Economics Bulletin

Appendix and Supplemental material not intended for publication-Round 2

Submission Number:EB-16-00499

Here include supplementary appendix to discuss more details about the identification and calculation issues involved in this paper.

I would like to thank the anonymous referee, Alan Fernihough, Ka-fu Wong, Feng Yao and participants of the Econometric Society Asian Meeting 2016 for their helpful comments. All remaining errors are mine.
Submitted: Jul 03 2016. **Revised:** October 05, 2017.

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 $\sigma_j = var(e_j)$.

To satisfy this condition, the most plausible way is to have ρ_{12}, σ_1 and σ_2 as constants not depending on z_2 and taken out of the expectation. 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] Another possible way is to have $\rho_{12}(z_2)\sigma_1(z_2)\sigma_2(z_2)$ a function of even-ordered polynomial terms of z_2 and so the expectation is an odd-ordered polynomial terms of z_2. If z_2 is symmetric along zero, this will give a zero expectation. But this is not straightforward how to justify this case in empirical applications, so I do not focus on this case.

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 $\sigma_j = \alpha_j \sigma_\theta^2(z_2)$.

[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.

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

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

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 appear in the loadings.

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

$$\begin{aligned} E(uy_1|Z_2 = 1) - E(uy_1|Z_2 = 0) &= [E(uy_2|Z_2 = 1) - E(uy_2|Z_2 = 0)]\beta \\ &\quad + E(u\varepsilon|Z_2 = 1) - E(u\varepsilon|Z_2 = 0) \\ &= [Var(u|Z_1 = 1) - Var(u|Z_1 = 0)]\beta \\ &\quad + E(e_1 e_2|Z_2 = 1) - E(e_1 e_2|Z_2 = 0) \end{aligned} \tag{9}$$

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

$$\beta_{LB} = \frac{E(uy_1|Z_2 = 1) - E(uy_1|Z_2 = 0)}{Var(u|Z_1 = 1) - Var(u|Z_1 = 0)} = \beta + \frac{E(e_1 e_2|Z_2 = 1) - E(e_1 e_2|Z_2 = 0)}{Var(v_2|Z_2 = 1) - Var(v_2|Z_2 = 0)} \tag{11}$$

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

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

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.

[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).

C Implementation of Lewbel (2012) Estimator

Originally, I use the `ivlewbel` package in R to perform the estimation of the original Lewbel (2012) estimator. However, an issue is discovered in the process of the simulation exercises for this paper. The issue is that, using the two commonly available packages for the Lewbel estimators, I find that they tend to give an actual rejection rate (or actual size) of the J test for over-identifying restrictions under the null lower than the nominal one, and this under-rejection does not diminish when the sample size increases. The results are shown in Table 1 below. The specification of the data generating process is the same as the Lewbel case specified in the main text, so the rejection rate should be the same as the size of test besides simulation errors and approximation error from the asymptotic distribution to the finite sample. I have only shown a few cases with $K = 3$, but it also applies to cases of other numbers of regressors and heteroscedastic related variables.

The first column shows the case using the `ivlewbel` package in R, using the default (with robust turned `TRUE`). The second column uses the `ivreg2h` module with option `robust` and `gmm2s`. The third and fourth columns show the results of two modified implementation I have tried coded in R. Both of them estimate all parameters using all four sets of moment conditions in (3) in this paper (or Lewbel, 2012, Corollary 4). Modification 1 uses the coefficients estimated from GMM with identity weight matrix to form the optimal weight matrix[6], while Modification 2 uses the coefficients from the two-stage least squares approach to construct the optimal weight matrix[7].

The results in Table 1 show that even when sample size increases to 10000, the rejection rates using the specific package for the Lewbel (2012) estimator in R and Stata do not converge to the desired nominal value of 5%, but are only about 3%. The results from the modifications I have tried are instead close to the nominal value. As can be seen from the source code for `ivlewbel` package in R[8], it uses the estimated first stage residuals and mean of variables used to construct Lewbel instruments before fitting them into the GMM system. It is likely that these pre-estimation procedures have distorted the distribution of the final J statistics, something analogous to the reduction of degree of freedom of the quadratic form when some of the parameters are estimated within the system.[9] To have a more accurate size of the over-identification test, it is better to estimate all parameters together in the GMM system (or to adjust the J statistics in some ways that accounts for pre-estimated quantities, which is not explored here.)

Between modification 1 and modification 2, modification 2 seems to perform better. Further investigation can be done to determine whether iteration on or continuously updating the weight matrix can give more desirable finite sample results. In the main text of this paper, I have used the Modification 2 for the Lewbel GMM method.

---

[6]It is implemented by the default of `gmm` package in R with `vcov=”iid”`

[7]The weight matrix is constructed by calculating the inverse of the empirical variance of the moment conditions based on the two-stage least squares parameters, first stage parameters and means of heteroscedasticity related variables and then I feed the weight matrix into the `gmm` command by using `vcov="TrueFixed"` option in `gmm`.

