Economics Bulletin

Volume 37, Issue 4

Identification through Heteroscedasticity: What If We Have the Wrong Form?

Tak Wai Chau
The Hong Kong Polytechnic University

Abstract

Recent literature proposes estimators that utilize the heteroscedasticity in the error terms to identify the coefficient of the endogenous regressor in a standard linear model, while these estimators do not require extra exogenous variables as the excluded instruments. The assumed forms of heteroscedasticity differ across estimators, but it is often not straightforward how to justify the validity of such assumption a-priori. This simulation study investigates the robustness of the two most popular estimators under different forms of heteroscedasticity. The results show that both estimators can be substantially biased under wrong assumptions on the form of heteroscedasticity. Moreover, the overidentification test proposed for one estimator can have low power against the wrong form of heteroscedasticity. This study also explores the use of the maximum likelihood framework and the use of Akaike Information Criteria (AIC) to distinguish these two models. The simulation results show that it has good performance

I would like to thank the anonymous referee, Alan Fernihough, Ka-fu Wong, Feng Yao and participants of the Econometric Society Asian Meeting 2016 for their helpful comments. All remaining errors are mine.
Citation: Tak Wai Chau, (2017) "Identification through Heteroscedasticity: What If We Have the Wrong Form?", *Economics Bulletin*, Volume 37, Issue 4, pages 2413-2421
Contact: Tak Wai Chau - chau.takwai@gmail.com.
Submitted: July 03, 2016.   **Published:** October 26, 2017.

Economics Bulletin

Submission Number: EB-16-00499

Identification through Heteroscedasticity: What If We Have the Wrong Form?

Tak wai Chau
The Chinese University of Hong Kong

Abstract

Recent literature propose estimators that utilize the heteroscedasticity in the error terms to identify the coefficient of the endogenous regressor and do not require excluded instruments. The assumed forms of heteroscedasticity differ across estimators, while it is often not straightforward how to justify the validity of such assumption a-priori. This simulation study investigates the robustness of the two most popular estimators under different forms of heteroscedasticity. The results show that both estimators can be substantially biased under the wrong form of heteroscedasticity. Moreover, the overidentification test proposed for one estimator can have low power against the wrong form of heteroscedasticity. This study also explores the use of the maximum likelihood framework and the use of Akaike Information Criteria (AIC) to distinguish these two models. The simulation results show that it has good performance.

Submitted: July 03, 2016.

1 Introduction

One common difficulty for empirical researchers to consistently estimate the coefficient of a potentially endogenous regressor of interest in a standard linear model is that exogenous instruments are not always available. Klein and Vella (2010) and Lewbel (2012) respectively introduce methods to identify such coefficients by utilizing the heteroscedasticity of the error terms, where excluded instruments from variables outside the equation are not required. Klein and Vella (2010) assume that heteroscedasticity is multiplicative to the whole structural and first-stage error terms respectively with a constant correlation coefficient. Lewbel (2012) assumes that the covariance, instead of correlation, of the structural and first-stage error terms is a constant, which essentially requires that heteroscedasticity only exists in the uncorrelated components of these error terms.[1] Nevertheless, for empirical researchers, it is not often straightforward how to justify a-priori which form of heteroscedasticity is true. This study investigates whether these estimators are robust to misspecification of heteroscedasticity and whether the existing diagnostic tests are powerful enough to distinguish them. If not, justification of the right form of heteroscedasticity, alongside with the existence of heteroscedasticity, are needed for applying these estimators for consistency. I also propose putting these two estimators under the same maximum likelihood framework for estimation, and using the Akaike Information Criteria (AIC) to obtain one more piece of evidence to choose between models.

The two estimators, especially the Lewbel estimator, are becoming more popular because they are easy to implement[2] and heteroscedasticity is common in empirical data. Most of these studies use the estimators for robustness check against other proposed methods of identification. Not all of them have a-priori justification for the form of heteroscedasticity assumed[3], thus their estimators may be inconsistent if they assume the wrong form of heteroscedasticity without noticing.

This study simulates data from a standard linear model with one endogenous regressor from the forms of heteroscedasticity corresponding to the Klein and Vella (2010) and Lewbel (2012) specifications respectively, estimates the parameters using various methods, and investigates the sampling distribution of the estimators and the diagnostic statistics. The simulation results show that the two estimators are substantially biased when the assumed and actual forms of heteroscedasticity do not match. The power of the over-identification test for the Lewbel estimator can be low under misspecification, while AIC usually has a reasonably high probability in choosing the right model.

