Asymptotic variance under many instruments: numerical computations

Albert Abutaliev
New Economic School

Stanislav Anatolyev*
New Economic School

Abstract

In models with many instruments, the asymptotic variance of the LIML estimator contains four components. Apart from the traditional variance, one term is due to instrument numerosity, and the last two appear if the model errors are non-normal. For a stylized instrumental variables model, we compute numerical values of these components to uncover how the four components are related to each other in magnitude.

1 Introduction

In the original Bekker (1994) framework of an instrumental variables regression with many instruments and normal errors the asymptotic variance of the limited information maximum likelihood (LIML) estimator is composed of two components, one is the traditional asymptotic variance, and the other is due to numerosity of instruments. Later it was shown (Hansen et al, 2008; van Hasselt, 2010) that under error non-normality there are two additional terms involving third and fourth moments of errors. While the 'instrument numerosity' term may be appreciable (Newey, 2004), there is a perception in the literature (e.g., Anderson et al, 2010; Anatolyev and Gospodinov, 2011) that the 'higher order moments' terms are impalpable. We compute the asymptotic variance components of the LIML estimator for a stylized instrumental variables model and discover that in fact the correction terms may be quite noticeable.

2 Setup

The standard model is given by

$$\begin{aligned} Y &= X\beta + U, \\ X &= Z\Pi + V, \end{aligned}$$

*Corresponding author. Address: New Economic School, Nakhimovsky Prospekt, 47, room 1721, Moscow, 117418, Russia. Web: sanatoly@nes.ru; http://www.nes.ru/~sanatoly/. We are grateful to a referee whose comments improved the presentation.

with one endogenous variable, K instruments, N observations, and random sampling. The errors u and v are independent of instruments z, have mean zero and variances σ_u^2 and σ_v^2 and covariance σ_{vu}. We adopt the Bekker (1994) many instrument asymptotic framework: $K/N = \alpha + o(1/\sqrt{N})$. We consider the LIML estimator

$$\hat{\beta} = \arg\min_b \frac{(Y - Xb)'P(Y - Xb)}{(Y - Xb)'(Y - Xb)},$$

where P is the projector matrix associated with Z. Under homoskedasticity LIML, unlike 2SLS, provides consistent estimation under many instruments (Bekker, 1994; Newey, 2004), and is asymptotically efficient in a certain class (Anderson et al, 2010). Even though the LIML estimator is non-robust to heteroskedasticity in contrast to recently developed jackknife-type estimators (Hausman et al, 2012), it is very popular and easy to compute via a certain eigenvalue problem, while it is not unambiguously dominated by these estimators in terms of asymptotic efficiency (see Hausman et al, 2012). The structure of the asymptotic variance of the LIML estimator is

$$V_{\hat{\beta}} = W_1 + W_2 + W_3 + W_4,$$

where W_1, W_2, W_3, W_4 are the 'traditional' LIML variance, 'instrumental numerosity effect', 'skewness effect', and 'kurtosis effect', respectively (Hansen et al, 2008; van Hasselt, 2010). When the errors are normal, $W_3 = W_4 = 0$.

We model the distribution of each error of the pair (u, v) as Skewed Student (Azzalini and Capitanio, 2003) adjusted to have zero mean and unit variance (the latter is a convenient normalization) with ν degrees of freedom, setting ν to 9 in order for eighth moments to exist so that asymptotic variance estimation (Hansen et al, 2008) goes through. The degree of asymmetry and leptokurtocity of each error is described by parameters γ_u and γ_v. In our experiments, we set each γ to take values in the set $\Gamma = \{\pm 1,\ \pm 0.5,\ 0\}$, which correspond to the same values of skewness and values $\{2.6,\ 1.6,\ 1.2\}$ of excess kurtosis. Note that these numbers indicate deviations from normality that are not drastic.

To separately control the dependence between u and v, we impose a Gaussian copula on the pair which is given by $C(u, v) = \Phi_G(\Phi^{-1}(u), \Phi^{-1}(v); \rho)$, where $\Phi(\cdot)$ is the standard normal CDF, and $\Phi_G(\cdot, \cdot\cdot; \rho)$ is the bivariate standard normal CDF with correlation parameter $\rho \in (-1, 1)$. We set ρ to take values in the set $\mathrm{P} = \{\pm 0.9,\ \pm 0.6,\ \pm 0.1\}$ which correspond to values $\{\pm 0.88,\ \pm 0.59,\ \pm 0.10\}$ of the covariance σ_{vu}. Note that ρ and σ_{vu} are approximately equal.

