



AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
<http://ageconsearch.umn.edu>
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

THE STATA JOURNAL

Editors

H. JOSEPH NEWTON
Department of Statistics
Texas A&M University
College Station, Texas
editors@stata-journal.com

NICHOLAS J. COX
Department of Geography
Durham University
Durham, UK
editors@stata-journal.com

Associate Editors

CHRISTOPHER F. BAUM, Boston College
NATHANIEL BECK, New York University
RINO BELLOCCO, Karolinska Institutet, Sweden, and
University of Milano-Bicocca, Italy
MAARTEN L. BUIS, University of Konstanz, Germany
A. COLIN CAMERON, University of California–Davis
MARIO A. CLEVES, University of Arkansas for
Medical Sciences
WILLIAM D. DUPONT, Vanderbilt University
PHILIP ENDER, University of California–Los Angeles
DAVID EPSTEIN, Columbia University
ALLAN GREGORY, Queen's University
JAMES HARDIN, University of South Carolina
BEN JANN, University of Bern, Switzerland
STEPHEN JENKINS, London School of Economics and
Political Science
ULRICH KOHLER, University of Potsdam, Germany

FRAUKE KREUTER, Univ. of Maryland–College Park
PETER A. LACHENBRUCH, Oregon State University
JENS LAURITSEN, Odense University Hospital
STANLEY LEMESHOW, Ohio State University
J. SCOTT LONG, Indiana University
ROGER NEWSON, Imperial College, London
AUSTIN NICHOLS, Urban Institute, Washington DC
MARCELLO PAGANO, Harvard School of Public Health
SOPHIA RABE-HESKETH, Univ. of California–Berkeley
J. PATRICK ROYSTON, MRC Clinical Trials Unit,
London
PHILIP RYAN, University of Adelaide
MARK E. SCHAFFER, Heriot-Watt Univ., Edinburgh
JEROEN WEESIE, Utrecht University
IAN WHITE, MRC Biostatistics Unit, Cambridge
NICHOLAS J. G. WINTER, University of Virginia
JEFFREY WOOLDRIDGE, Michigan State University

Stata Press Editorial Manager

LISA GILMORE

Stata Press Copy Editors

DAVID CULWELL, SHELBI SEINER, and DEIRDRE SKAGGS

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>

Subscriptions are available from StataCorp, 4905 Lakeway Drive, College Station, Texas 77845, telephone 979-696-4600 or 800-STATA-PC, fax 979-696-4601, or online at

<http://www.stata.com/bookstore/sj.html>

Subscription rates listed below include both a printed and an electronic copy unless otherwise mentioned.

U.S. and Canada		Elsewhere	
Printed & electronic		Printed & electronic	
1-year subscription	\$115	1-year subscription	\$145
2-year subscription	\$210	2-year subscription	\$270
3-year subscription	\$285	3-year subscription	\$375
1-year student subscription	\$ 85	1-year student subscription	\$115
1-year institutional subscription	\$345	1-year institutional subscription	\$375
2-year institutional subscription	\$625	2-year institutional subscription	\$685
3-year institutional subscription	\$875	3-year institutional subscription	\$965
Electronic only		Electronic only	
1-year subscription	\$ 85	1-year subscription	\$ 85
2-year subscription	\$155	2-year subscription	\$155
3-year subscription	\$215	3-year subscription	\$215
1-year student subscription	\$ 55	1-year student subscription	\$ 55

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.



Copyright © 2015 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**, Stata Press, Mata, **mata**, and NetCourse are registered trademarks of StataCorp LP.

ctreatreg: Command for fitting dose–response models under exogenous and endogenous treatment

Giovanni Cerulli

CNR-IRCrES

National Research Council of Italy

Institute for Research on Sustainable Economic Growth

Rome, Italy

giovanni.cerulli@ircres.cnr.it

Abstract. In this article, I present `ctreatreg`, a command for estimating a dose–response function when i) treatment is continuous, ii) individuals may react heterogeneously to observable confounders, and iii) the selection into treatment may be endogenous. I implement two estimation procedures: ordinary least squares under conditional mean independence and instrumental variables under selection endogeneity. Finally, I present a Monte Carlo experiment to test the reliability of the proposed command.

