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

Chau, Tak Wai

Shanghai University of Finance and Economics

July 2015

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

Tak Wai Chau*

July 31, 2015.

Abstract

Klein and Vella (2010) and Lewbel (2012) respectively propose estimators that utilize the heteroscedasticity of the error terms to identify the coefficient of the endogenous regressor in a standard linear model, even when there are no exogenous excluded instruments. The assumptions on the form of heteroscedasticity are different for these two estimators, and whether they are robust to misspecification is an important issue because it is not straightforward how to justify which form of heteroscedasticity is true. This paper presents some simulation results for the finite-sample performance of the two estimators under various forms of heteroscedasticity. The results reveal that both estimators can be substantially biased when the form of heteroscedasticity is of the wrong type, meaning that they lack robustness to misspecification of the form of heteroscedasticity. Moreover, the J statistics of the over-identification test for the Lewbel (2012) estimator has low power under the wrong form of heteroscedasticity in the cases considered. The results suggest that it is not enough for researchers to justify only the existence of heteroscedasticity when using the proposed estimators.

JEL codes: C13, C31, C36

*School of Economics, Shanghai University of Finance and Economics, and Key Laboratory of Mathematical Economics (SUFE), Ministry of Education. 777 Guoding Road, Yangpu, Shanghai, 200433, China. Email: chau.takwai@gmail.com.

1 Introduction

A common problem for empirical researchers attempting to obtain a reliable estimate of the causal effect from a regressor of interest is that the regressor may be endogenous. One major solution is to make use of exogenous instruments, such as quasi-experimental variations, that are excluded from the structural equation. However, this is often not an easy task. Klein and Vella (2010) and Lewbel (2012) respectively propose methods to identify the coefficient of the endogenous regressor by using the heteroscedasticity of the error terms, even when there are no excluded instruments.[1] They impose different assumptions for identification. Klein and Vella (2010) assume that the heteroscedasticity is multiplicative to the whole structural and first-stage error terms. In contrast, Lewbel (2012) assumes that the heteroscedasticity only applies to the component of the first-stage error term that is uncorrelated to the structural error term.[2]

The different assumptions may be violated in subtle ways. On one hand, the assumptions of heteroscedasticity for the Lewbel (2012) estimator would be violated if the unobserved common factor affecting both the outcome variable and the endogenous regressor is heteroscedastic, or if the effect of the unobserved variable on either the endogenous regressor or outcome variable varies with regressors that are used to generate the Lewbel-type instrument. For example, if the endogenous regressor is an income measure in a year, but is taken as a proxy for a permanent measure of income, then the measurement error is reflected in the error terms, and gives rise to the common factor of the two error terms. Such a measurement error can have different variances at different ages due to different variances of income over the lifecycle. Another example is about the return to education. One unobserved common factor is ability, but ability can have different loadings on earnings and education over different ages because their ability may affect their human capital accumulation process and the job search process differently

[1]Klein and Vella (2009a) propose a method similar to Klein and Vella (2010), but that is for the case of binary endogenous regressor. One major difference is that for the binary endogenous regressor case, heteroscedasticity in the latent-variable first-stage equation leads to a change in the conditional mean (probability) of the observed variable, which is not the case for a continuous endogenous variable.

[2]In Lewbel (2012), this point is only clearly presented in the example of single-factor model, and so, it can be easily overlooked. More discussions are presented in this paper.

over the lifecycle, and age may also carry cohort information that affects the education distribution. It may not be straightforward to determine whether these kinds of problems exist in a particular context. On the other hand, it may also not be straightforward to justify the form of heteroscedasticity that applies to all of the components of the error term that the Klein and Vella (2010) estimator assumes.

If researchers want to test the assumptions on the form of heteroscedasticity directly with data, there is also difficulty since the error terms are unobservable. Lewbel (2012) suggests using the Breusch and Pagan (1979) test and the overidentifying restriction J test (Hansen, 1982) to check the validity of assumptions. However, the former tests for heteroscedasticity of the whole first-stage error term rather than only its uncorrelated component, which will then detect also the wrong type of heteroscedasticity. Moreover, if all the generated instruments lead to similar biases from the wrong form of heteroscedasticity, the overidentification test may have little power detecting such violations even when the estimator is biased. Some researchers are also tempted to use the extra instruments from heteroscedasticity to test the validity of an existing excluded instrument. Nevertheless, if the true form of heteroscedasticity is not clear, a wrong form of heteroscedasticity may also provide us with a wrong conclusion. We may not be able to reject the null hypothesis in the cases that produce biased estimators. When it can reject the null, we remain uncertain whether it is the problem of excluded instrument or the form of heteroscedasticity.

The two estimators, especially the Lewbel estimator, are becoming more popular because they are easy to implement[3] and heteroscedasticity is common in data. Some examples of recent applications include Emran and Shilpi (2012), Denny and Oppedisano (2013), Emran and Hou (2013), Chowdhury et al. (2014), and Millimet and Roy (2015). Most of these studies use these estimators for comparison alongside with the usual instruments from exclusion restrictions. Among these studies, only Emran and Shilpi (2012) and Millimet and Roy (2015) have some a priori justification for the right type of heteroscedasticity. Therefore, it is useful to investigate how good the estimators

[3]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).

are robust to the wrong forms of heteroscedasticity and whether diagnostic tests such as Hensen's (1982) J test are sufficient for detecting problems, particularly in finite samples. It would be good news if the bias is small and results are reasonably robust. If the bias is substantial, merely checking for heteroscedasticity is not enough to guarantee a reliable estimate from these two estimators.

This study investigates the finite sample properties of the above two estimators under various forms of heteroscedasticity through Monte Carlo simulation to study the robustness of misspecification in the form of heteroscedasticity. I simulated data from a standard linear model with one endogenous regressor with three different types of multiplicative heteroscedasticity: 1) applicable to the whole error terms (consistent with Klein-Vella); 2) applicable only to the uncorrelated parts of the error terms (consistent with Lewbel); 3) applicable only to the correlated part of the error terms (consistent with neither estimator), estimated with the above methods and investigated the size and direction of bias if the form of heteroscedasticity is incorrect. I also considered the case when one weak exogenous instrument is also included, alongside with the heteroscedasticity to identify the parameter. Lewbel (2012) has explicitly considered including also excluded instruments, while Klein and Vella (2010), and their accompanying applied papers, Klein and Vella (2009b) and Farre, Klein and Vella (2013), do not incorporate any information from excluded exogenous instruments. Nevertheless, it is straightforward to extend their framework to allow for excluded instruments in the first-stage equation.

The simulation results show that the two estimators are substantially biased when the form of heteroscedasticity in the true data generating process is not consistent to the form assumed for the estimator. This means that the estimators are not robust to misspecification of the form of heteroscedasticity. The Lewbel (2012) estimator tends to bias towards the OLS, but not always, and the power of the J test can be low even when the estimator is substantially biased. The Klein and Vella (2010) estimator can be biased severely under the wrong form of heteroscedasticity, especially when the maximum likelihood method is used. Its bias can be on either side of the true value.

In Section 2, the two methods and the underlying assumptions are discussed. In particular, more details about the potential sources of bias due to wrong form of heteroscedasticity for the Lewbel (2012) estimator are presented. Section 3 describes the simulation setting and presents the simulation results. Section 4 presents the replication of the example in Lewbel (2012) of Engel curve estimation and compares the results with the Klein and Vella (2010) method. Section 5 concludes.

2 Model and Estimators

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

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

where y_2 is an endogenous regressor and 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. In this context, Z can be empty, and we have to depend solely on the heteroscedasticity of u and ε to identify consistently the parameter β_1.

The following discusses the identification assumptions and the estimation procedures of the two estimators.

2.1 The Lewbel Estimator

For the Lewbel (2012) estimator, the key identifying assumptions of the above triangular system are that there exists some variables Z_2, which may be variables in X, such that

$$\begin{aligned} E(W\varepsilon) &= 0 \\ E(Wu) &= 0 \end{aligned}$$

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

where $W = [X, Z]$ are the available exogenous variables. The first two conditions simply require that the exogenous regressors X are uncorrelated to the structural and first stage error terms, which are true by definition of exogeneity. The third condition requires that the covariance between the product of errors $u\varepsilon$ and Z_2 is zero, while the fourth condition requires that the first stage error u is heteroscedastic in terms of Z_2. Put it in another way, the last two conditions imply that $(Z_2 - \mu_2)u$ can be used as an instrument.

The estimation can be carried out by two-stage-least-squares (2SLS) using $(Z_2 - \bar{Z}_2)\hat{u}$, and Z if available, as instruments. The model can also be estimated by Generalized Method of Moments (GMM) using the first three conditions in (3), which can improve efficiency by using the optimal weight matrix.[4] The J statistic, which is the value of the GMM objective function with optimal weight matrix, allows us to perform the overidentifying restriction test (Hensen, 1982) to see whether a single set of parameters can make all moments close to zero. If it fails to do so, the moment conditions may be invalid.

To clarify what is required for the last two moment conditions in (3) to hold, we can decompose the two mean-zero error terms in the following way:

$$
\begin{aligned}
\varepsilon &= e_1 + v_1 \\
u &= e_2 + v_2
\end{aligned}
\tag{4}
$$

where all correlation between u and ε is captured by the first component so that conditional on X and Z, $cov(e_1, e_2) \neq 0$ whenever $cov(\varepsilon, u) \neq 0$, while conditional on X and Z, $cov(v_1, v_2) = 0$ and $cov(e_i, v_j) = 0$ for all $i, j = 1, 2$.[5] Denote $z_2 = Z_2 - \mu_2$ to simplify

[4]This can be done by a two-step approach, where the first step is to obtain a consistent estimator either using 2SLS or GMM with identity weight matrix, then we can use the consistent estimator to construct the optimal weight matrix and apply GMM.

[5]The common factor example in Lewbel (2012) is a special case where $e_1 = \alpha_1\theta$ and $e_2 = \alpha_2\theta$ for some α_1, α_2.

notation. The third condition requires

$$\begin{aligned}
E(z_2 u\varepsilon) &= E(z_2(e_2+v_2)(e_1+v_1)) \\
&= E(z_2 e_1 e_2) + E(z_2 v_2 e_1) + E(z_2 e_2 v_1) + E(z_2 v_2 v_1) \qquad (5) \\
&= E(z_2 E(e_1 e_2|z_2)) + E(z_2 E(v_2 e_1|z_2)) + E(z_2 E(e_2 v_1|z_2)) + E(z_2 E(v_2 v_1|z_2)) = 0
\end{aligned}$$

The last three terms are zero because the two error components are not correlated conditional on z_2 within and across equations, except for e_1, e_2. The first term can be written as

$$\begin{aligned}
E(z_2 e_1 e_2) &= E(z_2 E(e_1 e_2|z_2)) \\
&= E(z_2 cov(e_1, e_2|z_2)) \qquad (6) \\
&= E(z_2 \rho_{12}(z_2)\sigma_1(z_2)\sigma_2(z_2))
\end{aligned}$$

Thus, since $E(z_2) = 0$ and $\rho_{12}(z_2), \sigma_1(z_2), \sigma_2(z_2) \neq 0$, one sufficient condition for zero expectation is to have the variance and correlation between e_1 and e_2 independent of z_2, so that $E(z_2 e_1 e_2) = \rho_{12}\sigma_1\sigma_2 E(z_2) = 0$. If there is heteroscedasticity in terms of z_2, σ_1 and σ_2 are then functions of z_2. The term then involves higher moments of z_2, which are likely non-zero, especially for even moments. This then leads to a non-zero expectation unless the terms can exactly be canceled out in very particular ways. This point is more explicit in Lewbel (2012) when he discusses the single factor model, where he states that z_2 has to be uncorrelated to the square of the common factor, but correlated to square of v_2 defined above. However, in other models of his paper, he discusses in terms of the whole error term u (ε_2 in his notation) which makes this distinction implicit.