In Section 2, the two methods and the underlying assumptions are discussed. Section 3 describes the simulation setting and presents the simulation results. Section 4 discusses the case of including exogenous instrument. Section 5 concludes.

2 Model and Estimators

This study considers the linear regression model with one endogenous regressor y_2. The structural (outcome) equation is specified as:

$$y_1 = y_2\beta_1 + X\beta_2 + \varepsilon \tag{1}$$

where X contains exogenous regressors and a constant. The first-stage equation is give by

$$y_2 = Z\gamma_1 + X\gamma_2 + u \tag{2}$$

where Z contains excluded exogenous instruments, which is not necessary for the methods investigated in this paper.

2.1 The Lewbel Estimator

For the Lewbel (2012) estimator, the key identifying assumptions for coefficients, especially β_1, are that there exists some variables Z_2, which may be variables in X, such that

$$E(W\varepsilon) = 0$$

[1] The required condition is more general than this, but other possibilities are harder to justify in practice. More discussions are in the later sections.

[2] Lewbel's estimator can now be implemented by user written procedures in Stata (ivreg2h, see Baum and Schaffer, 2012) and R (ivlewbel, see Fernihough, 2014).

[3] Among these studies, only Emran and Shilpi (2012) and Millimet and Roy (2016) provide some justifications.

$$\begin{aligned} E(Wu) &= 0 \\ E((Z_2 - \mu_2)u\varepsilon) &= 0 \\ E(Z_2 - \mu_2) &= 0 \\ E((Z_2 - \mu_2)u^2) &\neq 0 \end{aligned} \tag{3}$$

where $W = [X, Z]$ are the available exogenous variables and Z_2 is a subset of W. The first two assumptions are exogeneity of W. The third condition requires zero expectation for the product of errors $u\varepsilon$ and demeaned Z_2, where the fourth condition requires $\mu_2 = E(Z_2)$. The fifth condition requires that the first stage error u is heteroscedastic in demeaned Z_2. The last three conditions imply that $(Z_2 - \mu_2)u$ can be used as an instrument for y_2.

A sufficient condition for the third condition is that the covariance between u and ε conditional on Z_2 does not depend on Z_2, since $E((Z_2 - \mu_2)u\varepsilon) = E((Z_2 - \mu_2)cov(u, \varepsilon|Z_2)) = E((Z_2 - \mu_2)\sigma_{u,\varepsilon}(z_2))$, which is zero when $\sigma_{u,\varepsilon}(z_2)$ is a constant. Equivalently, any heteroscedasticity related to Z_2 does not appear in the correlated component or common factor of the error terms, otherwise the covariance would depend on Z_2. This is the major distinction from the Klein-Vella (2010) method. Also note that it is not the only way to satisfy the above conditions. For example, if $\sigma_{u,\varepsilon}(z_2)$ is a polynomial of even-ordered terms of $Z_2 - \mu_2$, then the whole term inside expectation is a polynomial of odd-ordered terms, and if $Z_2 - \mu_2$ is symmetric about zero, then this expectation is also zero. However, it is hard to justify that an empirical application can satisfy this specific relationship. Thus, I focus on the interpretation that the heteroscedasticity has to come from the uncorrelated components of the error terms.

The model can be estimated by the Generalized Method of Moments (GMM) using the first three conditions in (3). The J statistic, which is the normalized value of the GMM objective function with optimal weight matrix, can be used as a test of overidentifying restrictions (Hansen, 1982). If it is rejected, some of the moment conditions are likely to be invalid. Lewbel (2012) also proposes using the Breusch and Pagan (1979) test for testing existence of heteroscedasticity required in the first-stage error term u, but this test may capture the wrong form of heteroscedasticity from the correlated component.

2.2 The Klein-Vella Estimator

Klein and Vella (2010) propose using multiplicative heteroscedasticity of the two error terms with constant correlation coefficient ρ to identify the model. In particular,

$$\begin{aligned} \varepsilon &= S_\varepsilon(Z_2)\varepsilon^* \\ u &= S_u(Z_2)u^* \end{aligned} \tag{4}$$

where $S_\varepsilon(Z_2)$ and $S_u(Z_2)$ describe the conditional standard deviations of the error terms as a function of Z_2. ε^* and u^* are homoscedastic (generally with unit variance) with constant correlation,

$$corr(\varepsilon^*, u^*) = corr(\varepsilon, u|Z_2) = \rho. \tag{5}$$

So, the correlation between ε and u conditional on Z_2 is also a constant. This is in contrast with Lewbel (2012) who assumes constant covariance.

