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

Chau, Tak Wai

The Chinese University of Hong Kong, Shanghai University of Finance and Economics

July 2015

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

Tak Wai Chau*

Shanghai University of Finance and Economics

The Chinese University of Hong Kong

March 2016

Abstract

Recent literature propose estimators that utilize heteroscedasticity of the error terms to identify the coefficient of the endogenous regressor without using excluded instruments. The assumed forms of heteroscedasticity differ across estimators. This study investigates the robustness of the two most popular estimators under different forms of heteroscedasticity through simulations. 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 Alkaline Information Criteria (AIC) to distinguish these two models. The simulation results show that it has good performance.

JEL codes: C13, C31, C36

Keywords: Instrumental Variable Estimation, Endogeneity, Heteroscedasticity, Misspecification, Maximum Likelihood

*1. 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. 2. Current Affiliation: Hong Kong Institute of Asia-Pacific Studies, The Chinese University of Hong Kong, Shatin, the New Territories, Hong Kong. Email: twchau@cuhk.edu.hk.

1 Introduction

One common difficulty for empirical researchers in consistently estimating the coefficient of a regressor of interest is that the regressor may be endogenous while exogenous instruments are not available. Klein and Vella (2010) and Lewbel (2012) respectively introduce methods to identify the coefficients using heteroscedasticity of the error terms, even without excluded instruments. Klein and Vella (2010) assume that heteroscedasticity is multiplicative to the whole structural and first-stage error terms with a constant correlation coefficient. Lewbel (2012) assumes that the covariance, instead of correlation, of these error terms is a constant, which essentially requires that heteroscedasticity only exists in the uncorrelated components of these error terms. However, there is no straightforward way to justify a priori which form of heteroscedasticity is true. This study investigates whether the estimators are robust to misspecification of heteroscedasticity and whether diagnostic tests are powerful enough to distinguish them. I also propose using the maximum likelihood method to estimate the two models, and the Alkaline Information Criteria (AIC) to choose between models.

The two estimators, especially the Lewbel estimator, are becoming more popular because they are easy to implement[1] and heteroscedasticity is common in data. Most of these studies use the estimators for robustness check. Not all of them have a priori justification for the form of heteroscedasticity assumed.[2]

This study simulates data from a standard linear model with one endogenous regressor from the forms of heteroscedasticity related to Klein and Vella (2010) and Lewbel (2012) specifications, estimates the parameters with 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 forms of heteroscedasticity do not match. The power of the overidentification test 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 concludes.

[1]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).

[2]Among these studies, only Emran and Shilpi (2012) and Millimet and Roy (2015) provide some justifications.

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 may not exist in this setting.

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

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

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

The third condition essentially requires that 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 cannot enter through the correlated component or common factor of the error terms. This is the major distinction from the Klein-Vella (2010) method.

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

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

So, 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 can be used, 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[3] as

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

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

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

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 also introduces the maximum likelihood framework, which allows us to estimate the two models in one framework, and use the corresponding Alkaline 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). For the Klein and Vella (2010) estimator, constant correlation θ_{KV} implies

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

We can use the Alkaline Information Criteria (AIC) to choose the model that gives a large value.

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

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:

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

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.[4] The correlation between the first-

[4]Normalized Chi-square (5) errors are also considered in the online appendix.

stage and structural error terms is generated by the common factor.

Simulated data from the above models are used to estimate the structural parameters β using the methods described above.[5] Here I use all variables X as Z_2 variables.[6] 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.[7] I present the results for the J statistics to investigate the effectiveness of overidentifying tests to detect the wrong form of heteroscedasticity.

In this study, I take $\beta_1 = 0$, so the value of mean and median of bootstrap samples represents the corresponding bias. $\beta_0 = \alpha_0 = 0$, $\beta_{2k} = \gamma_{2k} = 1$ 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. The number of replications is 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 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[8], while the 2-step and ML Klein-Vella estimators have medians close to their true value. If we choose the estimator according to AIC, the correct rates are usually higher than 0.5, though not very close to 1 in the cases considered. The resulting estimator has a lower bias than the wrong ones. The overidentification J test has low power in detecting the misspecification of the form of heteroscedasticity, with rejection rate below 40%.

