Practice Pattern Instrumental Variables for Comparative Effectiveness
Carol Conell. Division of Research. Kaiser Permanente Northern California, Oakland, CA
Alexander C. Flint. Department of Neuroscience. Kaiser Permanente, Redwood City, CA

ABSTRACT
Selection bias is a major problem limiting comparative effective studies (CES) using observational data. Fortunately, instrumental variable regression provides an excellent way to address this problem if one can identify good instrumental variables. Since clinical data is routinely collected in many healthcare settings, variation in health care providers' practice patterns can often function as instruments to identify the relative efficacy of alternative practices. This can work well provided uncertainty about the utility of specific treatments induces variation among high quality practitioners.

This paper provides a practical guide to using electronic health record (EHR) data to create appropriate practice pattern instrumental variables. We illustrate the approach using variation during emergency hospitalizations for intracranial hemorrhage (ICH) to investigate whether inpatient statins improved outcomes.

A set of candidate practice pattern indicators are created using a macro. PROC QLIM is used to fit an instrumental variable model. All analyses are performed in SAS® 9.3.

The specific approach fits any situation with a binary outcome and a binary treatment of unknown efficacy used for some but not all cases. With minor adaptation it extends to other CES within healthcare or other contexts where information on detailed practices and outcomes is routinely collected.

INTRODUCTION
One of the greatest drawbacks of observational studies compared with randomized clinical trials (RCT) is the potential confounding of results due to selection of treatments by indication, often based on factors that are apparent to the clinician but poorly documented. Instrumental variable models can provide a consistent, conservative test for the relative effect of alternative treatments despite the presence of this unmeasured confounding, provided suitable instruments are available. "Practice pattern" variables, i.e., measures of how likely a specific alternative is to be chosen by a particular provider at a given time can be extracted from the EHR, and function well as instruments when there is substantial variation in practices by provider and sufficient uncertainty about the best practices to minimize any association between providers' overall quality and their choice between the alternatives (Burgess et al. 2013).

We examine a common situation that involves a binary outcome and a binary treatment with unknown efficacy used for some but not all cases. We illustrate this with data from a study of inpatient statin use for patients presenting with intracranial hemorrhage (ICH) where death within 30 days is the outcome. Throughout the paper, we begin by considering the key features of a problem that affect the best analytic approaches and then focus on the illustrative case in order to demonstrate what we did and how we used SAS®. We include brief discussions of key concepts like instrumental variables along with references to more detailed discussions and existing technical papers on how to perform instrumental variable regression using SAS®.

THE TECHNICAL CHALLENGE: WHEN AND HOW TO USE THIS APPROACH.
When there is selection by indication, analysts need an easy way to identify an appropriate practice pattern instrument in order to evaluate the relative efficacy of alternative treatments. This section lays out the questions you should ask to determine whether practice pattern instrumental variables are appropriate for a given problem and how best to apply the approach. In subsequent sections, we will ask and answer these questions for an exemplary case: evaluating whether inpatient statins benefit ICH patients.

First, decide whether to consider instrumental variables. If there are unmeasured individual factors that might affect whether patients receive the treatment of interest, the answer is yes. Evidence of substantial differences between the patients who do and do not receive the treatments ("imbalance") can be an important cue. If there is information available on key unmeasured potential confounders for a subpopulation, this can be used to check for imbalance, as we will illustrate.

Second, evaluate whether practice patterns are appropriate instruments. In situations where there is substantial consensus in the literature throughout the observation period about the best practices, using practice variation as an instrument is usually inappropriate, since the “Recommended practice” will tend to correlate with higher quality of treatment. In contrast, when the best practices are uncertain AND there is observed variation in practices by providers or over time, practice pattern indicators can provide excellent instruments. In these circumstances, practice patterns have high face validity as instrumental variable. Treatment practices as opposed to how the specific patient
was actually treated cannot affect outcomes directly. Consequently, any association between preferences for a specific treatment and outcomes is probably attributable to the relative efficacy of one treatment compared to the alternatives weighted by the relative probability of receiving that treatment. Put differently, when observational studies of comparative effectiveness are most useful, i.e., when there is little or no consensus about the best practices, practice patterns are likely to be a good source of instruments.

