Are You Sure About That? Using Bayesian Methods to Account for Parameter Uncertainty in Instrumental Variable Regression

Rick McAlexander

November 29, 2016

The problem

Traditional instrumental variables (IV) regression assumes to direct effect of instrument on dependent variable (aka the exclusion restriction)

In practice, it is unlikely that the direct effect is exactly equal to zero.

This means that our inferences from traditional IV regression are more uncertain than the ones reported.

The solution

Instead of assuming no direct effect of instrument on dependent variable, lets relax this (the exclusion restriction).
One option is to put bounds on size of violation. What if we are concerned not about size of violation, but uncertainty?
The solution: use a Bayesian approach. Express direct effect from insturment on DV as a parameter.
Specify a full prior distribution on that parameter. Allows us to make more accurate inferences and account for uncertainty.

IV review

What is IV regression? IV are used when there is an endogenous independent variable.

IV can make consistent and unbiased estimates of the effect of an endogenous independent variable.

It can even make causal claimes, like estimates of treatment effects, under certain assumptions.

Requires a good instrument: a variable that is exogenous to the endogenous independent variable, correlated with it, and has no direct effect on the dependent variable. (Hard to come by in many political science applications)

Assumptions and Terms

Let the instrument be denoted Z, the endogenous independent variable be X, and the independent variable be Y.

IV assumes Z is correlated with X, and conditional on X, Z is independent of Y.

However, let us introduce a parameter γ. This parameter represents the direct effect of Z on Y.

That is, conditional on X, how much is a one unit increase of Z related to Y?

The exclusion restriction assumes that γ is exactly equal to zero, always and forever.

Gamma

Frequentist IV assumes γ takes on a specific value, which just happens to be 0.

Some works allow γ to take on different values, as a way to measure the bounds of the violation.

My Bayesian approach: specify a full prior normal distribution on γ.

Allow it to be centered at 0, but incorporate our uncertainty about exclusion restriction violations via the standard deviation.

More Review

Why do we even need the exclusion restriction? Frequentist answer: identifiability.

Structual form of IV model:

$y_i = \beta * x_i + \varepsilon_i$

$x_i = \zeta * z_i + \upsilon_i$

What is relationship between covariances of data and parameters in the model?

$$E[xz] = \zeta$$
$$E[yz] = \beta \cdot \zeta + E[\varepsilon \cdot z]$$
$$E[yx] = \beta + \zeta \cdot E[\varepsilon \cdot z] + E[\varepsilon \cdot \upsilon]$$

Four parameters on right hand side, but three observable covariances.

Frequentist solution is to assume that $E[\varepsilon \cdot z] = 0$

Why go Bayesian?

Bayesian statistics turn problems of identifiability into problems of computation.

Frequentist statistics makes inferences about parameters from data.

When data have nothing to say about a parameter, there are identifiability problems.

Frequentist solution is to put bounds or assume a parameter equals zero.

Bayesian solution is to 'fill in' the parameter values with a prior distribution. The prior speaks when the data cannot.

IMPORTANT: An assumption that a parameter takes on a specific value is a dogmatic prior.

Computation

How do we compute posterior distribution of our IV model when we are uncertain about exclusion restriction violation?

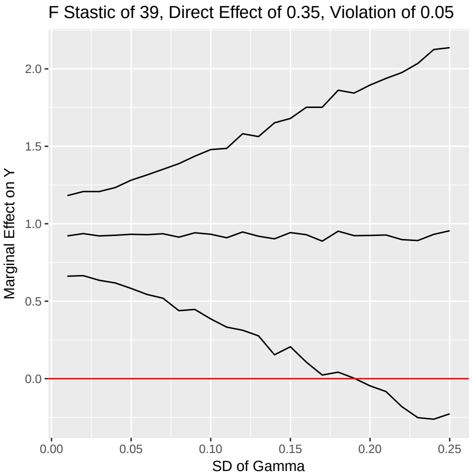
Traditional answer: Write your own gibbs sampler (uses complete conditionals of each parameter to sample from a stationary distribution that is the posterior distribution). Very painful to write for each problem we face. Same goes for variational inference.

Solution: use stan. Embarassingly easy to code up in R. Uses NUTS/HMC sampler. Fast and flexible.

