\nabla_\delta r_1|z_i,y_{2i}]'E[\nabla_\gamma r_1|y_{2i},z_i]] \\ E[\frac{1}{\sigma_1^2(y_{2i},z_i)}E[\nabla_\gamma r_1|z_i,y_{2i}]'E[\nabla_\delta r_1|y_{2i},z_i]], & E[\frac{1}{\sigma_1^2(y_{2i},z_i)}E[\nabla_\gamma r_1|z_i,y_{2i}]'\nabla_\gamma r_1 + \frac{1}{\sigma_2^2(z_i)}E[\nabla_\gamma r_2|z_i]'\nabla_\gamma r_2] \end{pmatrix}$$

For the lower right part, first divide the expectation into two parts. Thus, we have

$$\begin{aligned} &E[\frac{1}{\sigma_1^2(y_{2i},z_i)}E[\nabla_\gamma r_1|z_i,y_{2i}]'\nabla_\gamma r_1 + \frac{1}{\sigma_2^2(z_i)}E[\nabla_\gamma r_2|z_i]'\nabla_\gamma r_2] \\ &= E[E[\frac{1}{\sigma_1^2(y_{2i},z_i)}E[\nabla_\gamma r_1|z_i,y_{2i}]'\nabla_\gamma r_1] + E[\frac{1}{\sigma_2^2(z_i)}E[\nabla_\gamma r_2|z_i]'\nabla_\gamma r_2]] \end{aligned}$$

For the first term, we take the conditional expectation on z_i and y_{2i} and apply the law of iterated expectation. For the second part, we take the conditional expectation on z_i and apply the law of iterated expectation. Then, we have

$$\begin{aligned} &= E[E[E[\frac{1}{\sigma_1^2(y_{2i},z_i)}E[\nabla_\gamma r_1|z_i,y_{2i}]'\nabla_\gamma r_1|y_{2i},zi]] + E[E[\frac{1}{\sigma_2^2(z_i)}E[\nabla_\gamma r_2|z_i]'\nabla_\gamma r_2|z_i]]] \\ &= E[\frac{1}{\sigma_1^2(y_{2i},z_i)}E[\nabla_\gamma r_1|z_i,y_{2i}]'E[\nabla_\gamma r_1|y_{2i},zi]] + E[\frac{1}{\sigma_2^2(z_i)}E[\nabla_\gamma r_2|z_i]'E[\nabla_\gamma r_2|z_i]] \end{aligned}$$

Therefore, G_3 becomes

$$G_3 = \begin{pmatrix} E[\frac{1}{\sigma_1^2(y_{2i},z_i)} E[\nabla_\delta r_1|z_i, y_{2i}]' E[\nabla_\delta r_1|y_{2i}, z_i]], & E[\frac{1}{\sigma_1^2(y_{2i},z_i)} E[\nabla_\delta r_1|z_i, y_{2i}]' E[\nabla_\gamma r_1|y_{2i}, z_i]] \\ E[\frac{1}{\sigma_1^2(y_{2i},z_i)} E[\nabla_\gamma r_1|z_i, y_{2i}]' E[\nabla_\delta r_1|y_{2i}, z_i]], & E[\frac{1}{\sigma_1^2(y_{2i},z_i)} E[\nabla_\gamma r_1|z_i, y_{2i}]' E[\nabla_\gamma r_1|y_{2i}, zi]] + E[\frac{1}{\sigma_2^2(z_i)} E[\nabla_\gamma r_2|z_i]' E[\nabla_\gamma r_2|z_i]] \end{pmatrix}$$

Note that the above matrix is symmetric. It implies that $G_3 = G_3'$. Therefore, the asymptotic covariance of GMM can be written as

$$\begin{aligned} Avar(\sqrt{n}(\widehat{\theta}_3 - \theta_3^o)) &= [G_3' W_3 G_3]^{-1} G_3' W_3 E[q_{3i} q_{3i}'] W_3 G_3 [G_3' W_3 G_3]^{-1} \\ &= G_3^{-1} E[q_{3i} q_{3i}'] G_3^{-1} \end{aligned}$$