Keywords: st0412, `ctreatreg`, `boot_drf`, treatment effects, dose–response function, continuous treatment, Monte Carlo

1 Introduction

In this article, I present a command, `ctreatreg`, for estimating a dose–response function (DRF) through a regression approach when i) treatment is continuous, ii) individuals may react heterogeneously to observable confounders, and iii) the selection into treatment may be endogenous. In this context, the DRF is equal to the average treatment effect (ATE) given the level of treatment t [that is, $\text{ATE}(t)$], with t representing the continuous treatment variable. Other causal parameters of interest—such as the unconditional ATE, the ATE on the treated (ATET), and the ATE on the nontreated (ATENT)—are also estimated by `ctreatreg`, along with those effects conditional on the vector (\mathbf{x}, t) , where \mathbf{x} is a vector of predetermined variables.

In many socioeconomic and epidemiological contexts, interventions take the form of a continuous exposure to a certain type of treatment. From a program evaluation perspective, indeed, what is relevant in many settings is not only the binary treatment status but also the level of exposure (or dose) provided by a public agency. This is also in tune with the language of epidemiology, where DRFs are usually estimated to check patients’ resilience to different levels of drug administration (Robertson et al. 1994; Royston and Sauerbrei 2008).

Consider a policy program where the treatment is not assigned randomly (that is, it is assigned according to some structural rule) and where—after setting who is treated

and who is not—the program provides a different level or exposure to treatment ranging from 0 (no treatment) to 100 (maximum level of treatment). Two groups of units are thus formed: untreated, whose level of treatment (or dose) is 0, and treated, whose level of treatment is greater than 0.

We are interested in estimating the causal effect of the treatment variable t on an outcome y within the observed sample, assuming that treated and untreated units may respond differently both to specific observable confounders (that we collect in a vector \mathbf{x}) and to the intensity of the treatment t . We wish to estimate a DRF of y on t , when the treatment is assumed to be either exogenous (that is, selection into treatment depends only on factors observable to the analyst) or endogenous (that is, selection into treatment depends on factors both observable and unobservable to the analyst).

Compared with similar models—and in particular the one proposed by Hirano and Imbens (2004) implemented in Stata by Bia and Mattei (2008)—this model does not need a full normality assumption, and it is well-suited when many individuals have a treatment level of 0. Moreover, it may account for treatment endogeneity by exploiting an instrumental-variables (IV) estimation in a continuous treatment context.

When many units are not exposed to treatment, the distribution of t has a “spike” or non-nil probability mass at 0, that is, $p(t = 0) > 0$. So, assuming that the distribution of $t|\mathbf{x}$ comes from a normal distribution (or mixture of normal distributions), as assumed in the generalized propensity score (GPS) proposed by Hirano and Imbens (2004), is untenable, because in the presence of a spike at 0 this distribution is clearly discontinuous and thus nonnormal. Recently, however, Guardabascio and Ventura (2014) have proposed a modification of the Hirano–Imbens model extending the GPS approach to the case of a nonnormal continuous treatment variable. The authors consider a set of alternative distributions (binomial, Poisson, gamma, inverse Gaussian, etc.) derived from the exponential family of distributions.

Similarly, Bia et al. (2014) have proposed a generalization of the Hirano–Imbens model that allows one to estimate semiparametrically the DRF. In their article, the authors propose to estimate the GPS parametrically under various alternative distributional assumptions (such as the normal, inverse Gaussian, and gamma distributions) using as link functions the identity function, the logarithm, and the power function. Moreover, in the case of a treatment variable assuming values in the $(0, 1)$ interval, a two-parameter Beta distribution is also implemented.

Another interesting approach modeling a continuous treatment setting is the one proposed by Adorno, Bernini, and Pellegrini (2007). They propose a nonparametric two-step matching approach based on the generalized propensity function developed by Imai and van Dyk (2004). In the first step, they specify a participation decision rule and match units on the set of covariates that identify such process; in the second step, by considering only matched units from the first step, they perform another matching procedure pairing units having similar values only for the covariates explaining the treatment level assignment. Treated and untreated units are thus assumed to be balanced on both processes (binary selection and level assignment) because the only difference between the two groups should be in program participation. Although this model con-

siders the presence of “zeros” in the treatment level variable, it is suitable only under conditional independence.

The above approaches, relaxing some parametric assumptions, represent important improvements in the estimation of DRFs. Nevertheless, they still are unsuited to incorporate i) a zero-treatment probability mass and ii) the potential treatment endogeneity in the estimation of the DRF. I try to overcome both of these limitations in the present article.