A control function approach was proposed in Klein and Vella (2010), and the OLS estimator for the coefficients of the following equation is then consistent:

$$y_1 = y_2\beta_1 + X\beta_2 + \rho_0 \frac{\hat{S}_\varepsilon(Z_2)}{\hat{S}_u(Z_2)}\hat{u} + \tilde{\varepsilon} \tag{6}$$

Identification also requires that $S_\varepsilon(Z_2)/S_u(Z_2)$ depends on Z_2 and is not reduced to a constant or a linear function of Z_2.

In this paper, I first follow the parametric implementation of Farre, Klein and Vella (2013) by assuming the functional form of variance functions[4] as

$$\begin{aligned} S_{\varepsilon i} &= \sqrt{exp(Z'_{2i}\delta_\varepsilon)} \\ S_{ui} &= \sqrt{exp(Z'_{2i}\delta_u)} \end{aligned} \tag{7}$$

and their 2-step process, which is outlined in the online appendix.

[4]This also allows for a linear model for δ for easier estimation.

Klein and Vella (2009, 2010) do not explicitly propose specification tests for the existence of heteroscedasticity for identification or tests for validity of identifying restrictions. The Breusch and Pagan (1979) test can be used for detecting heteroscedasticity for the first-stage error, but it cannot test for other identification requirements.

2.3 The Maximum Likelihood Estimation

This study introduces the maximum likelihood framework, which allows us to estimate the two models in one single framework, and use the corresponding Akaike Information Criteria (AIC) to choose a better-fit model.

Assuming the structural error ε and the first-stage error u are distributed in bivariate normal, the log likelihood function is

$$L(\beta,\gamma,\delta,\theta)=\sum_{i=1}^{n}f(y_{1i},y_{2i}|W_i)=\sum_{i=1}^{n}\left[-ln(2\pi)-ln(s_{u,i}s_{\varepsilon,i})-\frac{1}{2}ln(1-\rho^2)-\frac{(\tilde{u}_i^2+\tilde{\varepsilon}_i^2-2\rho\tilde{u}_i\tilde{\varepsilon}_i)}{2(1-\rho^2)}\right] \tag{8}$$

where

$$\tilde{\varepsilon}_i = \frac{y_{1i}-y_{2i}\beta_1-X_i\beta_2}{s_{\varepsilon,i}} \tag{9}$$

$$\tilde{u}_i = \frac{y_{2i}-Z_i\gamma_1-X_i\gamma_2}{s_{u,i}} \tag{10}$$

$$s_{\varepsilon,i} = \sqrt{f_\varepsilon(Z'_{2i}\delta_\varepsilon)} \tag{11}$$

$$s_{u,i} = \sqrt{f_u(Z'_{2i}\delta_u)} \tag{12}$$

For a flexible specification of variance functions f_ε and f_u, a fourth order polynomial of a monotonic function of the single index is used, with certain restrictions to avoid spurious solutions. Details are available in the online appendix.

The key difference between the two models is the specification of correlation. For the Lewbel (2012) estimator, constant covariance θ_{LB} implies

$$\rho_{LB}=\frac{\theta_{LB}}{s_{\varepsilon,i}s_{u,i}} \tag{13}$$

where $s_{\varepsilon,i}$ and $s_{u,i}$ are specified in (11) and (12).[5]

For the Klein and Vella (2010) estimator, constant correlation θ_{KV} implies

$$\rho_{KV}=\theta_{KV} \tag{14}$$

We can use the Akaike Information Criteria (AIC) to choose the model that gives a larger value.[6]

$$AIC=2L(\beta,\gamma,\delta,\theta)-2K_p \tag{15}$$

where K_p is the total number of parameters in the model.