Third, identify appropriate practice pattern indicators based on a model of how treatments were selected. Start by determining who makes the decisions about what treatments to use. Most existing work in healthcare comparative effective studies (CES) focuses on individual practitioners, because data has been routinely collected for a relatively long period on outpatient prescriptions, including who prescribed the pharmaceutical for whom. Practice pattern variables require clear linkage to the entity responsible for a decision as well as to the recipient of the treatment. However, particularly for inpatient treatment, hospital or practice guidelines, including formulary rules, often dominate individual physicians’ decisions. Again, in some settings, specialists in particular diseases may be very influential even if decisions are officially ascribed to other physicians or healthcare works. In some situations, the linkage to a group is more reliable than the linkage to an individual. The legal documentation trail that treats individual physicians as responsible for decisions should not be confused with the effective source of decision-making. A combination of exploratory data analysis and information about the work flow behind decisions should be used to determine whether to look for individual level, clinic, practice group or hospital practice patterns. There are also at least three distinct types of practice pattern variables that can be created: providers can be classified by their observed propensity to use one of two treatments over the whole study period, providers can be classified as proponents of one or the other treatment during the period with the possibility of switching a limited number of times during the period, or providers’ preferences can be treated as varying and a short look back period can be used to indicate the current propensity.

Fourth, estimate the comparative efficacy of the treatment and alternatives with standard techniques for instrumental variable models using one or more practice pattern indicators as instruments.

**AN ILLUSTRATIVE CASE.**

We used practice pattern instrumental variable analysis to evaluate the efficacy of inpatient statins for health plan members hospitalized on an emergency basis for ICH at KPNC from 2002-11, comparing to no inpatient statins. During this period, uncertainty as to whether statins could be helpful for ICH patients was high, including some concern that they might actually be harmful. This encouraged variation in provider practices regarding inpatient statins, which could provide information on comparative efficacy. KPNC EHR provided rich data on treatment, outcome, and medical history.

**Figure 1. Interpreting the Relationship of Inpatient Statins to Mortality**

Figure 1 illustrates the problem that led to our developing practice pattern indicators to use in an instrumental variable model. Treated patients had better outcomes and this difference was robust to looking only at the death rate among those who survived overnight after admission. However, since there was no defined protocol determining who
received treatment, there was no reason to think that potential confounders were balanced. In particular, it seemed likely that at least part if not all of the association of treatment and outcome might reflect undocumented practices of not treating patients who were frailer or presented with higher severity. Moreover, we did not have information available on all the potential confounders.

WHY INSTRUMENTAL VARIABLE REGRESSION.

Since KPNC is a membership based health care provider and tracked diagnoses, utilization and pharmaceutical use systematically throughout this period, the initial analysis plan used a single equation multivariable logistic or probit regression technique to estimate the relative risk of death within 30 days for those who received inpatient statins, controlling for observable confounders. Specifically, we used the GENMOD procedure to estimate the relative risk of death for inpatient statins after controlling for the patient's age, relevant comorbidities, and receiving statins as an outpatient. (Throughout this presentation, we simplify the analysis by using only a few control variables—age, outpatient statin use, and two key comorbidities: for the full analysis see Flint et al (forthcoming), in which the instrumental variable analysis supplements a standard logistic regression analysis and adjusts for a much longer list of possible confounders.)

```plaintext
title "calculate relative risk of death within 30 days with inpatient statins";
title2 "controlling for outpatient statins, age and comorbidity";
PROC GENMOD data = dataset descending ;
class Treatment (Ref="No") studyid ;
model Outcome = Treatment controlvariables / dist = poisson link = log ;
   repeated subject = studyid / type = unstr ;
   estimate 'relative_risk_inpatient_statin ' Treatment 1 / exp;
   run;
```

The calculated relative risk (and 95 per cent confidence interval) of death as a result of receiving inpatient statins was .37 (.32-.43), using a single equation model controlling for measured confounders.

We were concerned that some of the positive association of survival with inpatient statins might reflect statins not being administered patients who were particularly frail or had a more severe bleed, i.e., negative selection bias among those with poorer mortality prospects. To check whether confounding by indication was plausible, we checked whether there were statistically significant differences in the distribution of observable potential confounders between those who were and were not treated (i.e., received inpatient statins), and found substantial differences in comorbidities. Although the multivariate model included these observed confounders, substantial imbalance in them is often viewed as an indirect indicator of imbalance in unobserved confounders. Enhancements of the EHR late in the study period also enabled us to systematically obtain information for about one quarter of the study subjects on the primary suspected unmeasured source of confounding: stroke severity. We found that treated subjects presented with less severe ICH than those who were untreated. I.e., there was substantial evidence that unmeasured clinical variation could be confounding the results of the analysis. Consequently, we decided to use an instrumental variable analysis to determine whether the effects we estimated from the single equation model would be robust.