However, the approach presented in this article has some of its own limitations. First, differently from the semiparametric approach adopted by Adorno, Bernini, and Pellegrini (2007) and Bia et al. (2014), the approach presented here assumes a parametric form of the potential outcomes with additive separability. Second, it does not need an estimation of the GPS, which might be interesting to analyze itself. Third, the proposed IV estimation uses a Heckman bivariate selection model, which requires additional distributional assumptions. Fourth, it only considers the presence of observable heterogeneity, thus neglecting the possible presence of unobservable heterogeneity. The model is, however, well suited to address DRF estimation when conditions (i) and (ii) above are present.

The reliability of the model and of its Stata implementation via `ctreatreg` is then checked by a Monte Carlo experiment, showing that the model and the command comply with the expected results and are also robust to departures from the joint normality of errors. The command also provides an interesting graphical representation of results by optionally plotting both the conditional effects’ distribution and the DRF along with its analytical and bootstrapped confidence intervals.

The article is organized as follows. Sections 2 and 3 present the model, its assumptions and formulas, and related estimation techniques. Section 4 presents and explains the use of `ctreatreg`. Section 5 shows an application of `ctreatreg` on real data. Section 6 sets out the results from a related Monte Carlo experiment to test the command’s reliability. And finally, section 7 concludes the article.

2 The model

Let us start with some notation. Consider two different and exclusive potential outcomes: one referring to the unit i when treated, y_{1i} , and one referring to the same unit when untreated, y_{0i} . Define w_i as the treatment indicator, taking value 1 for treated and 0 for untreated units, and define $\mathbf{x}_i = (x_{1i}, x_{2i}, x_{3i}, \dots, x_{Mi})$ as a row vector of M exogenous and observable characteristics (confounders) for unit $i = 1, \dots, N$. Let N be the total number of units, N_1 be the number of treated units, and N_0 be the number of untreated units, with $N = N_1 + N_0$.

Define two distinct functions, $g_1(\mathbf{x}_i)$ and $g_0(\mathbf{x}_i)$, as the unit i ’s responses to the vector of confounding variables \mathbf{x}_i when the unit is treated and untreated, respectively. Assume μ_1 and μ_0 to be two scalars, and assume e_1 and e_0 to be two random variables having 0 unconditional mean and constant variance. Finally, define t_i —taking values

within the continuous range $[0, 100]$ —as the continuous-treatment indicator, and define $h(t_i)$ as a general derivable function of t_i . In what follows, to simplify notation, we will get rid of the subscript i when defining population quantities and relations.

Given previous notation, we assume a specific population generating process for the two exclusive potential outcomes:¹

$$\begin{cases} w = 1 : & y_1 = \mu_1 + g_1(\mathbf{x}) + h(t) + e_1 \\ w = 0 : & y_0 = \mu_0 + g_0(\mathbf{x}) + e_0 \end{cases} \quad (1)$$

where the $h(t)$ function is different from 0 only in the treated status. Given this, we can also define the causal parameters of interest. Indeed, by defining the treatment effect (TE) as $TE = (y_1 - y_0)$, we define the causal parameters of interest as the population ATEs conditional on \mathbf{x} and t ; that is,

$$\begin{aligned} ATE(\mathbf{x}, t) &= E(y_1 - y_0 | \mathbf{x}, t) \\ ATET(\mathbf{x}, t > 0) &= E(y_1 - y_0 | \mathbf{x}, t > 0) \\ ATENT(\mathbf{x}, t = 0) &= E(y_1 - y_0 | \mathbf{x}, t = 0) \end{aligned} \quad (2)$$

where ATE indicates the overall average TE, ATET indicates the average TE on treated, and ATENT indicates the average TE on untreated units. By the law of iterated expectation, we know that the population unconditional ATEs are obtained as

$$\begin{aligned} ATE &= E_{(\mathbf{z}, t)} \{ATE(\mathbf{x}, t)\} \\ ATET &= E_{(\mathbf{x}, t > 0)} \{ATE(\mathbf{x}, t > 0)\} \\ ATENT &= E_{(\mathbf{x}, t = 0)} \{ATE(\mathbf{x}, t = 0)\} \end{aligned} \quad (3)$$

where $E_{\mathbf{z}}(\cdot)$ identifies the mean operator taken over the support of a generic vector of variables \mathbf{z} . By assuming a linear-in-parameters parametric form for $g_0(\mathbf{x}) = \mathbf{x}\boldsymbol{\delta}_0$ and $g_1(\mathbf{x}) = \mathbf{x}\boldsymbol{\delta}_1$, the ATE conditional on \mathbf{x} and t becomes

$$ATE(\mathbf{x}, t, w) = w \times \{\mu + \mathbf{x}\boldsymbol{\delta} + h(t)\} + (1 - w) \times (\mu + \mathbf{x}\boldsymbol{\delta})$$