3 Simulation Schemes and Results

3.1 Simulation Scheme

The simulation in this study follows (1) and (2) as:

$$\begin{aligned} y_{1i} &= \beta_0+\beta_1 y_{2i}+X_i'\beta_2+\varepsilon_i \\ y_{2i} &= \gamma_0+X_i'\gamma_2+u_i \end{aligned} \tag{16}$$

[5] Again, it is not the only interpretation of Lewbel's condition, but it is the most plausible one and I choose to focus on this interpretation.

[6] Other criteria such as Bayesian information criterion (BIC) / Schwarz criterion (SC) or Hannan–Quinn information criterion (HQC) can also be used. And besides choosing between Lewbel and Klein-Vella specification of heteroscedasticity, one may also use these criteria to choose the suitable level of complexity in the flexible approximation of the variance functions. In this paper, I choose to focus on the choice on the form of heteroscedasticity, and I have fixed the complexity of the variance functions so that the number of parameters for the two models are the same. These criteria essentially choose the model with a higher value of log likelihood.

without excluded instrument Z_i. There are K exogenous regressors x_i, which are independently distributed in standard normal. We consider the following two cases for the heteroscedastic error terms with a common factor θ_i.

Case 1: Klein-Vella Type

$$\begin{aligned} \varepsilon_i &= \sqrt{exp(X_i'\delta_\varepsilon)}\,(\alpha_1\theta_i + v_{1i}) \\ u_i &= \sqrt{exp(X_i'\delta_u)}\,(\alpha_2\theta_i + v_{2i}) \end{aligned} \tag{17}$$

where the heteroscedasticity affects the whole error term.

Case 2: Lewbel Type

$$\begin{aligned} \varepsilon_i &= \alpha_1\theta_i + \sqrt{exp(X_i'\delta_\varepsilon)}v_{1i} \\ u_i &= \alpha_2\theta_i + \sqrt{exp(X_i'\delta_u)}v_{2i} \end{aligned} \tag{18}$$

where the heteroscedasticity affects only the idiosyncratic component. θ_i, v_{1i} and v_{2i} follow independent standard normal distribution in the simulation.[7] The correlation between the first-stage and structural error term is generated by the common factor θ_i.

Simulated data from the above models are used to estimate the structural parameters β using the methods described above.[8] Here I use all variables X as Z_2 variables.[9] The focus is on the coefficient of the endogenous regressor β_1. Median, 10^{th} and 90^{th} percentiles for the point estimators are presented to assess the biasedness and skewness of the estimators.[10] I present the results for the J statistics to investigate the effectiveness of overidentifying tests to detect the wrong specification of heteroscedasticity.

In this study, I mainly take $\beta_1 = 0$[11], so the values of mean and median of bootstrap samples represent the corresponding biases. $\beta_0 = \alpha_0 = 0$, $\beta_{2k} = \gamma_{2k} = 1$ for all k. α_1 and α_2 are set to 1 and the associated correlation between ε and u is about 0.5. The number of observations for each sample considered is 500 for most cases and some are 1000 for comparison. The number of replications is at least 2000 for each design. To assess robustness, I allow different heteroscedastic parameters δ_{u1} and $\delta_{\varepsilon 1}$ for the first variable in X, and δ_{u2} and $\delta_{\varepsilon 2}$ for all remaining X variables. Here, $\delta_{\varepsilon 2}$ is always set to zero.

3.2 Simulation Results

Table 1 and 2 show the simulation results for the two forms of heteroscedasticity respectively. Results generally show that under the wrong form of heteroscedasticity, the estimators are generally biased. Table 1 shows the results for data generated from the Klein-Vella form of heteroscedasticity. The Lewbel estimators, both the original GMM and the ML, are biased upward in the cases considered[12], while the 2-step and ML Klein-Vella estimators have medians close to their true value. If we choose the estimator according to AIC[13], the correct rates are usually higher than 0.5, though not always close to 1 in the cases considered. The resulting estimator has a lower bias than the wrong ones. The over-identification J test has low power in detecting the misspecification of the form of heteroscedasticity, with rejection rates generally below 40%. The power is particularly weak when only one of the variables gives rise to the heteroscedasticity in the first-stage error term.

Table 2 shows the results for data generated from the Lewbel form heteroscedasticity. In this case, the two Klein-Vella estimators are biased downward in these cases, while the two Lewbel estimators have median close to the true value. If we choose the estimator according to AIC, the correct rates are generally higher than 0.5, and the resulting estimator is closer

[7] Results for the case with normalized Chi-square (5) errors and θ are shown in the online appendix.

