The Stata Journal publishes reviewed papers together with shorter notes or comments, regular columns, book reviews, and other material of interest to Stata users. Examples of the types of papers include 1) expository papers that link the use of Stata commands or programs to associated principles, such as those that will serve as tutorials for users first encountering a new field of statistics or a major new technique; 2) papers that go “beyond the Stata manual” in explaining key features or uses of Stata that are of interest to intermediate or advanced users of Stata; 3) papers that discuss new commands or Stata programs of interest either to a wide spectrum of users (e.g., in data management or graphics) or to some large segment of Stata users (e.g., in survey statistics, survival analysis, panel analysis, or limited dependent variable modeling); 4) papers analyzing the statistical properties of new or existing estimators and tests in Stata; 5) papers that could be of interest or usefulness to researchers, especially in fields that are of practical importance but are not often included in texts or other journals, such as the use of Stata in managing datasets, especially large datasets, with advice from hard-won experience; and 6) papers of interest to those who teach, including Stata with topics such as extended examples of techniques and interpretation of results, simulations of statistical concepts, and overviews of subject areas.

The Stata Journal is indexed and abstracted by CompuMath Citation Index, Current Contents/Social and Behavioral Sciences, RePEc: Research Papers in Economics, Science Citation Index Expanded (also known as SciSearch), Scopus, and Social Sciences Citation Index.

For more information on the Stata Journal, including information for authors, see the webpage http://www.stata-journal.com
Subscription rates listed below include both a printed and an electronic copy unless otherwise mentioned.

<table>
<thead>
<tr>
<th></th>
<th>U.S. and Canada</th>
<th>Elsewhere</th>
</tr>
</thead>
<tbody>
<tr>
<td>Printed &amp; electronic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-year subscription</td>
<td>$98</td>
<td>$138</td>
</tr>
<tr>
<td>2-year subscription</td>
<td>$165</td>
<td>$245</td>
</tr>
<tr>
<td>3-year subscription</td>
<td>$225</td>
<td>$345</td>
</tr>
<tr>
<td>1-year student subscription</td>
<td>$75</td>
<td>$99</td>
</tr>
<tr>
<td>1-year institutional subscription</td>
<td>$245</td>
<td>$285</td>
</tr>
<tr>
<td>2-year institutional subscription</td>
<td>$445</td>
<td>$525</td>
</tr>
<tr>
<td>3-year institutional subscription</td>
<td>$645</td>
<td>$765</td>
</tr>
<tr>
<td>Electronic only</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-year subscription</td>
<td>$75</td>
<td>$75</td>
</tr>
<tr>
<td>2-year subscription</td>
<td>$125</td>
<td>$125</td>
</tr>
<tr>
<td>3-year subscription</td>
<td>$165</td>
<td>$165</td>
</tr>
<tr>
<td>1-year student subscription</td>
<td>$45</td>
<td>$45</td>
</tr>
</tbody>
</table>

Back issues of the *Stata Journal* may be ordered online at

http://www.stata.com/bookstore/sjj.html

Individual articles three or more years old may be accessed online without charge. More recent articles may be ordered online.

http://www.stata-journal.com/archives.html

The *Stata Journal* is published quarterly by the Stata Press, College Station, Texas, USA.

Address changes should be sent to the *Stata Journal*, StataCorp, 4905 Lakeway Drive, College Station, TX 77845, USA, or emailed to sj@stata.com.

© 2014 by StataCorp LP

**Copyright Statement:** The *Stata Journal* and the contents of the supporting files (programs, datasets, and help files) are copyright © by StataCorp LP. The contents of the supporting files (programs, datasets, and help files) may be copied or reproduced by any means whatsoever, in whole or in part, as long as any copy or reproduction includes attribution to both (1) the author and (2) the *Stata Journal*.

The articles appearing in the *Stata Journal* may be copied or reproduced as printed copies, in whole or in part, as long as any copy or reproduction includes attribution to both (1) the author and (2) the *Stata Journal*.

Written permission must be obtained from StataCorp if you wish to make electronic copies of the insertions. This precludes placing electronic copies of the *Stata Journal*, in whole or in part, on publicly accessible websites, file servers, or other locations where the copy may be accessed by anyone other than the subscriber.

Users of any of the software, ideas, data, or other materials published in the *Stata Journal* or the supporting files understand that such use is made without warranty of any kind, by either the *Stata Journal*, the author, or StataCorp. In particular, there is no warranty of fitness of purpose or merchantability, nor for special, incidental, or consequential damages such as loss of profits. The purpose of the *Stata Journal* is to promote free communication among Stata users.

The *Stata Journal* (ISSN 1536-867X) is a publication of Stata Press. Stata, Stata Press, Mata, and NetCourse are registered trademarks of StataCorp LP.
Abstract. In this article, I present \texttt{ivtreatreg}, a command for fitting four different binary treatment models with and without heterogeneous average treatment effects under selection-on-unobservables (that is, treatment endogeneity). Depending on the model specified by the user, \texttt{ivtreatreg} provides consistent estimation of average treatment effects by using instrumental-variables estimators and a generalized two-step Heckman selection model. The added value of this new command is that it allows for generalization of the regression approach typically used in standard program evaluation by assuming heterogeneous response to treatment. It also serves as a sort of toolbox for conducting joint comparisons of different treatment methods, thus readily permitting checks on the robustness of results.

Keywords: st0346, ivtreatreg, microeconometrics, treatment models, instrumental variables, unobservable selection, treatment endogeneity, heterogeneous treatment response

1 Introduction

It is increasingly recognized as good practice to perform ex-post evaluation of economic and social programs through counterfactual evidence-based statistical analysis. Such analysis is particularly important at the policy-making level. The statistical approach is usually applied to measuring the causal effects of an intervention on part of an external authority, such as local or national government, on a set of subjects targeted by a given program, such as individuals and companies. Similar analysis is also becoming popular in reassessing causal relations among factors identified under modern microeconometric theory from a counterfactual perspective but not necessarily regarding policy implications.

Several official Stata commands and new user-written commands have been applied to enlarge the set of available statistical tools for conducting these counterfactual analyses. Table 1 contains a list of commands for estimating binary treatment effects. However, the most recent release of Stata, version 13, provides a new far-reaching suite called \texttt{teffects}, which can be used to estimate treatment effects from observational data.
Table 1. Commands for performing econometric program evaluation: ordinary least-squares (OLS) estimation using a control function; heckit, Heckman-type selection model; difference-in-differences (DID); instrumental variables (IV); regression discontinuity design

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>regress</td>
<td>OLS estimation based on a control function, linear reweighting, DID</td>
<td>StataCorp</td>
</tr>
<tr>
<td></td>
<td>(panel data)</td>
<td></td>
</tr>
<tr>
<td>ivregress</td>
<td>Basic IV, local average treatment effect</td>
<td>StataCorp</td>
</tr>
<tr>
<td>etregress</td>
<td>Selection model (heckit)</td>
<td>StataCorp</td>
</tr>
<tr>
<td>pscore*</td>
<td>Matching (propensity score)</td>
<td>Becker and Ichino (2002)</td>
</tr>
<tr>
<td>nnmatch*</td>
<td>Matching (nearest neighbor on covariates)</td>
<td>Abadie et al. (2004)</td>
</tr>
<tr>
<td>rd*</td>
<td>Regression discontinuity design (&quot;sharp&quot; and &quot;fuzzy&quot;)</td>
<td>Austin (2007)</td>
</tr>
<tr>
<td>treatrew*</td>
<td>Reweighting on propensity score</td>
<td>Cerulli (2014)</td>
</tr>
<tr>
<td>diff*</td>
<td>DID (repeated cross-section)</td>
<td>Villa (2009)</td>
</tr>
</tbody>
</table>

* User-written command downloadable from the Statistical Software Components archive