[8]The source code can be found at https://github.com/cran/ivlewbel/blob/master/R/lewbel.est.R

[9]I have no information about the procedures used for the Stata module, but based on the results in the table, it is plausible that pre-estimated values are also used.

Table 1: Simulated Actual Rejection Rate of the J Test With 5% Nominal Size Under Various Estimators

n	K	δ_{u1}	δ_{u2}	$\delta_{\varepsilon 1}$	$\delta_{\varepsilon 2}$	Rejection Rate of J Test at 5% Level			
						ivlewbel in R	ivreg2h in Stata	Modification 1	Modification 2
500	3	0.3	0.3	0.3	0	0.0461	0.0468	0.0223	0.0687
2000						0.0353	0.0354	0.0440	0.0561
5000						0.0306	0.0322	0.0475	0.0520
10000						0.0327	0.0294	0.0510	0.0530
500	3	0.5	0.3	0.3	0	0.0339	0.0390	0.0457	0.0588
2000						0.0317	0.0328	0.0484	0.0512
5000						0.0331	0.0315	0.0522	0.0532
10000						0.0308	0.0321	0.0473	0.0479
500	3	0.8	0.8	0.3	0	0.0341	0.0380	0.0565	0.0561
2000						0.0360	0.0350	0.0528	0.0529
5000						0.0371	0.0353	0.0526	0.0529
10000						0.0333	0.0360	0.0471	0.0469

Note: Please refer to the text for specification details. Number of repetition is at least 10000.

D 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

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

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, an unboundedness likelihood problem may occur, similar to the case of likelihood of a mixture distribution model[10]. 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

[10] 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.

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 the normal distribution function, evaluated at mean and variance of the empirical 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 Akaike 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 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.

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)	$\beta_{LB,GMM}$ median (q10,q90)	J median (% $p<0.05$)	$\beta_{KV,2\text{-step}}$ 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	0.2632	2.560	0.0215	0.2860	0.0256	0.1205	0.630
					(0.373,0.496)	(0.151,0.373)	(0.161)	(-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	0.1516	2.567	0.0134	0.1635	-0.0066	0.0366	0.689
					(0.356,0.468)	(0.058,0.238)	(0.159)	(-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	0.3246	2.672	0.0420	0.3355	0.0819	0.1641	0.598
					(0.418,0.545)	(0.157,0.495)	(0.176)	(-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	0.3205	1.545	0.0367	0.3430	0.0527	0.2511	0.493
					(0.385,0.507)	(0.199,0.439)	(0.056)	(-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	0.2320	11.01	0.0364	0.2545	0.0387	0.1723	0.514
					(0.350,0.467)	(0.139,0.325)	(0.140)	(-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	0.3303	9.449	0.0943	0.3458	0.2045	0.2928	0.361
					(0.384,0.505)	(0.213,0.446)	(0.066)	(-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	0.2993	3.975	0.0160	0.3166	0.0103	0.0562	0.749
					(0.426,0.511)	(0.199,0.399)	(0.305)	(-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	0.3171	1.490	0.0047	0.3463	0.0143	0.1817	0.562
					(0.400,0.489)	(0.234,0.401)	(0.049)	(-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)	$\beta_{LB,GMM}$ 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	0.0105	1.550	-0.5397	-0.0004	-0.5504	-0.0343	0.820
					(0.343,0.475)	(-0.154,0.157)	(0.063)	(-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	0.0122	1.545	-0.2961	0.0186	-0.3041	-0.0714	0.611
					(0.350,0.475)	(-0.143,0.147)	(0.060)	(-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	0.0604	1.575	-0.2195	-0.0087	0.0776	-0.0131	0.675
					(0.404,0.526)	(-0.254,0.306)	(0.077)	(-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	0.0174	1.522	-0.5775	-0.0184	-0.6978	-0.0405	0.766
					(0.355,0.487)	(-0.174,0.183)	(0.065)	(-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	0.0418	9.992	-0.2448	0.0395	-0.2641	0.0113	0.813
					(0.330,0.458)	(-0.092,0.173)	(0.100)	(-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	0.0731	10.109	-0.2594	0.0798	-0.0961	0.0394	0.829
					(0.359,0.488)	(-0.099,0.240)	(0.113)	(-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	0.0038	1.520	-0.6243	-0.0132	-0.7847	-0.0372	0.881
					(0.396,0.484)	(-0.111,0.108)	(0.056)	(-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	0.0089	1.487	-0.8255	-0.0266	-1.1570	-0.0498	0.876
					(0.374,0.468)	(-0.127,0.127)	(0.058)	(-1.727,-0.285)	(-0.195,0.113)	(-1.287,2.194)	(-0.383,0.111)	

Refer to the notes for Table A.1.