Another way that the identifying condition (5) is violated is when the effect of the common factor is related to z_2. Let $e_1 = a_1(z_2)\theta$ and $e_2 = a_2(z_2)\theta$, then this term becomes

$$E(z_2(a_1(z_2)\theta)(a_2(z_2)\theta)) = E(z_2 a_1(z_2) a_2(z_2) E(\theta^2|z_2)) \qquad (7)$$

and then even when θ is homoscedastic, with $E(\theta^2|z_2) = \sigma_\theta^2$ which does not depends on z_2, the term $E(z_2 a_1(z_2) a_2(z_2))$ involves higher moments of z_2 which are non-zero. This leads to the violation of identification assumption.

For the fourth condition, we require

$$\begin{aligned} E(z_2 u^2) &= E(z_2(e_2 + v_2)^2) \\ &= E(z_2 e_2^2) + E(z_2 v_2^2) + 2E(z_2 e_2 v_2) \neq 0 \end{aligned} \tag{8}$$

The third term is zero because e_2 and v_2 are uncorrelated conditional on z_2. The condition (5) requires e_2 to be homoscedastic with respect to z_2, then the first term is also zero. Therefore, the only way for the above expectation to be non-zero is to have the second term non-zero, which requires v_2 to be heteroscedastic in terms of z_2.

For illustration, I have derived the probability limit of the estimator for the simple case where there is a binary Z_2 variable and no X variable. As shown in the Appendix, the probability limit can be expressed as

$$\beta_{LB} = \frac{E(uy_1|Z_2 = 1) - E(uy_1|Z_2 = 0)}{Var(u|Z_1 = 1) - Var(u|Z_1 = 0)} = \beta + \frac{E(e_1 e_2|Z_2 = 1) - E(e_1 e_2|Z_2 = 0)}{Var(v_2|Z_2 = 1) - Var(v_2|Z_2 = 0)} \tag{9}$$

This expression shows that for consistency of the estimator, the variances of the first-stage error for the two groups defined by Z_2 have to be different, with the difference driven by the idiosyncratic component v_2, while at the same time, the covariances between the the correlated components e_1 and e_2 have to be the same for two groups.

We may also assess the direction of bias with (9) if there is a violation of the identification condition. The numerator of the bias is given by

$$E(e_1 e_2|Z_2 = 1) - E(e_1 e_2|Z_2 = 0) = \rho_1 \sigma_{e_1,1} \sigma_{e_2,1} - \rho_2 \sigma_{e_1,0} \sigma_{e_2,0} \tag{10}$$

where the second subscript represents the group defined by value of z_2. The denominator of the bias is given by

$$\begin{aligned} Var(u|Z_1 = 1) - Var(u|Z_1 = 0) &= Var(e_2 + v_2|Z_2 = 1) - Var(e_2 + v_2|Z_2 = 0) \\ &= (\sigma^2_{e_2,1} - \sigma^2_{e_2,0}) + (\sigma^2_{v_2,1} - \sigma^2_{v_2,0}) \end{aligned} \tag{11}$$

As a whole, the sign of the bias depends on how the variance of correlated and idiosyncratic components are correlated to z. Under the assumptions of the Klein and Vella (2010) estimator, ρ is a constant, then the numerator becomes $\rho(\sigma_{e_1,1}\sigma_{e_2,1} - \sigma_{e_1,0}\sigma_{e_2,0})$ and if the standard deviation of e_1 and e_2 are both correlated to Z_2 in the same direction, then its sign is given by the product of ρ and the correlation between σ_{e_2} and Z_2. On the other hand, since e and v are under the same form of heteroscedasticity, the sign of the denominator is given by the sign of correlation between σ_{e_2} and Z_2. As a result, in this case, the bias is of the same sign as ρ. Since the sign of ρ is also the sign of bias for the OLS estimator, the bias is then in the same direction as the OLS. However, if heteroscedasticity in e_1 are correlated to Z_2 in a different direction than that for e_2, the sign of bias will then depend on the resulting sign of the difference in (10).

In summary, one subtle requirement for consistency of the Lewbel estimator is that the heteroscedasticity of u should come from the idiosyncratic component v_2, which is uncorrelated to the structural error ε, while the component correlated to the structural error has to have no heteroscedasticity with z_2. The identification condition is also violated when the common factor in the two error terms has loadings that vary with z_2. For example, unobserved ability can be the common factor that it can affect outcomes and endogenous regressors differently at different ages. Then, age cannot be used as z_2 variable even it seems to induce heteroscedasticity in u.

It is also worth noting that in case of discrete (in particular, binary) endogenous regressors, the first-stage error is heteroscedastic by construction if we use a linear model for the first stage, but such heteroscedasticity applies to the whole error term, and so, from the above results, the consistency assumption for the Lewbel (2012) estimator are not satisfied and it should not be used.

Lewbel (2012) proposes using Breusch and Pagan (1979) test to test for the heteroscedasticity required in the first-stage error term u. But as shown above, if heteroscedasticity in the correlated component e_2 can lead to inconsistency, but this test does not distinguish the correlated component and idiosyncratic component v_2, and so, this test may identify wrong type of heteroscedasticity. Lewbel (2012) also proposes using Hansen (1982) J test to test if the validity of assumptions, but as will be shown in the simulation results, this test can be far from powerful when all Z_2 carry wrong form of heteroscedasticity.

2.2 The Klein-Vella Estimator

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

$$\begin{aligned} \varepsilon &= S_\varepsilon(X)\varepsilon^* \\ u &= S_u(X)u^* \end{aligned} \tag{12}$$

where $S_\varepsilon(X), S_u(X)$ describe how the standard deviations of the error terms depend on X.[6] ε^* and u^* are homoscedastic with constant correlation,

$$cov(\varepsilon^*, u^*) = \rho. \tag{13}$$

This in turns implies that the correlation coefficient of ε and u conditional on X is a constant ρ. Under these assumptions, a control function approach can be used, and OLS estimator for the coefficients of the following equation is then consistent:

$$y_1 = y_2\beta_1 + X\beta_2 + \rho_0 \frac{\hat{S}_\varepsilon(X)}{\hat{S}_u(X)}\hat{u} + \tilde{\varepsilon} \tag{14}$$

The term $[\hat{S}_\varepsilon(X)/\hat{S}_u(X)]\hat{u}$ is added as a control function so that β_1 and β_2 can be consistently estimated. Therefore, for this estimator to be valid, the heteroscedasticity should

[6] We can consider a subset of X for the variance function if we are confident that other variables does not enter this function.

be multiplicative and applies to the whole structural and first-stage error terms. Identification also requires that $S_\varepsilon(X)/S_u(X)$ depends on X and is not reduced to a constant or a linear function of X, and the correlation ρ is a constant that does not depend on X. However, justifying this structure by economic theory or by the nature of the problem is not a straightforward task.[7]

For estimation, Klein and Vella (2009b, 2010) propose a semiparametric approach in estimating $S_\varepsilon(X)$ and $S_u(X)$. First they assume single index functions $S_\varepsilon(X\delta_\varepsilon)$ and $S_u(X\delta_u)$ and then, they estimate the parameters δ_ε and δ_u as well as non-parametrically estimate the functions S_ε and S_u. Farre, Klein and Vella (2013) consider a parametric implementation, which is also used in this study and is detailed below. First, they assume the functional form of variance functions as

$$\begin{aligned} S_{\varepsilon i} &= \sqrt{exp(X_i'\delta_\varepsilon)} \\ S_{ui} &= \sqrt{exp(X_i'\delta_u)} \end{aligned} \tag{15}$$

These have the advantage that they are by definition positive and monotonic to the single linear indices. Following Farre, Klein and Vella (2013), the model can be estimated in multiple steps:

1. Use OLS on the first-stage regression (2) and obtain the residuals $\hat{u}$.

2. Regress $ln(\hat{u}^2)$ on X and obtain the coefficient $\hat{\delta}_u$. Construct $\hat{S}_u = exp(X\hat{\delta}_u)$. [8]

3. To improve efficiency, we may repeat step 1 and 2 using FGLS with $\hat{S}_u$ obtained above.

4. Estimate non-linearly the parameters β_1, β_2 and ρ by minimizing

$$\sum_{i=1}^{n}(y_{1i} - \beta_1 y_{2i} - X_i'\beta_2 - \rho\frac{\sqrt{exp(X_i'\hat{\delta}_\varepsilon)}}{\sqrt{exp(X_i'\hat{\delta}_u)}}\hat{u}_i)^2 \tag{16}$$

[7] It is not straightforward to deduce the bias term under wrong type of heteroscedasticity for this estimator, and thus I leave it to the simulation exercise to investigate the direction of bias.

[8] The constant term is not used in constructing S_u here, because it is not consistently estimated by the log-linear regression, while the functional form assumption implies that the constant term is multiplicative, allowing the constant terms to be combined with ρ.

For estimation of δ_ε, define $\hat{\varepsilon}_i = y_{1i} - \beta_1 y_{2i} - X_i'\beta_2$ for the given trial values of β_1 and β_2, then regress $ln(\hat{\varepsilon}_i^2)$ on X, obtain the coefficients $\hat{\delta}_\varepsilon$ and then put back into the expression (16) to calculate the value of the objective function. [9]

5. Use the minimized value of β_1 and β_2 to construct the control function term and perform an OLS by regressing y_{1i} on y_{2i}, X_i and the control function $\left(\sqrt{exp(X_i'\hat{\delta}_\varepsilon)}/\sqrt{exp(X_i'\hat{\delta}_u)}\right)\hat{u}_i$ and then obtain the final estimate.[10]

Thereafter, this estimator is called the two-step estimator because we estimate the first-stage equation first and then the structural equation separately. Though not considered by Klein and Vella (2010), it is straight-forward to include excluded instruments Z in step 1 and 2 above. In this study, if an excluded instrument Z is present, Z is also present in the first stage error variance function S_u, but not in the structural variance function S_ε.

In this study, I also implement this method using maximum likelihood and compare the results with the above two-step method. Assuming the error terms ε and u in (1) and (2) are distributed in bivariate normal, the log-likelihood function is given by

$$L(\beta,\gamma,\delta,\rho) = \sum_{i=1}^{n}\left[-ln(2\pi) - ln(s_{u,i}s_{\varepsilon,i}) - \frac{1}{2}ln(1-\rho^2) - \frac{1}{2(1-\rho^2)}(\tilde{u}_i^2 + \tilde{\varepsilon}_i^2 - 2\rho\tilde{u}_i\tilde{\varepsilon}_i)\right] \quad (17)$$

where

$$\tilde{\varepsilon}_i = \frac{y_{1i} - y_{2i}\beta_1 - X_i\beta_2}{\sqrt{exp(X_i\delta_\varepsilon)}} \quad (18)$$

$$\tilde{u}_i = \frac{y_{2i} - Z_i\gamma_1 - X\gamma_2}{\sqrt{exp(X_i\delta_u + Z\delta_{u,z})}} \quad (19)$$

$$s_{\varepsilon,i} = \sqrt{exp(X_i\delta_\varepsilon)} \quad (20)$$

$$s_{u,i} = \sqrt{exp(X_i\delta_u + Z\delta_{u,z})} \quad (21)$$

One advantage of this approach is that it is fully efficient under the assumptions and

[9] The constant term is again omitted and combined with ρ. A computational point to note is that, since some residuals are likely to be close to zero, I find that the calculated log squared residuals are rather sensitive to the parameter values and the objective function is not smooth. I smooth the objective function by using $ln(\hat{\varepsilon_i}^2 + 1/n)$ to avoid logarithm of very small numbers.

[10] This step is recommended by Farre, Klein and Vella (2013).

it can be performed in one step. The simulation results in this study will provide us evidence about its strength and weaknesses.