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 again higher than 0.5, and the resulting estimator is closer to be median unbiased. The results for J test agrees with the nominal power as the Lewbel is the true data

[5] I use the 'ivlewbel' package by Fernihough (2014) to estimate with the Lewbel estimator in R, while other estimators are coded in R by the author.

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

[7] Mean and standard deviation are not used since some estimators may not have moments.

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

generating process.

4 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), and 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. Further research may focus on studying the identification conditions for the cases in between these two models. For example, when the covariance or correlation of the two error terms depends on some variables instead of a constant, what are restrictions required for identification?

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

[4] 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.

[5] 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

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

[7] 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.

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

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

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

Tables

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)	β_{LB} median (q10,q90)	J median (% $p < 0.05$)	$\beta_{KV,\text{2-step}}$ median (q10,q90)	$\beta_{LB,ML}$ median (q10,q90)	$\beta_{KV,ML}$ median (q10,q90)	β_{AIC} median (q10,q90)	AIC correct rate
500	3	0.4	0.4	0.3	0.4362	0.2609	2.690	0.0206	0.2873	0.0076	0.0761	0.698
					(0.385,0.485)	(0.162,0.360)	(0.160)	(-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	0.1490	2.498	0.0073	0.1566	-0.0107	0.0231	0.742
					(0.363,0.459)	(0.066,0.223)	(0.149)	(-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	0.2817	4.681	0.0268	0.3198	0.0050	0.0173	0.878
					(0.400,0.507)	(0.177,0.393)	(0.362)	(-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	0.3313	2.872	0.0597	0.3323	0.0438	0.1171	0.675
					(0.431,0.531)	(0.178,0.477)	(0.190)	(-0.428,0.362)	(0.165,0.489)	(-0.512,0.387)	(-0.457,0.398)	
500	3	0.7	0	0.3	0.4443	0.3177	1.397	0.0385	0.3423	0.0198	0.2465	0.486
					(0.394,0.494)	(0.211,0.417)	(0.047)	(-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	0.4054	1.420	0.1584	0.4284	0.0957	0.3361	0.617
					(0.414,0.525)	(0.283,0.523)	(0.043)	(-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	0.2247	9.760	0.0508	0.2504	0.0141	0.1209	0.587
					(0.362,0.462)	(0.139,0.307)	(0.068)	(-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	0.3217	8.587	0.1105	0.3135	0.1127	0.2681	0.393
					(0.394,0.495)	(0.210,0.431)	(0.034)	(-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	0.4039	8.743	0.2049	0.4249	0.2660	0.3553	0.535
					(0.415,0.524)	(0.276,0.539)	(0.038)	(-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	0.2604	4.249	0.0156	0.2924	0.0044	0.0207	0.854
					(0.400,0.468)	(0.192,0.321)	(0.333)	(-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	0.3178	1.467	0.0202	0.3463	0.0002	0.1273	0.592
					(0.406,0.479)	(0.244,0.390)	(0.044)	(-0.319,0.223)	(0.285,0.409)	(-0.246,0.203)	(-0.197,0.379)	

The number of repetition is 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.

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

n	K	δ_{u1}	δ_{u2}	$\delta_{\varepsilon1}$	β_{OLS} median (q10,q90)	β_{LB} median (q10,q90)	J median (% $p < 0.05$)	$\beta_{KV,\text{2-step}}$ median (q10,q90)	$\beta_{LB,ML}$ median (q10,q90)	$\beta_{KV,ML}$ median (q10,q90)	β_{AIC} median (q10,q90)	AIC correct rate
500	3	0.5	0.5	0.3	0.4086	0.0026	1.228	-0.5354	-0.0114	-0.7692	-0.0398	0.846
					(0.354,0.462)	(-0.145,0.134)	(0.032)	(-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	0.0098	1.311	-0.2887	0.0152	-0.3349	-0.0878	0.579
					(0.355,0.460)	(-0.129,0.127)	(0.040)	(-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	0.0066	1.274	-0.5598	0.0065	-0.6655	-0.0026	0.930
					(0.355,0.461)	(-0.162,0.151)	(0.030)	(-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	0.0306	1.259	-0.1969	-0.0304	-0.1251	-0.0525	0.720
					(0.414,0.520)	(-0.248,0.259)	(0.045)	(-0.956,1.339)	(-0.320,0.256)	(-1.228,2.192)	(-2.259,0.260)	
500	3	0.8	0	0.3	0.4231	0.0089	1.172	-0.5741	-0.0294	-1.0829	-0.0549	0.825
					(0.366,0.477)	(-0.164,0.160)	(0.032)	(-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	0.0146	1.277	-0.5613	-0.0049	2.0812	0.0284	0.752
					(0.362,0.479)	(-0.182,0.177)	(0.042)	(-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	0.0238	7.970	-0.2512	0.0220	-0.3710	-0.0054	0.795
					(0.337,0.447)	(-0.108,0.149)	(0.029)	(-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	0.0523	7.977	-0.2474	0.0146	-0.3215	0.0051	0.821
					(0.368,0.478)	(-0.114,0.214)	(0.033)	(-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	0.0566	7.989	-0.1920	0.0176	1.1201	0.0444	0.786
					(0.363,0.479)	(-0.133,0.234)	(0.038)	(-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	0.0021	1.234	-0.6784	-0.0083	-0.9981	-0.0165	0.948
					(0.369,0.449)	(-0.107,0.100)	(0.027)	(-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	0.0074	1.239	-0.8069	-0.0151	-1.1894	-0.0280	0.922
					(0.383,0.460)	(-0.115,0.112)	(0.030)	(-1.583,-0.320)	(-0.167,0.102)	(-1.280,2.185)	(-0.248,0.096)	