It is very important how to design the reduced form. We have tried several designs for instruments that can be encountered in various simulation studies. Many of them result in W_3 and/or W_4 equaling zero (while the ratio W_2/W_1 is approximately as reported below). For instance, if $z \sim IID\ \mathcal{D}(0, I_K)$ and $\Pi = \pi\iota_K/\sqrt{K}$, where $\mathcal{D}$ is standard normal, scaled lognormal or scaled recentered chi-squared, then $W_3 = W_4 = 0$, and if $z = (1, \zeta, \zeta^2, \zeta^3, \zeta^4, \zeta D_1, \ldots, \zeta D_{K-5})'$, where $\zeta \sim \mathcal{D}(0, 1)$ and $D_k \sim IID\ B(\frac{1}{2})$, then $W_4 \neq 0$ but $W_3 = 0$. Therefore, we follow Bekker and van Ploeg (2005) and assume that the observations are split into groups, whose number K grows linearly with the sample size. Denote the size of the k^{th} group size by N_k and the corresponding group mean by π_k. The instrument matrix is comprised of dummy variables: $Z = \text{diag}\left\{\iota_{N_k}\right\}_{k=1}^{K}$, and the reduced form coefficients are $\Pi = (\pi_1, \pi_2, \ldots, \pi_K)'$. Bekker and van Ploeg (2005, p. 251) show that when group sizes are equal, $W_3 = W_4 = 0$, but these components are non-zero otherwise.

We generate the group means π_k as $IID\ U[0, \pi]$ and group sizes N_k as $IID\ U\{1, \ldots, n\}$, where $n = 2N/K - 1$, assuming that K and N are such that n is integer, and induce the dependence between N_k and π_k via a copula (without dependence $W_3 = 0$). Technically, we let the pair $(n_k, p_k) \sim IID\ C_G(\kappa)$, the Gaussian copula with correlation parameter κ, and set $N_k = \lceil nn_k \rceil$ and $\pi_k = \pi p_k$. The random design of group sizes induces non-zero impact of skewness and excess kurtosis, while the parameter κ is set to 0.99 to amplify the effect of the skewness term.

For the described design,

$$V_{\hat{\beta}} = Q^{-2} \left(Q + \Theta_2 \Sigma_2 + \Theta_3 \Sigma_3 + \Theta_4 \Sigma_4 \right).$$

Here, the quantities $Q = \lim N^{-1}(Z\Pi)'(Z\Pi) = \alpha\pi^2 E\left[N_k p_k^2\right]$, $\Theta_2 = \alpha(1-\gamma)/(1-\alpha)^2$, $\Theta_3 = 2\mu\alpha/(1-\alpha)$ and $\Theta_4 = \alpha(\gamma-\alpha)/(1-\alpha)^2$, where $\gamma = \lim K^{-1}\Sigma_{i=1}^N P_{ii}^2 = n^{-1}\sum_{j=1}^n j^{-1}$ and $\mu = \pi\left(\frac{1}{2} - \alpha E\left[N_k p_k\right]\right)$, depend on the reduced form design, while the quantities $\Sigma_2 = 1 - \sigma_{vu}^2$, $\Sigma_3 = E\left[u^2 w\right]$ and $\Sigma_4 = E\left[u^2 w^2\right]$, where $w = v - \sigma_{vu} u$, depend only on error moments.

3 Results

In evaluating various moments, we used Monte Carlo integration. Relative magnitudes of instrument numerosity, error skewness and error kurtosis corrections are given in Table 1. We report W_2/W_1, $\max_{\gamma_u \in \Gamma, \gamma_v \in \Gamma} |W_3|/W_1$ and $\max_{\gamma_u \in \Gamma, \gamma_v \in \Gamma} W_4/W_1$ as functions of α and the reduced form R^2, for the values of ρ from P.

