Instrumental variable (IV)

IV regression:

- Bayesian approach
- Bayes-Stein shrinkage
- Decision-theoretic methods
- Method of moments
- Dirichlet process mixtures
- Monte Carlo simulation

Bayesian versus classical:

- Lindley and El-Sayyad (1968)
- Kleibergen and Zivot (2003)

Related problems:

- Errors-in-the-variables models
- Co-integration
- Reduced rank regression

Bayesian advances

Drèze (1976) Bayesian limited information analysis of the simultaneous equations model. *Econometrica*, 44, 1045-1075.

Drèze and Richard (1983) Bayesian analysis of simultaneous equations systems. In *Handbook of Econometrics*, volume 1, edited by Z. Griliches and M.D. Intrilligator. Amsterdam: Elsevier Science.

Geweke (1996) Bayesian reduced rank regression. *Journal of Econometrics*, 75, 121-146.

Kleibergen and van Dijk (1998) Bayesian simultaneous equations analysis using reduced rank structures. *Econometric Theory*, 14, 701-743.

Chao and Phillips (2002) Posterior distribution in limited information analysis of the simultaneous equations model using Jeffreys prior. *Journal of Econometrics*, 87, 49-86.

Kleibergen and Zivot (2003) Bayesian and classical approaches to instrumental variable regression. *Journal of Econometrics*, 114, 29-72.

Chamberlain and Imbens (2004) Random effects estimators with many instrumental variables. *Econometrica*, 72, 295-306.

Hoogerheide, Kaashoek and van Dijk (2007) On the shape of posterior densities and credible sets in instrumental variable regression models with reduced rank. *Journal of Econometrics*, 139, 154-180.

Sims (2007) Thinking about instrumental variables. *Technical report*, Princeton University.

Conley, Hansen, McCulloch and Rossi (2008) A semi-parametric Bayesian approach to the instrumental variable problem. *Journal of Econometrics*, 144, 276-305.

Textbook material

Lancaster (2004).
An Introduction to Modern Bayesian Econometrics.
Chapter 8

Rossi, Allenby and McCulloch (2005).
Bayesian Statistics and Marketing.
Chapter 7

Koop, Poirier and Tobias (2007).
Bayesian Econometric Methods.
Chapter 14

Greenberg (2008).
Introduction to Bayesian Econometrics.
Chapter 11

Li and Tobias (2011).
Bayesian Methods in Microeconometrics.
Handbook of Bayesian Econometrics,
Chapter 6, Section 2.5.

Lopes, McCulloch and Carvalho (2012).
A First Course in Bayesian Econometrics.
Wiley. Chapter ?

Literature
IV
Bayes
Textbooks

IV regression
Endogeneity
Identification
Reduced form
Posterior inference
Illustration I

IV regression

Let y_i be the response variable and x_i the (endogenous) regressor obeying, for $i = 1, \ldots, n$, the system of recursive equations:

$$
\begin{aligned}
x_i &= z_i'\delta + \varepsilon_{1i} \\
y_i &= \gamma + \beta x_i + \varepsilon_{2i},
\end{aligned}
$$

with z_i a p-dimensional vector of instruments, related to x_i but independent of ε_{2i}.

For simplicity an intercept is included in z, such that there are, in fact, only $p - 1$ instrumental variables in the above structure.

The coefficient β is the causal effect parameter.

Endogeneity

Distinction to standard multivariate regression:

$$cov(\varepsilon_{1i}, \varepsilon_{2i}) \neq 0$$

For simplicity, let us assume that $\varepsilon_i = (\varepsilon_{1i}, \varepsilon_{2i}) \sim N(0, \Sigma)$:

$$\Sigma = \begin{pmatrix} \sigma_{11} & \rho\sqrt{\sigma_{11}\sigma_{22}} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}.$$

Therefore,

$$\begin{aligned} \varepsilon_{2i} &= \frac{\sigma_{12}}{\sigma_{11}}\varepsilon_{1i} + \sqrt{\sigma_{22}(1-\rho^2)}u_i \\ &= \frac{\sigma_{12}}{\sigma_{11}}(x_i - z_i'\delta) + \sqrt{\sigma_{22}(1-\rho^2)}u_i \end{aligned}$$