Refer to the notes for Table 1.

Online Appendix

A Identification of the Lewbel (2012) Estimator

A.1 More details on identification conditions

To clarify the requirements for identification, we decompose the two error terms as follows:

$$\begin{aligned} \varepsilon &= e_1 + v_1 \\ u &= e_2 + v_2 \end{aligned} \tag{1}$$

where the correlation between u and ε is captured by the correlated component e_1, e_2 so that $cov(e_1, e_2|W) \neq 0$ whenever $\rho_{12} = cov(\varepsilon, u|W) \neq 0$, while $cov(v_1, v_2|W) = 0$ and $cov(e_i, v_j|W) = 0$ for all $i, j = 1, 2$.[1] Denote $z_2 = Z_2 - \mu_2$ to simplify notation. Then, the third condition requires

$$\begin{aligned} E(z_2 u\varepsilon) &= E(z_2 E(u\varepsilon|W)) \\ &= E(z_2 cov(u, \varepsilon|W)) \\ &= E(z_2 cov(e_1, e_2|W)) \\ &= E(z_2 \rho_{12}(z_2)\sigma_1(z_2)\sigma_2(z_2)) = 0 \end{aligned} \tag{2}$$

where $\sigma_j = var(e_j)$.

Since the covariance between u and ε depends only on the correlated components e_1 and e_2, the lack of dependence of the covariance between u and ε on z_2 means that the covariance between the correlated components e_1 and e_2 should not depend on z_2, which also means that there should not be heteroscedasticity in these correlated components. Therefore, the required heteroscedasticity should only be associated to the uncorrelated component v_2 in order to satisfy the third and fourth conditions.[2]

A.2 Common Factor Model

In Lewbel (2012), a common factor model is used as an example. This can be represented by the setting of this paper with $e_1 = \alpha_1\theta$ and $e_2 = \alpha_2\theta$ for some α_1, α_2. When θ is heteroscedastic in variables in Z_2, the Lewbel estimator is also biased since $\sigma_j = \alpha_j \sigma_\theta^2(z_2)$.

Similarly, the identification condition is violated if the common factor has a loading that varies with z_2, or equivalently, the heteroscedasticity can be expressed in terms of factor loading,

$$\begin{aligned} E(z_2 E(e_1 e_2|W)) &= E(z_2(a_1(z_2)\theta)(a_2(z_2)\theta)) \\ &= E(z_2 a_1(z_2) a_2(z_2) E(\theta^2|z_2)) \neq 0 \end{aligned} \tag{3}$$

[1] 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.

[2] This point is only 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.

When θ is homoscedastic, with $E(\theta^2|z_2) = \sigma_\theta^2$ not depending on z_2, the term $E(z_2 a_1(z_2) a_2(z_2))$ still involves moments of z_2 other than the first moment, in which some of them are likely to be non-zero.