The `teffects` command can be used to estimate potential outcome means and average treatment effects (ATEs). As shown in table 2, the `teffects` suite covers a large set of methods, such as regression adjustment; inverse-probability weights; doubly robust methods, including inverse-probability-weighted regression adjustment; augmented inverse-probability weights; and matching on the propensity score or covariates (with nearest neighbors). Other subcommands can be used for postestimation purposes and for testing reliability of results; for example, `overlap` allows for plotting the estimated densities of the probability of getting each treatment level.
Table 2. Stata 13 `teffects` subcommands for estimating treatment effects from observational data

<table>
<thead>
<tr>
<th>Subcommand</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>aipw</td>
<td>Augmented inverse-probability weighting</td>
</tr>
<tr>
<td>ipw</td>
<td>Inverse-probability weighting</td>
</tr>
<tr>
<td>ipwra</td>
<td>Inverse-probability-weighted regression adjustment</td>
</tr>
<tr>
<td>nnmatch</td>
<td>Nearest-neighbor matching</td>
</tr>
<tr>
<td>overlap</td>
<td>Overlap plots</td>
</tr>
<tr>
<td>psmatch</td>
<td>Propensity-score matching</td>
</tr>
<tr>
<td>ra</td>
<td>Regression adjustment</td>
</tr>
</tbody>
</table>

When applying `teffects`, the outcome models can be continuous, binary, count, or nonnegative. Binary outcomes can be modeled using logit, probit, or heteroskedastic probit regression, and count and nonnegative outcomes can be modeled using Poisson regression. The treatment model can be binary or multinomial. Binary treatments can be modeled using logit, probit, or heteroskedastic probit regression. For multinomial treatments, one can use pairwise comparisons and then exploit binary treatment approaches.¹

While the `teffects` command deals mainly with estimation methods suitable under selection-on-observables, Stata 13 presents two further commands to deal with endogenous binary treatment (occurring in the case of selection-on-unobservables): `etregress` and `etpoisson`. `etregress` estimates the ATE and the other parameters of a linear regression model augmented with an endogenous binary treatment variable. Basically, `etregress` is an improvement on Stata’s `treatreg` command, whose estimation is based on the Heckman (1978) selection model. Because such a model is fully parametric, estimation can be performed either by full maximum likelihood or, less parametrically, by a two-step consistent estimator. Similarly, `etpoisson` estimates an endogenous binary treatment model when the outcome is a count variable by using a Poisson regression. Both the ATE and the ATE on the treated (ATET) can be estimated by `etpoisson`.

Although Stata 13 offers the above commands for dealing with endogenous treatment, the commands suffer from two important limitations. First, they assume joint normality of errors, meaning that they are not robust to violation of this hypothesis. Second, they do not allow—at least by default—for calculation of causal effects under observable heterogeneity, meaning that they assume causal effects to be the same in the subpopulation of treated and untreated units. This second limitation might be partially

¹. For multinomial treatment, readers can refer to the user-written command `poparms`, which estimates multivalued treatment effects under conditional independence by using the efficient semiparametric estimation of multivalued treatment effects. See Cattaneo (2010) and Cattaneo, Drukker, and Holland (2013) for tutorials.
overcome by introducing interactions between the binary treatment and the covariates in the outcome equation, but this requires further user programming to recover all the parameters of interest.

The gsem command, also new in Stata 13, can estimate the causal parameters of models with selection-on-unobservables, implemented as unobserved components, and heterogeneous effects, implemented as random coefficients. However, gsem uses full-information maximum likelihood (ML), thus assuming a fully specified parametric model, which in some contexts could present questionable reliability.

The ivtreatreg command I present in this article implements a series of methods for treatment-effects estimation under treatment endogeneity that use only conditional-moment restrictions. These methods are more robust than those implemented by etregress or gsem. ML estimators would be naturally more efficient under correct specification, and this means that a trade-off may arise between robustness and efficiency. On the one hand, assuming some parametric distributive form for the error terms allows one to use ML estimation reaching the Cramér–Rao lower variance bound. On the other hand, when these distributive assumptions are questionable, ML may be less reliable than less efficient (but consistent) estimation procedures, and the latter ones become more robust. Thus it seems useful to adopt distribution-free methods for dealing with treatment endogeneity, which the ivtreatreg command makes possible.

ivtreatreg fits four binary treatment models with and without idiosyncratic or heterogeneous ATEs. Depending on the model specified by the user, ivtreatreg provides consistent estimation of ATEs under the hypothesis of selection-on-unobservables by using IV and a generalized Heckman-style selection model.

Conditional on a prespecified subset of exogenous variables, x—thought of as driving the heterogeneous response to treatment—ivtreatreg calculates the ATE, the ATET, and the ATE on the nontreated (ATENT) for each called model, as well as the estimates of these parameters conditional on the observable factors x.

Specifically, the four models fit by ivtreatreg are direct-2sls (IV regression fit by direct two-stage least squares), probit-ols (IV two-step regression fit by probit and OLS), probit-2sls (IV regression fit by probit and two-stage least squares), and heckit (Heckman two-step selection model).

Extensive discussion of the conditions under which previous methods provide consistent estimation of ATE, ATET, and ATENT can be found in Wooldridge (2010).

ivtreatreg provides value by allowing for generalization of the regression approach typically employed in standard program evaluation by assuming heterogeneous response to treatment and treatment endogeneity. It is also a sort of toolbox for conducting joint comparisons of different treatment methods, thus readily permitting the researcher to run checks on the robustness of results.

In sections 2 and 3 of this article, I briefly present the statistical framework and estimation methods implemented by ivtreatreg. In section 4, I present the syntax

---

2. To my knowledge, no previous Stata command has addressed this objective.
with a description of the help file, and in section 5, I conduct a Monte Carlo experiment to test the reliability of `ivtreatreg`. In section 6, I demonstrate the command applied to real data from a study of the relationship between education and fertility. I conclude with section 7, where I provide a brief summary and affirm the value of `ivtreatreg`. In the appendix, I derive the formulas for the selection model.

# 2 Statistical framework

Our hypothetical evaluation objective is to estimate the effect of binary treatment \( w \) (taking value 1 for treated and 0 for untreated units) on scalar outcome \( y \). We suppose that the assignment to treatment is not random but instead due to some form of the unit’s self-selection or external selection. For each unit, \((y_1, y_0)\) denotes the two potential outcomes, where the outcome is \( y_1 \) when the individual is treated and \( y_0 \) when the individual is not treated. We then collect an independent and identically distributed sample of observations \((y_i, w_i, x_i)\) with \( i = 1, \ldots, N \), where \( x \) is a row vector of covariates hypothesized as driving the observable nonrandom assignment to treatment (confounders).

Here we are interested in estimating the ATE, defined as

\[
ATE = E(y_1 - y_0)
\]

If we rely on observational data alone, we cannot identify the ATE because, for the same individual and at the same time, we can observe just one out of the two quantities needed to calculate the ATE (Holland 1986). By restricting the analysis on the group of treated units, we can also define a second causal parameter, the ATET, as

\[
ATET = E(y_1 - y_0 \mid w = 1)
\]

Similarly, the ATENT, meaning the ATE calculated within the subsample of untreated units, is

\[
ATENT = E(y_1 - y_0 \mid w = 0)
\]

An interesting relationship links these three parameters:

\[
ATE = ATET p(w = 1) + ATENT p(w = 0)
\]

---

3. This section draws on the substantial literature on econometrics of program evaluation, such as Rubin (1974), Angrist (1991), Angrist, Imbens, and Rubin (1996), Heckman, LaLonde, and Smith (1999), Wooldridge (2010), and Cattaneo (2010). For a recent survey, see also Imbens and Wooldridge (2009).