CREATING INSTRUMENTAL VARIABLE INDICATORS.

The approach we adopted involves creating a family of elementary practice pattern indicators: the use of the treatment in a single relatively recent case treated by the same provider. This approach was pioneered in pharmacological studies, defining the physician as the provider (Brookhart et al 2007; Rassen et al 2008). Formally, the technique resembles “hot-deck” imputation for missing data. Specifically, we created 10 elementary practice pattern indicators (IVL1…IVLn) reflecting the immediately preceding case, the case before that, and so on up to the tenth preceding case, defining the hospital as provider.

Individually, each elementary practice pattern indicator was a “weak” instrument, i.e., accounted for only a small fraction of the variation in the likelihood that an individual received the treatment. By including all of these as instruments in the instrumental variable analysis, we implicitly created an instrument that used the best linear combination of treatment in the preceding 10 similar cases to predict treatment likelihood in the present case. The macro PRACPATTINSTRUMENTS (APPENDIX A) creates a family of instrumental variables with as many members as desired and evaluates the predictive power of the family of instruments. Optionally, one can select “similar” cases within the population based on control variables. In particular, in the substantive analysis we actually used instruments that reflected the most recent case where outpatient statin use matched the present subject, due to concerns about the interaction of inpatient and outpatient statin use and the impact of continuity of statin treatment on survival.
First, create a practice patterns database sorted by provider and time with the practice pattern indicator based on the lagged value of the treatment where there is a preceding case. In order to avoid missing information on the practice pattern at the beginning of the series, the macro uses the following case if necessary to fill in missing information.

Where information is available on the practice for a longer period or a larger set of subjects than the study population in the analysis, this is not necessary. One important point is that this logic requires that cases can be ordered by the time of treatment decisions with no ties. Since EHR times are usually recorded in minutes or even seconds, this is not likely to be a problem. In the absence of a unique time ordering, the macro would require modification to ensure that cases had a clear order imposed within providers. However, the underlying logic requires only obtaining a nearest neighbor for each case so that all cases can be used as sources of information and case weights, if not exactly equal, are close to equality.

Second, merge the instrumental variables IVL1…IVLn onto the analysis data set by Case. Note that Case must be identified using time as well as subject if study subjects can contribute more than one observation to the analysis.

The code below creates a single indicator based on the immediately preceding case with no control variables:

```sas
title "Create a practice pattern indicator";
title2 "Based on the previous case by the same provider ";

PROC SORT data= dataset (keep=Time Provider Treatment Case) nodupkey force
out=practicepatterns;
by Provider Time;
DATA practicepatterns;
set practicepatterns;
by provider time;
IVL0 = treatment ;
prvdrL0 = provider;
IVL1 = Lag(treatment );
prvdrL1 = Lag(provider);
if not( prvdrL0 = prvdrL1) then IVL1 = .;
```

Once we had the potential instruments available, we used the LOGISTIC procedure to verify that the family of practice pattern instruments accounted for a substantial fraction of variation in Treatment: the F-test was 69.7 with 10 df ($\alpha<.001$). The macro uses stepwise selection to check how many lags to include and, in this situation, indicated using nine. We also verified that substituting a practice pattern indicator for the treatment substantially reduced the imbalance in severity (the unobserved confounder) and in the measured potential confounders.

![Figure 2a. Severity by Treatment](image-url)
Figure 2a. Distribution of severity by Treatment for the provider’s previous patient

Figure 2 provides detail on the relationship of severity, the suspected source of unmeasured confounding, to the Treatment and to an Instrument, specifically, to the elementary instrumental variable indicating the treatment in the immediately proceeding case. The relationships were estimated using the GLM procedure. The F statistics show that there is a strong relationship between the severity of the ICH and whether a patient received treatment. Moreover, it is clear from Figure 2a that most treated subjects presented with minor or moderate severity, if minor is defined as 0–5 and moderate as 6–15. (The shaded area in each plot indicates the interquartile range and the range shown is based on the Tukey convention, which identifies outliers as outside the “normal range” if their distance from the nearest interquartile boundary exceeds 1.5 times the interquartile range (MgGill 1978). The diamonds indicate the means and the horizontal line inside the shaded area the means.) In contrast, figure 2b shows no imbalance in severity between those who were positive and negative for the instrumental variable based on treatment in the healthcare provider’s previous patient. Although we only examine the distribution of severity relative to one of the instruments, the construction of this family of instruments ensures that all are identically distributed and have the same joint distribution with the actual treatment.