Klein and Vella (2009b, 2010) do not explicitly propose specification tests for the existence of heteroscedasticity for identification or tests for validity of overidentifying restrictions. However, the Breusch and Pagan (1979) test can be useful in detecting heteroscedasticity for the first stage error, but it cannot test for heteroscedasticity of the structural error as well as to test whether there is the collinearity of control function term with regressors X.

3 Simulation Schemes and Results

3.1 Simulation Scheme

The simulation in this study follows the data generating process below. The basic model is

$$\begin{aligned} y_{1i} &= \beta_0 + \beta_1 y_{2i} + x_i'\beta_2 + \varepsilon_i \qquad (22) \\ y_{2i} &= \gamma_0 + \gamma_1 Z_i + x_i'\gamma_2 + u_i \end{aligned}$$

where in some cases there is one excluded instrument Z_i, which is distributed in standard normal, and in some cases there are no excluded instrument. There are K exogenous regressors x_i, which are independent and distributed in standard normal. Since the key is to study the finite sample property of the estimators under wrong forms of heteroscedasticity, we consider three forms of heteroscedasticity with a common factor that generates the correlation between structural and first-stage errors.

Case 1: Klein-Vella Type

$$\begin{aligned} \varepsilon_i &= \sqrt{exp(x_i'\delta_\varepsilon)}\,(\alpha_1\theta_i + v_{1i}) \qquad (23) \\ u_i &= \sqrt{exp(x_i'\delta_u)}\,(\alpha_2\theta_i + v_{2i}) \end{aligned}$$

where the heteroscedasticity affects the whole error term.

Case 2: Lewbel Type

$$\begin{aligned} \varepsilon_i &= \alpha_1\theta_i + \sqrt{exp(x_i'\delta_\varepsilon)}v_{1i} \\ u_i &= \alpha_2\theta_i + \sqrt{exp(x_i'\delta_u)}v_{2i} \end{aligned} \tag{24}$$

where the heteroscedasticity affects only the idiosyncratic component.

Case 3: Correlated Component Only

$$\begin{aligned} \varepsilon_i &= \sqrt{exp(x_i'\delta_\varepsilon)}\alpha_1\theta_i + v_{1i} \\ u_i &= \sqrt{exp(x_i'\delta_u)}\alpha_2\theta_i + v_{2i} \end{aligned} \tag{25}$$

where it affects only the correlated component. δ_ε and δ_u control the degree of heteroscedasticity. θ_i, v_{1i} and v_{2i} are distributed in independent standard normal distribution in the simulation.

Simulated data from the above models are used to estimate the structural parameters β using the methods described above.[11] Here I use all X as Z_2 variables without using other selection criteria.[12] I will also present the distribution of Breusch and Pagan (1979) (BP) and J statistics in order to investigate whether they are effective tests for relevant and valid instruments. The BP test is conducted by regressing the squared first-stage residual on X and Z (if available) and making use of the associated nR^2, which is distributed χ^2 of degree of freedom equal to number of non-constant regressors.[13] I also present the robust first-stage F statistic for the endogenous regressor using the Lewbel generated instruments and other exogenous instrument Z.[14] Median, 10^{th} and 90^{th} percentiles for the point estimators are presented to assess the biasedness and skewness of the estimators. [15]

[11] I code the Klein-Vella estimator using R, while I use the 'ivlewbel' package by Fernihough (2014) to estimate with the Lewbel estimator.

[12] The 'ivreg2h' in Stata default to use all exogenous regressors for Z_2.

[13] Here the standard one is used rather than using the log of squared residuals more directly related to the Klein-Vella estimator.

[14] This is provided by the 'ivlewbel' R package.

[15] Mean and standard deviation are not used since some estimators do not have moments. Also, the

Without further specification, I take $\beta_1 = 0$, and therefore the value of mean and median of bootstrap samples represent mean and median bias respectively. $\beta_0 = \alpha_0 = 0$, $\beta_{2k} = \gamma_{2k} = 1$ for all k. α_1 and α_2 are set to 1 and the associated correlation between ε and u is about 0.5.[16] The number of observation for each sample considered is 500.[17] The number of replications is 5000 for each design. To distinguish the cases where heteroscedasticity is equally spread across variables versus being concentrated on one variable, I allow different heteroscedastic parameters δ_{u1} and $\delta_{\varepsilon 1}$ for the first variable, and δ_{u2} and $\delta_{\varepsilon 2}$ for all remaining variables. Here, $\delta_{\varepsilon 2}$ is always set to zero. For the case of exogenous instruments, the associated parameter γ_1 in the first stage equation is set so that by itself the first-stage F (non-robust) is about 3, which is moderately weak.

3.2 Simulation Results

Table 1 and Table 2 show the results for Case 1, the Klein-Vella type heteroscedasticity, with and without an exogenous excluded instrument, respectively. The results show that the Lewbel estimator is severely biased in this case, while at the same time, the rejection rate of the overidentifying restriction J test is only close to its nominal size, meaning that it does not have power in detecting such misspecification. The bias tends to be in the same direction as the OLS when the heteroscedasticity in the structural error is small. The bias is smaller if the δ_u and δ_ε are of different signs because their effects may cancel out each other. If $\delta_{\varepsilon 1}$ is very negative and δ_{u1} is positive, the bias can be in the opposite direction of bias of the OLS. These are consistent with the discussions in Section 2.1. For the Klein and Vella estimator, the maximum likelihood method provides estimators with very small median bias and a smaller spread between percentiles, while the two-step estimator has a finite sample bias and a larger spread. They both show some skewness towards the negative side. The finite sample (median) bias is larger when we have more variables in X, and Z_2. The bias is substantially larger when there are only one variable among the ten variables in Z_2 that is truly heteroscedastic. The finite sample bias is also

high and low percentiles provide information on the skewness of the estimators, besides their spread.

[16]The actual correlation is smaller than 0.5 because the form of heteroscedasticity implies a mean of the multiplicative factor slightly higher than 1.

[17]Results for $n = 100$ are also available in the Appendix.

larger when heteroscedasticity is weak.

With a weak exogenous excluded instrument, as shown in Table 2, the results are very similar to the case without the excluded instrument. In comparison, the bias and spread are slightly smaller than those without the excluded instrument, while the J statistic is still small and so, the rate of rejection of overidentifying restrictions is still close to its nominal size. When the excluded instrument is stronger, the bias is smaller and the rejection rate of overidentification test increases. However, the bias is still substantial and the power of J test is still small. Therefore, these results imply that Lewbel estimator is not robust to misspecification of the form of heteroscedasticity, and failing to reject the overidentifying restriction J test is insufficient to conclude that there is no substantial bias, even when a valid excluded instrument is available. On the other hand, it may also be problematic to use J test to justify the validity of the excluded instrument by also using the Lewbel's heteroscedasticity instrument when reseachers are uncertain about the true form of heteroscedasticity, because they can both lead to a similar bias, where the J test lacks the power to detect.

Table 3 and Table 4 show the results for Case 2, the Lewbel type heteroscedasticity, with and without the exogenous instrument, respectively. In this case, the Lewbel estimator provides a median unbiased estimator with approximately symmetric distribution. The finite sample bias is larger when the heteroscedasticity for identification is weak, or when we include many weakly heteroscedastic variables, similar to the Klein-Vella estimator for Case 1. As the specification is correct, it is desirable that the rejection rate of the J test is a close to its nominal size. In contrast, the Klein-Vella estimators are seriously biased, mainly in the opposite direction of the bias of the OLS estimator for the specifications considered. Under misspecification, the maximum likelihood estimator is more seriously biased with wider confidence interval, while the two-step estimator has a lower bias and a smaller spread. This is reasonable because the maximum likelihood approach takes all assumptions together, while the two-step method breaks down the estimation of the two equations separately, making the latter more robust to misspecification. Similar to the Lewbel estimator in Case 1, if the directions of heteroscedasticity for first-stage and structural errors are different, the bias tends to be smaller or even in another direction, as biases from different directions tend to cancel out each other. Similar to Case 1, if a weak valid excluded instrument is included, as shown in Table 4, the bias for the Klein-Vella estimators are reduced, so as the spread for all estimators.

Table 5 and Table 6 show the results for Case 3, where heteroscedasticity only applies to the correlated component of the error terms. The results show that both estimators are biased substantially, sometimes the bias is even larger than that of the OLS, while both estimators are biased in the same direction as the OLS in the specifications considered. Thus, the Lewbel estimator has the same direction of bias as in Case 1 while the Klein-Vella estimator has an opposite direction of bias as in Case 2. Similar to the cases before, the bias goes into the opposite directions if the direction of heteroscedasticity for first-stage and structural errors are opposite to each other and the latter is stronger than the former. The overidentifying J test again has low power in detecting the related source of bias for the Lewbel estimator. Inclusion of a weak excluded instrument, as shown in Table 6, reduces the bias and spread slightly.

In summary, the simulation results show that both Lewbel and Klein-Vella estimators are not robust to misspecification of the form of heteroscedasticity, while the overidentifying restriction test, the Hansen's J test, for Lewbel's estimator, has low power in detecting such violations. The sign of bias is not always the same, depending on the nature of violation in the form of heteroscedasticity and on the direction of heteroscedasticity of structural and first-stage errors with respect to Z_2 variables. Including many Z_2 variables with weak heteroscedasticity also increases finite sample bias even when the form of heteroscedasticity is correct. Including some excluded instruments may not always help us detect if we have the wrong form of heteroscedasticity.

4 Empirical Example

Lewbel (2012) has demonstrated an empirical example where he estimates the Engel curve: the effect of log total expenditure on food share. This study replicates his results and further estimates the model using the Klein-Vella approach for comparison. As Lewbel (2012) describes, the data consist of the same set of demographically homogeneous households, which were used by Banks, Blundell, and Lewbel (1997) to analyze Engel curves. These are all households in the United Kingdom Family Expenditure Survey 1980–1982, composed of two married adults without children, living in the Southeast (including London).[18] Other control variables include age, spouse's age, squared ages, seasonal dummies, and dummies for whether the spouse working, whether having gas central heating, whether owning a washing machine, whether owning one car, and whether owning two cars. The total number of observations is 854. The total expenditure can potentially be endogenous because of measurement errors, in particular from infrequently purchased items (see, e.g. Meghir and Robin 1992).

Table 7 reports the results of this exercise. The results for OLS, 2SLS using the excluded instrument log total income, and the Lewbel's GMM estimator using both heteroscedasticity-generated instruments together with log total income as excluded instrument are close to what Lewbel (2012) reports. My result for the specification where only the Lewbel's heteroscedasticity-generated instruments are used have a noticeable discrepancy from the results reported in Lewbel (2012). My result, however, is very close to what Baum (2013) reports when he demonstrates the use of the related Stata command. Therefore, my results are consistent to the best estimates using these methods. The coefficients for the total expenditure are about -0.05 to -0.09 and they are statistically significant except for the Lewbel's approach with only heteroscedasticity related instruments. Reported J statistics do not lead us to reject the null of validity of overidentifying restrictions. One concern is that the related first-stage F statistics for Lewbel's approach are rather small[19], indeed they are even smaller than the statistical significance level. Moreover, the BP statistics are also low and fail to reject the null of homoscedasticity when no excluded instrument is used. One possible source of the problem is that it uses all exogenous regressors to generate heteroscedastic instruments, but it may give rise to

[18] The data is obtained from the database directed from Baum and Schaffer (2012) Stata user-written command ivreg2h.

[19] The statistic is heteroscedasticity robust F statistic on excluded and generated instruments reported by ivlewbel package for R.

many weak instruments problem in this case. This may lead to a larger finite sample bias and unreliable asymptotic standard errors. It may be advisable to reduce the number of heteroscedasticity generated instruments by using only the variables with the highest level of heteroscedasticity.

Table 7 also reports the results for the Klein and Vella estimator, both two-stage and maximum likelihood estimators in parametric form. The results for the two-stage method are close to what the Lewbel estimator provides, which is close to the pattern for Case 3 in the simulation results above. However, the maximum likelihood estimator provides a very noisy estimator for the case without excluded instrument. This feature is close to those in Case 2 in the simulation results above, where the Lewbel's form of heteroscedasticity is correct. Since the forms of heteroscedasticity investigated above are not exhaustive, and there is mixed evidence, so I cannot draw conclusion from the results about which form of heteroscedasticity is more likely to be true.