[8] The estimators are coded in R by the author. See Online Appendix for more details.

[9] The default of 'ivreg2h' in Stata uses all exogenous regressors for Z_2.

[10] Mean and standard deviation are not used here because there is a concern that some estimators may not have moments like LIML.

[11] This follows from Davidson and MacKinnon (2010) where the value of β_1 is set to zero without loss of generality, and the corresponding sampling distribution should just translate horizontally. I have also included two cases with $\beta_1 = 1$, and the results also show only a horizontal translation. The discrepancy is likely due to sampling errors.

[12] More discussions about the sign of the bias are in the appendix. The direction relative to the OLS bias is not always the same.

[13] As the number of parameters in the two models are the same, this essentially compares the log likelihood values.

to be median unbiased. The results for J test agrees with the nominal power as the Lewbel is the true data generating process[14].

Table 1: Simulation Results for Data from the Klein and Vella Form of Heteroscedasticity, Normal Errors

n	K	δ_{u1}	δ_{u2}	$\delta_{\varepsilon 1}$	β_{OLS} median (q10,q90)	$\beta_{LB,GMM}$ median (q10,q90)	J median (% $p < .05$)	$\beta_{KV,\text{2-step}}$ median (q10,q90)	$\beta_{LB,ML}$ median (q10,q90)	$\beta_{KV,ML}$ median (q10,q90)	β_{AIC} median (q10,q90)	AIC correct rate
$\beta_1 = 0$												
500	3	0.4	0.4	0.3	0.4362	0.2639	2.965	0.0206	0.2873	0.0076	0.0761	0.698
					(0.385,0.485)	(0.166,0.360)	(0.198)	(-0.309,0.210)	(0.192,0.379)	(-0.249,0.201)	(-0.216,0.315)	
500	3	0.4	0.4	-0.3	0.4099	0.1500	2.857	0.0073	0.1566	-0.0107	0.0231	0.742
					(0.363,0.459)	(0.071,0.226)	(0.193)	(-0.182,0.132)	(0.066,0.236)	(-0.165,0.106)	(-0.147,0.174)	
500	3	0.4	0.4	0.5	0.4532	0.2906	5.392	0.0268	0.3198	0.0050	0.0173	0.878
					(0.400,0.507)	(0.180,0.401)	(0.442)	(-0.341,0.218)	(0.226,0.409)	(-0.250,0.193)	(-0.242,0.266)	
500	3	0.25	0.25	0.3	0.4824	0.3253	3.092	0.0597	0.3323	0.0438	0.1171	0.675
					(0.431,0.531)	(0.177,0.471)	(0.213)	(-0.428,0.362)	(0.165,0.489)	(-0.512,0.387)	(-0.457,0.398)	
500	3	0.7	0.7	0.3	0.3112	0.1782	2.668	0.0108	0.2205	-0.0118	-0.0001	0.853
					(0.260,0.363)	(0.118,0.239)	(0.167)	(-0.118,0.104)	(0.165,0.277)	(-0.115,0.079)	(-0.110,0.200)	
500	3	0.7	0	0.3	0.4443	0.3209	1.482	0.0385	0.3423	0.0198	0.2465	0.486
					(0.394,0.494)	(0.216,0.422)	(0.051)	(-0.415,0.303)	(0.244,0.432)	(-0.399,0.317)	(-0.291,0.403)	
500	3	0.7	0	0.5	0.4682	0.4034	1.506	0.1584	0.4284	0.0957	0.3361	0.617
					(0.414,0.525)	(0.280,0.526)	(0.052)	(-0.510,0.857)	(0.335,0.517)	(-1.110,0.998)	(-0.855,0.837)	
500	10	0.25	0.25	0.3	0.4100	0.2290	10.98	0.0508	0.2504	0.0141	0.1209	0.587
					(0.362,0.462)	(0.146,0.312)	(0.141)	(-0.126,0.179)	(0.176,0.324)	(-0.172,0.164)	(-0.136,0.286)	
500	10	0.7	0	0.3	0.4446	0.3275	9.171	0.1105	0.3135	0.1127	0.2681	0.393
					(0.394,0.495)	(0.221,0.431)	(0.061)	(-0.176,0.353)	(0.200,0.424)	(-0.298,0.389)	(-0.124,0.407)	
500	10	0.7	0	0.5	0.4693	0.4087	9.126	0.2049	0.4249	0.2660	0.3553	0.535
					(0.415,0.524)	(0.287,0.530)	(0.054)	(-0.173,0.808)	(0.325,0.528)	(-0.945,1.019)	(-0.520,0.872)	
1000	3	0.4	0.4	0.3	0.4339	0.2610	4.562	0.0156	0.2924	0.0044	0.0207	0.854
					(0.400,0.468)	(0.193,0.329)	(0.367)	(-0.226,0.155)	(0.228,0.351)	(-0.151,0.134)	(-0.146,0.264)	
1000	3	0.7	0	0.3	0.4419	0.3184	1.454	0.0202	0.3463	0.0002	0.1273	0.592
					(0.406,0.479)	(0.245,0.391)	(0.053)	(-0.319,0.223)	(0.285,0.409)	(-0.246,0.203)	(-0.197,0.379)	
1000	10	0.25	0.25	0.3	0.4083	0.2226	12.83	0.0290	0.2487	0.0045	0.0252	0.800
					(0.374,0.443)	(0.164,0.280)	(0.244)	(-0.102,0.127)	(0.192,0.302)	(-0.119,0.104)	(-0.112,0.240)	
$\beta_1 = 1$												
500	3	0.4	0.4	0.3	1.4339	1.2633	2.978	1.0167	1.2842	1.0104	1.0830	0.676
					(1.383,1.487)	(1.168,1.358)	(0.205)	(0.703,1.209)	(1.186,1.376)	(0.762,1.203)	(0.791,1.309)	
500	3	0.7	0	0.3	1.4443	1.3203	1.498	1.0431	1.3436	1.0216	1.2390	0.516
					(1.391,1.498)	(1.216,1.422)	(0.053)	(0.579,1.308)	(1.239,1.439)	(0.591,1.313)	(0.708,1.408)	