We also examined whether measured potential confounders were significantly imbalanced by Treatment and whether they were balanced relative to the Instrument. There was significant imbalance by Treatment. The distribution of age was balanced for Treatment groups, but outpatient statin treatment and all relevant comorbidities were imbalanced. Substituting the instrument for the observed treatment substantially improved the balance of all the measured comorbidities. In fact, except for coronary artery disease, where there was a much reduced imbalance the instrument balanced all key covariates, there was, unsurprisingly, negligible imbalance between groups defined by treatment in previous cases. Consequently, we did not bother calculating a formal measure of total improvement in balance from using the instruments. Where there is more ambiguity, the recommended practice measures the reduction in imbalance as change in the Mahalanobis distance when groups defined by instruments are substituted for groups defined by treatment as a measure of the reduction in imbalance (Huybrechts 2014).

ESTIMATING THE TREATMENT EFFECT: INSTRUMENTAL VARIABLE REGRESSION.

We used the QLIM procedure to estimate the relative risk associated with the treatment based on instrumental variable regression that incorporated the instrumentals created by the PRACPATTINSTRUMENTS macro. PROC QLIM automatically allows for a possible correlation between the error terms in the equations predicting Treatment and Outcome. A macro PracticePatt_iv is available as Appendix B.

Using an instrumental variable model requires specifying two equations. The first equation predicts Treatment as a function of the InstrumentalVariables plus the ControlVariables. The macro specifies a linear regression model...
predicting the probability of Treatment, which previous studies have shown often works better than using a probit model for Treatment. If the likelihood of treatment was very close to 0 or 1 for a large fraction of the cases, this specification would be problematic. The macro would then need modification to specify a probit model for the first equation, specifying that the dependent variable in the Treatment equation (Treatment) is a discrete variable with a normal distribution, as well as appropriate modification where effects are calculated.

In the ICH inpatient statin investigation, treatment estimates were bounded well away from both 0 and 1 in all but a very small fraction of cases enabling us to use the continuous prevalence estimation. (For nine tenths of the cases, the estimated likelihood of treatment was between .08 and .75, and only .7% had estimates under 0, the lowest being -.008).

The second equation predicts Outcome based on the probability of Treatment, as estimated in the instrumental variable model, and the Control variables. This is specified as a probit equation, by indicating that the dependent variable (Outcome) is a discrete variable with a normal distribution.

The output statement creates a dataset (Estimates) containing predicted values and residuals for each case. By default the EXPECTED parameter creates a variable Expct_Treatment for the endogenous variable Treatment. This makes it easy to check that the resulting estimates of Treatment probabilities are well-behaved. In particular, we use the UNIVARIATE procedure to examine the distribution and to determine whether there were cases where the predicted values for the probability fall outside the legal range of [0,1].

The ODS OUTPUT statement creates a dataset ParameterEstimates. We use this to calculate the point value and a 95 per cent confidence interval for relative risk associated with Treatment.

The Test statements provide a measure of the strength of the instrumental variables as a family and test whether the errors for the Treatment and Outcome equations are correlated.

title "Estimate relative risk of Outcome given Treatment ";
title2 "controlling for Controlvariables ";
PROC QLIM data= Dataset ;
ods output ParameterEstimates = ParameterEstimates;
model Treatment = Controlvariables Instrumentalvariables ;
model Outcome = Treatment Controlvariables /discrete (distribution=normal) ;
output out= Estimates PREDICTED EXPECTED XBETA ;

TREATMENT_F_IV : test Treatment.Ivs=0 / LR ;
title3 "Evaluate the strength of the statistical relationship between the practice pattern instrument set and the treatment ";
Uncorrelated Errors: test _rho=0 / LR;
PROC UNIVARIATE data= Estimates
var exptCTreatment; run;
run;
title3 "Calculate relative risk and confidence interval for effect of Treatment"

DATA treatmenteffect;
set Parameterestimates(where={parameter='Outcome.Treatment'});
195=estimate - 1.96* stderr;
u95= estimate + 1.96 * stderr;
rel_riskest = exp(estimate);
rel_risk_195 = exp(195);
rel_risk_u95 = exp(u95);
run;
PROC PRINT data=treatmenteffect; run;
run;

Table 1 shows key output. Bold is used to indicate the most important features. Always check for convergence before examining the parameter estimates, which are only interpretable if convergence occurred. Beyond that, check whether the distribution of the estimates from the instrumental variable estimation justifies using linear regression to predict the probability of Treatment. If not, the first equation, like the second should be specified as a probit equation,
by indicating that the dependent variable (Treatment) is a discrete variable with a normal distribution, and the calculation of the treatment effect will need to be revised appropriately.