Besides asymptotic standard errors, the standard errors obtained by bootstrap are also presented. Pair bootstrap at observation level with 500 repetitions is used. It shows that the bootstrap standard errors are very close to the asymptotic standard errors for OLS and 2SLS, and are slightly larger for Lewbel's GMM estimators. The asymptotic standard errors for the Klein-Vella maximum likelihood estimator seem to have underestimated the true standard errors more substantially, especially for the cases that use only heteroscedasticity for identification.[20] I only obtain the bootstrap standard errors for the Klein-Vella two-step estimator, and the bootstrap standard errors are also close to those from the Lewbel estimator.

5 Conclusion

In this paper, simulation exercises are performed to investigate the degree of bias and other finite sample properties for the Lewbel (2012) and the Klein and Vella (2010) estimators under various form of heteroscedasticity. The results clearly show that the estimators are not robust to misspecification of the form of heteroscedasticity: whether it

[20]The problem may be that the global maximum fluctuates between two very different local minima.

applies to the whole error term or only to the idiosyncratic components of the structural and first-stage error terms do matter. Moreover, the over-identification test, in particular, Hansen's (1982) J test, has low power to reject the null when there is a wrong form of heteroscedasticity, with or without other exogenous variables as excluded instruments.

Therefore, researchers should be cautious when using these estimators. It is not sufficient to justify only the existence of heteroscedasticity in the error term. We should also justify which form of heteroscedasticity appears in the error terms. Moreover, since the biases of both estimators can be on the both sides of the true value and are not always on the same side of the bias of OLS, it is also not safe to conclude that these estimators are bounds for the true value without examining further conditions. Further theoretical investigations for appropriate tests about the form of heteroscedasticity existed in the data or the conditions required to take the estimates from these two methods as bounds on the true value are valuable future research directions.

References

[1] Banks, J., Blundell, R., and Lewbel, A. (1997), "Quadratic Engel Curves and Consumer Demand," *Review of Economics and Statistics*, 79, 527– 539.

[2] 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.

[3] Baum, B.F. (2013) "Implementing new econometric tools in Stata," Mexican Stata Users' Group Meetings 2013-09, Stata Users Group.

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

[5] Chowdhury, M., Bhattacharya, P., Mallick D., and Ulubasoglu, M. (2014) “An Empirical Inquiry into the Role of Sectoral Diversification in Exchange Rate Regime Choice,” *European Economic Review*, 67, 210-227.

[6] Denny, K. and Oppedisano, V. (2013) “The Surprising Effect of Larger Class Sizes: Evidence Using Two Identification Strategies,” *Labour Economics*, 23, 57-65.

[7] Emran, M. and Hou, Z. (2013) “Access to Markets and Rural Poverty: Evidence From Household Consumption in China,” *Review of Economics and Statistics*, 95(2), 682-697.

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

[9] 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.

[10] 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

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

[12] Klein, R. and Vella, F. (2009a) “A Semiparametric Model for Binary Response and Continuous Outcomes Under Index Heteroscedasticity,” *Journal of Applied Econometrics*, 24, 735-762.

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

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

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

[16] Meghir, C., and Robin, J.-M. (1992), “Frequency of Purchase and the Estimation of Demand Systems,” *Journal of Econometrics*, 53, 53–86.

[17] Millimet, D. and Roy, J. (2015) “Empirical Tests of the Pollution Haven Hypothesis When Environmental Regulation is Endogenous,” *Journal of Applied Econometrics*, forthcoming.

Appendix

A Simplified Case

To illustrate the conditions required for consistency, consider a simplified case where there is no covariates X, and y_1 and y_2 are mean zero and also the heteroscedasticity related variable Z_2 is a binary variable. We may consider y_1 and y_2 as their residuals of the regression on other covariates. The model can be expressed in terms of variables with mean zero

$$\begin{aligned} y_1 &= y_2\beta + \varepsilon \qquad (26) \\ y_2 &= u \end{aligned}$$

Then, the probability limit of the Lewbel's IV estimator using $(Z_2 - \mu_2)u$ as instrument is given by

$$\beta_L = \frac{cov((Z_2-\mu_2)u, y_1)}{cov((Z_2-\mu_2)u, y_2)} = \frac{E((Z_2-\mu_2)uy_1)}{E((Z_2-\mu_2)uy_2)} = \frac{E((Z_2-\mu_2)E(uy_1|Z_2))}{E((Z_2-\mu_2)E(uy_2|Z_2))} \qquad (27)$$

where $\mu_2 = E(Z_2)$. Since Z_2 is a dummy variable, $\mu_2 = E(z_2) = Pr(Z_2 = 1)$. Denoting this probability as p, we have

$$\begin{aligned} E((Z_2 - \mu_2)E(uy_1|Z_2)) &= p(1-p)E(uy_1|z_2 = 1) + (1-p)(-p)E(uy_1|Z_2 = 0) \quad (28) \\ &= p(1-p)\left[E(uy_1|z_2 = 1) - E(uy_1|Z_2 = 0)\right] \end{aligned}$$

Similarly, the denominator can also be expressed as

$$E((Z_2-\mu_2)E(uy_2|Z_2)) = p(1-p)\left[E(uy_2|Z_2 = 1) - E(uy_2|Z_2 = 0)\right] = Var(u|Z_1 = 1)-Var(u|Z_1 = 0) \tag{29}$$

since $u = y_2$.

As a result, the Lewbel estimator has a probability limit

$$\beta_{LB} = \frac{E(uy_1|Z_2 = 1) - E(uy_1|Z_2 = 0)}{E(uy_2|Z_2 = 1) - E(uy_2|Z_2 = 0)} = \frac{E(uy_1|Z_2 = 1) - E(uy_1|Z_2 = 0)}{Var(u|Z_1 = 1) - Var(u|Z_1 = 0)} \tag{30}$$

which is the ratio of the differences in covariance between two groups for u and y and difference in variance of u between the two groups defined by Z_2. Further, putting $y_1 = y_2\beta + \varepsilon$, the numerator becomes

$$\begin{aligned} E(uy_1|Z_2 = 1) - E(uy_1|Z_2 = 0) &= [E(uy_2|Z_2 = 1) - E(uy_2|Z_2 = 0)]\beta \quad (31) \\ &\quad +E(u\varepsilon|Z_2 = 1) - E(u\varepsilon|Z_2 = 0) \\ &= [E(u^2|Z_2 = 1) - E(u^2|Z_2 = 0)]\beta \\ &\quad +E(e_1e_2|Z_2 = 1) - E(e_1e_2|Z_2 = 0) \end{aligned}$$

The last equality holds because conditional on Z_2, $cov(v_1, v_2) = 0$ and $cov(e_i, v_j) = 0$ for all $i, j = 1, 2$. On the other hand, the denominator becomes

$$\begin{aligned} Var(u|Z_1 = 1) - Var(u|Z_1 = 0) &= Var(e_2 + v_2|Z_2 = 1) - Var(e_2 + v_2|Z_2 = 0) \\ &= [Var(e_2|Z_2 = 1) - Var(e_2|Z_2 = 0)] \quad (32) \\ &\quad +[Var(v_2|Z_1 = 1) - Var(v_2|Z = 0)] \end{aligned}$$

If we require the covariance between e_1 and e_2 to be independent of Z_2, then it is very unlikely we can have heteroscedasticity in e_2 itself. Therefore, the difference in variance has to be driven by any difference in conditional variance in v_2.

Therefore,

$$\beta_{LB} = \beta + \frac{E(e_1 e_2|Z_2 = 1) - E(e_1 e_2|Z_2 = 0)}{Var(v_2|Z_2 = 1) - Var(v_2|Z_2 = 0)} \tag{33}$$

This implies $\beta_L = \beta$ only when $E(e_1 e_2|Z_2 = 1) = E(e_1 e_2|Z_2 = 0)$. That means, the two groups defined by the dummy variable z_2 must have different conditional variances for v_2, but they must have the same covariance for the correlated component in the two error terms. This highlights the requirement that heteroscedasticity can only appear in the uncorrelated component of the error terms.

Table 1: Simulation Results for Data from Klein and Vella Form of Heteroscedasticity without Excluded Instruments

n	K	δ_{u1}	δ_{u2}	$\delta_{\varepsilon 1}$	β_{OLS} median (q10,q90)	BP median (BP/K)	β_{LB} median (q10,q90)	F median (% F>10)	J median (% $p < 0.05$)	$\beta_{KV,\text{2-step}}$ median (q10,q90)	$\beta_{KV,ML}$ median (q10,q90)
500	2	0.4	0.4	0	0.4452	52.35	0.2246	61.03	0.4398	0.0120	0.0053
					(0.396,0.493)	(26.18)	(0.118,0.323)	(0.999)	(0.045)	(-0.356,0.202)	(-0.227,0.185)
500	2	0.4	0.4	0.2	0.4547	52.27	0.2812	61.23	0.8484	0.0138	0.0067
					(0.406,0.507)	(26.14)	(0.176,0.391)	(1.00)	(0.114)	(-0.507,0.258)	(-0.291,0.230)
500	2	0.4	0.4	-0.2	0.4360	52.13	0.1613	61.09	0.7380	0.0093	0.0014
					(0.389,0.486)	(26.07)	(0.061,0.253)	(1.00)	(0.097)	(-0.257,0.160)	(-0.167,0.141)
500	2	-0.4	-0.4	0.2	0.4384	52.51	0.1635	61.23	0.7373	0.0043	-0.0013
					(0.389,0.486)	(26.25)	(0.058,0.251)	(1.00)	(0.097)	(-0.257,0.161)	(-0.165,0.132)
500	2	-0.4	-0.4	-0.2	0.4543	52.41	0.2803	61.90	0.8150	0.0152	-0.0022
					(0.406,0.506)	(26.20)	(0.168,0.386)	(1.00)	(0.113)	(-0.487,0.254)	(-0.307,0.228)
500	2	0.4	-0.4	0.2	0.4562	52.41	0.2796	62.10	0.8320	0.0110	-0.0039
					(0.406,0.505)	(26.20)	(0.170,0.385)	(1.00)	(0.110)	(-0.468,0.253)	(-0.292,0.232)
500	2	0.2	0.2	0.2	0.4930	18.23	0.3669	9.5920	0.8106	0.1116	0.0084
					(0.442,0.542)	(9.12)	(0.151,0.570)	(0.472)	(0.114)	(-0.601,0.986)	(-0.998,0.592)
500	2	0.5	0	0.2	0.4700	43.90	0.3291	43.310	0.4435	0.0477	-0.0028
					(0.420,0.520)	(21.95)	(0.198,0.454)	(1.00)	(0.0402)	(-0.639,0.418)	(-0.548,0.358)
500	2	0.5	0	-0.8	0.4457	44.65	-0.1222	44.84	0.3683	-0.0085	-0.0005
					(0.392,0.503)	(22.33)	(-0.284,-0.004)	(1.00)	(0.034)	(-0.165,0.107)	(-0.070,0.063)
500	10	0.2	0.2	0.2	0.4379	65.16	0.2458	19.33	9.094	0.0629	0.0024
					(0.391,0.488)	(6.516)	(0.143,0.344)	(0.969)	(0.052)	(-0.140,0.233)	(-0.235,0.188)
500	10	0.5	0	0.2	0.4683	50.84	0.3427	11.07	8.3720	0.1737	0.0134
					(0.419,0.520)	(5.084)	(0.212,0.469)	(0.601)	(0.034)	(-0.131,0.574)	(-0.726,0.512)

The total number of repetition is 5000. The correlation between the first stage and structural error is set at about 0.5. δ_{u1} is the coefficient for the variance function for the first variable of X, while δ_{u2} is the coefficient for all remaining X variables. Similarly, $\delta_{\varepsilon 1}$ is the coefficient of the variance function for the first variable in X, while that for all remaining X variables are zero. BP test is the nR^2 statistic of regressing squares of first-stage residuals on all K exogenous regressors X. BP/K is then having $F_{K,\infty}$ distribution for comparison. F is the first-stage (robust) F statistics using Lewbel generated instruments, and the J statistic is the corresponding statistic under Lewbel GMM method. For estimators, median, 10th and 90th percentiles are presented to understand the bias, spread and symmetry of the estimators.