In the second step, we calculate $E[q_{3i} q_{3i}']$. When the optimal instrument Z_{3i}^* is used, $q_{3i} = Z_{3i}^{*'} r_{3i}$. Thus, $E[q_{3i} q_{3i}']$ becomes

$$\begin{aligned} & E[q_{3i} q_{3i}'] \\ &= E[Z_{3i}^{*'} r_{3i} r_{3i}' Z_{3i}^*] \\ &= E[R(z_i, y_{2i})' \Omega(z_i, y_{2_i})^{-1} r_{3i} r_{3i}' \Omega(z_i, y_{2_i})^{-1} R(z_i, y_{2i})] \end{aligned}$$

Taking the conditional expectation on y_{2i} and z_i and applying the law of the iterated expectation, we have

$$E[R(z_i,y_{2i})'\Omega(z_i,y_{2_i})^{-1}E[r_{3i}r_{3i}'|y_{2i},z_i]\Omega(z_i,y_{2_i})^{-1}R(z_i,y_{2i})]$$

$$= E[R(z_i,y_{2i})'\Omega(z_i,y_{2_i})^{-1}\begin{pmatrix} E[r_{1i}^2|y_{2i},z_i] & 0 \\ 0 & E[r_{2i}^2|z_i,y_{2i}] \end{pmatrix}\Omega(z_i,y_{2_i})^{-1}R(z_i,y_{2i})]$$

$$= E[R(z_i,y_{2i})'\begin{pmatrix} \frac{1}{\sigma_1^2(y_{2i},z_i)} & 0 \\ 0 & E[r_{2i}^2|z_i]^{-1}E[r_2^2|z_i,y_{2i}]E[r_2^2|z]^{-1} \end{pmatrix}R(z_i,y_{2i})]$$

$$= E[\begin{pmatrix} E[\nabla_\delta r_1|z_i,y_{2i}]' & \mathbf{0}_{(2+k)\times 1} \\ E[\nabla_\gamma r_1|z_i,y_{2i}]' & E[\nabla_\gamma r_2|z_i]' \end{pmatrix}\begin{pmatrix} \frac{1}{\sigma_1^2(y_{2i},z_i)} & 0 \\ 0 & \frac{1}{\sigma_2^2(z_i)}E[r_2^2|z_i,y_{2i}]\frac{1}{\sigma_2^2(z_i)} \end{pmatrix}$$

$$\times \begin{pmatrix} E[\nabla_\delta r_1|z_i,y_{2i}] & E[\nabla_\gamma r_1|z_i,y_{2i}] \\ \mathbf{0}_{1\times(2+k)} & E[\nabla_\gamma r_2|z_i] \end{pmatrix}]$$

$$= E[\begin{pmatrix} E[\nabla_\delta r_1|z_i,y_{2i}]' & \mathbf{0}_{(2+k)\times 1} \\ E[\nabla_\gamma r_1|z_i,y_{2i}]' & E[\nabla_\gamma r_2|z_i]' \end{pmatrix}$$

$$\times \begin{pmatrix} \frac{1}{\sigma_1^2(y_{2i},z_i)}E[\nabla_\delta r_1|z_i,y_{2i}] & \frac{1}{\sigma_1^2(y_{2i},z_i)}E[\nabla_\gamma r_1|z_i,y_{2i}] \\ 0 & \sigma_2^2(z_i)^{-2}E[r_2^2|z_i,y_{2i}]E[\nabla_\gamma r_2|z_i] \end{pmatrix}]$$