where $\mu = (\mu_1 - \mu_0)$ and $\boldsymbol{\delta} = (\boldsymbol{\delta}_1 - \boldsymbol{\delta}_0)$. The unconditional ATE related to model (1) is equal to

$$ATE = p(w = 1) \times (\mu + \bar{\mathbf{x}}_{t>0}\boldsymbol{\delta} + \bar{h}_{t>0}) + p(w = 0) \times (\mu + \bar{\mathbf{x}}_{t=0}\boldsymbol{\delta})$$

where $p(\cdot)$ is a probability and $\bar{h}_{t>0}$ is the average of the response function taken over $t > 0$. Because we have, by law of iterated expectation, that $ATE = p(w = 1) \times ATET + p(w = 0) \times ATENT$, we obtain from the previous formula that

$$\begin{cases} ATE &= p(w = 1) (\mu + \bar{\mathbf{x}}_{t>0}\boldsymbol{\delta} + \bar{h}_{t>0}) + p(w = 0) (\mu + \bar{\mathbf{x}}_{t=0}\boldsymbol{\delta}) \\ ATET &= \mu + \bar{\mathbf{x}}_{t>0}\boldsymbol{\delta} + \bar{h}_{t>0} \\ ATENT &= \mu + \bar{\mathbf{x}}_{t=0}\boldsymbol{\delta} \end{cases} \quad (4)$$

1. Such a model is the representation of a treatment random coefficient regression as shown by Wooldridge (1997, 2003). See also Wooldridge (2010, chap. 21).

where the DRF is given by averaging $\text{ATE}(\mathbf{x}, t)$ over \mathbf{x} ,

$$\text{ATE}(t) = \begin{cases} \text{ATE}_T + \{h(t) - \bar{h}_{t>0}\} & \text{if } t > 0 \\ \text{ATE}_T & \text{if } t = 0 \end{cases} \quad (5)$$

that is, the DRF is a function of the treatment intensity t . The estimation of (5) under different identification assumptions is the main purpose of section 3.

3 The regression approach

In this section, we will consider the conditions for a consistent estimation of the causal parameters defined in (2) and (3) and thus of the DRF in (5). What is needed first, however, is a consistent estimation of the parameters of the potential outcomes in (1)—the “basic” parameters—because both the ATEs and the DRF are functions of these parameters.

Under previous definitions and assumptions, and in particular the form of the potential outcomes in model (1), to be substituted into Rubin’s (1974) potential outcome equation $y_i = y_{0i} + w(y_{1i} - y_{0i})$, the following baseline random-coefficient regression can be obtained (Wooldridge 1997, 2003):

$$y_i = \mu_0 + w_i \times \text{ATE} + \mathbf{x}_i \boldsymbol{\delta}_0 + w_i \times (\mathbf{x}_i - \bar{\mathbf{x}}) \boldsymbol{\delta} + w_i \times \{h(t_i) - \bar{h}\} + \eta_i \quad (6)$$

where $\eta_i = e_{0i} + w_i \times (e_{1i} - e_{0i})$.

Equation (6) provides the baseline regression for estimating the basic parameters $(\mu_0, \mu_1, \boldsymbol{\delta}_0, \boldsymbol{\delta}_1, \text{ATE})$ and then all the remaining ATEs. A semiparametric or parametric approach can be used as soon as a parametric or nonparametric form of the function $h(t)$ is assumed. In both cases, however, to get a consistent estimation of basic parameters, we need some additional assumptions. We start by first assuming unconfoundedness or conditional mean independence (CMI), showing that it is sufficient to provide parameters’ consistent estimation. We then remove this assumption and introduce other identifying assumptions.

3.1 Estimation under unconfoundedness

According to Rubin (1974), unconfoundedness states that, conditional on the knowledge of the true exogenous confounders \mathbf{x} , the conditions for randomization² are restored and causal parameters become identifiable. Given the set of random variables $\{y_{0i}, y_{1i}, \mathbf{x}_i, w_i, t_i\}$ as defined above, unconfoundedness (or CMI) implies in this specific case that

$$E(y_{ji}|w_i, t_i, \mathbf{x}_i) = E(y_{ji}|\mathbf{x}_i) \quad \text{with } j = \{0, 1\}$$

This form of the CMI assumption is a sufficient condition for identifying ATEs and the DRF in this context. Indeed, this assumption entails that, given the observable variables

2. In program evaluation literature, assuming randomization is equivalent to assuming probabilistic independence between potential outcomes and treatment (Rubin 1974).