The recursive system of equations

$$\begin{aligned} x_i &= z_i'\delta + \varepsilon_{1i} \\ y_i &= \gamma + \beta x_i + \varepsilon_{2i}, \end{aligned}$$

can now be rewritten as

$$\begin{aligned} x_i &= z_i'\delta + \varepsilon_{1i} \\ y_i &= \gamma + \left(\beta + \frac{\sigma_{12}}{\sigma_{11}}\right) x_i + z_i'\left(-\frac{\sigma_{12}}{\sigma_{11}}\delta\right) + u_i \end{aligned}$$

with u_is iid $N(0, \sigma_{22}(1-\rho^2))$ and independent of ε_{1i}s.

"Endogeneity" bias:

Information of x_i that is correlated with ε_{2i} should not be used when learning about β.

Identification

Let $\sigma_{2|1} = \sigma_{22}(1-\rho^2)$.
Learn (δ, σ_{11}) from $x_i|z_i \sim N(z_i'\delta, \sigma_{11})$.
Learn $(\gamma, \beta^*, \delta^*, \psi)$ from $y_i|x_i, z_i \sim N(\gamma + \beta^* x_i + z_i'\delta^*, \psi)$.

Finally, let

$$
\begin{aligned}
\sigma_{12} &= -\sigma_{11}\delta_j^*/\delta_j \qquad \text{(for any } j\text{)} \\
\beta &= \beta^* - \sigma_{12}/\sigma_{11} \\
\sigma_{22} &= \sigma_{2|1} + \sigma_{12}^2/\sigma_{11}
\end{aligned}
$$

The model is not identified in the limiting case of $\delta = 0$. More generally, if the instruments z explain only a small portion of the variability of x (weak instrument case), then the likelihood function is concentrated around $\beta^* = \beta + \sigma_{12}/\sigma_{11}$ for some estimable constant β^*.

Reduced form

The recursive system of equations

$$\begin{aligned} x_i &= z_i'\delta + \varepsilon_{1i} \\ y_i &= \gamma + \beta x_i + \varepsilon_{2i}, \end{aligned}$$

can also be written as (reduced form)

$$\begin{aligned} x_i &= z_i'\pi_x + \nu_{1i} \\ y_i &= \gamma + z_i'\pi_y + \nu_{2i}, \end{aligned}$$

where $\pi_x = \delta$, $\nu_{1i} = \varepsilon_{1i}$, $\pi_y = \beta\delta$ and $\nu_{2i} = \beta\varepsilon_{1i} + \varepsilon_{2i}$.

The (triangular) relation between ε_i and ν_i is

$$\nu_i = \begin{pmatrix} 1 & 0 \\ \beta & 1 \end{pmatrix} \varepsilon_i = B\varepsilon_i$$

Posterior inference

The reduced form errors ν_i are iid $N(0, \Omega)$

$$\Omega = B\Sigma B' = \begin{pmatrix} \sigma_{11} & \beta\sigma_{11} + \sigma_{12} \\ \beta\sigma_{11} + \sigma_{12} & \beta^2\sigma_{11} + 2\beta\sigma_{12} + \sigma_{22} \end{pmatrix}$$

β and Σ are intertwined in the reduced form and independent priors for both parameters would be counter-intuitive.

Rossi *et al.* (2005) suggests

$$\begin{aligned} \delta &\sim N(d_0, D_0) \\ (\gamma, \beta)' &\sim N(b_0, B_0) \\ \Sigma &\sim IW(v_0, \Sigma_0)^1 \end{aligned}$$

for known hyperparameters d_0, D_0, b_0, B_0, v_0 and Σ_0.

[1] $p(\Sigma) \propto |\Sigma|^{-(v_0+p+1)/2} \exp\{-0.5 \text{tr} \Sigma_0 \Sigma^{-1}\}$, for $v_0 > p$. If $v_0 > p + 3$, $E(\Sigma) = \Sigma_0/(v_0 - p - 1)$.