The number of repetition is at least 2000. The correlation between the first stage and structural error is set at about 0.5. δ_{u1} is the coefficient for the variance function of the first stage error for the first variable of X, while δ_{u2} is the coefficient for all remaining X variables. Similar for $\delta_{\varepsilon 1}$ and $\delta_{\varepsilon 2}$ and I set $\delta_{\varepsilon 2} = 0$. The J statistic is the corresponding statistic under the Lewbel GMM method. β_{AIC} reports the estimate when the one with higher AIC is chosen between the two ML estimators.

[14]Some computational issues for the J statistic are discussed in the Online Appendix. Also note that, the finite sample bias can take a rather large sample size to remove, as also shown in Stock and Wright (2000).

Table 2: Simulation Results for Data from the Lewbel Form of Heteroscedasticity, Normal Errors

n	K	δ_{u1}	δ_{u2}	$\delta_{\varepsilon 1}$	β_{OLS} median (q10,q90)	$\beta_{LB,GMM}$ median (q10,q90)	J median (% $p<.05$)	$\beta_{KV,\text{2-step}}$ median (q10,q90)	$\beta_{LB,ML}$ median (q10,q90)	$\beta_{KV,ML}$ median (q10,q90)	β_{AIC} median (q10,q90)	AIC correct rate
$\beta_1=0$												
500	3	0.5	0.5	0.3	0.4086	0.0083	1.483	-0.5354	-0.0114	-0.7692	-0.0398	0.846
					(0.354,0.462)	(-0.145,0.139)	(0.061)	(-1.210,-0.110)	(-0.177,0.130)	(-1.242,-0.143)	(-0.694,0.120)	
500	3	0.5	0.5	-0.3	0.4067	0.0118	1.451	-0.2887	0.0152	-0.3349	-0.0878	0.579
					(0.355,0.460)	(-0.131,0.129)	(0.059)	(-0.799,-0.021)	(-0.135,0.141)	(-0.703,-0.086)	(-0.554,0.101)	
500	3	0.5	0.5	0.5	0.4092	0.0087	1.480	-0.5598	0.0065	-0.6655	-0.0026	0.930
					(0.355,0.461)	(-0.152,0.149)	(0.059)	(-1.263,0.068)	(-0.164,0.154)	(-1.243,2.118)	(-0.215,0.152)	
500	3	0.3	0.3	0.3	0.4689	0.0400	1.494	-0.1969	-0.0304	-0.1251	-0.0525	0.720
					(0.414,0.520)	(-0.233,0.258)	(0.067)	(-0.956,1.339)	(-0.320,0.256)	(-1.228,2.192)	(-2.259,0.260)	
500	3	0.8	0.8	0.3	0.2830	-0.0004	1.530	-0.4740	-0.0036	-0.5085	-0.0127	0.894
					(0.222,0.342)	(-0.081,0.074)	(0.058)	(-1.000,-0.211)	(-0.085,0.076)	(-0.965,-0.256)	(-0.282,0.073)	
500	3	0.8	0	0.3	0.4231	0.0104	1.485	-0.5741	-0.0294	-1.0829	-0.0549	0.825
