An Alternative Way of Computing Efficient Instrumental Variable Estimators

Xiaohong Chen (Yale University)
David T. Jacho-Chávez
Oliver B. Linton (London School of Economics)

Department of Economics
Indiana University
http://mypage.iu.edu/~djachoch/
Outline

Introduction

Estimation Idea

Large Sample Properties

Numerical Results
Framework

We observe i.i.d. sample \( \{Z_i\}_{i=1}^n \), where \( Z_i^\top = (Y_i^\top, X_i^\top) \). We suppose that there is a unique \( \theta_0 \in \Theta \subseteq \mathbb{R}^p \) satisfying the conditional moment conditions

\[
E[\rho(Z_i, \theta_0) \mid X_i] = 0 \text{ w.p.} 1.
\]

This implies the unconditional moment conditions

\[
E[A(X_i)\rho(Z_i, \theta_0)] = 0,
\]

Then an estimator can be constructed based on

\[
\frac{1}{n} \sum_{i=1}^n A(X_i)\rho(Z_i, \hat{\theta}) = o_P(n^{-1/2}).
\]
Framework

We observe i.i.d. sample \(\{Z_i\}_{i=1}^n\), where \(Z_i^\top = (Y_i^\top, X_i^\top)\). We suppose that there is a unique \(\theta_0 \in \Theta \subseteq \mathbb{R}^p\) satisfying the conditional moment conditions

\[
E[\rho(Z_i, \theta_0) | X_i] = 0 \text{ w.p.1.}
\]

This implies the unconditional moment conditions

\[
E[A(X_i)\rho(Z_i, \theta_0)] = 0,
\]

Hansen (1985), Chamberlain (1987) and Newey (1990, 1993): If one chooses \(A_{oiv}(X_i) \propto D_0(X_i)\sigma_0^{-2}(X_i)\), then

\[
\frac{1}{n} \sum_{i=1}^n A_{oiv}(X_i)\rho(Z_i, \tilde{\theta}_{oiv}) = o_p(n^{-1/2}).
\]
Framework

We observe i.i.d. sample \( \{Z_i\}_{i=1}^n \), where \( Z_i^\top = (Y_i^\top, X_i^\top) \). We suppose that there is a unique \( \theta_0 \in \Theta \subseteq \mathbb{R}^p \) satisfying the conditional moment conditions

\[
E[\rho(Z_i, \theta_0) | X_i] = 0 \quad \text{w.p.1.}
\]

This implies the unconditional moment conditions

\[
E[A(X_i)\rho(Z_i, \theta_0)] = 0,
\]

Hansen (1985), Chamberlain (1987) and Newey (1990, 1993): If one chooses \( \hat{A}(x) = \sum_{j=1}^{\tau(n)} \hat{\beta}_j \phi_j(x) \), then

\[
\frac{1}{n} \sum_{i=1}^{n} \hat{A}(X_i)\rho(Z_i, \tilde{\theta}_{oiv}) = o_p(n^{-1/2}).
\]
Framework

We observe i.i.d. sample \( \{Z_i\}_{i=1}^n \), where \( Z_i^\top = (Y_i^\top, X_i^\top) \). We suppose that there is a unique \( \theta_0 \in \Theta \subseteq \mathbb{R}^p \) satisfying the conditional moment conditions

\[
E[\rho(Z_i, \theta_0) | X_i] = 0 \text{ w.p.1.}
\]

This implies the unconditional moment conditions

\[
E[A(X_i)\rho(Z_i, \theta_0)] = 0,
\]

Hansen (1985), Chamberlain (1987) and Newey (1990, 1993): If one chooses \( \hat{A}(x) = \sum_{j=1}^{\tau(n)} \hat{\beta}_j \phi_j(x) \), then \( \hat{\theta}_{oiv} \) has Asy. Var.

\[
\Sigma_{oiv} = \left\{ E[\sigma_0^{-2}(X_i)D_0(X_i)D_0(X_i)^\top] \right\}^{-1}.
\]
Our Estimation Idea