Table 2: Simulation Results for Data from Klein and Vella Form of Heteroscedasticity with Excluded Instruments

n	K	δ_{u1}	δ_{u2}	$\delta_{\varepsilon 1}$	γ_1	β_{OLS} median (q10,q90)	β_{2SLS} median (q10,q90)	BP median (BP/K+1)	β_{LB} median (q10,q90)	F median (% F>10)	J median (% $p<0.05$)	β_{KV},2-step median (q10,q90)	$\beta_{KV,ML}$ median (q10,q90)
500	2	0.4	0.4	0	0.13	0.4402	0.0203	53.19	0.2180	43.36	1.556	0.0400	-0.0005
						(0.393,0.489)	(-1.043,0.600)	(17.73)	(0.118,0.315)	(1.00)	(0.062)	(-0.223,0.211)	(-0.217,0.178)
500	2	0.4	0.4	0.2	0.13	0.4528	0.0261	53.34	0.2735	43.07	2.190	0.0542	-0.0006
						(0.403,0.502)	(-0.995,0.615)	(17.78)	(0.162,0.377)	(1.00)	(0.117)	(-0.260,0.258)	(-0.293,0.222)
500	2	0.4	0.4	-0.2	0.13	0.4331	0.0271	52.76	0.1596	42.84	1.872	0.0243	-0.0017
						(0.386,0.480)	(-1.041,0.630)	(17.59)	(0.063,0.250)	(1.00)	(0.089)	(-0.182,0.162)	(-0.158,0.130)
500	2	-0.4	-0.4	0.2	0.13	0.4347	0.0271	52.66	0.1614	42.61	1.815	0.0300	0.0030
						(0.387,0.481)	(-0.986,0.609)	(17.55)	(0.062,0.253)	(1.00)	(0.080)	(-0.175,0.173)	(-0.155,0.139)
500	2	-0.4	-0.4	-0.2	0.13	0.4522	0.0242	52.97	0.2717	43.10	2.186	0.0506	-0.0022
						(0.404,0.502)	(-0.985,0.614)	(17.66)	(0.162,0.380)	(1.00)	(0.111)	(-0.259,0.258)	(-0.274,0.214)
500	2	0.4	-0.4	0.2	0.13	0.4519	0.0235	52.96	0.2717	42.87	2.169	0.0517	-0.0021
						(0.400,0.502)	(-1.061,0.600)	(17.65)	(0.162,0.381)	(0.999)	(0.129)	(-0.271,0.263)	(-0.296,0.223)
500	2	0.4	0.4	0.2	0.5	0.4102	0.0055	53.10	0.1790	65.34	4.755	0.0194	0.0002
						(0.362,0.459)	(-0.173,0.151)	(17.70)	(0.080,0.273)	(1.00)	(0.382)	(-0.139,0.147)	(-0.149,0.129)
500	2	0.2	0.2	0.2	0.13	0.4896	0.0120	19.25	0.3262	8.169	2.340	0.1140	-0.0047
						(0.439,0.540)	(-1.112,0.600)	(6.417)	(0.112,0.518)	(0.336)	(0.137)	(-0.350,0.490)	(-0.717,0.416)
500	2	0.5	0	0.2	0.13	0.4657	0.0218	45.46	0.3148	31.62	1.813	0.0821	-0.0062
						(0.415,0.515)	(-1.037,0.595)	(15.15)	(0.186,0.434)	(0.995)	(0.090)	(-0.336,0.354)	(-0.455,0.308)
500	2	0.5	0	-0.8	0.13	0.4434	0.0235	45.69	-0.1027	31.77	1.304	0.0086	0.0004
						(0.391,0.500)	(-1.055,0.793)	(15.23)	(-0.259,0.009)	(0.994)	(0.043)	(-0.138,0.122)	(-0.068,0.063)
500	10	0.2	0.2	0.2	0.13	0.4346	0.0372	66.31	0.2375	18.12	10.29	0.0776	-0.0069
						(0.385,0.484)	(-1.100,0.639)	(6.028)	(0.139,0.339)	(0.957)	(0.062)	(-0.112,0.231)	(-0.229,0.175)
500	10	0.5	0	0.2	0.13	0.4664	0.0255	51.77	0.3262	10.630	9.90	0.1561	0.0034
						(0.414,0.517)	(-1.014,0.626)	(4.706)	(0.198,0.456)	(0.562)	(0.049)	(-0.101,0.444)	(-0.597,0.391)

Refer to the notes of Table 1. The first stage coefficient on excluded instrument is set so that the first stage F is about 3. The first-stage (robust) F statistics include also the excluded instrument z. First-stage F statistics also include the effect of the excluded instrument z.

Table 3: Simulation Results for Data from Lewbel Form of Heteroscedasticity without Excluded Instruments

n	K	δ_{u1}	δ_{u2}	$\delta_{\varepsilon 1}$	β_{OLS}	BP	β_{LB}	F	J	$\beta_{KV,\text{2-step}}$	$\beta_{KV,ML}$
					median	median	median	median	median	median	median
					(q10,q90)	(BP/K)	(q10,q90)	(% F>10)	(% $p < 0.05$)	(q10,q90)	(q10,q90)
500	2	0.5	0.5	0	0.4393	30.98	0.0103	18.16	0.4292	-0.48039	-0.8238
					(0.387,0.490)	(15.49)	(-0.175,0.161)	(0.845)	(0.038)	(-1.442,-0.033)	(-1.941,-0.267)
500	2	0.5	0.5	0.2	0.4374	31.18	0.0069	18.35	0.4128	-0.5825	-1.3098
					(0.386,0.489)	(15.59)	(-0.185,0.169)	(0.852)	(0.036)	(-1.704,0.134)	(-2.902,-0.282)
500	2	-0.5	-0.5	0.2	0.4375	31.11	0.0057	18.17	0.4064	-0.3539	-0.4497
					(0.387,0.490)	(15.56)	(-0.180,0.156)	(0.853)	(0.040)	(-1.138,0.002)	(-1.040,-0.114)
500	2	-0.5	-0.5	-0.2	0.4376	31.07	0.0061	18.50	0.3783	-0.5784	-1.3340
					(0.385,0.491)	(15.53)	(-0.187,0.166)	(0.859)	(0.034)	(-1.675,0.116)	(-2.912,-0.325)
500	2	0.5	-0.5	0.2	0.4394	30.70	0.0018	18.06	0.3936	-0.5816	-1.3423
					(0.386,0.492)	(15.35)	(-0.198,0.165)	(0.835)	(0.037)	(-1.691,0.126)	(-2.956,-0.310)
500	2	0.3	0.3	0.2	0.4784	12.13	0.0318	5.102	0.3605	-0.2106	-0.7134
					(0.425,0.530)	(6.064)	(-0.323,0.283)	(0.152)	(0.043)	(-1.259,1.605)	(-2.660,3.148)
500	2	0.7	0	0.2	0.4405	30.43	0.0050	17.61	0.4003	-0.6173	-1.7284
					(0.388,0.492)	(15.22)	(-0.190,0.168)	(0.831)	(0.037)	(-1.893,1.357)	(-3.135,3.264)
500	2	0.5	0.5	-0.2	0.4385	31.07	0.0077	18.30	0.3981	-0.3514	-0.4294
					(0.387,0.491)	(15.53)	(-0.177,0.157)	(0.857)	(0.038)	(-1.146,0.003)	(-0.980,-0.106)
500	2	0.7	0	-1.0	0.4388	30.32	0.0170	17.70	0.3815	0.1169	0.0476
					(0.384,0.492)	(15.16)	(-0.201,0.183)	(0.832)	(0.036)	(-0.094,0.287)	(-0.094,0.172)
500	10	0.3	0.3	0.2	0.3910	55.31	0.0278	11.44	7.775	-0.2519	-0.7950
					(0.336,0.446)	(5.531)	(-0.100,0.146)	(0.614)	(0.033)	(-0.531,-0.032)	(-1.861,-0.327)
500	10	0.7	0	0.2	0.4393	37.15	0.0716	5.075	8.051	-0.1685	-1.4321
					(0.386,0.492)	(3.715)	(-0.102,0.237)	(0.086)	(0.047)	(-0.507,0.981)	(-3.209,3.454)

Refer to the notes of Table 1.

Table 4: Simulation Results for Data from Lewbel Form of Heteroscedasticity with Excluded Instruments

n	K	δ_{u1}	δ_{u2}	$\delta_{\varepsilon 1}$	α_1	β_{OLS}	β_{2SLS}	BP	β_{LB}	F	J	$\beta_{KV,\text{2-step}}$	$\beta_{KV,ML}$
						median	median	median	median	median	median	median	median
						(q10,q90)	(q10,q90)	(BP/K+1)	(q10,q90)	(% F>10)	(% $p < 0.05$)	(q10,q90)	(q10,q90)
500	2	0.5	0.5	0	0.13	0.4340	0.0377	31.94	0.0105	14.24	1.346	-0.2286	-0.6357
						(0.385,0.485)	(-0.979,0.629)	(10.65)	(-0.163,0.160)	(0.744)	(0.042)	(-0.657,0.026)	(-1.544,-0.203)
500	2	0.5	0.5	0.2	0.13	0.4374	0.0271	31.39	0.0169	13.75	1.245	-0.2544	-0.9751
						(0.385,0.488)	(-1.020,0.595)	(10.46)	(-0.170,0.168)	(0.738)	(0.045)	(-0.723,0.067)	(-2.620,-0.247)
500	2	0.5	0.5	-0.2	0.13	0.4363	0.0252	31.79	0.0169	14.01	1.324	-0.1640	-0.3724
						(0.386,0.488)	(-1.036,0.610)	(10.60)	(-0.156,0.158)	(0.747)	(0.038)	(-0.530,0.062)	(-0.864,-0.063)
500	2	-0.5	-0.5	0.2	0.13	0.4359	0.0211	31.60	0.0192	14.07	1.276	-0.1653	-0.3588
						(0.386,0.487)	(-1.014,0.601)	(10.53)	(-0.154,0.160)	(0.745)	(0.044)	(-0.543,0.061)	(-0.852,-0.071)
500	2	-0.5	-0.5	-0.2	0.13	0.4337	0.0430	31.98	0.0121	14.00	1.304	-0.2570	-0.9994
						(0.382,0.486)	(-0.995,0.627)	(10.66)	(-0.171,0.166)	(0.753)	(0.045)	(-0.721,0.060)	(-2.692,-0.277)
500	2	0.5	-0.5	0.2	0.13	0.4360	0.0258	31.47	0.0162	13.98	1.329	-0.2640	-0.9514
						(0.384,0.489)	(-1.002,0.596)	(10.49)	(-0.166,0.164)	(0.744)	(0.045)	(-0.712,0.066)	(-2.646,-0.254)
500	2	0.5	0.5	0.2	0.5	0.3959	0.0024	31.43	0.0065	34.25	1.222	-0.0754	-0.1468
						(0.345,0.446)	(-0.184,0.152)	(10.48)	(-0.126,0.116)	(1.00)	(0.038)	(-0.245,0.068)	(-0.374,0.037)
500	2	0.3	0.3	0.2	0.13	0.4728	0.0149	12.85	0.0440	5.104	1.281	-0.0720	-0.5503
						(0.423,0.524)	(-1.067,0.578)	(4.284)	(-0.260,0.277)	(0.097)	(0.055)	(-0.579,0.844)	(-2.616,0.821)
500	2	0.7	0	0.2	0.13	0.4382	0.0278	31.08	0.0170	13.35	1.310	-0.2799	-1.3348
						(0.385,0.490)	(-0.989,0.628)	(10.36)	(-0.173,0.176)	(0.719)	(0.042)	(-0.765,0.096)	(-3.075,-0.316)
500	2	0.7	0	-1.0	0.13	0.4376	0.0211	30.82	0.0232	13.49	1.325	0.1181	0.0464
						(0.384,0.493)	(-1.138,0.741)	(10.27)	(-0.171,0.185)	(0.722)	(0.037)	(-0.075,0.281)	(-0.091,0.172)
500	10	0.3	0.3	0.2	0.13	0.3884	0.0329	56.11	0.0309	10.95	8.787	-0.1831	-0.6878
						(0.335,0.442)	(-0.997,0.672)	(5.101)	(-0.096,0.146)	(0.590)	(0.032)	(-0.414,-0.0003)	(-1.546,-0.280)
500	10	0.7	0	0.2	0.13	0.4357	0.0251	37.93	0.0768	5.13	9.108	-0.1087	-1.0880
						(0.383,0.491)	(-1.046,0.626)	(3.448)	(-0.092,0.241)	(0.067)	(0.047)	(-0.390,0.279)	(-2.767,0.938)

Refer to the notes of Table 1 and 2.