					(0.366,0.477)	(-0.166,0.160)	(0.060)	(-1.352,1.205)	(-0.219,0.141)	(-1.274,2.186)	(-0.968,0.177)	
500	3	0.8	0	0.5	0.4222	0.0113	1.483	-0.5613	-0.0049	2.0812	0.0284	0.752
					(0.362,0.479)	(-0.181,0.184)	(0.058)	(-1.445,1.533)	(-0.199,0.163)	(-1.221,2.267)	(-0.213,2.139)	
500	10	0.3	0.3	0.3	0.3905	0.0337	9.467	-0.2512	0.0220	-0.3710	-0.0054	0.795
					(0.337,0.447)	(-0.097,0.151)	(0.084)	(-0.518,-0.030)	(-0.093,0.137)	(-0.851,-0.047)	(-0.483,0.124)	
500	10	0.8	0	0.3	0.4229	0.0564	9.495	-0.2474	0.0146	-0.3215	0.0051	0.821
					(0.368,0.478)	(-0.114,0.208)	(0.088)	(-0.583,0.910)	(-0.140,0.188)	(-1.146,1.602)	(-0.268,0.227)	
500	10	0.8	0	0.5	0.4210	0.0584	9.551	-0.1920	0.0176	1.1201	0.0444	0.786
					(0.363,0.479)	(-0.126,0.231)	(0.089)	(-0.624,1.171)	(-0.156,0.207)	(-1.139,2.121)	(-0.169,1.546)	
1000	3	0.5	0.5	0.3	0.4087	0.0052	1.439	-0.6784	-0.0083	-0.9981	-0.0165	0.948
					(0.369,0.449)	(-0.102,0.099)	(0.055)	(-1.358,-0.285)	(-0.128,0.093)	(-1.251,-0.443)	(-0.167,0.087)	
1000	3	0.8	0	0.3	0.4216	0.0040	1.422	-0.8069	-0.0151	-1.1894	-0.0280	0.922
					(0.383,0.460)	(-0.117,0.113)	(0.051)	(-1.583,-0.320)	(-0.167,0.102)	(-1.280,2.185)	(-0.248,0.096)	
1000	10	0.3	0.3	0.3	0.3900	0.0153	9.006	-0.3531	-0.0003	-0.4846	-0.0081	0.926
					(0.352,0.430)	(-0.075,0.099)	(0.068)	(-0.609,-0.154)	(-0.092,0.079)	(-0.948,-0.237)	(-0.145,0.074)	
$\beta_1=1$												
500	3	0.5	0.5	0.3	1.4071	1.0093	1.473	0.4742	0.9918	0.2510	0.9629	0.851
					(1.354,1.462)	(0.856,1.139)	(0.059)	(-0.173,0.897)	(0.835,1.134)	(-0.234,0.856)	(0.362,1.124)	
500	3	0.8	0	0.3	1.4226	1.0115	1.518	0.4131	0.9784	-0.09827	0.9425	0.818
					(1.370,1.477)	(0.835,1.163)	(0.057)	(-0.382,2.214)	(0.776,1.145)	(-0.277,3.191)	(-0.047,1.152)	

Refer to the notes for Table 1.

4 Issues of Including Exogenous Excluded Instruments

The two estimators considered in this paper can be adjusted to include exogenous excluded instruments Z. Lewbel (2012) has shown this in his GMM formulation. For the two-step Klein and Vella (2010) estimator and the ML formulation of this paper, it is also straightforward to include Z in the first-stage equation and the variance functions.