5. For simplicity, I avoid writing the subscript form of the unit \( i \) when referring to population parameters.
where $p(w = 1)$ is the probability of being treated and $p(w = 0)$ is the probability of being untreated. Where $x$ is known, we can also define the previous parameters “conditional on $x$” as follows:

\[
\begin{align*}
\text{ATE}(x) &= E(y_1 - y_0 | x) \\
\text{ATET}(x) &= E(y_1 - y_0 | w = 1, x) \\
\text{ATENT}(x) &= E(y_1 - y_0 | w = 0, x)
\end{align*}
\]

These quantities are functions of $x$, which means that they can be seen as individual-specific ATEs because each individual owns a specific value of $x$. Furthermore, by law of iterated expectation, we have

\[
\begin{align*}
\text{ATE} &= E_x\{\text{ATE}(x)\} \\
\text{ATET} &= E_x\{\text{ATET}(x)\} \\
\text{ATENT} &= E_x\{\text{ATENT}(x)\}
\end{align*}
\]

The analyst needs to recover consistent (and, when possible, efficient) estimators of the previous parameters from observational data. Before going on, note that throughout this article we assume that the “stable unit treatment value assumption” (Rubin 1978) holds. This assumption states that “the treatment received by one unit does not affect other units’ outcome” (Cox 1958). We thus restrict the analysis to a “no-interference” setting. Indeed, when the stable unit treatment value assumption does not hold, treatment externality effects between units may occur and pose severe problems in identifying effects.\(^6\)

### 3 Estimation methods

The new command `ivtreatreg` implements four models to consistently estimate previous parameters, and three of these are IV estimators. These methods are direct-2sls (IV regression estimated by direct two-stage least squares), probit-ols (IV two-step regression estimated by probit and OLS), probit-2sls (IV regression estimated by probit and two-stage least squares), and heckit (Heckman two-step selection model). Each of these can be estimated by assuming either homogeneous or heterogeneous response to treatment (for a total of eight models). Before presenting how `ivtreatreg` works, I briefly set out the formulas, conditions, and procedures of each model (see Wooldridge [2010, chap. 21]). We start by assuming that

\[
\begin{align*}
y_0 &= \mu_0 + x\beta_0 + e_0, \quad E(e_0) = 0, \quad E(e_0 | x) = 0, \quad \mu_0 = \text{parameter} \\
y_1 &= \mu_1 + x\beta_1 + e_1, \quad E(e_1) = 0, \quad E(e_1 | x) = 0, \quad \mu_1 = \text{parameter} \\
y &= y_0 + w(y_1 - y_0)
\end{align*}
\]

\(^6\) Treatment-effects estimation under interference between units is a challenging field of study. Sobel (2006), Rosenbaum (2007), and Hudgens and Halloran (2008) offer important contributions on correct inferences within such a setting.
Equations (1) and (2) represent the potential outcome equations assumed to be linear in parameters, while the vector $x$ can also contain nonlinear functions of the various covariates. Equation (3) is the so-called “potential outcome model” and expresses the observational rule of the model, because $y$ is the observed outcome. We do not need to explicitly specify an equation for $w$ (that is, a selection equation) in this model; however, we could specify an equation. We could assume, for instance, that a linear probability model for the propensity to be selected into treatment is

$$w = \theta_0 + x\theta_1 + a$$

(4)

where $a$ is an error component. As soon as we hold that $a$ is uncorrelated with $(e_1; e_0)$, then (4) is redundant and not needed to identify causal parameters. However, we must know $w$ to identify the causal parameters, as we will discuss later. By substituting (1)−(2) into (3), we get

$$y = \mu_0 + (\mu_1 - \mu_0)w + x\beta_0 + w(x\beta_1 - x\beta_0) + e_0 + w(e_1 - e_0)$$

where $\beta_0 \neq \beta_1$ implies observable heterogeneity and $e_1 \neq e_0$ implies unobservable heterogeneity.

Next, we define $\eta = e_0 + w(e_1 - e_0)$. We can distinguish two cases: 1) $e_1 = e_0$ and 2) $e_1 \neq e_0$, which can in turn be split into the following subcases:

Case 1.1. $e_1 = e_0 = e$, $\beta_0 = \beta_1 = \beta$, $E(e \mid x, w) = 0$: unobservable homogeneity, homogeneous reaction function of $y_0$ and $y_1$ to $x$, treatment exogeneity.

In this case, we can show that

$$E(y \mid w, x) = \mu_0 + w \text{ATE} + x\beta$$

$$\text{ATE} = \text{ATE}(x) = \text{ATET} = \text{ATENT}(x) = \mu_1 - \mu_0$$

Thus no heterogeneous $\text{ATE}$ (over $x$) exists. Furthermore, OLS consistently estimates $\text{ATE}$.

Case 1.2. $e_1 = e_0 = e$, $\beta_0 \neq \beta_1$; $E(e \mid x, w) = 0$: unobservable homogeneity, heterogeneous reaction function of $y_0$ and $y_1$ to $x$, treatment exogeneity.

In this case, we can show that

$$E(y \mid w, x) = \mu_0 + w \text{ATE} + x\beta_0 + w(x - \mu_x)\beta$$

(5)

$$\text{ATE} \neq \text{ATET} \neq \text{ATENT}$$

where $\beta = \beta_1 - \beta_0$ and $\mu_x = E(x)$ is the sample mean of $x$. In this case, heterogeneous $\text{ATE}$ (over $x$) exist, and the population causal parameters take the forms

$$\text{ATE} = (\mu_1 - \mu_0) + \mu_x\beta$$
$$\text{ATE}(x) = \text{ATE} + (x - \mu_x)\beta$$
$$\text{ATET} = \text{ATE} + E_x(x - \mu_x \mid w = 1)\beta$$
$$\text{ATET}(x) = \text{ATE} + \{(x - \mu_x)\beta \mid w = 1\}$$
$$\text{ATENT} = \text{ATE} + E_x(x - \mu_x \mid w = 0)\beta$$
$$\text{ATENT}(x) = \text{ATE} + \{(x - \mu_x)\beta \mid w = 0\}$$
whose sample equivalents are

\[
\text{ATE} = \hat{\alpha}_{\text{OLS}} \\
\text{ATE}(x) = \hat{\alpha}_{\text{OLS}} + (x - \bar{x})\hat{\beta}_{\text{OLS}} \\
\text{ATET} = \hat{\alpha}_{\text{OLS}} + \frac{1}{N} \sum_{i=1}^{N} w_i (x_i - \bar{x})\hat{\beta}_{\text{OLS}} \\
\text{ATET}(x) = \{\hat{\alpha}_{\text{OLS}} + (x - \bar{x})\hat{\beta}_{\text{OLS}}\}_{(w=1)} \\
\text{ATENT} = \hat{\alpha}_{\text{OLS}} + \frac{1}{N} \sum_{i=1}^{N} (1 - w_i) (x_i - \bar{x})\hat{\beta}_{\text{OLS}} \\
\text{ATENT}(x) = \{\hat{\alpha}_{\text{OLS}} + (x - \bar{x})\hat{\beta}_{\text{OLS}}\}_{(w=0)}
\]

where it is clear that, under treatment exogeneity, these parameters can be consistently estimated by plugging-in the parameters from an OLS of (5).

But what happens when treatment exogeneity fails and \(w\) becomes endogenous? We then have three subcases.

Case 2.1. \(e_1 = e_0 = e, \beta_0 = \beta_1 = \beta, E(e | x, w) \neq 0\): unobservable homogeneity, homogeneous reaction function of \(y_0\) and \(y_1\) to \(x\), treatment endogeneity.

In this case, we can show that

\[
E(y | w, x) \neq \mu_0 + w \text{ATE} + x\beta_0 \\
\text{ATE} \neq \text{ATET} \neq \text{ATENT}
\]

However, if an IV \(z\) is available, we can consistently estimate ATE by exploiting an IV approach.