Table 5: Simulation Results for Data from Heteroscedasticity on Correlated Component without Excluded Instruments

n	K	δ_{u1}	δ_{u2}	$\delta_{\varepsilon 1}$	β_{OLS} median (q10,q90)	BP median (BP/K)	β_{LB} median (q10,q90)	F median (% F>10)	J median (% $p < 0.05$)	β_{KV},2-step median (q10,q90)	$\beta_{KV,ML}$ median (q10,q90)
500	2	0.5	0.5	0	0.4673	30.83	0.4190	17.99	0.4765	0.4140	0.3930
					(0.420,0.513)	(15.41)	(0.285,0.552)	(0.847)	(0.042)	(0.151,0.634)	(0.188,0.598)
500	2	0.5	0.5	0.2	0.4814	31.16	0.5128	18.30	0.9609	0.5381	0.5231
					(0.436,0.526)	(15.58)	(0.379,0.665)	(0.851)	(0.137)	(0.234,0.862)	(0.295,0.796)
500	2	0.5	0.5	-0.2	0.4573	30.72	0.3289	17.99	0.8143	0.2840	0.2559
					(0.411,0.505)	(15.36)	(0.188,0.469)	(0.846)	(0.126)	(-0.056,0.495)	(0.044,0.451)
500	2	-0.5	-0.5	0.2	0.4580	30.95	0.3319	18.19	0.9206	0.2848	0.2559
					(0.411,0.504)	(15.48)	(0.194,0.470)	(0.849)	(0.125)	(-0.035,0.499)	(0.055,0.452)
500	2	-0.5	-0.5	-0.2	0.4803	30.84	0.5164	18.15	0.9443	0.5470	0.5332
					(0.435,0.527)	(15.42)	(0.383,0.662)	(0.849)	(0.141)	(0.242,0.858)	(0.308,0.795)
500	2	0.5	-0.5	0.2	0.4798	30.95	0.5134	18.31	0.9420	0.5403	0.5259
					(0.436,0.526)	(15.47)	(0.381,0.661)	(0.849)	(0.137)	(0.232,0.860)	(0.295,0.785)
500	2	0.3	0.3	0.2	0.4979	12.05	0.6275	5.024	0.7460	0.6710	0.7052
					(0.449,0.545)	(6.025)	(0.382,0.918)	(0.149)	(0.1116)	(-0.221,1.379)	(0.103,1.493)
500	2	0.7	0	0.2	0.4863	30.42	0.5579	17.69	0.4489	0.6246	0.6235
					(0.440,0.532)	(15.21)	(0.424,0.703)	(0.835)	(0.046)	(0.364,0.959)	(0.394,0.896)
500	2	0.7	0	-1.0	0.4438	30.33	-0.1476	17.71	0.3589	-0.0676	-0.0992
					(0.390,0.502)	(15.17)	(-0.414,0.031)	(0.823)	(0.031)	(-0.381,0.124)	(-0.266,0.042)
500	10	0.3	0.3	0.2	0.4459	55.39	0.4037	11.59	9.552	0.3788	0.3524
					(0.403,0.491)	(5.539)	(0.303,0.516)	(0.625)	(0.071)	(0.175,0.583)	(0.188,0.538)
500	10	0.7	0	0.2	0.4855	37.29	0.5470	5.088	8.188	0.6107	0.6230
					(0.439,0.531)	(3.729)	(0.414,0.686)	(0.084)	(0.033)	(0.277,0.870)	(0.344,1.005)

Refer to the notes of Table 1.

Table 6: Simulation Results for Data from Heteroscedasticity on Correlated Component with Excluded Instruments

n	K	δ_{u1}	δ_{u2}	$\delta_{\varepsilon 1}$	α_1	β_{OLS}	β_{2SLS}	BP	β_{LB}	F	J	$\beta_{KV,\text{2-step}}$	$\beta_{KV,ML}$
						median	median	median	median	median	median	median	median
						(q10,q90)	(q10,q90)	(BP/K+1)	(q10,q90)	(% F>10)	(% $p < 0.05$)	(q10,q90)	(q10,q90)
500	2	0.5	0.5	0	0.13	0.4628	0.0458	31.90	0.3983	14.04	2.055	0.3671	0.3420
						(0.418,0.511)	(-0.946,0.618)	(10.63)	(0.269,0.527)	(0.748)	(0.121)	(0.083,0.564)	(0.132,0.544)
500	2	0.5	0.5	0.2	0.13	0.4766	0.0434	31.99	0.4890	13.98	3.167	0.4583	0.4607
						(0.431,0.523)	(-0.946,0.618)	(10.66)	(0.355,0.625)	(0.739)	(0.232)	(0.048,0.695)	(0.206,0.716)
500	2	0.5	0.5	-0.2	0.13	0.4541	0.0296	31.88	0.3138	14.07	2.277	0.2612	0.2271
						(0.408,0.499)	(-0.977,0.599)	(10.63)	(0.177,0.447)	(0.750)	(0.139)	(-0.023,0.449)	(0.031,0.411)
500	2	-0.5	-0.5	0.2	0.13	0.4541	0.0314	32.16	0.3150	14.48	2.341	0.2608	0.2280
						(0.408,0.500)	(-1.004,0.608)	(10.72)	(0.180,0.445)	(0.753)	(0.144)	(-0.022,0.448)	(0.0264,0.411)
500	2	-0.5	-0.5	-0.2	0.13	0.4762	0.0139	31.74	0.4851	13.98	3.261	0.4563	0.4555
						(0.431,0.521)	(-1.072,0.582)	(10.58)	(0.356,0.623)	(0.748)	(0.226)	(0.026,0.691)	(0.206,0.706)
500	2	0.5	-0.5	0.2	0.13	0.4775	0.0347	31.66	0.4902	13.83	3.133	0.4615	0.4616
						(0.429,0.525)	(-1.062,0.619)	(10.55)	(0.354,0.623)	(0.737)	(0.229)	(0.046,0.691)	(0.210,0.700)
500	2	0.5	0.5	0.2	0.5	0.4325	0.0031	32.23	0.2641	34.39	10.95	0.0909	0.1315
						(0.388,0.477)	(-0.186,0.151)	(10.74)	(0.128,0.384)	(1.00)	(0.837)	(-0.101,0.270)	(-0.031,0.273)
500	2	0.3	0.3	0.2	0.13	0.4934	0.0087	13.16	0.5274	5.22	3.449	0.3642	0.4039
						(0.446,0.543)	(-1.034,0.567)	(4.386)	(0.296,0.754)	(0.0928)	(0.257)	(-0.289,0.806)	(-0.382,1.001)
500	2	0.7	0	0.2	0.13	0.4830	0.0272	31.09	0.5248	13.66	2.743	0.5224	0.5336
						(0.437,0.528)	(-0.964,0.594)	(10.36)	(0.397,0.657)	(0.738)	(0.185)	(0.136,0.741)	(0.285,0.803)
500	2	0.7	0	-1.0	0.13	0.4410	0.0226	31.20	-0.1151	13.66	1.310	-0.0386	-0.0929
						(0.388,0.498)	(-1.117,0.712)	(10.40)	(-0.341,0.048)	(0.734)	(0.048)	(-0.272,0.138)	(-0.251,0.043)
500	10	0.3	0.3	0.2	0.13	0.4417	0.0393	56.36	0.3901	11.12	11.26	0.3502	0.3264
						(0.400,0.487)	(-1.011,0.614)	(5.124)	(0.291,0.500)	(0.594)	(0.096)	(0.152,0.534)	(0.162,0.494)
500	10	0.7	0	0.2	0.13	0.4830	0.0260	37.99	0.5193	10.82	5.23	0.5218	0.5286
						(0.436,0.529)	(-1.038,0.614)	(3.454)	(0.391,0.657)	(0.087)	(0.071)	(0.190,0.760)	(0.223,0.869)

Refer to the notes of Table 1 and 2.

Table 7: Engel Curve Estimation

	OLS	2SLS1	Lewbel GMM2	Lewbel GMM3	K-V 2-stage	K-V 2-stage	K-V ML	K-V ML
With external IV		x		x		x		x
Use heteroscedasticity			x	x	x	x	x	x
coefficient on Y_2:γ_1	-0.127	-0.0858	-0.0523	-0.0868	-0.0516	-0.0898	-1.231	-0.0793
Asymptotic SE	(0.0084)	(0.0198)	(0.0550)	(0.0180)			(0.284)	(0.0186)
Boostrap SE	[0.0084]	[0.0205]	[0.0609]	[0.0207]	[0.0590]	[0.0219]	[0.6166]	[0.0237]
J statistics			12.48	16.279				
p-value			0.328	0.179				
First-stage F			0.906	2.486				
BP statistic					15.018	22.943		
p-value					0.2404	0.0424		

The data is the same as Lewbel (2012), which is obtained through the Stata package ivreg2h (Baum and Schaffer, 2012). Following Lewbel (2012), all exogenous regressors are used to construct Lewbel type instruments. Asymptotic standard errors are White's robust standard errors for OLS, 2SLS and GMM. K-V ML asymptotic standard errors are obtained by the Hessian of negative of the log-likelihood function. Bootstrap standard errors are obtained by pair boostrap with 500 repetitions. BP test refers to the test for heteroscedasticity of residuals of first stage regression using all first stage regressors. First-stage F is the robust F statistics including the Lewbel-type instruments and excluded instruments.