$$= E[\begin{pmatrix} E[\nabla_\delta r_1|z_i,y_{2i}]'\frac{1}{\sigma_1^2(y_{2i},z_i)}E[\nabla_\delta r_1|z_i,y_{2i}], & E[\nabla_\delta r_1|z_i,y_{2i}]'\frac{1}{\sigma_1^2(y_{2i},z_i)}E[\nabla_\gamma r_1|z_i,y_{2i}] \\ E[\nabla_\gamma r_1|z_i,y_{2i}]'\frac{1}{\sigma_1^2(y_{2i},z_i)}E[\nabla_\delta r_1|z_i,y_{2i}], & E[\nabla_\gamma r_1|z_i,y_{2i}]'\frac{1}{\sigma_1^2(y_{2i},z_i)}E[\nabla_\gamma r_1|z_i,y_{2i}] + E[\nabla_\gamma r_2|z_i]'\sigma_2^2(z_i)^{-2}E[r_2^2|z_i,y_{2i}]E[\nabla_\gamma r_2|z_i] \end{pmatrix}]$$

$$= \begin{pmatrix} E\left[E[\nabla_\delta r_1|z_i,y_{2i}]'\frac{1}{\sigma_1^2(y_{2i},z_i)}E[\nabla_\delta r_1|z_i,y_{2i}]\right], & E\left[E[\nabla_\delta r_1|z_i,y_{2i}]'\frac{1}{\sigma_1^2(y_{2i},z_i)}E[\nabla_\gamma r_1|z_i,y_{2i}]\right] \\ E\left[E[\nabla_\gamma r_1|z_i,y_{2i}]'\frac{1}{\sigma_1^2(y_{2i},z_i)}E[\nabla_\delta r_1|z_i,y_{2i}]\right], & E\left[E[\nabla_\gamma r_1|z_i,y_{2i}]'\frac{1}{\sigma_1^2(y_{2i},z_i)}E[\nabla_\gamma r_1|z_i,y_{2i}] + E[\nabla_\gamma r_2|z_i]'\sigma_2^2(z_i)^{-2}E[r_2^2|z_i,y_{2i}]E[\nabla_\gamma r_2|z_i]\right] \end{pmatrix}$$

Now, focus on the lower right part. Divide the expectation on two terms. Then, we have

$$E[E[\nabla_\gamma r_1|z_i,y_{2i}]'\frac{1}{\sigma_1^2(y_{2i},z_i)}E[\nabla_\gamma r_1|z_i,y_{2i}]]+E\left[E[\nabla_\gamma r_2|z_i]'\sigma_2^2(z_i)^{-2}E[r_2^2|z_i,y_{2i}]E[\nabla_\gamma r_2|z_i]\right]$$

For the second term, taking the conditional expectation on z_i and apply the law of iterated expectation. Note that $E[E[r_2^2|z_i, y_{2i}]|z_i] = E[r_2^2|z_i]$. Then, we have

$$
\begin{aligned}
&E[E[\nabla_\gamma r_1|z_i, y_{2i}]' \frac{1}{\sigma_1^2(y_{2i}, z_i)} E[\nabla_\gamma r_1|z_i, y_{2i}]] + E[E\left[E[\nabla_\gamma r_2|z_i]'\sigma_2^2(z_i)^{-2}E[r_2^2|z_i, y_{2i}]E[\nabla_\gamma r_2|z_i]|z_i\right]] \\
&= E[E[\nabla_\gamma r_1|z_i, y_{2i}]' \frac{1}{\sigma_1^2(y_{2i}, z_i)} E[\nabla_\gamma r_1|z_i, y_{2i}]] + E[E\left[E[\nabla_\gamma r_2|z_i]'\sigma_2^2(z_i)^{-2}E[r_2^2|z_i]E[\nabla_\gamma r_2|z_i]\right]] \\
&= E[E[\nabla_\gamma r_1|z_i, y_{2i}]' \frac{1}{\sigma_1^2(y_{2i}, z_i)} E[\nabla_\gamma r_1|z_i, y_{2i}]] + E[E[\nabla_\gamma r_2|z_i]'\sigma_2^2(z_i)^{-1}E[\nabla_\gamma r_2|z_i]] \\
&= E[E[\nabla_\gamma r_1|z_i, y_{2i}]' \frac{1}{\sigma_1^2(y_{2i}, z_i)} E[\nabla_\gamma r_1|z_i, y_{2i}] + E[\nabla_\gamma r_2|z_i]' \frac{1}{\sigma_2^2(z_i)} E[\nabla_\gamma r_2|z_i]]
\end{aligned}
$$