A.3 A Simplified Case

To illustrate the conditions required for consistency of the Lewbel (2012) estimator, here I consider a simplified case where there is no covariates X, and y_1 and y_2 are mean zero and 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

$$\begin{aligned} y_1 &= y_2\beta + \varepsilon \qquad (4) \\ y_2 &= u \end{aligned}$$

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

$$\beta_{LB} = \frac{cov((Z_2-\mu_2)u, y_1)}{cov((Z_2-\mu_2)u, y_2)} = \frac{E((Z_2-\mu_2)uy_1)}{E((Z_2-\mu_2)uy_2)} = \frac{E((Z_2-\mu_2)E(uy_1|Z_2))}{E((Z_2-\mu_2)E(uy_2|Z_2))} \qquad (5)$$

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

$$\begin{aligned} E((Z_2-\mu_2)E(uy_1|Z_2)) &= p(1-p)E(uy_1|z_2=1) + (1-p)(-p)E(uy_1|Z_2=0) \qquad (6) \\ &= p(1-p)\left[E(uy_1|z_2=1) - E(uy_1|Z_2=0)\right] \end{aligned}$$

Similarly, the denominator can also be expressed as

$$\begin{aligned} E((Z_2-\mu_2)E(uy_2|Z_2)) &= p(1-p)\left[E(uy_2|Z_2=1) - E(uy_2|Z_2=0)\right] \\ &= Var(u|Z_1=1) - Var(u|Z_1=0) \qquad (7) \end{aligned}$$

since $u = y_2$.

As a result, the Lewbel estimator has a probability limit

$$\beta_{LB} = \frac{E(uy_1|Z_2=1) - E(uy_1|Z_2=0)}{E(uy_2|Z_2=1) - E(uy_2|Z_2=0)} = \frac{E(uy_1|Z_2=1) - E(uy_1|Z_2=0)}{Var(u|Z_1=1) - Var(u|Z_1=0)} \qquad (8)$$

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

$$\begin{aligned} E(uy_1|Z_2=1) - E(uy_1|Z_2=0) &= [E(uy_2|Z_2=1) - E(uy_2|Z_2=0)]\beta \qquad (9) \\ &\quad + E(u\varepsilon|Z_2=1) - E(u\varepsilon|Z_2=0) \\ &= [Var(u|Z_1=1) - Var(u|Z_1=0)]\beta \\ &\quad + E(e_1e_2|Z_2=1) - E(e_1e_2|Z_2=0) \end{aligned}$$

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

$$\begin{aligned} Var(u|Z_1 = 1) - Var(u|Z_1 = 0) &= Var(e_2 + v_2|Z_2 = 1) - Var(e_2 + v_2|Z_2 = 0) \\ &= [Var(e_2|Z_2 = 1) - Var(e_2|Z_2 = 0)] \\ &\quad +[Var(v_2|Z_1 = 1) - Var(v_2|Z = 0)] \\ &= [Var(e_2|Z_2 = 1) - Var(e_2|Z_2 = 0)] \end{aligned} \tag{10}$$

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, the probability limit can be expressed as

$$\beta_{LB} = \frac{E(uy_1|Z_2 = 1) - E(uy_1|Z_2 = 0)}{Var(u|Z_1 = 1) - Var(u|Z_1 = 0)} = \beta + \frac{E(e_1e_2|Z_2 = 1) - E(e_1e_2|Z_2 = 0)}{Var(v_2|Z_2 = 1) - Var(v_2|Z_2 = 0)} \tag{11}$$

This expression shows that for consistency of the estimator, the variances of the first-stage error u 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 correlated components e_1 and e_2 have to be the same for the two groups.

We may also assess the direction of bias with (11) if there is a violation of the identification condition. The numerator of the bias is given by

$$E(e_1e_2|Z_2 = 1) - E(e_1e_2|Z_2 = 0) \quad = \quad \rho_1\sigma_{e_1,1}\sigma_{e_2,1} - \rho_2\sigma_{e_1,0}\sigma_{e_2,0} \tag{12}$$

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

$$\begin{aligned} Var(u|Z_1 = 1) - Var(u|Z_1 = 0) &= Var(e_2 + v_2|Z_2 = 1) - Var(e_2 + v_2|Z_2 = 0) \\ &= (\sigma^2_{e_2,1} - \sigma^2_{e_2,0}) + (\sigma^2_{v_2,1} - \sigma^2_{v_2,0}) \end{aligned} \tag{13}$$

As a whole, the sign of the bias depends on how the variances of correlated and idiosyncratic components are correlated to z. Under the assumptions of the Klein and Vella (2010) estimator, that ρ is a constant, then the numerator of the bias term 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 the sign of the numerator of bias is given by the sign of the product of ρ and the correlation between σ_{e_2} and Z_2. However, 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 the heteroscedasticity in e_1 are correlated to Z_2 in a different direction from that for e_2, the sign of bias will then depend on the resulting sign of the difference in (12). Therefore in general, we cannot sign the direction of bias.