Unpublished Appendix: Extra Tables

Table 8: Simulation Results for Data from Klein and Vella Form of Heteroscedasticity without Excluded Instruments

n	K	δ_{u1}	δ_{u2}	$\delta_{\varepsilon 1}$	β_{OLS}	BP	β_{LB}	F	J	β_{KV},2-step	$\beta_{KV,ML}$
					median	median	median	median	median	median	median
					(q10,q90)	(BP/K)	(q10,q90)	(% F>10)	(% $p < 0.05$)	(q10,q90)	(q10,q90)
100	2	0.7	0.7	0	0.3545	18.32	0.1813	47.37	0.5465	0.0208	0.0067
					(0.255,0.465)	(9.159)	(0.037,0.326)	(0.991)	(0.0432)	(-0.439,0.234)	(-0.296,0.220)
100	2	0.7	0.7	0.2	0.3693	18.53	0.2153	47.59	0.5972	0.0214	0.0070
					(0.260,0.481)	(9.263)	(0.0501,0.378)	(0.990)	(0.053)	(-0.496,0.293)	(-0.370,0.267)
100	2	0.5	0.5	0.2	0.4307	14.23	0.2600	23.65	0.5717	0.0642	0.0076
					(0.319,0.545)	(7.113)	(0.032,0.473)	(0.859)	(0.051)	(-0.615,0.513)	(-0.754,0.439)
100	2	0.5	0.5	-0.2	0.4089	14.13	0.1672	23.64	0.5262	0.0227	0.0031
					(0.306,0.522)	(7.063)	(-0.033,0.351)	(0.865)	(0.0432)	(-0.467,0.286)	(-0.372,0.267)
100	2	-0.5	-0.5	0.2	0.4115	14.04	0.1665	23.66	0.5161	0.0178	-0.0020
					(0.304,0.521)	(7.020)	(-0.039,0.339)	(0.856)	(0.044)	(-0.509,0.291)	(-0.385,0.266)
100	2	-0.5	-0.5	-0.2	0.4324	14.06	0.2605	22.99	0.5662	0.0641	-0.0006
					(0.321,0.547)	(7.030)	(0.038,0.476)	(0.859)	(0.050)	(-0.632,0.514)	(-0.752,0.440)
100	2	0.5	-0.5	0.2	0.4333	14.11	0.2558	23.65	0.5491	0.0569	-0.0018
					(0.322,0.548)	(7.055)	(0.027,0.476)	(0.861)	(0.051)	(-0.640,0.493)	(-0.738,0.422)
100	2	1.0	0	0.2	0.3704	18.64	0.2253	48.09	0.5717	0.0266	-0.0014
					(0.266,0.482)	(9.319)	(0.062,0.394)	(0.993)	(0.0432)	(-0.549,0.311)	(-0.430,0.284)
100	2	1.0	0	-1.5	0.3242	18.71	-0.0642	48.51	0.4306	-0.0375	0.0024
					(0.211,0.469)	(9.355)	(-0.249,0.041)	(0.992)	(0.0274)	(-0.229,0.102)	(-0.050,0.049)
100	10	0.5	0.5	0.2	0.2226	26.71	0.1303	49.20	8.661	0.0138	0.0071
					(0.130,0.331)	(2.671)	(0.029,0.254)	(0.989)	(0.0118)	(-0.164,0.151)	(-0.204,0.157)
100	10	1.0	0	0.2	0.3694	24.18	0.2521	17.690	8.762	0.0874	0.0131
					(0.262,0.487)	(2.418)	(0.074,0.446)	(0.800)	(0.020)	(-0.236,0.463)	(-0.941,0.520)

Total number of repetition is 5000. The correlation between the first stage and structural error is set at about 0.5. δ_{u1} is the coefficient for the variance function for the first variable of X, while δ_{u2} is the coefficient for all remaining X variables. Similarly, $\delta_{\varepsilon 1}$ is the coefficient of the variance function for the first variable in X, while that for all remaining X variables are zero. BP test is the nR^2 statistic of regressing squares of first-stage residuals on all K exogenous regressors X. BP/K is then having $F_{K,\infty}$ distribution for comparison. F is the first-stage (robust) F statistics using Lewbel generated instruments, and the J statistic is the corresponding statistic under Lewbel GMM method. For estimators, median, 10th and 90th percentiles are presented to understand the bias, spread and symmetry of the estimators.

Table 9: Simulation Results for Data from Klein and Vella Form of Heteroscedasticity with Excluded Instruments

n	K	δ_{u1}	δ_{u2}	$\delta_{\varepsilon 1}$	γ_1	β_{OLS} median (q10,q90)	β_{2SLS} median (q10,q90)	BP median (BP/K+1)	β_{LB} median (q10,q90)	F median (% F>10)	J median (% $p < 0.05$)	β_{KV},2-step median (q10,q90)	$\beta_{KV,ML}$ median (q10,q90)
100	2	0.7	0.7	0	0.3	0.3446	0.0522	19.29	0.1775	34.680	1.563	0.0571	0.0038
						(0.244,0.454)	(-0.923,0.681)	(6.430)	(0.034,0.317)	(0.977)	(0.049)	(-0.218,0.251)	(-0.275,0.216)
100	2	0.7	0.7	0.2	0.3	0.3571	0.0523	19.19	0.2025	34.600	1.738	0.0596	0.0013
						(0.256,0.468)	(-0.983,0.671)	(6.397)	(0.047,0.359)	(0.982)	(0.060)	(-0.263,0.288)	(-0.349,0.252)
100	2	0.5	0.5	0.2	0.3	0.4180	0.0240	14.75	0.2367	18.24	1.789	0.0941	-0.0034
						(0.311,0.527)	(-0.940,0.614)	(4.917)	(0.016,0.439)	(0.810)	(0.064)	(-0.306,0.409)	(-0.557,0.367)
100	2	0.5	0.5	-0.2	0.3	0.3956	0.0276	14.76	0.1605	17.97	1.605	0.0640	0.0061
						(0.295,0.501)	(-1.004,0.620)	(4.92)	(-0.031,0.329)	(0.816)	(0.051)	(-0.266,0.290)	(-0.342,0.250)
100	2	-0.5	-0.5	0.2	0.3	0.3945	0.0037	14.68	0.1551	18.12	1.543	0.0552	-0.0068
						(0.295,0.502)	(-1.040,0.612)	(4.893)	(-0.038,0.325)	(0.807)	(0.049)	(-0.281,0.286)	(-0.346,0.243)
100	2	-0.5	-0.5	-0.2	0.3	0.4174	0.0343	14.91	0.2321	18.23	1.731	0.0897	-0.0027
						(0.309,0.531)	(-0.971,0.626)	(4.969)	(0.021,0.433)	(0.81)	(0.062)	(-0.309,0.402)	(-0.548,0.366)
100	2	0.5	-0.5	0.2	0.3	0.4152	0.0121	14.73	0.2356	18.12	1.784	0.0914	-0.0072
						(0.308,0.524)	(-1.059,0.595)	(4.909)	(0.021,0.427)	(0.813)	(0.065)	(-0.325,0.395)	(-0.582,0.348)
100	2	0.5	0.5	0.2	1.0	0.3120	0.0020	14.84	0.1050	37.08	2.434	0.0236	0.0024
						(0.217,0.405)	(-0.204,0.173)	(4.946)	(-0.049,0.247)	(0.998)	(0.105)	(-0.161,0.179)	(-0.189,0.163)
100	2	1.0	0	0.2	0.3	0.3627	0.0511	19.21	0.2204	35.12	1.713	0.0686	0.0024
						(0.258,0.474)	(-1.003,0.684)	(6.405)	(0.059,0.375)	(0.986)	(0.065)	(-0.279,0.304)	(-0.388,0.274)
100	2	1.0	0	-1.5	0.3	0.3121	0.0450	19.37	-0.0477	35.22	1.456	-0.0174	0.0019
						(0.197,0.457)	(-1.565,1.261)	(6.458)	(-0.226,0.055)	(0.983)	(0.029)	(-0.200,0.131)	(-0.051,0.049)
100	10	0.5	0.5	0.2	0.3	0.2226	0.0808	27.37	0.1346	45.97	9.693	0.0286	0.0052
						(0.132,0.324)	(-0.984,0.901)	(2.488)	(0.031,0.255)	(0.988)	(0.014)	(-0.127,0.163)	(-0.194,0.156)
100	10	1.0	0	0.2	0.3	0.3602	0.0523	24.69	0.2410	16.58	9.880	0.1028	-0.0094
						(0.251,0.474)	(-1.038,0.753)	(2.245)	(0.058,0.422)	(0.779)	(0.023)	(-0.165,0.393)	(-0.748,0.418)

Refer to the notes of Table 1. The first stage coefficient on excluded instrument is set so that the first stage F is about 3. ($\alpha_1 = \alpha_2 = 1$) The first-stage (robust) F statistics include also the excluded instrument z.

Table 10: Simulation Results for Data from Lewbel Form of Heteroscedasticity without Excluded Instruments

n	K	δ_{u1}	δ_{u2}	$\delta_{\varepsilon 1}$	β_{OLS}	BP	β_{LB}	F	J	β_{KV},2-step	$\beta_{KV,ML}$
					median	median	median	median	median	median	median
					(q10,q90)	(BP/K)	(q10,q90)	(% F>10)	(% $p < 0.05$)	(q10,q90)	(q10,q90)
100	2	0.9	0.9	0	0.3236	14.89	0.0043	23.25	0.4735	-0.3192	-0.5251
					(0.200,0.453)	(7.446)	(-0.224,0.189)	(0.798)	(0.037)	(-1.165,0.090)	(-1.955,-0.039)
100	2	0.9	0.9	0.2	0.3214	15.04	-0.0007	23.44	0.4359	-0.3410	-0.6708
					(0.199,0.447)	(7.521)	(-0.239,0.197)	(0.804)	(0.041)	(-1.276,0.119)	(-2.547,0.0199)
100	2	0.9	0.9	-0.2	0.3203	14.95	0.0025	23.35	0.4431	-0.2758	-0.3875
					(0.199,0.448)	(7.475)	(-0.213,0.181)	(0.804)	(0.036)	(-1.067,0.088)	(-1.338,-0.003)
100	2	-0.9	-0.9	0.2	0.3200	15.21	0.0059	23.52	0.4740	-0.2644	-0.3743
					(0.201,0.444)	(7.606)	(-0.222,0.176)	(0.798)	(0.036)	(-1.060,0.080)	(-1.313,-0.011)
100	2	-0.9	-0.9	-0.2	0.3191	14.98	-0.0012	24.04	0.4650	-0.3397	-0.6983
					(0.199,0.449)	(7.489)	(-0.227,0.196)	(0.808)	(0.043)	(-1.230,0.127)	(-2.565,-0.016)
100	2	0.9	-0.9	0.2	0.3243	14.86	0.0015	22.41	0.4904	-0.3369	-0.6606
					(0.200,0.455)	(7.432)	(-0.230,0.202)	(0.792)	(0.035)	(-1.233,0.154)	(-2.570,0.011)
100	2	0.6	0.6	0.2	0.4150	8.720	0.0256	6.608	0.4502	-0.1775	-0.5596
					(0.295,0.541)	(4.360)	(-0.369,0.334)	(0.346)	(0.043)	(-1.136,1.252)	(-2.812,2.796)
100	2	1.2	0	0.2	0.3395	14.06	-0.0036	18.76	0.4598	-0.3548	-0.7772
					(0.219,0.469)	(7.032)	(-0.259,0.218)	(0.742)	(0.038)	(-1.362,0.252)	(-2.795,0.910)
100	2	1.2	0	-1.5	0.3339	14.04	0.0217	19.83	0.4726	0.0573	-0.0025
					(0.210,0.476)	(7.02)	(-0.253,0.245)	(0.740)	(0.032)	(-0.287,0.376)	(-0.196,0.177)
100	10	0.6	0.6	0.2	0.1777	24.25	0.0151	32.78	8.605	-0.1527	-0.3222
					(0.071,0.304)	(2.425)	(-0.115,0.148)	(0.898)	(0.0172)	(-0.414,0.027)	(-1.457,-0.007)
100	10	1.2	0	0.2	0.3383	20.04	0.0979	8.215	8.928	-0.0900	-0.4132
					(0.209,0.470)	(2.004)	(-0.144,0.358)	(0.403)	(0.0234)	(-0.449,0.645)	(-2.992,2.795)

Refer to the notes of Table 1.