Therefore, $E[q_{3i}q_{3i}']$ becomes

$$
\begin{aligned}
&E[q_{3i}q_{3i}'] \\
&= \begin{pmatrix} E\left[E[\nabla_\delta r_1|z_i, y_{2i}]' \frac{1}{\sigma_1^2(y_{2i}, z_i)} E[\nabla_\delta r_1|z_i, y_{2i}]\right] & E\left[E[\nabla_\delta r_1|z_i, y_{2i}]' \frac{1}{\sigma_1^2(y_{2i}, z_i)} E[\nabla_\gamma r_1|z_i, y_{2i}]\right] \\ E\left[E[\nabla_\gamma r_1|z_i, y_{2i}]' \frac{1}{\sigma_1^2(y_{2i}, z_i)} E[\nabla_\delta r_1|z_i, y_{2i}]\right] & E[E[\nabla_\gamma r_1|z_i, y_{2i}]' \frac{1}{\sigma_1^2(y_{2i}, z_i)} E[\nabla_\gamma r_1|z_i, y_{2i}] + E[\nabla_\gamma r_2|z_i]' \frac{1}{\sigma_2^2(z_i)} E[\nabla_\gamma r_2|z_i]] \end{pmatrix} \\
&= G_3
\end{aligned}
$$

Therefore, we have

$$G_3 = G_3' = E[q_{3i}q_{3i}']$$

This implies the asymptotic covariance becomes

$$Avar(\sqrt{n}(\widehat{\theta}_3 - \theta_3^o)) = G_3^{-1}$$

In the third step, we calculate $E[R(y_{2i},z_i)^{'}\Omega(z_i,y_{2_i})^{-1}R(y_{2i},z_i)]$.

$$
\begin{aligned}
&E[R(y_{2i},z_i)^{'}\Omega(z_i,y_{2_i})^{-1}R(y_{2i},z_i)] \\
&= E[\begin{pmatrix} \frac{1}{\sigma_1^2(y_{2i},z_i)}E[\nabla_\delta r_1|z_i,y_{2i}]' & \mathbf{0}_{(2+k)\times 1} \\ \frac{1}{\sigma_1^2(y_{2i},z_i)}E[\nabla_\gamma r_1|z_i,y_{2i}]' & \frac{1}{\sigma_2^2(z_i)}E[\nabla_\gamma r_2|z_i]' \end{pmatrix} R(y_{2i},z_i)] \\
&= E[\begin{pmatrix} \frac{1}{\sigma_1^2(y_{2i},z_i)}E[\nabla_\delta r_1|z_i,y_{2i}]' & \mathbf{0}_{(2+k)\times 1} \\ \frac{1}{\sigma_1^2(y_{2i},z_i)}E[\nabla_\gamma r_1|z_i,y_{2i}]' & \frac{1}{\sigma_2^2(z_i)}E[\nabla_\gamma r_2|z_i]' \end{pmatrix} \begin{pmatrix} E[\nabla_\delta r_1|z_i,y_{2i}] & E[\nabla_\gamma r_1|z_i,y_{2i}] \\ \mathbf{0}_{1\times(2+k)} & E[\nabla_\gamma r_2|z_i] \end{pmatrix}] \\
&= E\begin{pmatrix} \frac{1}{\sigma_1^2(y_{2i},z_i)}E[\nabla_\delta r_1|z_i,y_{2i}]'E[\nabla_\delta r_1|z_i,y_{2i}], & \frac{1}{\sigma_1^2(y_{2i},z_i)}E[\nabla_\delta r_1|z_i,y_{2i}]'E[\nabla_\gamma r_1|z_i,y_{2i}] \\ \frac{1}{\sigma_1^2(y_{2i},z_i)}E[\nabla_\gamma r_1|z_i,y_{2i}]'E[\nabla_\delta r_1|z_i,y_{2i}], & \frac{1}{\sigma_1^2(y_{2i},z_i)}E[\nabla_\gamma r_1|z_i,y_{2i}]'E[\nabla_\gamma r_1|z_i,y_{2i}] + \frac{1}{\sigma_2^2(z_i)}E[\nabla_\gamma r_2|z_i]'E[\nabla_\gamma r_2|z_i] \end{pmatrix} \\
&= G_3
\end{aligned}
$$