One purpose of including both types of instruments is to increase the precision of the estimator by using both sources of identification, especially when the excluded instrument is weak. Another purpose is to test the validity of the excluded instrument at hand, especially under the Lewbel estimator through the over-identification J test. If we want to test the validity of excluded instruments, we need the instruments from heteroscedasticity to be valid. However, when we are usually not clear about the correct form of heteroscedasticity, the results of this study show that the power of rejecting the null under a wrong form of heteroscedasticity (Klein-Vella form) can be low even when the estimator is substantially biased. When we reject the null hypothesis of valid over-identifying restrictions, it is not clear whether it is the problem of the form of heteroscedasticity, or the endogeneity of the excluded instrument. Similarly, if we cannot reject the null of valid overidentifying restrictions, it can be that all instruments are valid, but it is also possible that the biases happen to be similar from the two sources. Therefore, the J test alone cannot really provide us a clear conclusion.

Combining the maximum likelihood and model selection with AIC, the model chosen by

AIC should have more support from data, and we are then more confident about the true form of heteroscedasticity. Then, if the Lewbel model is chosen, the corresponding over-identification test can be more reliable. However, it should also be noted that there is still a substantial probability that we would conclude a wrong form of heteroscedasticity from AIC, and so the conclusion is still not totally reliable. Moreover, there can also be other possible forms of heteroscedasticity.

5 Conclusion

The simulation results in this study show that the Lewbel (2012) and the Klein and Vella (2010) estimators are not robust to misspecification of the form of heteroscedasticity. Moreover, the over-identification test proposed by Lewbel (2012), the Hansen's (1982) J test, has low power to reject the null under the Klein-Vella form of heteroscedasticity. The use of AIC under maximum likelihood is more capable of distinguishing these two models. One potential path for further research is to study the identification conditions for the cases in between these two models. In particular, we should study what restrictions on variance and covariance functions are necessary for identification and what procedure would enable us to search for the right form of heteroscedasticity and the right variables to be included in each variance and covariance function.

Empirical researchers should be cautious when using these two estimators. It is not sufficient to justify only the existence of heteroscedasticity in the error term to apply these estimators for consistency. We should also justify which form of heteroscedasticity appears in the error terms. The use of the proposed maximum likelihood method and choosing the model with largest AIC is a possible way to give more confidence in the chosen form of heteroscedasticity against the other one.

References

[1] Baum, C.F and Schaffer, M. E., (2012) "IVREG2H: Stata module to perform instrumental variables estimation using heteroskedasticity-based instruments," Statistical Software Components S457555, Boston College Department of Economics, revised 02 Apr 2015.

[2] Breusch, T. and Pagan A. (1979) "A Simple Test for Heteroscedasticity and Random Coefficient Variation," *Econometrica*, 47, 1287-1294.

[3] Davidson, R. and MacKinnon, J.G. (2010) "Wild Bootstrap Tests for IV Regression", *Journal of Business & Economic Statistics*, 28(1), 128-144

[4] Emran, M. and Shilpi, S. (2012) "The Extent of the Market and Stages of Agricultural Specialization," *Canadian Journal of Economics*, 12, 1125-1153.

[5] Farre, L, Klein, R. and Vella, F. (2013) "A parametric control function approach to estimating the returns to schooling in the absence of exclusion restrictions: an application to the NLSY," *Empirical Economics*, 44, 111-133.

[6] Fernihough, A. (2014). "ivlewbel: Uses heteroscedasticity to estimate mismeasured and endogenous regressor models." R package version 1.1. http://CRAN.R-project.org/package=ivlewbel

[7] Hansen, L (1982) "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50,1029-1054.

[8] Klein, R. and Vella, F. (2009) "Estimating the Return to Endogenous Schooling Decisions via Conditional Second Moments," *Journal of Human Resources*, 44(4), 1047-1065.

[9] Klein, R. and Vella, F. (2010) "Estimating a class of triangular simultaneous equations models without exclusion restrictions," *Journal of Econometrics*, 154, 154-164.

[10] Lewbel, A. (2012) "Using Heteroscedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models," *Journal of Business and Economic Statistics*, 30(1), 67-80.

[11] Millimet, D. and Roy, J. (2016) "Empirical Tests of the Pollution Haven Hypothesis When Environmental Regulation is Endogenous," *Journal of Applied Econometrics*, 31, 652-677.

[12] Stock, J.H. and Wright J. H. (2000) "GMM with weak identification," *Econometrica*, 68(5), 1055-1096,