Case 2.2. \(e_1 = e_0 = e, \beta_0 \neq \beta_1, E(e | x, w) \neq 0\): unobservable homogeneity, heterogeneous reaction function of \(y_0\) and \(y_1\) to \(x\), treatment endogeneity.

In this case, we can show that

\[
E(y | w, x) \neq \mu_0 + w \text{ATE} + x\beta_0 + w(x - \mu_x)\beta \\
\text{ATE} \neq \text{ATET} \neq \text{ATENT}
\]
Even in this case, if an IV \( z \) is available, we can consistently estimate ATE by exploiting an IV approach. Observe, however, that we have two endogenous variables: \( w \) and \( w(x - \mu_x) \). However, once IV estimations of parameters in (6) are available, we can consistently recover all the causal parameters of interest as follows:

\[
\text{ATE} = \hat{\alpha}_{IV}
\]
\[
\text{ATE}(x) = \hat{\alpha}_{IV} + (x - \bar{x})\hat{\beta}_{IV}
\]
\[
\text{ATET} = \hat{\alpha}_{IV} + \frac{1}{N} \sum_{i=1}^{N} w_i (x_i - \bar{x})\hat{\beta}_{IV}
\]
\[
\text{ATET}(x) = \left\{ \hat{\alpha}_{IV} + (x - \bar{x})\hat{\beta}_{IV} \right\}_{(w=1)}
\]
\[
\text{ATENT} = \hat{\alpha}_{IV} + \frac{1}{N} \sum_{i=1}^{N} (1 - w_i)(x_i - \bar{x})\hat{\beta}_{IV}
\]
\[
\text{ATENT}(x) = \left\{ \hat{\alpha}_{IV} + (x - \bar{x})\hat{\beta}_{IV} \right\}_{(w=0)}
\]

Case 2.3. \( e_1 \neq e_0, \beta_0 \neq \beta_1, E(\eta | x, w) \neq 0 \): unobservable heterogeneity, heterogeneous reaction function of \( y_0 \) and \( y_1 \) to \( x \), treatment endogeneity.

In this case, we can show that

\[
E(y | w, x) \neq \mu_0 + w \text{ATE} + x\beta_0 + w(x - \mu_x)\beta
\]
\[
\text{ATE} \neq \text{ATET} \neq \text{ATENT}
\]

To apply IV and get consistent estimation, this case requires a further orthogonal condition,

\[
E\{w(e_1 - e_0) | x, z\} = E\{w(e_1 - e_0)\}
\]  
\[
(7)
\]

Given this condition, estimation may proceed as in Case 2.2.

Next, I present the methods implemented by \textit{ivtreatreg} by referring to the case of heterogeneous reaction.

\subsection{3.1 Control-function regression}

Control-function regression consistently estimates the previously defined causal effects under selection-on-observables, that is, when conditional mean independence (CMI) holds. CMI implies treatment exogeneity by restricting the independence between potential outcomes and treatment to the mean once covariates \( x \) are fixed at a certain level. The control-function estimation protocol is as follows:
1. Estimate \( y_i = \mu_0 + w_i \alpha + x_i \beta_0 + w_i(x_i - \mu_x) \beta + \text{error}_i \) by OLS, thus getting consistent estimates of \( \mu_0, \alpha, \beta_0, \) and \( \beta, \) with \( \alpha = \text{ATE}. \)

2. Plug these estimated parameters into the sample formulas and recover all the causal effects.

3. Obtain standard errors for \( \text{ATE} \) and \( \text{ATEN} \) via bootstrap.

However, \texttt{ivtreatreg} does not fit such a model, because it can be more robustly obtained by using the regression-adjustment estimator implemented in the \texttt{teffects} command of Stata 13 (with the suboption \texttt{ra}). This command handles many functional forms other than the linear one, and an estimation of \( \text{ATEN} \) can also be obtained using the \texttt{margins} command after running the regression in step 1. For this reason, \texttt{ivtreatreg} concentrates on the endogenous treatment-effect case, for which it adds new tools.

### 3.2 Instrumental variables

When the CMI hypothesis does not hold, control-function regression causes biased estimates of causal effects. This happens when the selection-into-treatment is due to both observable and unobservable factors. In this case, \( w \) becomes endogenous, that is, correlated with the regression error term. This is the case when the error term of (4) is correlated with \( \varepsilon_0 \) in (1) or with \( \varepsilon_1 \) in (2). IV can also restore consistency under the selection-on-unobservables. Nevertheless, applying IV requires the availability of at least one variable \( z \)—the instrumental variable—assumed to be directly correlated with the treatment \( w \) and directly uncorrelated with the outcome \( y \). This implies an exclusion restriction under which IV identifies causal parameters. \texttt{ivtreatreg} implements the following three consistent but differently efficient IV methods: \texttt{direct-2sls}, \texttt{probit-ols}, and \texttt{probit-2sls}.

**direct-2sls**

By using \texttt{direct-2sls}, the analyst does not consider the binary nature of \( w \). This method follows the typical IV steps:

1. Run an OLS regression of \( w \) on \( x \) and \( z \), thus getting the predicted values of \( w_i \), indicated by \( w_{fv,i} \).

2. Run a second OLS of \( y \) on \( \{x, w_{fv,i}, w_{fv,i}(x - \mu_x)\} \). The coefficient of \( w_{fv,i} \) is a consistent estimation of \( \text{ATE} \).

3. Plug these estimated parameters into the sample formulas, recover all the other causal effects, and obtain standard errors for \( \text{ATE} \) and \( \text{ATEN} \) via \texttt{bootstrap}.
In this case, the analyst exploits the binary nature of $w$ by fitting a probit regression in the first step. Operationally, probit-ols follows these three steps:

1. Apply a probit of $w$ on $x$ and $z$, thus getting $p_w$, the predicted probability of $w$.
2. Run an OLS of $y$ on $\{1, x, p_w, p_w(x - \mu_x)\}$.
3. Follow step 3 above.

The coefficient of $p_w$ is a consistent and more efficient estimator of ATE (compared with direct-2sls) given that the process generating $w$ is correctly specified. It has higher efficiency because the propensity score is the orthogonal projection of $w$ in the vector space generated by $(x, z)$. However, with this method, standard errors must be corrected for the presence of a generated regressor and heteroskedasticity.

probit-2sls

Operationally, probit-2sls follows these four steps:

1. Apply a probit of $w$ on $x$ and $z$, thus getting $p_w$, the predicted probability of $w$.
2. Run an OLS of $w$ on $(1, x, p_w)$, thus getting the fitted values $w_{2fv,i}$.
3. Run a second OLS of $y$ on $\{1, x, w_{2fv,i}, w_{2fv,i}(x - \mu_x)\}$.
4. Follow step 3 above.

The coefficient of $w_{2fv,i}$ is a more efficient estimator of ATE compared with direct-2sls. Furthermore, to achieve consistency, this procedure does not require that the process generating $w$ is correctly specified; thus, it is more robust than probit-ols.

3.3 heckit

ivtreatreg considers a generalized heckit model to consistently estimate previous parameters without using an IV. The price is that of relying on a trivariate normality assumption between the error terms of the potential outcomes and the error term of the treatment. However, this model has the advantage of fitting Case 2.3 without invoking (7). The reference model is again the system of (1–4), where we also assume that $(e_0, e_1, a)$ are trivariate normal. Such a model, as implemented by ivtreatreg, generalizes the two-step option of the official Stata command treatreg.