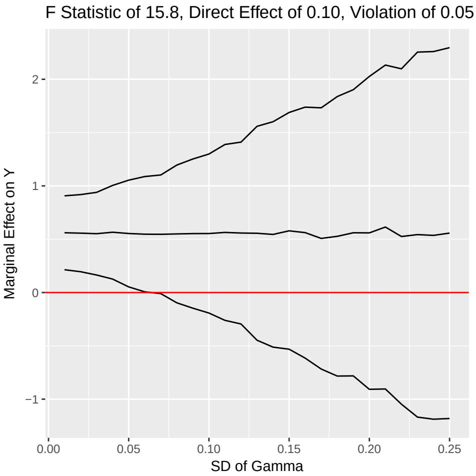
stan code

```
data{
  int n;
  vector[n] x;
  vector[n] y;
  vector[n] z;
  real gamma_sig;
}
parameters{
  real beta;
  real pi;
  real<lower=0> gamma;
}
model{
  y ~ normal(z * pi * beta + z*gamma , 1);
  x ~ normal(z * pi, 1);
  gamma ~ normal( 0 , gamma_sig);
}
```

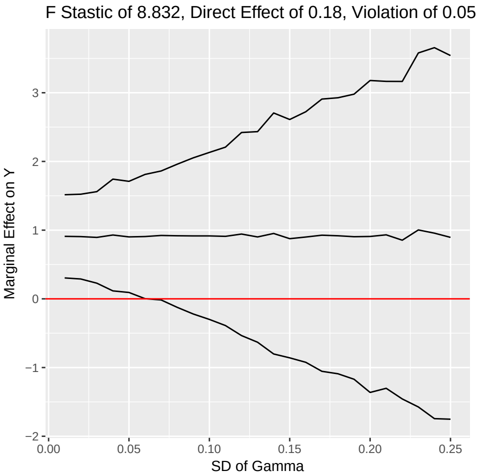
How do our inferences change?



How do our inferences change? Cont.



How do our inferences change? Cont.



Wait

Are you telling me that if there is uncertainty about whether the exclusion restriction is violated, this does not bias our (frequentist) parameter estimates?

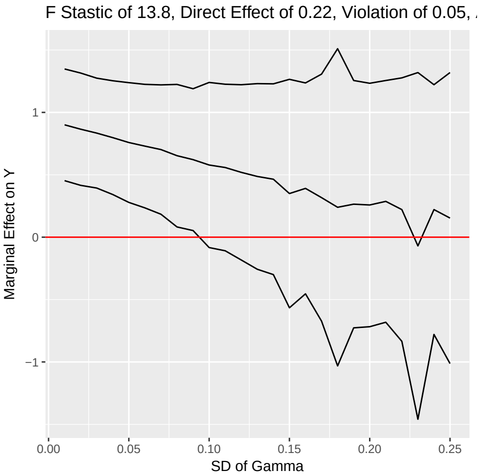
For Bayesians, uncertainty about the exclusion restriction has no effect on the mode of the posterior distribution?

Yes!

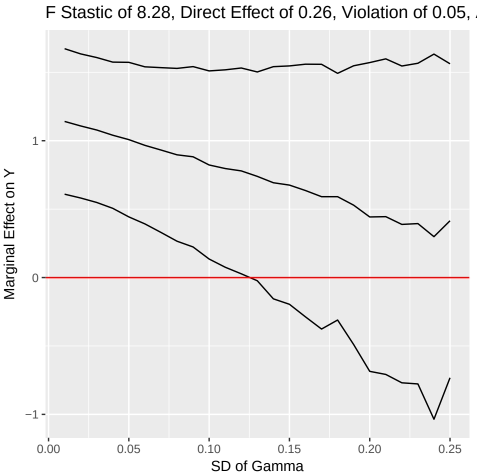
However

We are rarely concerned with symmetric uncertainty.
Practically, we care about whether there is a direct effect in the same direction as the indirect effect.
What happens when we force γ to be strictly positive?

How do our inferences change when γ is positive?



How do our inferences change when γ is positive? Cont.



What's the moral of the story?

Uncertainty symmetric around a mean of zero about whether the exclusion restriction is violated has no effect on point estimate/mode of posterior distribution.

Using informative priors that reflect our actual knowledge of the world is more useful, and shows how inferences change as uncertainty increases.

Parameters should not be assumed to be zero purely for convenience. Informative priors work better.

Important factor in how reliable IV estimates are: strength of second stage effect.

Go full Bayesian!