Table 1: Relative (to W_1) magnitudes of the other asymptotic variance components

K/N	R^2	$\rho = \pm 0.1$			$\rho = \pm 0.6$			$\rho = \pm 0.9$		
		W_2	W_3	W_4	W_2	W_3	W_4	W_2	W_3	W_4
0.05	0.001	48.84	0.043	3.455	32.39	0.241	3.646	10.89	0.383	3.462
	0.010	4.840	0.014	0.342	3.209	0.076	0.361	1.079	0.121	0.343
	0.100	0.440	0.004	0.031	0.292	0.023	0.033	0.098	0.036	0.031
	0.500	0.049	0.001	0.003	0.032	0.008	0.004	0.011	0.012	0.003
0.10	0.001	99.34	0.090	11.31	65.90	0.496	11.93	22.16	0.789	11.33
	0.010	9.845	0.028	1.120	6.529	0.156	1.182	2.196	0.248	1.123
	0.100	0.895	0.008	0.102	0.594	0.047	0.107	0.200	0.075	0.102
	0.500	0.099	0.003	0.011	0.066	0.016	0.012	0.022	0.025	0.011
0.20	0.001	212.0	0.190	37.72	140.6	1.056	39.81	47.28	1.678	37.80
	0.010	21.01	0.060	3.738	13.93	0.332	3.945	4.686	0.528	3.746
	0.100	1.910	0.018	0.340	1.267	0.100	0.359	0.426	0.159	0.341
	0.500	0.212	0.006	0.038	0.141	0.033	0.040	0.047	0.053	0.038

The ratio W_2/W_1 is increasing in K/N and decreasing in R^2 and ρ. It is indeed sizeable even for relatively strong relatively few instruments. For instance, W_2 is half the size of the leading term when $R^2 = 0.1$, $\rho = 0.1$ and $K/N = 0.05$. Further, the instrument numerosity correction is of different order of magnitude for pathologically weak instruments than for relatively strong ones. This agrees with simulation resuls in Hahn and Inoue (2002) who show that the Bekker asymptotic approximation is of most value for weaker instruments, and with the theoretical results given in Chao and Swanson (2006).

The excess kurtosis correction turns out to be quite sizable. It varies from roughly one tenth (for smaller ρ) of instrument numerosity correction to comparable magnitudes (for larger ρ). It remains perceptible even for relatively strong instruments when compared to the leading term. In contrast, the error skewness has, on average, a much smaller impact, but still should not be excluded from consideration, especially when K/N and/or ρ are high. Moreover, observe that when instruments are pathologically weak, i.e., when $R^2 = 0.001$, the relative size of W_3 is negligible, which is also consistent with the many weak instrument results of Hansen et al (2008). Interestingly, the 'kurtosis effect' hardly varies with ρ, unlike the 'skewness effect' which is roughly proportional to ρ.

References

Anatolyev, S. and N. Gospodinov (2011): "Specification testing in models with many instruments," *Econometric Theory*, 27, 427–441.

Anderson, T.W., N. Kunimoto and Y. Matsushita (2010): "On the asymptotic optimality of the LIML estimator with possibly many instruments," *Journal of Econometrics*, 157, 191–204.

Azzalini, A. and A. Capitanio (2003): "Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution," *Journal of the Royal Statistical Society Series B*, 65, 367–389.

Bekker, P. A. (1994): "Alternative approximations to the distributions of instrumental variable estimators," *Econometrica*, 62, 657–681.

Bekker, P. A. and J. van der Ploeg (2005): "Instrumental variable estimation based on grouped data," *Statistica Neerlandica*, 59, 239–267.

Chao, J.C. and N.R. Swanson (2006): "Asymptotic normality of single-equation estimators for the case with a large number of weak instruments," in: D. Corbae, S.N. Durlauf, and B.E.Hansen (eds.) *Frontiers of Analysis and Applied Research: Essays in Honor of Peter C. B. Phillips*, 82–124. Cambridge: Cambridge University Press.

Hahn, J., and A. Inoue (2002): "A Monte Carlo comparison of various asymptotic approximations to the distribution of instrumental variables estimators," *Econometric Reviews*, 21, 309–336.

Hansen, C., J. Hausman and W.K. Newey (2008): "Estimation with many instrumental variables," *Journal of Business & Economics Statistics*, 26, 398–422.

Hausman, J.A., W.K. Newey, T. Woutersen, J.C. Chao, and N.R. Swanson (2012): "Instrumental variable estimation with heteroskedasticity and many instruments," *Quantitative Economics*, 3, 211–255.

Newey, W. K. (2004): "Many instrument asymptotics," manuscript, MIT.

van Hasselt, M. (2010): "Many instruments asymptotic approximations under nonnormal error distributions," *Econometric Theory*, 26, 633–645.