Full conditionals

Error variance Σ

$$(\Sigma|\gamma, \beta, \delta, \text{data}) \sim IW(v_0 + n, \Sigma_0 + S)$$

where

$$S = \sum_{i=1}^{n} \varepsilon_i \varepsilon_i'$$

and data $= \{(x_i, y_i, z_i); i = 1, \ldots, n\}$.

Full conditionals

Regression parameters (γ, β)

$$(\gamma, \beta | \delta, \Sigma, \text{data}) \sim N(b_1, B_1),$$

where

$$
\begin{aligned}
B_1^{-1} &= B_0^{-1} + \sum_{i=1}^{n} \tilde{x}_i \tilde{x}_i' \\
B_1^{-1} b_1 &= B_0^{-1} b_0 + \sum_{i=1}^{n} \tilde{x}_i \tilde{y}_i \\
\tilde{x}_i &= \frac{1}{\sigma_{2|1}} (1, x_i)' \\
\tilde{y}_i &= \frac{1}{\sigma_{2|1}} \left(y_i - \frac{\sigma_{12}}{\sigma_{11}} (x_i - z_i' \delta) \right) \\
\sigma_{2|1} &= \sigma_{22}(1 - \rho^2).
\end{aligned}
$$

Full conditionals

Regression parameter δ

$$(\delta|\gamma, \beta, \Sigma, \text{data}) \sim N(d_1, D_1),$$

where

$$\begin{aligned}
D_1^{-1} &= D_0^{-1} + \sum_{i=1}^n \tilde{z}_i \tilde{z}_i' \\
d_1 &= D_1 (D_0^{-1} d_0 + \sum_{i=1}^n \tilde{z}_i \tilde{x}_i) \\
\tilde{x}_i &= \frac{1}{\sigma_{1|2}} \left(x_i - \frac{\sigma_{12}}{\sigma_{22}} (y_i - \gamma - \beta x_i) \right) \\
\tilde{z}_i &= \frac{1}{\sigma_{1|2}} z_i \\
\sigma_{1|2} &= \sigma_{11}(1 - \rho^2).
\end{aligned}$$

Illustration I

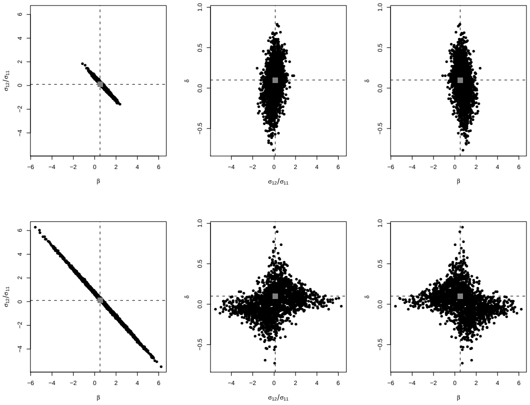
$n = 200$ simulated observations based on $\gamma = 1.0$, $\beta = 0.5$, $\sigma_{11} = \sigma_{22} = 1$, $\delta = (1.0, 0.1)'$ (one weak instrument) and $\rho = 0.1$ (low degree of endogeneity).

The prior hyperparameters are $d_0 = b_0$ and $D_0 = B_0 = 25I_2$, suggesting relatively low prior information about δ, β and γ.

For Σ two priors are entertained:
$v_0 = 3.00000$ and $\Sigma_0 = 3.00000I_2$ (relatively vague prior)
$v_0 = 0.00001$ and $\Sigma_0 = 0.00001I_2$ (non-informative prior).

The independent priors on γ, β and the components of δ are all $N(0, 25)$.

Top row: relatively vague prior.
Bottom row: non-informative prior.



MCMC run for 200000 draws and keeping every 100th.

Findings

The marginal posterior distributions of β and σ_{12}/σ_{11} becomes flatter as the values of δ approach zero.

The relatively vague prior turns out to be rather informative when compared to the non-informative prior:

Relatively vague prior: posterior of β roughly between $(-2, 2)$
Non-informative prior: posterior of β roughly between $(-6, 6)$.

The variability of σ_{12}/σ_{11} is also greatly affected by the choice of the prior on Σ.