Table 11: Simulation Results for Data from Lewbel Form of Heteroscedasticity with Excluded Instruments

n	K	δ_{u1}	δ_{u2}	$\delta_{\varepsilon 1}$	α_1	β_{OLS}	β_{2SLS}	BP	β_{LB}	F	J	$\beta_{KV,\text{2-step}}$	$\beta_{KV,ML}$
						median	median	median	median	median	median	median	median
						(q10,q90)	(q10,q90)	(BP/K+1)	(q10,q90)	(% F>10)	(% $p<0.05$)	(q10,q90)	(q10,q90)
100	2	0.9	0.9	0	0.3	0.3113	0.0294	15.65	0.0072	17.91	1.414	-0.1304	-0.3916
						(0.194,0.430)	(-0.935,0.685)	(5.216)	(-0.200,0.188)	(0.749)	(0.046)	(-0.519,0.125)	(-1.408,-0.008)
100	2	0.9	0.9	0.2	0.3	0.3119	0.0353	15.58	0.0039	17.83	1.453	-0.1500	-0.4959
						(0.190,0.437)	(-0.951,0.708)	(5.195)	(-0.212,0.195)	(0.744)	(0.045)	(-0.569,0.152)	(-1.993,0.009)
100	2	0.9	0.9	-0.2	0.3	0.3112	0.0490	15.75	0.0144	18.34	1.427	-0.1015	-0.2968
						(0.196,0.433)	(-0.941,0.718)	(5.251)	(-0.182,0.184)	(0.762)	(0.047)	(-0.483,0.139)	(-1.027,0.0306)
100	2	-0.9	-0.9	0.2	0.3	0.3166	0.0259	15.60	0.0106	17.71	1.409	-0.1093	-0.3083
						(0.200,0.433)	(-0.985,0.696)	(5.201)	(-0.191,0.183)	(0.748)	(0.039)	(-0.488,0.132)	(-1.014,0.0281)
100	2	-0.9	-0.9	-0.2	0.3	0.3128	0.0292	15.74	0.0103	17.75	1.447	-0.1407	-0.4920
						(0.193,0.436)	(-1.080,0.669)	(5.246)	(-0.209,0.192)	(0.749)	(0.046)	(-0.549,0.157)	(-2.080,-0.007)
100	2	0.9	-0.9	0.2	0.3	0.3159	0.0459	15.58	0.0093	18.09	1.512	-0.1438	-0.4857
						(0.193,0.444)	(-1.022,0.686)	(5.194)	(-0.208,0.200)	(0.745)	(0.049)	(-0.546,0.162)	(-1.997,0.001)
100	2	0.9	0.9	0.2	1.0	0.2433	-0.0006	15.48	0.0010	34.32	1.397	-0.056	-0.1125
						(0.144,0.342)	(-0.206,0.177)	(5.159)	(-0.148,0.139)	(0.989)	(0.040)	(-0.234,0.109)	(-0.355,0.081)
100	2	0.6	0.6	0.2	0.3	0.4004	0.0329	9.716	0.0447	6.719	1.421	-0.0378	-0.3822
						(0.283,0.522)	(-0.965,0.624)	(3.239)	(-0.272,0.317)	(0.308)	(0.045)	(-0.523,0.503)	(-2.484,0.381)
100	2	1.2	0	0.2	0.3	0.3318	0.0333	14.96	0.0136	15.49	1.464	-0.1346	-0.5268
						(0.207,0.458)	(-1.006,0.651)	(4.985)	(-0.221,0.221)	(0.689)	(0.048)	(-0.550,0.209)	(-2.286,0.048)
100	2	1.2	0	-1.5	0.3	0.3260	0.0310	14.88	0.0339	15.52	1.408	0.0694	-0.0025
						(0.204,0.462)	(-1.284,0.984)	(4.96)	(-0.216,0.248)	(0.694)	(0.035)	(-0.210,0.369)	(-0.193,0.174)
100	10	0.6	0.6	0.2	0.3	0.1691	0.0445	25.23	0.0153	31.33	9.664	-0.1057	-0.2828
						(0.065,0.294)	(-1.023,0.912)	(2.294)	(-0.106,0.148)	(0.894)	(0.018)	(-0.320,0.054)	(-1.219,-0.005)
100	10	1.2	0	0.2	0.3	0.3307	0.0320	20.81	0.0984	8.308	9.971	-0.0419	-0.4056
						(0.206,0.461)	(-0.983,0.718)	(1.892)	(-0.134,0.338)	(0.401)	(0.025)	(-0.327,0.405)	(-2.621,0.834)

Refer to the notes of Table 1 and 2.

Table 12: Simulation Results for Data from Heteroscedasticity on Correlated Component without Excluded Instruments

n	K	δ_{u1}	δ_{u2}	$\delta_{\varepsilon 1}$	β_{OLS}	BP	β_{LB}	F	J	β_{KV},2-step	$\beta_{KV,ML}$
					median	median	median	median	median	median	median
					(q10,q90)	(BP/K)	(q10,q90)	(% F>10)	(% $p < 0.05$)	(q10,q90)	(q10,q90)
100	2	0.9	0.9	0	0.3865	14.81	0.3006	23.65	0.5023	0.2732	0.2451
					(0.290,0.486)	(7.407)	(0.140,0.491)	(0.796)	(0.047)	(-0.188,0.585)	(-0.003,0.530)
100	2	0.9	0.9	0.2	0.4049	14.84	0.3549	23.31	0.5789	0.3422	0.3216
					(0.310,0.501)	(7.421)	(0.187,0.551)	(0.794)	(0.051)	(-0.118,0.682)	(0.070,0.647)
100	2	0.9	0.9	-0.2	0.3706	15.03	0.2574	23.95	0.5418	0.2204	0.1899
					(0.270,0.474)	(7.514)	(0.100,0.452)	(0.804)	(0.052)	(-0.218,0.535)	(-0.051,0.457)
100	2	-0.9	-0.9	0.2	0.3714	14.82	0.2563	23.48	0.5451	0.2187	0.1889
					(0.274,0.472)	(7.409)	(0.100,0.446)	(0.796)	(0.051)	(-0.246,0.520)	(-0.052,0.453)
100	2	-0.9	-0.9	-0.2	0.4039	15.01	0.3531	23.56	0.5796	0.3332	0.3151
					(0.309,0.501)	(7.505)	(0.188,0.551)	(0.807)	(0.060)	(-0.126,0.680)	(0.055,0.637)
100	2	0.9	-0.9	0.2	0.4028	15.05	0.3485	23.65	0.5635	0.3282	0.3135
					(0.308,0.501)	(7.524)	(0.187,0.545)	(0.806)	(0.050)	(-0.106,0.667)	(0.059,0.628)
100	2	0.6	0.6	0.2	0.4655	8.817	0.4704	6.671	0.5321	0.4651	0.4646
					(0.364,0.566)	(4.409)	(0.214,0.768)	(0.346)	(0.0498)	(-0.174,1.055)	(-0.053,1.067)
100	2	1.2	0	0.2	0.4235	14.18	0.3954	20.07	0.5248	0.3974	0.3813
					(0.327,0.520)	(7.091)	(0.225,0.602)	(0.740)	(0.043)	(-0.029,0.769)	(0.101,0.739)
100	2	1.2	0	-1.5	0.3441	14.13	-0.0312	19.71	0.4427	-0.0184	-0.0537
					(0.212,0.488)	(7.067)	(-0.344,0.183)	(0.740)	(0.035)	(-0.411,0.295)	(-0.275,0.129)
100	10	0.6	0.6	0.2	0.2674	24.40	0.2060	33.75	8.912	0.1288	0.1269
					(0.167,0.371)	(2.44)	(0.089,0.356)	(0.901)	(0.016)	(-0.080,0.377)	(-0.030,0.387)
100	10	1.2	0	0.2	0.4219	20.16	0.4003	8.351	8.660	0.3505	0.3674
					(0.324,0.524)	(2.016)	(0.212,0.607)	(0.415)	(0.014)	(-0.013,0.809)	(-0.262,1.251)

Refer to the notes of Table 1.

Table 13: Simulation Results for Data from Heteroscedasticity on Correlated Component with Excluded Instruments

n	K	δ_{u1}	δ_{u2}	$\delta_{\varepsilon 1}$	α_1	β_{OLS}	β_{2SLS}	BP	β_{LB}	F	J	$\beta_{KV,\text{2-step}}$	$\beta_{KV,ML}$
						median	median	median	median	median	median	median	median
						(q10,q90)	(q10,q90)	(BP/K+1)	(q10,q90)	(% F>10)	(% $p<0.05$)	(q10,q90)	(q10,q90)
100	2	0.9	0.9	0	0.3	0.3748	0.0504	15.38	0.2842	17.73	1.765	0.2381	0.2074
						(0.279,0.474)	(-0.965,0.644)	(5.127)	(0.131,0.456)	(0.742)	(0.071)	(-0.138,0.497)	(-0.048,0.472)
100	2	0.9	0.9	0.2	0.3	0.3944	0.0421	15.90	0.3269	18.29	2.026	0.2780	0.2590
						(0.301,0.488)	(-1.070,0.625)	(5.302)	(0.168,0.499)	(0.758)	(0.098)	(-0.123,0.542)	(-0.013,0.531)
100	2	0.9	0.9	-0.2	0.3	0.3626	0.0507	15.65	0.2452	17.72	1.777	0.2012	0.1614
						(0.266,0.463)	(-1.028,0.683)	(5.217)	(0.085,0.422)	(0.747)	(0.067)	(-0.154,0.453)	(-0.089,0.401)
100	2	-0.9	-0.9	0.2	0.3	0.3629	0.0346	15.87	0.2464	18.07	1.728	0.2005	0.1622
						(0.266,0.461)	(-0.927,0.673)	(5.29)	(0.090,0.421)	(0.761)	(0.075)	(-0.149,0.444)	(-0.080,0.400)
100	2	-0.9	-0.9	-0.2	0.3	0.3924	0.0472	15.62	0.3283	17.91	2.106	0.2750	0.2571
						(0.299,0.485)	(-1.006,1.163)	(5.206)	(0.171,0.501)	(0.747)	(0.086)	(-0.123,0.534)	(-0.013,0.527)
100	2	0.9	-0.9	0.2	0.3	0.3931	0.0465	15.64	0.3254	17.84	2.049	0.2670	0.2566
						(0.299,0.487)	(-0.950,0.674)	(5.213)	(0.161,0.510)	(0.747)	(0.098)	(-0.135,0.544)	(-0.024,0.535)
100	2	0.9	0.9	0.2	1.0	0.3040	-0.0035	15.64	0.1840	34.23	4.332	0.0690	0.0935
						(0.222,0.390)	(-0.217,0.162)	(5.213)	(0.025,0.317)	(0.992)	(0.315)	(-0.133,0.243)	(-0.079,0.240)
100	2	0.6	0.6	0.2	0.3	0.4489	0.0262	9.633	0.3923	6.830	2.169	0.2965	0.2880
						(0.351,0.550)	(-0.961,0.586)	(3.211)	(0.122,0.632)	(0.309)	(0.099)	(-0.222,0.666)	(-0.343,0.724)
100	2	1.2	0	0.2	0.3	0.4127	0.0499	14.66	0.3627	15.08	2.051	0.3044	0.2955
						(0.320,0.507)	(-1.021,0.647)	(4.885)	(0.198,0.544)	(0.680)	(0.096)	(-0.106,0.583)	(-0.012,0.598)
100	2	1.2	0	-1.5	0.3	0.3291	0.0226	14.69	-0.0203	15.16	1.409	0.0099	-0.0522
						(0.205,0.470)	(-1.299,1.046)	(4.896)	(-0.282,0.185)	(0.683)	(0.039)	(-0.290,0.280)	(-0.263,0.125)
100	10	0.6	0.6	0.2	0.3	0.2633	0.0791	25.04	0.1989	31.81	9.947	0.1316	0.1222
						(0.164,0.366)	(-0.965,0.848)	(2.276)	(0.084,0.339)	(0.900)	(0.022)	(-0.064,0.342)	(-0.039,0.343)
100	10	1.2	0	0.2	0.3	0.4126	0.0551	20.90	0.3737	8.301	10.18	0.2919	0.2891
						(0.316,0.511)	(-0.989,0.678)	(1.900)	(0.196,0.563)	(0.405)	(0.027)	(-0.041,0.664)	(-0.368,0.870)

Refer to the notes of Table 1 and 2.