Consider a sequence of pre-specified basis ($p \times 1$ vector-valued) functions $\{A_j(\cdot)\}$ such that $E[||A_j(X_i)||^2] < \infty$. We define the estimators $\hat{\theta}_j, j = 1, 2, \ldots$, as any sequence that satisfies

$$G_{nj}(\hat{\theta}_j) = \frac{1}{n} \sum_{i=1}^{n} A_j(X_i) \rho(Z_i, \hat{\theta}_j) = o_p(n^{-1/2}).$$

We combine these estimators in a linear fashion to produce a new estimator

$$\hat{\theta} = \sum_{j=1}^{\tau(n)} W_{nj} \hat{\theta}_j, \text{ with } \sum_{j=1}^{\tau(n)} W_{nj} = I_p$$
Our Estimation Idea

This defines a class of estimators $\mathcal{E}$ indexed by the weighting matrices $\{W_{nj}, j = 1, \ldots, \tau(n)\}$. 

(i) It is shown that, by an appropriate choice of weights, the semiparametrically efficient estimator is a member of $\mathcal{E}$. 

(i) Smoothness conditions on the residual function $\rho(Z_i, \theta)$ are not required.

We combine these estimators in a linear fashion to produce a new estimator 

$$\hat{\theta} = \sum_{j=1}^{\tau(n)} W_{nj} \hat{\theta}_j, \text{ with } \sum_{j=1}^{\tau(n)} W_{nj} = I_p$$
**Example:** Classical two stage least squares in simultaneous equations. Suppose that

\[ y_{1i} = \theta y_{2i} + \varepsilon_i; \quad y_{2i} = \pi_2^\top X_i + u_i, \]

where \((\varepsilon_i, u_i)^\top\) are i.i.d. error terms, \(E[\varepsilon_i|X_i] = 0, E[u_i|X_i] = 0\) and \(X_i \in \mathbb{R}^k\).

The two stage least squares (oiv) estimator is

\[ \hat{\theta}_{oiv} = \frac{\sum_{i=1}^{n} \hat{y}_{2i} y_{1i}}{\sum_{i=1}^{n} \hat{y}_{2i}^2}. \]

Our estimator is \(\hat{\theta} = \sum_{j=1}^{k} W_{nj} \hat{\theta}_j\), where \(\hat{\theta}_j = \frac{\sum_{i=1}^{n} \hat{y}_{2i}^j y_{1i}}{\sum_{i=1}^{n} \hat{y}_{2i}^j y_{2i}}\), and \(\hat{y}_{2i}^j = \hat{\pi}_2^j X_{ji}\).
Main Results

Theorem (Consistency)

(i) Suppose that Assumption A holds. Then $\hat{\theta} - \theta_0 = o_p(1)$.
(ii) Suppose that Assumption A* holds. Then $\hat{\theta} - \theta_0 = o_p(n^{-1/4})$. 
Main Results

Theorem (Asymptotic Normality)

Suppose that Assumption B holds. Then \( \sqrt{n}(\hat{\theta} - \theta_0) \rightarrow N(0, \Sigma) \), where

\[
\sum_{j=1}^{\tau(n)} \sum_{l=1}^{\tau(n)} W_{nj}^{-1} \Gamma_j^{-1} E \left[ g_j(Z_i, \theta_0) g_l(Z_i, \theta_0)^\top \right] \Gamma_l^{-1\top} W_{nl}^{0\top} \rightarrow \Sigma \quad \text{as} \quad n \rightarrow \infty.
\]

\[
g_j(Z_i, \theta) = A_j(X_i) \rho(Z_i, \theta)
\]

\[
\Gamma_j = E[A_j(X_i)D_0(X_i)^\top]
\]

\[
D_0(X_i) = \left\{ \partial E[\rho(Z_i, \theta)|X_i]/\partial \theta \right\}_{\theta=\theta_0}
\]
**Optimal Weights**