By default, the treatreg command assumes neither observable heterogeneity (because it holds that $\beta_0 = \beta_1$) nor unobservable heterogeneity (because it holds that $e_1 = e_0$). When these two assumptions are removed, the model leads to the following
baseline regression function, which can be consistently estimated by \textit{OLS} (see Wooldridge [2010, 949]):

$$E(y \mid x, z, w) = \mu_0 + \alpha w + x\beta_0 + w(x - \mu_x)\beta + \rho_1 w \frac{\phi(q\theta)}{\Phi(q\theta)} + \rho_0 (1 - w) \frac{\phi(q\theta)}{1 - \Phi(q\theta)}$$

where $\alpha$ is the ATE, $\rho_1$ and $\rho_0$ are the correlations between the two potential outcomes’ errors and the treatment’s error, and $\phi(x)$ and $\Phi(x)$ are the density and cumulative normal distribution, respectively. To estimate the previous regression, \textit{ivtreatreg} performs the following two-step procedure:

1. Run a probit regression of $w_i$ on $(1, x_i, z_i)$ and get $(\hat{\phi}_i, \hat{\Phi}_i)$.
2. Run an \textit{OLS} of $y_i$ on $\{1, w_i, x_i, w_i(x_i - \mu_x)_i, w_i\hat{\phi}_i/\hat{\Phi}_i, (1 - w_i)\hat{\phi}_i/1 - \hat{\Phi}_i\}$.

After estimation, one can also test the hypothesis of no selection-on-unobservables by testing the null:

$$H_0: \rho_1 = \rho_0 = 0$$

More importantly, it is easy to show that

$$\text{ATE} = \alpha$$

$$\text{ATE}(x) = \alpha + (x - \bar{x})\beta$$

although $\text{ATET}(x)$, $\text{ATET}$, $\text{ATENT}(x)$, and $\text{ATENT}$ assume different forms compared with previous models, specifically

$$\text{ATET}(x) = \{\alpha + (x - \bar{x})\beta + (\rho_0 + \rho_1) \times \lambda_1(q\theta)\}_{(w=1)}$$

and

$$\text{ATENT}(x) = \{\alpha + (x - \bar{x})\beta + (\rho_1 + \rho_0) \times \lambda_0(q\theta)\}_{(w=0)}$$

Given the estimates of $\alpha$, $\rho_1$, $\rho_0$, $\beta$, $\lambda_1$, and $\lambda_0$ from the previous two-step procedure, one can easily calculate all the causal effects. Here bootstrapping can again be used to obtain standard errors for $\text{ATET}$ and $\text{ATENT}$.

\textit{7. See the appendix for the derivation of these formulas.}
The `ivtreatreg` command fits the four binary treatment models presented above, with and without idiosyncratic or heterogeneous ATEs. The command calculates the ATE, ATET, and ATENT, as well as the estimates of these parameters conditional on the observable factors $x$ [that is, $\text{ATE}(x)$, $\text{ATET}(x)$, and $\text{ATENT}(x)$].

### 4.1 Syntax

\[
\text{ivtreatreg \ outcome \ treatment \ [varlist] \ [if] \ [in] \ [weight], \ model(modeltype) \ [hetero(varlist_h) \ iv(varlist_iv) \ conf(#) \ graphic \ vce(vcetype) \ beta \ const(noconstant) \ head(noheader)]}
\]

where `outcome` specifies the target variable that is the object of the evaluation, `treatment` specifies the binary treatment variable (that is, 1 = treated or 0 = untreated), and `varlist` defines the list of exogenous variables that are considered as observable confounders.

fweights, iweights, and pweights are allowed; see [U] 11.1.6 weight.

### 4.2 Options

`model(modeltype)` specifies the treatment model to be fit, where `modeltype` must be one of the following four models (described in sections 3.3 and 3.4 above): `direct-2sls`, `probit-2sls`, `probit-ols`, or `heckit`. `model()` is required.

<table>
<thead>
<tr>
<th><code>modeltype</code></th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>direct-2sls</td>
<td>IV regression fit by direct two-stage least squares</td>
</tr>
<tr>
<td>probit-2sls</td>
<td>IV regression fit by probit and two-stage least squares</td>
</tr>
<tr>
<td>probit-ols</td>
<td>IV two-step regression fit by probit and OLS</td>
</tr>
<tr>
<td>heckit</td>
<td>Heckman two-step selection model</td>
</tr>
</tbody>
</table>

`hetero(varlist_h)` specifies the list of variables over which to calculate the idiosyncratic ATE($x$), ATET($x$), and ATENT($x$), where $x = varlist_h$. When this option is not specified, the command fits the specified model without heterogeneous ATE. `varlist_h` should be the same set or a subset of the variables specified in `varlist`.

`iv(varlist_iv)` specifies the variables to be used as instruments. This option is required with `model(direct-2sls)`; it is optional with other `modeltypes`.

`conf(#)` sets the confidence level to the specified number. The default is `conf(95)`.
graphic requests a graphical representation of the density distributions of ATE(x), ATET(x), and ATENT(x). graphic gives an outcome only if specified with hetero().

vce(vcetype) specifies the type of standard error reported.

vce(robust) specifies to report standard errors that are robust to some kinds of misspecification.

vce(bootstrap|jackknife|conventional) may be specified when hetero() is not specified with model(heckit) to report standard errors that use the bootstrap method, the jackknife method, or the conventionally derived variance estimator.

beta reports standardized beta coefficients.

const(noconstant) suppresses the regression constant term.

head(noheader) suppresses the display of summary statistics at the top of the output; only the coefficient table is displayed.

4.3 Remarks

The ivtreatreg command also creates several variables that can be used to further examine the data:

- _ws_varname_h are the additional regressors used in a model’s regression when hetero(varlist_h) is specified. _ws_varname_h are created for all models.

- _z_varname_h are the IVs used in a model’s regression when hetero(varlist_h) and iv(varlist_iv) are specified. _z_varname_h are created only for IV models.

- ATE_x is an estimate of the idiosyncratic ATE.

- ATET_x is an estimate of the idiosyncratic ATET.

- ATENT_x is an estimate of the idiosyncratic ATENT.

- G_fv is the predicted probability from the probit regression, conditional on the observable confounders used.

- wL0 and wL1 are the Heckman correction terms.

Finally, the treatment must be a 0/1 binary variable (1 = treated, 0 = untreated). The standard errors for ATET and ATENT can be obtained via bootstrapping. Also, when option hetero() is not specified, then ATE(x), ATET(x), and ATENT(x) are single numbers equal to ATE = ATET = ATENT.
4.4 Stored results

ivtreatreg stores the following in e():

Scalars
- e(N_tot): total number of used observations
- e(N_treated): number of used treated units
- e(N_untreated): number of used untreated units
- e(ate): value of the ATE
- e(atet): value of the ATET
- e(atent): value of the ATENT

5 A Monte Carlo experiment for testing ivtreatreg

In this section, I provide a Monte Carlo experiment to check whether ivtreatreg complies with predictions from the theory and to assess its correctness from a computational point of view. The first step is to define a data-generating process (DGP) as follows:

\[
\begin{align*}
    w &= 1(0.5 + 0.5x_1 + 0.3x_2 + 0.6z + a > 0) \\
y_0 &= 0.1 + 0.2x_1 + 0.2x_2 + e_0 \\
y_1 &= 0.3 + 0.3x_1 + 0.3x_2 + e_1
\end{align*}
\]

where

\[
\begin{align*}
x_1 &: \ln(h_1) \\
x_2 &: \ln(h_2) \\
z &: \ln(h_3) \\
h_1 &: \chi^2(1) + c \\
h_2 &: \chi^2(1) + c \\
h_3 &: \chi^2(1) + c \\
c &: \chi^2(1)
\end{align*}
\]

and

\[
(a, e_0, e_1) \sim N(\mathbf{0}, \Omega)
\]

\[
\Omega = \begin{pmatrix}
\sigma^2_a & \rho_{a,e_0}\sigma_a\sigma_{e_0} & \rho_{a,e_1}\sigma_a\sigma_{e_1} \\
\rho_{a,e_0}\sigma_a\sigma_{e_0} & \sigma^2_{e_0} & \rho_{e_0,e_1}\sigma_{e_0}\sigma_{e_1} \\
\rho_{a,e_1}\sigma_a\sigma_{e_1} & \rho_{e_0,e_1}\sigma_{e_0}\sigma_{e_1} & \sigma^2_{e_1}
\end{pmatrix}
\]

\[
\sigma^2_a = 1, \quad \sigma^2_{e_0} = 3, \quad \sigma^2_{e_1} = 6.5
\]

\[
\rho_{a,e_0} = 0.5, \quad \rho_{a,e_1} = 0.3, \quad \rho_{e_0,e_1} = 0
\]