B Details of Implementation for Klein and Vella (2010) Estimator

Following Farre, Klein and Vella (2013), the two-step approach in this paper is estimated in the following steps:

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

2. Regress $ln(\hat{u}^2)$ on X (and Z if available) and obtain the coefficient $\hat{\delta}_u$. Construct $\hat{S_u} = exp(Z_{2i}\hat{\delta_u})$. [3]

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, ρ and δ_ε by choosing β_1, β_2 and ρ to minimize

$$\sum_{i=1}^{n}\left[y_{1i} - \beta_1 y_{2i} - X_i'\beta_2 - \rho\frac{\sqrt{exp(Z_{2i}'\hat{\delta}_\varepsilon)}}{\sqrt{exp(Z_{2i}'\hat{\delta}_u)}}\hat{u}_i\right]^2 \tag{14}$$

   and for each set of (β_1,β_2,ρ), we regress $ln(\hat{\varepsilon}_i^2)$ on X, where $\hat{\varepsilon}_i = y_{1i} - \beta_1 y_{2i} - X_i'\beta_2$ to obtain $\hat{\delta}_\varepsilon$. Then, put back into the expression (14) to calculate the value of the objective function. [4]

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

In this paper, this estimator is called the two-step estimator because we estimate the first-stage equation first and then the structural equation separately. Although not considered by Klein and Vella (2010), it is straight-forward to include excluded instruments Z in steps 1 and 2 above. One may also freely include this Z in the variance functions for the two error terms.

C Details of Maximum Likelihood Estimator

We also consider the maximum likelihood method for estimation of the two setups. Assuming the two error terms follow bivariate normal distribution under the variance functions assumed, the log-likelihood function is given by

$$L(\beta,\gamma,\delta,\rho) = \sum_{i=1}^{n}\left[-ln(2\pi) - ln(s_{u,i}s_{\varepsilon,i}) - \frac{1}{2}ln(1-\rho^2) - \frac{1}{2(1-\rho^2)}(\tilde{u}_i^2 + \tilde{\varepsilon}_i^2 - 2\rho\tilde{u}_i\tilde{\varepsilon}_i)\right] \tag{15}$$

[3]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. We follow this functional form because it allows for log linear regression in estimation, which is straightforward and stable.

[4]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.

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

where

$$\tilde{\varepsilon}_i = \frac{y_{1i} - y_{2i}\beta_1 - X_i\beta_2}{s_{\varepsilon,i}} \quad (16)$$

$$\tilde{u}_i = \frac{y_{2i} - Z_i\gamma_1 - X_i\gamma_2}{s_{u,i}} \quad (17)$$

$$s_{\varepsilon,i} = \sqrt{f_\varepsilon(Z'_{2i}\delta_\varepsilon)} \quad (18)$$

$$s_{u,i} = \sqrt{f_u(Z'_{2i}\delta_u)} \quad (19)$$

where the scale of single index is fixed by taking 1 as the coefficient first variable in Z_2.

There are some computation issues. First, notice that under a free function of heteroscedasticity, and when the distribution of any of the Z_2 variables has a tail, an unboundedness likelihood problem may occur, similar to the case of likelihood of a mixture distribution model[6]. The problem is that for a tail observation of the single index $Z_i'\delta$ with no or few observations nearby, it is possible to give this observation very low error variances and a very high correlation, leading to a spuriously high likelihood value for this observation. As the correlation is set closer and closer to 1, the likelihood value will become larger and larger. To avoid this spuriously high likelihood value, I have adopted a few measures.

(1) Instead of directly using a fourth order polynomial of the single index, I apply a bounded transformation before forming the polynomial. In particular,

$$f_j(w) = P(\Phi(w), \gamma_j)$$

where Φ is a normal distribution function, evaluated at mean and variance of the empirial value of $w = \delta' z_2$. P represents a fourth order polynomial that is applied to the transformed value. In this way, the tail values will not be very far away from other observations, which can substantially reduce the possibility of fitting a very small variance value for a small number of observations. This may not be needed if Z_2 variables are discrete or do not have a long tail.