The panel headed Parameter Estimates in the output provides information on all the parameters that were estimated, including coefficients for the two equations plus summary parameters that indicate whether the error terms for the two equations were correlated and the overall explanatory power of the model. The coefficient for the impact of each Predictor variable on the two Predicted variables is designated as Predicted.Predictor. Thus, the first parameter reported for Outcome equation is Outcome.Intercept. The estimated effect of Treatment, which is the primary parameter of interest, is referred to as Outcome.Treatment.

Using standard normal approximation theory, we calculated the relative risk of death as a result of receiving inpatient statins as .32, with a 95 per cent confidence interval of (.15, -.69). The point estimate for the effect of the treatment using the instrumental variable model is very close to that estimated by the single equation multivariate probit model controlling for measured controls. However, the confidence interval is substantially wider. The agreement of the instrumental variable regression with the single equation multivariate risk model suggests that unmeasured confounding was not a major problem in this situation.

The panel headed Test Results contains the information produced by the two test statements. Treatment_Ivs=0 measures the strength of the instrumental variables. We find that collectively they explain substantial variation in Treatment, controlling for measured control variables. Uncorrelated_Errors tests whether there was a significant correlation between the errors in the Treatment equation and Outcome equations. We found no apparent correlation between these error terms given the specification, which indicates no endogeneity bias. This is consistent with the finding that the estimate for the impact of the treatment on the outcome obtained using instrumental variable regression was very close to the estimate from ordinary probit regression.
Table 1. Key Output from Instrumental Variable Model:

<table>
<thead>
<tr>
<th>Model Fit Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Endogenous Variables</td>
</tr>
<tr>
<td>Endogenous Variable</td>
</tr>
<tr>
<td>Number of Observations</td>
</tr>
<tr>
<td>Log Likelihood</td>
</tr>
<tr>
<td>Optimization Method</td>
</tr>
<tr>
<td>AIC</td>
</tr>
<tr>
<td>Algorithm converged.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
</tr>
<tr>
<td>Outcome.Intercept</td>
</tr>
<tr>
<td>Outcome.Treatment</td>
</tr>
<tr>
<td>Outcome.Outpatient Statin</td>
</tr>
<tr>
<td>Outcome.Ln(Age)</td>
</tr>
<tr>
<td>Outcome.Atrial Fibrillation</td>
</tr>
<tr>
<td>Outcome.Coronary Artery Disease</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Test Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test</td>
</tr>
<tr>
<td>TREATMENT_F_iv</td>
</tr>
<tr>
<td>UNCORRELATED_Errors</td>
</tr>
</tbody>
</table>

DISCUSSION
An important limitation on the use of practice pattern instrumental variables for comparative effectiveness involves the need to link decisions to providers. Another limitation is the fact that although the design of the provider practice measures described here basically ensures balancing individual case related unmeasured attributes, including unmeasured attributes of patients, it does not balance unmeasured quality differences among providers. For that reason, where there is agreement on comparative efficacy this approach will not work.

CONCLUSION
“Practice pattern” instrumental variables can provide a consistent, conservative test for relative treatment efficacy in many observational contexts where the concern that the estimates of ordinary multivariate generalized regression models are confounded by selection by indication. This is particularly useful when some potential confounders cannot be observed consistently. Using practice pattern instrumental variables enabled us to validate the finding from the multivariate single equation model that inpatient statin use improved 30 day survival in KPNC 2002-11.

Practice pattern based instrumental variable models are easy to implement using PROC QLIM in SAS® 9.3. They are an important tool for CES. However, care is required in interpreting the results, since the estimates only reflect the relative efficacy of treatments in situations where the treatment chosen varies substantially by provider.

REFERENCES


ACKNOWLEDGMENTS

We acknowledge support from the Kaiser Permanente Community Benefits Grant Program. The research used to illustrate the use of practice pattern instrumental variables was approved by the KFRI IRB.