Thus, $G_3 = G_3' = E[R(y_{2i},z_i)^{'}\Omega(z_i,y_{2_i})^{-1}R(y_{2i},z_i)]$. This implies that

$$Avar(\sqrt{n}(\widehat{\theta_3}-\theta_3)) = G_3^{-1} = E[R(y_{2i},z_i)^{'}\Omega(z_i,y_{2_i})^{-1}R(y_{2i},z_i)]^{-1}$$

At the final step, we show that when $Avar(\sqrt{n}(\widehat{\theta_3}-\theta_3))$ has this form, it achieves its minimum bound. To prove this result, let $m_A = G_3^{'}W_3q_{3i}$. Then,

$$
\begin{aligned}
&E[m_A m_A'] \\
&= E[G_3^{'}W_3q_{3i}q_{3i}W_3G_3] \\
&= G_3^{'}W_3E[q_{3i}q_{3i}']W_3G_3
\end{aligned}
$$

Let $m_B = Z_{3i}^{*}r_{3i}$.Then, $Avar(\sqrt{n}(\widehat{\theta_3}-\theta_3^o)) = E[m_B m_B']^{-1}$.

$$
\begin{aligned}
&E[m_A m_B']\\
&= E[G_3' W_3 Z_{3i}' r_{3i} r_{3i}' R_i]\\
&= G_3' W_3 E[Z_{3i}' E[r_{3i} r_{3i} | y_{2i} z_i] Z_{3i}^*]\\
&= G_3' W_3 E[Z_{3i}' \begin{pmatrix} E[r_1^2|y_{2i}, z_i] & 0 \\ 0 & E[r_2^2|y_{2i}, z_i] \end{pmatrix} \Omega(z_i, y_{2_i})^{-1} \times R(z_i, y_{2i})]\\
&= G_3' W_3 E[Z_{3i}' \begin{pmatrix} E[r_1^2|y_{2i}, z_i] E[r_1^2|y_{2i}, z_i]^{-1} & 0 \\ 0 & E[r_2^2|y_{2i}, z_i] E[r_1^2|z_i]^{-1} \end{pmatrix}\\
&\times \begin{pmatrix} E[\nabla_\delta r_1|z_i, y_{2i}] & E[\nabla_\gamma r_1|z_i, y_{2i}] \\ \mathbf{0}_{1\times(2+k)} & E[\nabla_\gamma r_2|z_i] \end{pmatrix}]\\
&= G_3' W_3 E[Z_{3i}' \begin{pmatrix} 1 & 0 \\ 0 & E[r_2^2|y_{2i}, z_i] E[r_1^2|z_i]^{-1} \end{pmatrix} \begin{pmatrix} E[\nabla_\delta r_1|z_i, y_{2i}] & E[\nabla_\gamma r_1|z_i, y_{2i}] \\ \mathbf{0}_{1\times(2+k)} & E[\nabla_\gamma r_2|z_i] \end{pmatrix}]\\
&= G_3' W_3 E[Z_{3i}' \begin{pmatrix} E[\nabla_\delta r_1|z_i, y_{2i}] & E[\nabla_\gamma r_1|z_i, y_{2i}] \\ 0 & E[r_2^2|y_{2i}, z_i] E[r_1^2|z_i]^{-1} E[\nabla_\gamma r_2|z_i] \end{pmatrix}]\\
&= G_3' W_3 E[[Z_{1i}' \ Z_{2i}'] \begin{pmatrix} E[\nabla_\delta r_1|z_i, y_{2i}] & E[\nabla_\gamma r_1|z_i, y_{2i}] \\ 0 & E[r_2^2|y_{2i}, z_i] E[r_1^2|z_i]^{-1} E[\nabla_\gamma r_2|z_i] \end{pmatrix}]\\
&= G_3' W_3 E[Z_{1i}' E[\nabla_\delta r_1|z_i, y_{2i}], \ Z_{1i}' E[\nabla_\gamma r_1|z_i, y_{2i}] + Z_{2i}' E[r_2^2|y_{2i}, z_i] E[r_1^2|z_i]^{-1} E[\nabla_\gamma r_2|z_i]]\\
&= G_3' W_3 E[E[Z_{1i}' \nabla_\delta r_1|z_i, y_{2i}], \ E[Z_{1i}' \nabla_\gamma r_1|z_i, y_{2i}] + E[r_2^2|y_{2i}, z_i] E[r_1^2|z_i]^{-1} E[Z_{2i}' \nabla_\gamma r_2|z_i]]
\end{aligned}
$$