By assuming that the correlation between \(a\) and \(e_0(\rho_{a,e_0})\) and the correlation between \(a\) and \(e_1(\rho_{a,e_1})\) are different from 0, the \(w\)—the selection binary indicator—is endogenous. We indicate the instrument with \(z\), which is directly correlated with \(w\) but directly uncorrelated with \(y_1\) and \(y_0\). Given these assumptions, the DGP is completed by the potential outcome means, \(y_i = y_{0i} + w_i(y_{1i} - y_{0i})\), generating the observable outcome \(y\).
The DGP is simulated 2,000 times using a sample size of 2,000. For each simulation, we get a different data matrix \((x_1, x_2, y, w, z)\) on which we apply the four models implemented by `ivtreatreg`. Table 3 and figure 1 set out the simulation results.

Table 3. Monte Carlo simulation output of `ivtreatreg`

<table>
<thead>
<tr>
<th>Estimator</th>
<th>(1) Bias%</th>
<th>(2) Mean</th>
<th>(3) Std. dev.</th>
<th>(4) Mean SE</th>
<th>(5) Rejection rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>direct-2sls</td>
<td>5.05</td>
<td>0.235</td>
<td>0.316</td>
<td>0.318</td>
<td>0.042</td>
</tr>
<tr>
<td>probit-ols</td>
<td>2.92</td>
<td>0.217</td>
<td>0.272</td>
<td>0.268</td>
<td>0.045</td>
</tr>
<tr>
<td>probit-2sls</td>
<td>1.16</td>
<td>0.227</td>
<td>0.267</td>
<td>0.267</td>
<td>0.045</td>
</tr>
<tr>
<td>heckit</td>
<td>0.87</td>
<td>0.226</td>
<td>0.248</td>
<td>0.240</td>
<td>0.045</td>
</tr>
</tbody>
</table>

True value of ATE 0.224

We see that the true value of ATE is 0.224. As expected, all the IV procedures consistently estimate the true ATE, with a slight bias of around 5% only for `direct-2sls`.

Figure 1 confirms these findings by jointly plotting the distributions of ATEs obtained by each single method over the 2,000 DGP simulations. All methods give similar results, though `direct-2sls` has a slightly different shape with fatter tails. This suggests that we should examine the estimation precision. Under our DGP assumptions, we expect model `heckit` to be the most efficient method, followed by model `probit-ols` and model `probit-2sls`, with model `direct-2sls` performing the worst. In fact, our DGP follows exactly the same assumptions on which the model `heckit` is based, as well as the joint trivariate normality of \(a, e_0\), and \(e_1\).
Monte Carlo for ATE – Comparison of methods under treatment endogeneity

Figure 1. Monte Carlo distributions of ATE; comparison of IV methods

Table 3 confirms the following theoretical predictions: the lowest standard deviation is achieved by model hecket (0.248) and the highest by model direct-2sls (0.316), with the other methods lying in the middle with no appreciable differences. Observe that the standard error means (mean SE in column 4) show that the values of the standard deviations of the estimators in column 3 are estimated precisely (values are much the same). This means that the asymptotic distribution of the ATE estimators approximates finite-sample distribution well.

Table 3 also shows simulation results for test size. The size of a test is the probability of rejecting a hypothesis $H_0$ when $H_0$ is true. In our DGP, we set the size level at 0.05 for a two-sided test where $H_0$: ATE = 0.224 against the alternative $H_1$: ATE $\neq$ 0.224. The results, under the heading “Rejection rate” (column 5), represent the proportion of simulations that lead to rejection of $H_0$. These values should be interpreted as the simulation estimate of the true test size (which we assumed to be 0.05). As expected, the rejection rates are all lower than the usual 5% significance.

As a conclusion, these results seem to confirm both our expected theoretical results and the computational reliability of the ivtreatreg command.

6 ivtreatreg in practice: An application to the relationship between education and fertility

To see how ivtreatreg works in practice, we consider an instructional dataset called fertil2.dta, which accompanies the book Introductory Econometrics: A Modern Approach by Wooldridge (2013) and is a collection of cross-sectional data on 4,361 women
of childbearing age in Botswana. This dataset contains 28 variables on various female and family characteristics. In this exercise, we are particularly interested in evaluating the impact of the variable educ7 (taking value 1 if a woman has seven years of education or more and 0 otherwise) on the number of family children (children). Several conditioning (or confounding) observable factors are included in the dataset, such as the age of the woman (age), whether or not the family owns a TV (tv), and whether or not the woman lives in a city (urban). To inquire about the relationship between education and fertility, following Wooldridge (2010), we estimate the following specification for each of the four models implemented by ivtreatreg:

```
. ivtreatreg children educ7 age agesq evermarr urban electric tv,
   > hetero(age agesq evermarr urban) iv(frsthalf) model(modeltype) graphic
```

This specification adopts the covariate frsthalf as the IV and takes value 1 if the woman was born in the first six months of the year and 0 otherwise. This variable is partially correlated with educ7, but it should not have any direct relationship with the number of family children.

The simple difference-in-mean estimator (the mean of the treated ones, which are the children in the group of more educated women, minus the mean of the untreated ones, which are the children in the group of less educated women) is $-1.77$ with a $t$-value of $-28.46$. This means that women with more education show about two children fewer than women with less education, without ceteris paribus conditions. By adding confounding factors in the regression specification, we get the OLS estimate of ATE as $-0.394$ with a $t$-value of $-7.94$, still in absence of heterogeneous treatment. This is still significant, but the magnitude, as expected, dropped considerably compared with the difference-in-mean estimation, thus showing that confounders are relevant. When we consider OLS estimation with heterogeneity, we get an ATE equal to $-0.37$, which is still significant at 1%.9

When we consider IV estimation, results change dramatically. As we did in our working example of how to use ivtreatreg, we estimate the previous specification for probit-2sls with heterogeneous treatment response. The main outcome is reported below, where results from both the probit first-step and the IV regression of the second step are set out. Results on the probit show that frsthalf is partially correlated with educ7, thus it can be reliably used as an instrument for this variable. Step 2 shows that the ATE (again, the coefficient of educ7) is no more significant and that it changes sign, becoming positive and equal to 0.30.