**Fixed $\tau$**

**Optimal Instrumental Variables**

Suppose that we know only that

$$E[A_j(X_i)\rho(Z_i, \theta_0)] = 0, \quad j = 1, \ldots, \tau,$$

where $\tau$ is fixed, and $A_j \in \mathbb{R}^p$. The oiv estimator $\tilde{\theta}_{oiv}^{\tau}$ of $\theta_0$ for this model is such that

$$\sqrt{n}(\tilde{\theta}_{oiv}^{\tau} - \theta_0) \overset{d}{\to} N(0, \Sigma_{oiv}^{\tau})$$

as $n \to \infty$, where the asymptotic variance is given by:

$$\Sigma_{oiv}^{\tau} = \left( E[A^\tau(X)D_0(X)^\top] \right)^\top \left[ E(\sigma_0^2(X)A^\tau(X)^\top) \right]^{-1} E[A^\tau(X)D_0(X)^\top]^{-1}$$

where $A^\tau = (A^\top_1, \ldots, A^\top_\tau)^\top \in \mathbb{R}^{\tau p}$.
Optimal Weights

Fixed $\tau$

Minimum Distance [Rothemberg (1973)]

Let $\hat{\theta}_{\text{omd}}^\tau$ minimize the criterion function

$$Q_n(\theta) = \left[ \begin{pmatrix} \hat{\theta}_1 \\ \vdots \\ \hat{\theta}_\tau \end{pmatrix} - \theta \otimes i_\tau \right]^T \Gamma^{-1} \left[ \begin{pmatrix} \hat{\theta}_1 \\ \vdots \\ \hat{\theta}_\tau \end{pmatrix} - \theta \otimes i_\tau \right],$$

where $i_\tau$ is a $\tau \times 1$ vector of ones, and $\Gamma$ is the $\tau p \times \tau p$ asymptotic (as $n \to \infty$ holding $\tau$ constant) variance matrix of the vector $(\sqrt{n}(\hat{\theta}_1 - \theta_0)^T, \ldots, \sqrt{n}(\hat{\theta}_\tau - \theta_0)^T)^T$, i.e., $\Gamma = [V_{j,l}]$, where $V_{j,l} = \Gamma_j^{-1} E[A_j(X_i)\sigma_0^2(X_i)A_l(X_i)^T] \Gamma_l^{-1}$ for all $j, l = 1, \ldots, \tau$. 

Alternative Efficient Instrumental Variable Estimator
**Optimal Weights**

*Fixed $\tau$*

**Minimum Distance [Rothemberg (1973)]**

The FOC implies that the optimal estimator $\hat{\theta}_{\text{omd}}^\tau$ is a linear combination of the $\hat{\theta}_j$ with $\hat{\theta}_{\text{omd}}^\tau = \sum_{j=1}^{\tau} W_{0j}^{\text{opt}} \hat{\theta}_j$, where

$$W_{0j}^{\text{opt}} = \left( \sum_{l=1}^{\tau} B_l \right)^{-1} B_j,$$

and $(B_1, \ldots, B_\tau) = (I_p \otimes i_\tau)^\top V^{-1}$.

Furthermore, $\sqrt{n}(\hat{\theta}_{\text{omd}}^\tau - \theta_0) \Rightarrow N(0, \Sigma_{\text{omd}}^\tau)$, where the asymptotic [as $n \to \infty$ and $\tau$ fixed] variance is $\Sigma_{\text{omd}}^\tau = W_{\text{opt}} V W_{\text{opt}}^\top = \left( (I_p \otimes i_\tau)^\top V^{-1}(I_p \otimes i_\tau) \right)^{-1}$. 