collected in \mathbf{x} , both w and t are exogenous in (6). Thus we can write the regression line of the response y simply as

$$E(y_i|w_i, t_i, \mathbf{x}_i) = \mu_0 + w_i \times \text{ATE} + \mathbf{x}_i \boldsymbol{\delta}_0 + w_i \times (\mathbf{x}_i - \bar{\mathbf{x}}) \boldsymbol{\delta} + w_i \times \{h(t_i) - \bar{h}\} \quad (7)$$

and ordinary least squares (OLS) can be used to retrieve consistent estimation of all parameters. Once a consistent estimation of the parameters in (7) is obtained, we can estimate ATE directly from this regression; we can estimate ATET, ATENT, and the DRF by plugging the estimated basic parameters into (4) and (5). This is possible because these parameters are functions of consistent estimates and thus are consistent themselves. Standard errors for ATET and ATENT can be correctly obtained via bootstrapping (see Wooldridge [2010, 911–919]).

To complete the identification of ATEs and the DRF, we finally assume a polynomial parametric form of degree m for $h(t)$:

$$h(t_i) = \lambda_1 t_i + \lambda_2 t_i^2 + \lambda_3 t_i^3 + \cdots + \lambda_m t_i^m$$

where $\lambda_i (i = 1, \dots, m)$ are among the parameters to be estimated in regression (7).

Under CMI, an OLS estimation of (7) produces consistent estimates of the parameters we indicate as $\hat{\mu}_0, \hat{\boldsymbol{\delta}}_0, \hat{\text{ATE}}, \hat{\boldsymbol{\delta}}, \hat{\lambda}_1, \dots, \hat{\lambda}_m$. For the sake of simplicity, assume that $m = 3$, $\lambda_1 = a$, $\lambda_2 = b$, and $\lambda_3 = c$. With a consistent estimation of these parameters at hand, we can finally estimate consistently the DRF as

$$\begin{aligned} \widehat{\text{ATE}}(t_i) = w \left\{ \widehat{\text{ATET}} + \hat{a} \left(t_i - \frac{1}{N} \sum_{i=1}^N t_i \right) + \hat{b} \left(t_i^2 - \frac{1}{N} \sum_{i=1}^N t_i^2 \right) + \hat{c} \left(t_i^3 - \frac{1}{N} \sum_{i=1}^N t_i^3 \right) \right\} \\ + (1 - w) \widehat{\text{ATENT}} \end{aligned}$$

where

$$\widehat{\text{ATET}}(t_i) = \widehat{\text{ATE}}(t_i)_{t_i > 0}$$

A simple plot of the curve $\widehat{\text{ATE}}(t_i)_{t_i > 0}$ over the support of t returns the pattern of the DRF. Moreover, for each level of the dose t , it is also possible to calculate the α -confidence interval around the dose-response curve. Indeed, by defining $T_1 = t - E(t)$, $T_2 = t^2 - E(t^2)$, and $T_3 = t^3 - E(t^3)$, the standard error of the DRF is equal to³

$$\hat{\sigma}_{\widehat{\text{ATE}}(t)} = (T_1^2 \hat{\sigma}_a^2 + T_2^2 \hat{\sigma}_b^2 + T_3^2 \hat{\sigma}_c^2 + 2T_1 T_2 \hat{\sigma}_{a,b} + 2T_1 T_3 \hat{\sigma}_{a,c} + 2T_2 T_3 \hat{\sigma}_{b,c})^{1/2}$$

The α -confidence interval of $\widehat{\text{ATE}}(t)$ for each t is then given by

$$\left\{ \widehat{\text{ATE}}(t) \pm Z_{\alpha/2} \times \hat{\sigma}_{\widehat{\text{ATE}}(t)} \right\}$$

which can be usefully plotted along the dose-response curve to visually detect the statistical significance of the TE along the support of the dose t .

3. This comes from the variance-covariance properties where T_1 , T_2 , and T_3 are taken as constant and a , b , and c are taken as random variables.

Of course, the derivative of the DRF also can be estimated. The formula of this function is

$$\frac{\partial \text{ATE}(t)}{\partial t} = \hat{a} + 2\hat{b}t + 3\hat{c}t^2$$

with

$$\hat{\sigma}_{\frac{\partial \text{ATE}(t)}{\partial t}} = (\hat{\sigma}_a^2 + 4t^2\hat{\sigma}_b^2 + 9t^4\hat{\sigma}_c^2 + 2t\hat{\sigma}_{a,b} + 3t^2\hat{\sigma}_{a,c} + 6t^3\hat{\sigma}_{b,c})^{1/2}$$

The α -confidence interval of $\widehat{\text{ATE}}(t)$ for the derivative of each t is then given by

$$\left\{ \frac{\partial \text{ATE}(t)}{\partial t} \pm Z_{\alpha/2} \times \hat{\sigma}_{\frac{\partial \text{ATE}(t)}{\partial t}} \right\}$$

which can be drawn as a function of t .