RECOMMENDED READING

Instrumental Variables:


Provider Preference Measures:

- Davies NM et al. 2013. Physicians' prescribing preferences were a potential instrument for patients' actual prescriptions of antidepressants. Journal of Clinical Epidemiology 66,12:1386–96. Illustrates the use of multiple “hot deck” style indicators based on recent cases and compares this type with alternative instruments.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Carol Conell
Division of Research, Kaiser Permanente
2000 Broadway
Oakland, CA 94612
510 891-3574
FAX: 510-891-3606
Carol.Conell@kp.org

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.
APPENDIX A: PRACPATTINSTRUMENTS

PRACPATTINSTRUMENTS creates 2 related families of practice pattern instrumental variables
Both have members with lags identified by ivlag running from minivlag to maxivlag
One family consists of elementary practice pattern indicators
For each value of ivlag these indicators assume the value of Treatment associated with a case offset from the
current case by ivlag.
E.g., the indicator for ivlag=1 is based on the immediately preceding case meeting the required similarity criteria.

Required Input:
Data is read from a data set with name specified by macro variable Inda from a library specified as Inlib, using macro
variables.
Output is written to a library specified by macro variable outlib, which defaults to
inlib if not specified.
The names of the treatment, and the list of optional control variables are specified by the controlvariable macro
variable.
The prefix for the instrumental variables must be specified as the macro variable ivmoniker.
Values for the minimum and maximum instrumental variable lags are
specified by macro variables: minivlag and
maxivlag.
A unique id variable must be specified by macro variable &id.
The name of the ordering variable must be specified by macro variable &Time.

The input Data set must contain the following variables:
A provider variable specified by the macro variable &provider indicating who or what is responsible for treatment
decisions.
A unique id variable specified by macro variable &id
A time variable creating a unique ordering within the set of patients treated by each provider.
A dichotomous treatment variable &Treatment
A series of dichotomous instrumental variables &ivmoniker&minivlag,...,&ivmoniker&maxivlag
Any optional control variables as specified by &Controlvariables.
The format N0Y1f must be included if the elementary instrumental variables are to use and report codes 0 for No and
1 for Yes. Alternatively option nofmterr must be specified.

Usage:
proc format;
  value n0y1f
  0="No"
  1="Yes"
  ;
%PRACPATTINSTRUMENTS(inlib=,inda=,prvdr=,prvdrindic=,Treatment=,
cntrlv=,ncntrlv=,cntrlfmtl=n0y1f,
time=,&id=,cntrlvpx=,
cntrlf=n0y1f,minivlag=1,maxivlag=5 ,vers= ,outlib=)
run;
*********************************************************************/
%macro PracpattInstruments(inlib=,inda=,prvdr=,prvdrindic=,Treatment=,
cntrlv=,ncntrlv=,cntrlfmtl=n0y1f,
time=,&id=,cntrlvpx=,
cntrlf=n0y1f,minivlag=1,maxivlag=5,vers=,outlib=);
%do cntrlv=1 %to &ncntrlv;
%let cntrlindic=;
%let cntrlf=ncntrlv;
%let cntrlf=%upcase(%substr(&&cntrlv&cntrlv,1,1))
%let cntrlindic=&&cntrlf&cntrlv
%end;
%end;
%if &ncntrlv=0 %then %let cntrlvpfx=U ; /* universe */
%if %length(&cntrlvpfx)=1 %then %let cltPfx=&cntrlvpfx ;
%if %length(&cntrlvpfx)=0 %then %let cltPfx = &cltindic ;
%if %upcase(&Treatment)= IPSTATIN %THEN %LET TreatmentIndic = IP ;
%if %length(&prvdrindic)= 0 %then %let prvdrindic = %substr(&prvdr,1,1);
%if %length(&outlib)= 0 %then %let outlib = &inlib ;
proc sort data=&inlib.in .&inda(keep=&Time &prvdr &Treatment &cntrlvl &id ) nodupkey force out=&ctlPfx ;
by &cntrlvl &prvdr &Time ;
data &ctlPfx ; set &ctlPfx ;
by &cntrlvl &prvdr &Time ;
PredvalL0 = &Treatment ;
prvdrl0 = &prvdr ;
%do IvLag=&minIvLag %to &maxIvLag ;
%let prvivlag=%eval(&IvLag -1) ;
PredvalL&IvLag = lag(PredvalL&prvivlag ) ;
prvdrL&IvLag = lag (prvdrL&prvIvLag ) ;
if not( prvdrL0 = prvdrl&IvLag ) then PredvalL&IvLag = . ;
%end ;
/* repeat if necessary to get following values to supplement IvLag values */
proc sort data=&ctlPfx ;
by &cntrlvl &prvdr descending &Time ;
data &ctlPfx ; length
%do IvLag=&minIvLag %to &maxIvLag ;
&ctlPfx&TreatmentIndic._&prvdrindic.L&IvLag %end ;
3 ;
set &ctlPfx ; by &cntrlvl &prvdr descending &Time ;
%do IvLag=&minIvLag %to &maxIvLag ;
label &ctlPfx&TreatmentIndic._&prvdrindic.L&IvLag = "&Treatment Lag(&IvLag) by &prvdrindic " ;
%end ;
format
%do IvLag=&minIvLag %to &maxIvLag ;
&ctlPfx&TreatmentIndic._&prvdrindic.L&IvLag %end ;
noy1f. ;
%DO CNTRLV=1 %TO &NCNTRLV ;
FORMAT &&CNTRLV&CNTRLFMT&CNTRLV &CNTRLFMT&CNTRLV.. ;
%PUT " FORMAT &&CNTRLV&CNTRLV &CNTRLFMT&CNTRLV.. ; ";
%END ;
nextvalA0 = lag(&Treatment) ;
prvdrA0 = &prvdr ;
%do IvLag=&minIvLag %to &maxIvLag ;
%let prvivlag=%eval(&IvLag -1) ;
NextvalA&IvLag = lag(NextvalA&prvivlag ) ;
prvdrA&IvLag = lag (prvdrA&prvIvLag ) ;
if not( prvdrL0 = prvdrA&IvLag ) then NextvalA&IvLag = . ;
%end ;
%do IvLag=&minIvLag %to &maxIvLag ;
&ctlPfx&TreatmentIndic._&prvdrindic.L&IvLag = predvalL&IvLag ;
if predvalL&IvLag = . then
&ctlPfx&TreatmentIndic._&prvdrindic.L&IvLag = nextvalA&IvLag ;
&ctlPfx&TreatmentIndic._&prvdrindic.avL&IvLag =
sum(0
%do entry=&minIvLag %to &IvLag ;
, &ctlPfx&TreatmentIndic._&prvdrindic.L&entry %end ;
)/ &IvLag ;
%end ;
runtime;
APPENDIX B: PRACTICEPATTERN_IV