Alternative Efficient Instrumental Variable Estimator
Optimal Weights

Fixed $\tau$

Proposition

For each fixed $\tau$, $\hat{\theta}_\text{omd}^\tau$ is asymptotically efficient, i.e. $\Sigma_\text{omd}^\tau = \Sigma_\text{oiv}^\tau$. Moreover the optimal weighting is simply

$$W_{0j}^{oiv} = -\left(\sum_{j=1}^{\tau} \alpha_j \Gamma_j^\top\right)^{-1} \alpha_j \Gamma_j^\top \text{ for } j = 1, \ldots, \tau, \text{ with}$$

$$(\alpha_1, \ldots, \alpha_\tau) = \Gamma^{\tau\top} \Psi_\tau^{-1}.$$
Example:
Recall the optimal GMM estimator in this model [i.e., under homoskedasticity, etc.] is simply the two stage least squares estimator

\[ \tilde{\theta} = (Y_2^\top P_X Y_2)^{-1} Y_2^\top P_X Y_1, \]

where \( P_X = X(X^\top X)^{-1}X^\top \), \( Y_1 = (y_{11}, \ldots, y_{1n})^\top \), \( Y_2 = (y_{21}, \ldots, y_{2n})^\top \), \( X = (X_1^\top, \ldots, X_n^\top) \), \( X_i = (X_{1i}, \ldots, X_{ki})^\top \).
Example:
Within our class of estimators $\mathcal{E}$, the optimal estimator is

$$
\hat{\theta} = \sum_{j=1}^{k} W_{nj}^{\text{opt}} \hat{\theta}_j = \left( \left( i_k^\top V^{-1} i_k \right)^{-1} i_k^\top \right) V^{-1} \begin{bmatrix} \hat{\theta}_1 \\ \vdots \\ \hat{\theta}_k \end{bmatrix},
$$

where $\hat{\theta}_j = \left( Y_2^\top P_j Y_2 \right)^{-1} Y_2^\top P_j Y_1$ for $j = 1, \ldots, k$, where $P_j = X_j (X_j^\top X_j)^{-1} X_j^\top$ and $V$ is the $k \times k$ covariance matrix with $V_{jl} = \text{asy. cov}(\hat{\theta}_j, \hat{\theta}_l)$. 
Optimal Weights
Increasing $\tau$

Let $\Sigma_{oiv}$ be the asymptotic variance of the optimal instrumental variable (oiv) estimator, and let $\Sigma_{omd}$ be the asymptotic variance as $n \to \infty$ and $\tau(n) \to \infty$ of the optimal minimum distance (omd) estimator.

**Theorem**

*Suppose that $E[\rho(Z_i, \theta_0) | X_i] = 0$ and that Assumption C hold. Then,*

$$
\Sigma_{omd} = \Sigma_{oiv} = \left( E[\sigma_0^{-2}(X_i)D_0(X_i)D_0(X_i)^\top] \right)^{-1}.
$$
**Optimal Weights**

**Increasing \( \tau \)**

Let \( \Sigma_{oiv} \) be the asymptotic variance of the optimal instrumental variable (oiv) estimator, and let \( \Sigma_{omd} \) be the asymptotic variance as \( n \to \infty \) and \( \tau(n) \to \infty \) of the optimal minimum distance (omd) estimator.

The optimal weights in this case are any sequence like

\[
W_{nj}^0 = \left( \sum_{l=1}^{\tau(n)} V_{ll}^{-1} \right)^{-1} V_{jj}^{-1},
\]

where \( V_{jj} \) is the asymptotic variance matrix of \( \sqrt{n}(\hat{\theta}_j - \theta_0) \).
Monte Carlo
Newey (1990)

Endogeneous dummy variable model:

\[ Y_i = \beta_{10} + \beta_{20}s_i + \varepsilon_i; \]

\[ DGP1: s_i = 1(\alpha_{10} + \alpha_{20}X_i + \eta_i > 0), \]
\[ X_i \sim N(0, 1); \quad \alpha_{10} = \alpha_{20} = \beta_{10} = \beta_{20} = 1, \]

where the errors \( \varepsilon_i \) and \( \eta_i \) are generated as

\[
\begin{bmatrix}
\varepsilon_i \\
\eta_i
\end{bmatrix} \sim N\left( \begin{bmatrix}
0 \\
0
\end{bmatrix}, \begin{bmatrix}
1 & \varphi \\
\varphi & 1
\end{bmatrix} \right),
\]

in which \( \varphi \in \{0.2, 0.5, 0.8\} \) indicate weak, medium and strong endogeneity respectively. The optimal instrument for \( s \) is