3.2 Estimation under treatment endogeneity

When w (and thus t) are endogenous, the CMI assumption no longer holds, and the OLS estimate of regression (7) becomes inconsistent. This occurs because the orthogonality condition implied by unconfoundedness fails, so

$$E(\eta_i | w_i, t_i, \mathbf{x}_i) = E\{e_{0i} + w_i \times (e_{1i} - e_{0i}) | w_i, t_i, \mathbf{x}_i\} \neq 0$$

where it is clear that inequality depends on the endogeneity of w_i (and t_i), with \mathbf{x}_i assumed to be predetermined. In such a case, however, an IV estimation may be implemented to restore consistency. To this aim, it is sufficient to express the previous model in a semistructural form; that is,

$$y_i = \mu_0 + \mathbf{x}_i \boldsymbol{\delta}_0 + w_i \text{ATE} + w_i(\mathbf{x}_i - \bar{\mathbf{x}})\boldsymbol{\delta} + w_i T_{1i} + b w_i T_{2i} + c w_i T_{3i} + \eta_i \quad (8.1)$$

$$w_i^* = \mathbf{x}_{w,i} \boldsymbol{\beta}_w + \epsilon_{w,i} \quad (8.2)$$

$$t_i' = \mathbf{x}_{t,i} \boldsymbol{\beta}_t + \epsilon_{t,i} \quad (8.3)$$

where $T_{1i} = t_i - E(t_i)$, $T_{2i} = t_i^2 - E(t_i^2)$, and $T_{3i} = t_i^3 - E(t_i^3)$; w_i^* represents the latent unobservable counterpart of the binary variable w_i (for instance, w_i^* might be seen as the net benefit—cost minus return—of an agency choosing to finance specific subjects); t_i is fully observed only when $w_i = 1$ (and $t_i = t_i'$) and otherwise is supposed to be unobserved (although put equal to 0); $\mathbf{x}_{w,i}$ and $\mathbf{x}_{t,i}$ are two sets of exogenous regressors; and $\epsilon_{w,i}$, $\epsilon_{t,i}$, and η_i are error terms that are supposed to be freely correlated with one another with 0 unconditional mean. Equation (8.2)—the selection equation—defines the regression explaining the net benefit indicator w^* . The vector of covariates $\mathbf{x}_{w,i}$ are the selection criteria used, for instance, by an agency to set the treated and untreated groups. In turn, (8.3)—the treatment-level equation—defines how the level of unit treatment is decided, and it only considers units that were eligible for treatment. Finally, the vector of covariates $\mathbf{x}_{t,i}$ are those exogenous variables thought of as determining the treatment level.

In (8.1), w_i , T_{1i} , T_{2i} , and T_{3i} are endogenous, with the latter three being functions of the endogenous t . In general, with two endogenous variables, the identification of system

(8.1–8.3) would require the availability of at least two IVs, $z_{w,i}$ and $z_{t,i}$, supposed to be directly correlated with w_i^* and t_i' but not with y_i (exclusion restriction), and supposed to be uncorrelated with $\epsilon_{w,i}$, $\epsilon_{t,i}$, and η_i (exogeneity). This leads naturally to the following specification of the exogenous confounding variables in system (8.1–8.3):

$$\begin{aligned}\mathbf{x}_{w,i} &= (\mathbf{x}_i, z_{w,i}) \\ \mathbf{x}_{t,i} &= (\mathbf{x}_i, z_{t,i})\end{aligned}$$

Practical estimation of system (8.1–8.3) starts from recognizing that (8.2) and (8.3) together represent a bivariate sample-selection model or type-2 tobit model (Heckman 1979).⁴ This model is usually fit by invoking some distributive assumptions regarding the error terms. We assume that the error terms in (8.2) and (8.3) are jointly normally distributed and homoskedastic:

$$\begin{bmatrix} \epsilon_{w,i} \\ \epsilon_{t,i} \end{bmatrix} \sim N \left[\begin{bmatrix} 0 \\ 0 \end{bmatrix} : \begin{bmatrix} 1 & \sigma_{w,t} \\ \sigma_{w,t} & \sigma_t^2 \end{bmatrix} \right]$$

where the normalization $\sigma_w = 1$ is used because only the sign of w_i^* is observed. Given this additional assumption, all the ingredients are available to provide a procedure for estimating system (8.1–8.3) consistently:

1. Fit (8.2)–(8.3) jointly by a type-2 tobit model.

Comment. This can be achieved by a Heckman two-step procedure (Heckman 1979). The Heckman two-step procedure performs a probit of w_i on $\mathbf{x}_{w,i}$ in the first step, using only the N_1 selected observations. In the second step, it performs an OLS regression of t_i' on $\mathbf{x}_{t,i}$, augmented by the Mills' ratio obtained from the probit in the first step, using all the N observations as predictions are also made for the censored data. However, because of the errors' joint normality, a maximum likelihood estimation also can be used, leading to more efficient estimates of β_w and β_t .

2. Compute the predicted values of w_i (that is, \hat{p}_{wi}) and t_i (that is, \hat{t}_i) from the previous type-2 tobit estimation. Then perform a two-stage least squares (2SLS) for (8.1) using as instruments the following exogenous variables: $\{\mathbf{x}_i, \hat{p}_{wi}, \hat{p}_{wi}(\mathbf{x}_i - \bar{\mathbf{x}}), \hat{p}_{wi}\hat{T}_{1i}, \hat{p}_{wi}\hat{T}_{2i}, \hat{p}_{wi}\hat{T}_{3i}\}$.

4. It is not clear which is the link between the definition of the local ATE as proposed by Imbens and Angrist (1994) and the IV approach as proposed in this article. The problem is that local ATE identifies the causal effect of w on y in a setting where the instrument z is binary. Although extensions have been provided to the case in which w can be multivalued and more than one binary z is available (Angrist and Pischke 2009, 173–186), no comparable findings have been found in the literature thus far for the case in which both the treatment and the instrument take values on a continuous support. The use of IV under heterogeneous effects will, however, identify a causal effect for a subpopulation and not for the (complete) population. This is irrespective of the precise characterization of the subpopulation (for example, compliers in the binary setting). This certainly represents a trade-off when moving to the use of IV when there is effect heterogeneity based on unobservables.

Comment. This 2SLS approach provides consistent estimation of the basic coefficients μ_0 , δ_0 , ATE, δ , a , b , and c (Wooldridge 2010, 937–951).⁵

3. Once the previous step consistently estimates the basic parameters in system (8.1–8.3), the causal parameters of interest—ATEs and the DRF—can be consistently estimated by the same plug-in approach used for the OLS case.

Because this third step uses generator regressors, the standard errors must be adjusted for this. Bootstrapping standard errors may be a valid alternative, and `ctreatreg` allows for calculating the DRF bootstrapped standard errors via the postestimation command `boot_drf`.

3.3 Estimation of comparative DRFs

Aside from the DRF and other causal parameters of interest as defined above, the previous model also allows for calculating the average comparative response at different levels of treatment (as in Hirano and Imbens [2004]).

$$\text{ATE}(\mathbf{x}, \Delta) = E\{y(t + \Delta) - y(t)\} \quad (9)$$

Equation (9) identifies the ATE between two states (or levels of treatment): t and $t + \Delta$. Given a level of $\Delta = \overline{\Delta}$, we can get a particular $\text{ATE}(t, \overline{\Delta})$ that can be seen as the treatment function at $\overline{\Delta}$.

4 The `ctreatreg` command

The command `ctreatreg` estimates DRFs under CMI and under treatment endogeneity.⁶ A description of all available options is provided in the `ctreatreg` help file. Here, I report the syntax and comment on the functionality of just the main options.

4.1 Syntax of `ctreatreg`

```
ctreatreg outcome treatment [varlist] [if] [in] [weight], model(modeltype)
      ct(treat_level) m(number) s(number) [hetero(varlist_h) estype(model)
      iv_t(instrument_t) iv_w(instrument_w) delta(number) ci(number) graphate
      graphdrf conf(number) vce(robust) heckvce(vcetype) const(noconstant)
      head(noheader) ]
```

`ctreatreg` is straightforward to use and provides suitable graphical representations of results. In particular, it provides a plot of the DRF (along with its confidence interval curves) and of the density of $\text{ATE}(\mathbf{x}, t)$, $\text{ATE}(\mathbf{x}, t)$, and $\text{ATE}(\mathbf{x}, t)$.