---

8. The data are downloadable at http://fmwww.bc.edu/ec-p/data/wooldridge/fertil2.dta.
9. OLS results on ATE are obtained by estimating the baseline regression set out in section 3.1 with OLS.
. use fertil2.dta
. ivtreatreg children educ7 age agesq evermarr urban electric tv,
>   > hetero(age agesq evermarr urban) iv(frsthalf) model(probit-2sls) graphic
> (output omitted)

Probit regression
Number of obs = 4358
LR chi2(7) = 1130.84
Prob > chi2 = 0.0000
Log likelihood = -2428.384 Pseudo R2 = 0.1889

|       | Coef.     | Std. Err. |      z |  P>|z|    |   [95% Conf. Interval] |
|-------|-----------|-----------|-------|--------|-----------------------|
| educ7 |           |           |       |        |                       |
| frsthalf | -.2206627 | .0418563  | -5.27 | 0.000  | -.3026995 -.1386259   |
| age   | -.0150337 | .0174845  | -0.86 | 0.390  | -.0493027 .0192354    |
| agesq | -.0007325 | .0002897  | -2.53 | 0.011  | -.0013003 -.0001647   |
| evermarr | -.2972879 | .0486734  | -6.11 | 0.000  | -.392686 -.2018898    |
| urban  | .2998122  | .0432321  | 6.93  | 0.000  | .2150789 .3845456     |
| electric | .4246668  | .0751255  | 5.65  | 0.000  | .2774235 .57191       |
| tv    | .9281707  | .0977462  | 9.50  | 0.000  | .7365915 1.11975      |
| _cons | 1.13537   | .2440057  | 4.65  | 0.000  | .6571273 1.613612     |

(output omitted)

Instrumental variables (2SLS) regression
Number of obs = 4358
F( 11, 4346) = 448.51
Prob > F = 0.0000
R-squared = 0.4741
Adj R-squared = 0.4728

<table>
<thead>
<tr>
<th></th>
<th>SS</th>
<th>df</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>10198.4139</td>
<td>11</td>
<td>927.128534</td>
</tr>
<tr>
<td>Residual</td>
<td>11311.6182</td>
<td>4346</td>
<td>2.60276536</td>
</tr>
<tr>
<td>Total</td>
<td>21510.0321</td>
<td>4357</td>
<td>4.93689055</td>
</tr>
</tbody>
</table>

|       | Coef.     | Std. Err. |      t |  P>|t|    |   [95% Conf. Interval] |
|-------|-----------|-----------|-------|--------|-----------------------|
| children |           |           |       |        |                       |
| educ7 | .3004007  | .4995617  | 0.60  | 0.548  | -.6789951 1.279797    |
| _ws_age | -.8428913 | .1368854  | -6.16 | 0.000  | -1.111256 -.5745262  |
| _ws_agesq | .011469 | .0019061  | 6.02  | 0.000  | .007732 .0152059     |
| _ws_evermarr | -.8979833 | .2856655  | -3.14 | 0.002  | -1.458033 -.3379333  |
| _ws_urban | .4167504 | .2316103  | 1.80  | 0.072  | -.037324 .8708247    |
| age   | .859302   | .0966912  | 8.89  | 0.000  | .669738 1.048866     |
| agesq | -.01003   | .0012496  | -8.03 | 0.000  | -.012499 -.0075801   |
| evermarr | 1.253709 | .1586299  | 7.90  | 0.000  | .9427132 1.564704    |
| urban  | -.5313325 | .1379893  | -3.85 | 0.000  | -.801862 -.260803    |
| electric | -.2392104 | .1010705  | -2.37 | 0.018  | -.43736 -.0410608    |
| tv    | -.2348937 | .1478488  | -1.59 | 0.112  | -.5247528 .0549653   |
| _cons | -13.7584  | 1.876365  | -7.33 | 0.000  | -17.43704 -10.07977   |

Instrumented: educ7 _ws_age _ws_agesq _ws_evermarr _ws_urban
Instruments: age agesq evermarr urban electric tv G_fv _z_age _z_agesq
   _z_evermarr _z_urban

(output omitted)
This result is in line with the IV estimation obtained by Wooldridge (2010). Nevertheless, having assumed heterogeneous response to treatment, we can now also calculate the ATET and ATENT, and inspect the cross-unit distribution of these effects. First, `ivtreatreg` returns these parameters as scalars (along with treated and untreated sample size).

```
. ereturn list
scalars:
(output omitted)
  e(ate) = .300400740951661
  e(atet) = .898290019586237
  e(atent) = -.4468834318294228
  e(N_tot) = 4358
  e(N_treat) = 2421
  e(N_untreat) = 1937
(output omitted)
```

To get the standard errors for testing ATET and ATENT significance, we can easily implement a bootstrap procedure as follows:

```
. bootstrap atet=e(atet) atent=e(atent), rep(100):
> ivtreatreg children educ7 age agesq evermarr urban electric tv,
>    > hetero(age agesq evermarr urban) iv(frsthalf) model(probit-2sls)
```

<table>
<thead>
<tr>
<th></th>
<th>Observed</th>
<th>Bootstrap</th>
<th>Normal-based</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef.</td>
<td>Std. Err.</td>
<td></td>
</tr>
<tr>
<td>atet</td>
<td>0.3004</td>
<td>0.5488</td>
<td>0.1774</td>
</tr>
<tr>
<td>atent</td>
<td>-0.4469</td>
<td>0.4124</td>
<td>-1.2553</td>
</tr>
</tbody>
</table>

The results show that both ATET and ATENT are not significant and show quite different values, but the values are not far from that of ATE. Furthermore, a simple check shows that ATE = ATETp(w = 1) + ATENTp(w = 0), for example,

```
. di "ATE= " (e(N_treat)/e(N_tot))*e(atet)+(e(N_untreat)/e(N_tot))*e(atent)
ATE= .30040086
```

which confirms the expected result. Finally, we analyze the distribution of ATE(x), ATET(x), and ATENT(x). Figure 2 shows the result.
Figure 2. Distribution of $\text{ATE}(x)$, $\text{ATET}(x)$, and $\text{ATENT}(x)$ in model \texttt{probit-2sls}

We see that $\text{ATET}(x)$ presents a substantially uniform distribution, while both $\text{ATE}(x)$ and $\text{ATENT}(x)$ show a distribution more concentrated on negative values. In particular, $\text{ATENT}(x)$ shows the highest modal value around $-2.2$ children, thus predicting that less-educated women would have been less fertile if they had been more educated.

ATE results for all four models and for the simple difference-in-mean test ($t$ test) are shown below. The $\text{ATE}$ obtained by IV methods is consistently not significant, but it has a positive value for only \texttt{probit-2sls}. The rest of the $\text{ATEs}$ consistently show negative values—meaning that more-educated women would have been more fertile if they had been less educated. \texttt{heckit} is a little more puzzling because the result is significant and very close to the difference-in-mean estimation that is highly suspected as biased. This could be because the identification conditions of \texttt{heckit} are not met in this dataset.
. regress children educ7
   (output omitted)
. estimates store ttest
. ivtreatreg children educ7 age agesq evermarr urban electric tv,
   > hetero(age agesq evermarr urban) iv(frsthalf) model(heckit) graphic
   (output omitted)
. estimates store heckit
. ivtreatreg children educ7 age agesq evermarr urban electric tv,
   > hetero(age agesq evermarr urban) iv(frsthalf) model(probit-ols) graphic
   (output omitted)
. estimates store probit_ols
. ivtreatreg children educ7 age agesq evermarr urban electric tv,
   > hetero(age agesq evermarr urban) iv(frsthalf) model(direct-2sls) graphic
   (output omitted)
. estimates store direct_2sls
. ivtreatreg children educ7 age agesq evermarr urban electric tv,
   > hetero(age agesq evermarr urban) iv(frsthalf) model(probit-2sls) graphic
   (output omitted)
. estimates store probit_2sls
. estimates table ttest probit_ols direct_2sls probit_2sls heckit,
   > b(%9.2f) keep(educ7 G_fv) star

<table>
<thead>
<tr>
<th>Variable</th>
<th>ttest</th>
<th>probit_ols</th>
<th>direct_2sls</th>
<th>probit_2sls</th>
</tr>
</thead>
<tbody>
<tr>
<td>educ7</td>
<td>-1.77***</td>
<td>-1.04</td>
<td>0.30</td>
<td></td>
</tr>
<tr>
<td>G_fv</td>
<td>-0.11</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

legend: * p<0.05; ** p<0.01; *** p<0.001

<table>
<thead>
<tr>
<th>Variable</th>
<th>heckit</th>
</tr>
</thead>
<tbody>
<tr>
<td>educ7</td>
<td>-1.92***</td>
</tr>
<tr>
<td>G_fv</td>
<td></td>
</tr>
</tbody>
</table>

legend: * p<0.05; ** p<0.01; *** p<0.001

Finally, figure 3 shows the plot of the ATE distribution for each method. These distributions largely follow a similar pattern, although direct-2sls and heckit show some appreciable differences. Heckit, in particular, shows a very different pattern with a strong demarcation between the plot of treated and untreated units. Consequently, it appears to not be a reliable estimation procedure, an observation that deserves further inspection.
Figure 3. Distribution of ATE(x), ATET(x), and ATENT(x) for the four models fit by ivtreatreg

7 Conclusion

In this article, I presented a new user-written Stata command, ivtreatreg, for fitting four different binary treatment models with and without idiosyncratic or heterogeneous ATEs. Depending on the model specified, ivtreatreg consistently estimates ATEs under the hypothesis of selection-on-unobservables exploiting IV estimators and a generalized two-step Heckman selection model.