Note that $E[E[Z_{1i}' \nabla_\delta r_1|z_i, y_{2i}]] = E[Z_{1i}' \nabla_\delta r_1]$.Also note that

$$
\begin{aligned}
&E[E[r_2^2|y_{2i}, z_i] E[r_1^2|z_i]^{-1} E[Z_{2i}' \nabla_\gamma r_2|z_i]]\\
&= E[E[r_1^2|z_i]^{-1} E[Z_{2i}' \nabla_\gamma r_2|z_i] E[E[r_2^2|y_{2i}, z_i]|z_i]]\\
&= E[E[r_1^2|z_i]^{-1} E[Z_{2i}' \nabla_\gamma r_2|z_i] E[r_2^2||z_i]]\\
&= E[E[Z_{2i}' \nabla_\gamma r_2|z_i]]\\
&= E[Z_{2i}' \nabla_\gamma r_2]
\end{aligned}
$$

Thus, the above equation becomes

$$
\begin{aligned}
&= G_3^{'} W_3 [E[Z_{1i}^{'} \nabla_\delta r_1],\ E[Z_{1i}^{'} \nabla_\gamma r_1] + E[Z_{2i}' \nabla_\gamma r_2]] \\
&= G_3^{'} W_3 G_3
\end{aligned}
$$

Therefore, $E[m_A m_B'] = G_3^{'} W_3 G_3$

Thus,

$$
\begin{aligned}
&[G_3' W_3 G_3]^{-1} G_3^{'} W_3 E[q_{3i} q_{3i}'] W_3 G_3 [G_3' W_3 G_3]^{-1} - E[m_B m_B']^{-1}. \\
&= E[m_A m_B']^{-1} E[m_A m_A'] E[m_B m_A']^{-1} - E[m_B m_B']^{-1} \\
&= E[m_A m_B']^{-1} \{E[m_A m_A'] E[m_B m_A']^{-1} - E[m_A m_B'] E[m_B m_B']^{-1}\} \\
&= E[m_A m_B']^{-1} \{E[m_A m_A'] - E[m_A m_B'] E[m_B m_B']^{-1} E[m_B m_A']\} E[m_B m_A']^{-1} \\
&= E[RR']
\end{aligned}
$$

where $R = E[m_A m_B']^{-1} \{m_A - E[m_A m_B'] E[m_B m_B']^{-1} m_B\}$

$[G_3' W_3 G_3]^{-1} G_3^{'} W_3 E[q_{3i} q_{3i}'] W_3 G_3 [G_3' W_3 G_3]^{-1} - E[m_B m_B']^{-1}$ is positive semi definite. This implies that $E[m_B m_B']^{-1}$ is lower bound of the asymptotic covariance. Since $E[R(y_{2i}, z_i)^{'} \Omega(z_i, y_{2_i})^{-1} R(y_{2i}, z_i)]^{-1} = E[m_B m_B']^{-1}$, it archives the minim of the asymptotic covariance.