(2) I have essentially restricted the value of parameters so that the variances of errors are not below 0.15 while the correlation coefficients of all observations are not above 0.90 in absolute value. These two parameters should be set according to what values are likely to be valid and what values are unlikely in the actual situation. I impose this by adding a large penalty term for any violations:

$$L_p = L + 10000\sum_{i=1}^{n}(min(0, s_{\varepsilon,i} - 0.15))^2 + 10000\sum_{i=1}^{n}(min(0, s_{u,i} - 0.15))^2 + 10000\sum_{i=1}^{n}(max(0, |\rho_i| - 0.9))^2$$

Since these restrictions are sometimes binding, numerical hessian sometimes fails to be negative definite. For inference, bootstrap standard errors and tests are more appropriate.

The use of Alkaline Information Criteria (AIC) for model selection can also be extended to the choice of complexity of the approximating functions, such as the degree of polynomial, or comparing with other

[6] In that case, one component of the mixture may fit one point exactly, leading to an unbounded likelihood, while the other components fit other points as if there is no first component.

forms of approximating functions (such as splines.) Here I focus on the selection between Klein-Vella and Lewbel models and fix the order of polynomial at 4.

Though no formal proof is provided here, similar to the usual LIML, when the model is basically identified by the first two moments, the normality assumption in the likelihood is probably not leading to substantial bias when the true error terms are non-normal. Simulation results with asymmetric distributions, under the normalized chi-square errors and common factor, that is if $\chi^2 \sim \chi^2(p)$,

$$v_{ji} = \frac{\chi^2 - p}{\sqrt{2p}},$$

with $p = 5$ are presented in the last part of the appendix, and the finite sample medians are similar to the case of normal errors.

D 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. In the ML formulation of both estimators, it is straightforward to include Z in the first-stage equation and in any of the variance functions. For the two-step Klein and Vella (2010) estimator, it is also straightforward to include it in the first-stage equation and variance functions.

Concerning the purpose of including both types of instruments, one 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 for the Lewbel estimator using the overidentification J test. If we want to test the validity of excluded instruments, we need the instruments from heteroscedasticity to be valid. However, we are usually not clear about the correct form of heteroscedasticity, and 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. So, 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 for excluded instruments and from heteroscedasticity. 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 overidentification 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.

E Extra Tables of Results

Here I present the results for the case where the error terms are Chi-square distributed with 5 degrees of freedom, normalized to mean zero and variance one.

Table A.1: Simulation Results for Data from the Klein and Vella Form of Heteroscedasticity, Chi-square(5) Errors

n	K	δ_{u1}	δ_{u2}	$\delta_{\varepsilon 1}$	β_{OLS} median (q10,q90)	β_{LB} median (q10,q90)	J median (% $p < 0.05$)	β_{KV},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
500	3	0.4	0.4	0.3	0.4353	0.2614	2.494	0.0215	0.2860	0.0256	0.1205	0.630
					(0.373,0.496)	(0.155,0.372)	(0.134)	(-0.351,0.237)	(0.175,0.388)	(-0.270,0.239)	(-0.219,0.327)	
500	3	0.4	0.4	-0.3	0.4116	0.1527	2.246	0.0134	0.1635	-0.0066	0.0366	0.689
					(0.356,0.468)	(0.057,0.243)	(0.106)	(-0.191,0.150)	(0.055,0.262)	(-0.181,0.136)	(-0.158,0.198)	
500	3	0.25	0.25	0.3	0.4827	0.3289	2.403	0.0420	0.3355	0.0819	0.1641	0.598
					(0.418,0.545)	(0.149,0.492)	(0.138)	(-0.509,0.402)	(0.123,0.523)	(-0.546,0.484)	(-0.452,0.473)	
500	3	0.7	0	0.3	0.4458	0.3184	1.536	0.0367	0.3430	0.0527	0.2511	0.493
					(0.385,0.507)	(0.194,0.433)	(0.040)	(-0.495,0.358)	(0.221,0.448)	(-0.523,0.389)	(-0.369,0.422)	
500	10	0.25	0.25	0.3	0.4100	0.2256	9.564	0.0364	0.2545	0.0387	0.1723	0.514
					(0.350,0.467)	(0.129,0.324)	(0.069)	(-0.167,0.195)	(0.163,0.345)	(-0.193,0.223)	(-0.142,0.316)	
500	10	0.7	0	0.3	0.4445	0.3276	8.481	0.0943	0.3458	0.2045	0.2928	0.361
					(0.384,0.505)	(0.203,0.447)	(0.033)	(-0.213,0.384)	(0.237,0.451)	(-0.205,0.522)	(0.021,0.460)	
1000	3	0.3	0.3	0.3	0.4680	0.2964	3.684	0.0160	0.3166	0.0103	0.0562	0.749
					(0.426,0.511)	(0.194,0.399)	(0.266)	(-0.324,0.226)	(0.203,0.411)	(-0.281,0.227)	(-0.255,0.321)	
1000	3	0.7	0	0.3	0.4429	0.3160	1.412	0.0047	0.3463	0.0143	0.1817	0.562
					(0.400,0.489)	(0.233,0.400)	(0.05)	(-0.411,0.251)	(0.265,0.424)	(-0.280,0.248)	(-0.212,0.389)	

