

Re: Logistic regression with instrumental variables, or...?

Source: <http://sci.tech-archive.net/Archive/sci.stat.math/2004-12/0648.html>

From: AK (andreas1973_at_hotmail.com)

Date: 12/15/04

Date: Wed, 15 Dec 2004 14:48:36 +0100

On Tue, 14 Dec 2004 17:53:33 -0800, Dick Startz
<richardstartz1@comcast.net> wrote:

>On 14 Dec 2004 17:37:37 -0800, andreas1973@hotmail.com (AK) wrote:

>

>>I have some problems with a paper where the main aim is to find out
>>whether unemployed people have a higher risk of abstaining from
>>consulting a physician than employed, when the perceived need for a
>>consultation is accounted for.

>>

>>In a logistic regression analysis with "unmet care needs" as
>>dependent variable (binary, 0/1) are employment status and symptoms
>>of disease explanatory variables.

>>

>>A reviewer comments:

>>

>>"It seems reasonable to suggest that these symptoms both have an
>>impact on the probability of being unemployed and an impact on the
>>probability of having unmet care needs. Hence, it is problematic to
>>consider employment status as an exogenous variable in the analysis,
>>as the author does. There are actually two equations involved: one
>>equation determining the probability of being unemployed and a second
>>equation determining the probability of having unmet care needs. The
>>variable describing employment status is then likely to be correlated
>>with the random term in the second equation and the assumptions of the
>>regression model are violated. It seems to me that the author needs to
>>take the endogeneity into account by for instance introducing some
>>kind of
>>instrumental variable for employment status."

>>

>>First, is it really a problem? In an OLS regression are the
>>assumptions of the regression model violated if an explanatory variable
>>is correlated with the error term. But a logistic regression is not
>>estimated by least squares but by maximum likelihood. Then is should
>>not be a problem if an explanatory variable in a logistic regression is
>>correlated with the error term. Or am I wrong?

>>
>>*And if it is a problem, how should I solve this problem? Especially if*
>>*I don't find any instrumental variables? Are there software available*
>>*to use for this?*
>>
>>*Could it be estimated in a generalized linear model, using GENMOD in*
>>*SAS?*
>>
>>*Please help!*
>
>*The problem is fundamentally the same in logit as in OLS.*
>*In principle, you can write out the two equations as the referee has*
>*suggested and estimate them jointly by maximum-likelihood. The problem*
>*is that you do need instruments to get this to work.*
>*-Dick Startz*
>
>-----
>*Richard Startz RichardStartz@comcast.net*
>*Lundberg Startz Associates*

Thanks for your answer. Is it correct that you mean that for a logistic regression estimated by maximum likelihood it is a requirement that an explanatory variable should not be correlated with the error term (just as it is for linear regression with least squares)? Do you have any references for this that I could check?

Is it possible to solve this problem without using instrumental variables?

What software could I use?

Thanks again.

Andreas