After presenting the statistical framework, I provided evidence on the reliability of ivtreatreg by using a Monte Carlo experiment. To familiarize the reader with the command, I also applied it to a real dataset. Results from both the Monte Carlo experiment and the real dataset encourage one to use the command when the empirical and theoretical setting suggests that treatment endogeneity and heterogeneous response to treatment are present. In such cases, performing more than one method may be a useful robustness check. The ivtreatreg command makes such checks possible and easy to perform.
References


**About the author**

Giovanni Cerulli is a researcher at Ceris-CNR, National Research Council of Italy, Institute for Economic Research on Firms and Growth. He received a degree in statistics and a PhD in economic sciences from Sapienza University of Rome and is editor-in-chief of the *International Journal of Computational Economics and Econometrics*. His research interests are mainly on applied microeconometrics, with a focus on counterfactual treatment-effects models for program evaluation. Stata programming and simulation- and agent-based methods are also among his related fields of study. He has published articles in high-quality, refereed economics journals.
Appendix

Derivation of ATET(x), ATET, ATENT(x), and ATENT in the heckit model

Proof.
The heckit model with observable and unobservable heterogeneity relies on these assumptions:

1. $y = \mu_0 + \alpha w + x_\beta_0 + w(x - \mu_x)\beta + u$
2. $E(e_1 \mid x, z) = E(e_0 \mid x, z) = 0$
3. $w = 1(\theta_0 + \theta_1 x + \theta_2 z + a \geq 0) = 1(q\theta \geq 0)$
4. $E(a \mid x, z) = 0$
5. $(a, e_0, e_1) \sim 3N$
6. $a \sim N(0, 1) \rightarrow \sigma_a = 1$
7. $u = e_0 + w(e_1 - e_0)$

By definition, we know that

$ATET(x) = E(y_1 - y_0 \mid x, w = 1) = (\mu_1 - \mu_0) + \{g_1(x) - g_0(x)\} + E(e_1 - e_0 \mid x, w = 1)$

At the same time, because $e_1$ and $e_0$ are independent of $x$, we also have

$E(e_1 - e_0 \mid x, w = 1) = E(e_1 - e_0 \mid w = 1)$

The value of the last expectation is easy to compute; indeed, by putting $e_1 - e_0 = \eta$ it follows that $\eta$ still has a normal distribution. This means that, from the property of truncated normal distributions,

$E(\eta \mid w = 1) = E(\eta \mid \theta_0 + \theta_1 x + \theta_2 z + a \geq 0) = E(\eta \mid q\theta \geq 0) = \sigma_{\eta a} \frac{\phi(q\theta)}{\Phi(q\theta)}$

From the linearity property of the covariance, we get

$\sigma_{\eta a} = \text{Cov}(\eta; a) = \text{Cov}(e_1 - e_0; a) = \text{Cov}(e_1; a) = \text{Cov}(e_0; a) = \sigma_{e_1 a} - \sigma_{e_0 a} = \rho_1 + \rho_0$

because $\rho_0 = -\sigma_{e_0 a}$ and $\rho_1 = \sigma_{e_1 a}$. This implies that

$ATET(x) = \{\alpha + (x - \overline{x})\beta + (\rho_1 + \rho_0) \times \lambda_1(q\theta)\}_(w=1)$

$ATET = \alpha + \frac{1}{N} \sum_{i=1}^{N} w_i (x_i - \overline{x})\beta + (\rho_1 + \rho_0) \times \frac{1}{N} \sum_{i=1}^{N} w_i \times \lambda_1(q\theta)$
where
\[ \lambda_1(q\theta) = \frac{\phi(q\theta)}{\Phi(q\theta)} \]

As for ATET, applying a similar procedure, it is immediate to get
\[
ATENT(x) = \{\alpha + (x - \bar{x})\beta + (\rho_1 + \rho_0) \times \lambda_1(q\theta)\}_{(w=1)}
\]

\[
\begin{align*}
ATENT &= \alpha + \frac{1}{N} \sum_{i=1}^{N} (1 - w_i) (x_i - \bar{x})\beta + (\rho_1 + \rho_0) \\
& \times \frac{1}{N} \sum_{i=1}^{N} (1 - w_i) \times \lambda_0(q\theta)
\end{align*}
\]

where
\[ \lambda_0(q\theta) = \frac{\phi(q\theta)}{1 - \Phi(q\theta)} \]

**Showing that ATE = ATET p(w=1) + ATENT p(w=0) in the heckit model**

*Proof.*
Consider formulas for ATE(x) and ATENT(x) in the heckit model:
\[
ATET(x) = \{\alpha + (x - \bar{x})\beta + (\rho_1 + \rho_0) \times \lambda_1(q\theta)\}_{(w=1)}
\]

\[
\begin{align*}
ATENT(x) &= \{\alpha + (x - \bar{x})\beta + (\rho_1 + \rho_0) \times \lambda_0(q\theta)\}_{(w=0)}
\end{align*}
\]

It follows that
\[
ATE(x) = p(w = 1) \{\alpha + (x - \bar{x})\beta + (\rho_1 + \rho_0) \times \lambda_1(q\theta)\} \\
+ p(w = 0) \{\alpha + (x - \bar{x})\beta + (\rho_1 + \rho_0) \times \lambda_0(q\theta)\}
\]

This implies that
\[
ATE(x) = \{\alpha + (x - \bar{x})\beta\} \{p(w = 1) + p(w = 0)\} + p(w = 1) \{(\rho_1 + \rho_0) \times \lambda_1(q\theta)\} \\
+ p(w = 0) \{(\rho_1 + \rho_0) \times \lambda_0(q\theta)\}
\]

\[
= \{\alpha + (x - \bar{x})\beta\} + p(w = 1) \left\{(\rho_1 + \rho_0) \times \frac{\phi(q\theta)}{\Phi(q\theta)} \right\} \\
+ p(w = 0) \left\{(\rho_1 + \rho_0) \times \frac{\phi(q\theta)}{1 - \Phi(q\theta)} \right\}
\]

because \(p(w = 1) + p(w = 0) = 1\). However, we saw that \(E(\eta | q\theta \geq 0) = (\rho_1 + \rho_0) \times \phi(q\theta)/\Phi(q\theta)\) and \(E(\eta | q\theta < 0) = (\rho_1 + \rho_0) \times \phi(q\theta)/\{1 - \Phi(q\theta)\} \).
For the law of iterated expectations, we get $E(\eta) = p(w = 1)E(\eta \mid q\theta \geq 0) + p(w = 0)E(\eta \mid q\theta \geq 0) = 0$, because $E(\eta) = E(e_1 - e_0) = 0$, proving that

$$\text{ATE}(x) = \alpha + (x - \bar{x})\beta$$

and finally

$$\text{ATE} = E_x\{\text{ATE}(x)\} = \alpha$$