PracticePattern_IV fits an instrumental variable model using QLIM.

**Required Input:**

Data is read from a data set with name specified as Inda from library specified as Inlib, using macro variables.

Output is written to library specified by macro variable outlib, which defaults to inlib if not specified.

The names of the outcome, treatment, and the list of control variables specified by the controlvariables macro variable.

The preface for the instrumental variables must be specified as the macro variable ivmoniker.

The minimum and maximum lags for the instrumental variables are specified by macro variables: minivlag and maxivlag.

The input dataset must contain the following variables:

Control variables as specified by &Controlvariables;

A dichotomous variable &outcome

A dichotomous treatment variable &Treatment

A series of dichotomous instrumental variables &ivmoniker&minivlag,.....,&ivmoniker&maxivlag

proc sort nodupkey force
data=&ctlPfx (keep=&prvdr &id &Treatment
%do IvLag=&minIvLag %to &maxIvLag;
&ctlPfx&TreatmentIndic._&prvdrindic.L&IvLag
&ctlPfx&TreatmentIndic._&prvdrindic.avL&IvLag
%end;
)
out=&outlib..&ctlPfx&TreatmentIndic._&prvdrindic&vers ;
by &id &prvdr &Treatment ;
run;
title "check association of possible ivs and original measure ";
title2 "for hotdeck lagged indicators backward selection";
proc logist data=&outlib..&ctlPfx&TreatmentIndic._&prvdrindic&vers ;
model &Treatment ( event=" Yes") =
%do IvLag=&minIvLag %to &maxIvLag;
&ctlPfx&TreatmentIndic._&prvdrindic.L&IvLag
%end;
/selection = backward sequential ; run;
title "check association of possible ivs and original measure ";
title2 "for hotdeck lagged indicators ";
proc logist data=&outlib..&ctlPfx&TreatmentIndic._&prvdrindic&vers ;
model &Treatment ( event=" Yes") =
%do IvLag=&minIvLag %to &maxIvLag;
&ctlPfx&TreatmentIndic._&prvdrindic.L&IvLag
%end;
/selection = forward sequential ; run;
%mend PracpattInstrume
**Estimation Details:**

The model assumes that &Treatment and &Outcome are endogenous and &controlvariables and &ivmoniker&minivlag,...&ivmoniker&maxivlag are exogenous.