\( \pi(x) = \Pr[s = 1|X = x] \), which makes \( D(x) = (1, \pi(x)). \)
Monte Carlo

\[ A_j(x) = x^{j-1}, \; j = 2, 3, \ldots, 7; \; \varphi = 0.5; \; n = 100 \]

<table>
<thead>
<tr>
<th>( \tau )</th>
<th>Bias (oiv)</th>
<th>Std. Dev. (oiv)</th>
<th>RMSE (oiv)</th>
<th>Bias (omd)</th>
<th>Std. Dev. (omd)</th>
<th>RMSE (omd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.014</td>
<td>0.470</td>
<td>0.47</td>
<td>0.014</td>
<td>0.470</td>
<td>0.470</td>
</tr>
<tr>
<td>3</td>
<td>0.027</td>
<td>0.442</td>
<td>0.441</td>
<td>0.011</td>
<td>0.465</td>
<td>0.465</td>
</tr>
<tr>
<td>4</td>
<td>0.042</td>
<td>0.431</td>
<td>0.431</td>
<td>0.033</td>
<td>0.515</td>
<td>0.514</td>
</tr>
<tr>
<td>5</td>
<td>0.057</td>
<td>0.422</td>
<td>0.423</td>
<td>0.010</td>
<td>0.643</td>
<td>0.644</td>
</tr>
<tr>
<td>6</td>
<td>0.074</td>
<td>0.418</td>
<td>0.420</td>
<td>-0.040</td>
<td>0.915</td>
<td>0.916</td>
</tr>
<tr>
<td>7</td>
<td>0.092</td>
<td>0.414</td>
<td>0.419</td>
<td>-0.051</td>
<td>1.102</td>
<td>1.103</td>
</tr>
</tbody>
</table>

Alternative Efficient Instrumental Variable Estimator
Monte Carlo

\[ A_j(x) = x^{j-1}, \quad j = 2, 3, \ldots, 7; \quad \varphi = 0.5; \quad n = 200 \]

<table>
<thead>
<tr>
<th>( \tau )</th>
<th>oiv Bias</th>
<th>oiv Std. Dev.</th>
<th>oiv RMSE</th>
<th>omd Bias</th>
<th>omd Std. Dev.</th>
<th>omd RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.000</td>
<td>0.327</td>
<td>0.327</td>
<td>0.000</td>
<td>0.327</td>
<td>0.327</td>
</tr>
<tr>
<td>3</td>
<td>-0.001</td>
<td>0.309</td>
<td>0.309</td>
<td>0.001</td>
<td>0.324</td>
<td>0.324</td>
</tr>
<tr>
<td>4</td>
<td>0.003</td>
<td>0.302</td>
<td>0.302</td>
<td>0.000</td>
<td>0.371</td>
<td>0.371</td>
</tr>
<tr>
<td>5</td>
<td>0.006</td>
<td>0.296</td>
<td>0.296</td>
<td>0.009</td>
<td>0.471</td>
<td>0.471</td>
</tr>
<tr>
<td>6</td>
<td>0.014</td>
<td>0.294</td>
<td>0.294</td>
<td>-0.051</td>
<td>0.642</td>
<td>0.643</td>
</tr>
<tr>
<td>7</td>
<td>0.024</td>
<td>0.293</td>
<td>0.294</td>
<td>-0.009</td>
<td>0.749</td>
<td>0.750</td>
</tr>
</tbody>
</table>

Alternative Efficient Instrumental Variable Estimator