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

\[ \begin{align*}
Y & = X\beta + U, \\
X & = Z\Pi + V,
\end{align*} \]

*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 $\sigma_u^2$ and $\sigma_v^2$ and covariance $\sigma_{uv}$. 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 $\nu$ degrees of freedom, setting $\nu$ 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 leptokurticity of each error is described by parameters $\gamma_u$ and $\gamma_v$. In our experiments, we set each $\gamma$ 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 $\Phi_G(\cdot, \ldots; \rho)$ is the bivariate standard normal CDF with correlation parameter $\rho \in (-1, 1)$. We set $\rho$ to take values in the set $\mathcal{P} = \{\pm 0.9, \pm 0.6, \pm 0.1\}$ which correspond to values $\{\pm 0.88, \pm 0.59, \pm 0.10\}$ of the covariance $\sigma_{uv}$. Note that $\rho$ and $\sigma_{uv}$ 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 $\Pi = \pi(t_k)K/\sqrt{K}$, 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 $D_k \sim IID \mathcal{B}(\frac{1}{2})$, 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 $\bar{\pi}_k$. The instrument matrix is comprised of dummy variables: $Z = \text{diag}\{t_{N_k}\}_{k=1}^K$, 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 $\pi_k$ 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 $\pi_k$ 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 $\kappa$, and set $N_k = \lfloor \kappa n_k \rfloor$ and $\pi_k = \pi_\kappa k$. The random design of group sizes induces non-zero impact of skewness and excess kurtosis, while the parameter $\kappa$ is set to 0.99 to amplify the effect of the skewness term.

For the described design,

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

Here, the quantities $Q = \lim N^{-1}(Z_1)'(Z_1) = \alpha \pi^2 E[N_k p_k^2]$, $\Theta_2 = \alpha (1 - \gamma)/(1 - \alpha)^2$, $\Theta_3 = 2\mu \alpha/(1 - \alpha)$ and $\Theta_4 = \alpha (\gamma - \alpha)/(1 - \alpha)^2$, where $\gamma = \lim K^{-1} \sum_{i=1}^K P_i^2 = n^{-1} \sum_{j=1}^n j^{-1}$ and $\mu = \pi (\frac{1}{2} - \alpha E[N_k p_k])$, depend on the reduced form design, while the quantities $\Sigma_2 = 1 - \sigma_{wv}$, $\Sigma_3 = E[u^2 w]$ and $\Sigma_4 = E[u^2 w^2]$ 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 $W_2/W_1$, $\max_{\gamma_i \in \Gamma, \gamma_v \in \Gamma} |W_3|/W_1$ and $\max_{\gamma_i \in \Gamma, \gamma_v \in \Gamma} W_4/W_1$ as functions of $\alpha$ and the reduced form $R^2$, for the values of $\rho$ from $P$.

<table>
<thead>
<tr>
<th>$K/N$</th>
<th>$R^2$</th>
<th>$\rho = \pm 0.1$</th>
<th>$\rho = \pm 0.6$</th>
<th>$\rho = \pm 0.9$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$W_2$</td>
<td>$W_3$</td>
<td>$W_4$</td>
</tr>
<tr>
<td>0.05</td>
<td>0.001</td>
<td>48.84</td>
<td>0.043</td>
<td>3.455</td>
</tr>
<tr>
<td></td>
<td>0.010</td>
<td>4.840</td>
<td>0.014</td>
<td>0.342</td>
</tr>
<tr>
<td></td>
<td>0.100</td>
<td>0.440</td>
<td>0.004</td>
<td>0.031</td>
</tr>
<tr>
<td></td>
<td>0.500</td>
<td>0.049</td>
<td>0.001</td>
<td>0.003</td>
</tr>
<tr>
<td>0.10</td>
<td>0.001</td>
<td>99.34</td>
<td>0.090</td>
<td>11.31</td>
</tr>
<tr>
<td></td>
<td>0.010</td>
<td>9.845</td>
<td>0.028</td>
<td>1.120</td>
</tr>
<tr>
<td></td>
<td>0.100</td>
<td>0.895</td>
<td>0.008</td>
<td>0.102</td>
</tr>
<tr>
<td></td>
<td>0.500</td>
<td>0.099</td>
<td>0.003</td>
<td>0.011</td>
</tr>
<tr>
<td>0.20</td>
<td>0.001</td>
<td>212.0</td>
<td>0.190</td>
<td>37.72</td>
</tr>
<tr>
<td></td>
<td>0.010</td>
<td>21.01</td>
<td>0.060</td>
<td>3.738</td>
</tr>
<tr>
<td></td>
<td>0.100</td>
<td>1.910</td>
<td>0.018</td>
<td>0.340</td>
</tr>
<tr>
<td></td>
<td>0.500</td>
<td>0.212</td>
<td>0.006</td>
<td>0.038</td>
</tr>
</tbody>
</table>

The ratio $W_2/W_1$ is increasing in $K/N$ and decreasing in $R^2$ and $\rho$. 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 results 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 \( \rho \)) of instrument numerosity correction to comparable magnitudes (for larger \( \rho \)). 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 \( \rho \) 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 \( \rho \), unlike the ‘skewness effect’ which is roughly proportional to \( \rho \).

References


NEWEY, W. K. (2004): “Many instrument asymptotics,” manuscript, MIT.