The probability of &treatment is estimated, using a linear regression model.

The relative risk of &Outcome given &Treatment is estimated using a probit model.

Key output and data sets used in describing results are saved.

**Usage:**

```%Practice_Pattern_Iv (ivmoniker=&wussippfx,minivlag=1,maxivlag=10, treatment=&wusstreatment ,project=_ichss , controlvariables = &wussctlv , outcome= &wussoutcome , inlib=&wussinlib,inda=&wussinda , outlib=&wussoutlib,vers=&version) run;```

```%macro Practice_Pattern_Iv (ivmoniker=&wussippfx,minivlag=1,maxivlag=10, treatment=&wusstreatment ,project=_ichss , controlvariables = &wussctlv , outcome= &wussoutcome , inlib=&wussinlib,inda=&wussinda , outlib=&wussoutlib,vers=&version); /* Initialize macro variables */ %let dataset = &&inlib...&inda ; %put "data read from &dataset "; footnote "initialize instrumental variables "; %do ivlag=&minivlag %to &maxivlag ; %let iv&ivlag = &&ivmoniker..&ivlag ; %end; /* set up outlib by default to save info in same library as input */ %if %length(outlib)=0 %then %let outlib=&inlib ; title2 "version &vers iv variables: iv&minivlag ... iv&maxivlag "; PROC QLIM data=&dataset (keep= &outcome &treatment %do ivlag=&minivlag %to &maxivlag ; &&iv&ivlag %end; &controlvariables ) ; format &outcome &treatment ; /* write all output of interest to data sets in &outlib */ ods output SummaryContResponse = &OUTLIB...&treatment.._ qlimiv&treatment..&project&vers TestResults = &OUTLIB...TESTS qlimiv&treatment..&project&vers```
ResponseProfile = &outlib...&outcome..._qlimiv&&treatment...&project&vers
FitSummary = &outlib...FITSUM qlimiv&&treatment...&project&vers
ConvergenceStatus = &outlib...converge qlimiv&&treatment...&project&vers
ParameterEstimates = &outlib...par qlimiv&&treatment...&project&vers

/* specify equations */
   model &treatment =
   %do ivlag=&minivlag %to &maxivlag ;
      &&ivmoniker..&ivlag
   %end;
   &controlvariables
   / ;
   model &outcome  =   &treatment &controlvariables / discrete (distribution = normal);
/* in version 13.1 we will also be able to use the following test: endotest(&treatment)*/
   title3 "Evaluate the strength of the statistical relationship between the practice pattern instrument set and the treatment ";

   TREATMENT_F_IV : test
      &TREATMENT...&iv1
      %do ivlag=%eval(&minivlag+1) %to &maxivlag ;
      , &TREATMENT...&iv&ivlag
      %end;
      / LR
   ;
   title3 "Evaluate whether errors in the 2 equations are uncorrelated ";
   Uncorrelated_Errors: test _rho = 0 / LR;

   /* in version 13.1 we will also be able to test for overidentification using overid &treatment (oip_fl10) for example */
   /* create an output data set with Estimates for each observation */
   /* Used to check whether estimates from the first equation justify the use of linear model */
   output out=&outlib...&treatment...&ivmoniker..&minivlag..&maxivlag&vers
   PREDICTED
   RESIDUAL
   XBETA
   ERRSTD
   PROB
   PROBALL
   EXPECTED
CONDITIONAL

MARGINAL

;
run;
title3 "Calculate relative risk and confidence interval for effect of Treatment from Parameter Estimates ";

DATA treatmenteffect;
set &outlib...par_qlimiv&treatment..&project&vers (where=(parameter='Outcome.Treatment'));

l95=estimate - 1.96 * stderr;
u95= estimate + 1.96 * stderr;
rel_riskest = exp(estimate);
rel_risk_l95 = exp(l95);
rel_risk_u95 = exp(u95);
run;

PROC PRINT data=treatmenteffect; run;
run;
title3 "Check that Distribution for Estimates of Probability of Treatment is consistent with form of the IV Model ";
PROC UNIVARIATE data= &outlib...&treatment..&Ivmoniker..&minivlag..&maxivlag&vers ;
var expct_Treatment; run;
run;
%mend Practice_Pattern_Iv ;