The number of repetitions is 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.

Table A.2: Simulation Results for Data from the Lewbel Form of Heteroscedasticity, Chi-square(5) Errors

n	K	δ_{u1}	δ_{u2}	$\delta_{\varepsilon 1}$	β_{OLS} median (q10,q90)	β_{LB} median (q10,q90)	J median (% $p < 0.05$)	$\beta_{KV,\text{2-step}}$ median (q10,q90)	$\beta_{LB,ML}$ median (q10,q90)	$\beta_{KV,ML}$ median (q10,q90)	β_{AIC} median (q10,q90)	AIC correct rate
500	3	0.5	0.5	0.3	0.4081	0.0076	1.310	-0.5397	-0.0004	-0.5504	-0.0343	0.820
					(0.343,0.475)	(-0.147,0.152)	(0.034)	(-1.254,-0.096)	(-0.164,0.153)	(-1.199,0.527)	(-0.620,0.146)	
500	3	0.5	0.5	-0.3	0.4124	0.0160	1.364	-0.2961	0.0186	-0.3041	-0.0714	0.611
					(0.350,0.475)	(-0.140,0.151)	(0.038)	(-0.864,-0.001)	(-0.141,0.161)	(-0.694,-0.039)	(-0.512,0.117)	
500	3	0.3	0.3	0.3	0.4671	0.0535	1.299	-0.2195	-0.0087	0.0776	-0.0131	0.675
					(0.404,0.526)	(-0.269,0.303)	(0.055)	(-1.032,1.395)	(-0.303,0.330)	(-1.175,2.114)	(-0.553,1.276)	
500	3	0.8	0	0.3	0.4223	0.0169	1.351	-0.5775	-0.0184	-0.6978	-0.0405	0.766
					(0.355,0.487)	(-0.179,0.182)	(0.043)	(-1.417,1.156)	(-0.217,0.171)	(-1.259,2.180)	(-0.823,0.311)	
500	10	0.3	0.3	0.3	0.3918	0.0395	8.084	-0.2448	0.0395	-0.2641	0.0113	0.813
					(0.330,0.458)	(-0.094,0.171)	(0.037)	(-0.552,-0.021)	(-0.082,0.166)	(-0.717,0.097)	(-0.297,0.153)	
500	10	0.8	0	0.3	0.4235	0.0756	8.408	-0.2594	0.0798	-0.0961	0.0394	0.829
					(0.359,0.488)	(-0.093,0.233)	(0.037)	(-0.609,0.861)	(-0.152,0.258)	(-0.958,1.401)	(-0.170,0.298)	
1000	3	0.5	0.5	0.3	0.4420	0.0145	1.292	-0.6243	-0.0132	-0.7847	-0.0372	0.881
					(0.396,0.484)	(-0.141,0.146)	(0.036)	(-1.369,-0.072)	(-0.194,0.138)	(-1.268,1.406)	(-0.450,0.130)	
1000	3	0.8	0	0.3	0.4203	0.0046	1.334	-0.8255	-0.0266	-1.1570	-0.0498	0.876
					(0.374,0.468)	(-0.122,0.123)	(0.042)	(-1.727,-0.285)	(-0.195,0.113)	(-1.287,2.194)	(-0.383,0.111)	

Refer to the notes for Table A.1.