5. Observe that the instruments used in 2SLS are based on the orthogonal projection of w_i and t_i on the vector space generated by all the exogenous variables of system (8.1–8.3).

6. For a Stata implementation when the treatment is binary, see Cerulli (2014).

4.2 Options

model(*modeltype*) specifies the treatment model to be fit, where *modeltype* must be one of **ct-ols** or **ct-iv**. **model()** is required.

ct(*treat_level*) specifies the treatment level (or dose). This variable takes values in the $[0, 100]$ interval, where 0 is the treatment level of nontreated units. The maximum dose is thus 100.⁷ **ct()** is required.

m(*number*) sets the polynomial degree of the DRF equal to *number*. **m()** is required.

s(*number*) sets a specific value of the continuous treatment variable where the DRF is evaluated. The value of $ATE(s)$ is reported in the return scalar **e(ate_s)**. **s()** is required.

hetero(*varlist_h*) specifies the variables for which to calculate the idiosyncratic $ATE(\mathbf{x})$, $ATET(\mathbf{x})$, and $ATENT(\mathbf{x})$, where $\mathbf{x} = \text{varlist}_h$. **hetero()** is optional for all models. When this option is not specified, the command fits the specified model without the heterogeneous average effect. *varlist_h* should be the same set or a subset of the variables specified in *varlist*, and only numerical variables may be considered.

estype(*model*) specifies the type of estimation method to use for fitting the type-2 tobit model in the endogenous treatment case. *model* may be **twostep** to implement a Heckman two-step procedure or **ml** to implement a maximum likelihood estimation. **estype()** is required only for **ct-iv**.

iv.t(*instrument_t*) specifies the variable to use as the instrument for the continuous treatment variable *t* in the type-2 tobit model. **iv.t()** is required only with **ct-iv**.

iv.w(*instrument_w*) specifies the variable to use as the instrument for the binary treatment variable *w* in the type-2 tobit model. **iv.w()** is required only with **ct-iv**.

delta(*number*) identifies the ATE between two states: *t* and *t* + delta. For any chosen delta, we can obtain the response function $ATE(t, \text{delta}) = E\{y(t) - y(t + \text{delta})\}$.

ci(*number*) sets the significance level for the DRF, where *number* may be 1, 5, or 10.

graphate requests a graphical representation of the density distributions of $ATE(\mathbf{x}, t)$, $ATET(\mathbf{x}, t)$, and $ATENT(\mathbf{x}, t)$. It provides an outcome only if variables are specified in **hetero()**.

graphdrf requests a graphical representation of the DRF and its derivative. By default, it also plots the 95% confidence interval of the DRF and its derivative over the dose levels. When **graphdrf** is specified, you must also specify **ci()**.

Finally, **ctreatreg** generates some useful variables for postestimation analysis and returns the estimated TES in scalars to get, for instance, bootstrapped standard errors for $ATET$ and $ATENT$ that do not have a standard analytical form (see the **ctreatreg** help file).

7. If the continuous treatment variable does not naturally lie between 0 and 100, one can use a suitable transformation to ensure that it falls within this interval.

4.3 Bootstrapped standard errors with `boot_drf`

Sometimes, it might be useful to also obtain bootstrapped standard errors for the DRF. To this aim, I provide a `ctreatreg` postestimation command called `boot_drf`, which calculates such standard errors and graphs the result automatically after running `ctreatreg` normally.

Syntax of `boot_drf`

```
boot_drf, rep(number) [size(number) saving(filename) bca]
```

Options

`rep(number)` specifies the number of bootstrap replications. `rep()` is required.

`size(number)` specifies the sample size of a single bootstrap replication. By default, it is equal to the current sample size.

`saving(filename)` specifies to save the resulting graph in *filename.gph*.

`bca` specifies to estimate confidence intervals by the bias-corrected and accelerated method.

The statistical significance level assumed by `boot_drf` is the same as the one declared in the `ci()` option of `ctreatreg`.

5 An illustrative application

Let us consider the Stata 14 example dataset `nls88.dta` collecting data from the National Longitudinal Survey of Young Women 1988, containing information on women's labor conditions such as wages, educational level, race, and marital status. We wish to study the impact of the variable `tenure` (job tenure) on `wage` (wages in dollars per hour) conditional on a series of other covariates (that is, observable confounders) referring to each single woman. The variable `tenure` is a good candidate to be exploited as continuous treatment (that is, dose) for such a model, having a (small) spike at 0:

```
. sysuse nls88.dta
(NLSW, 1988 extract)
. summarize tenure
```

Variable	Obs	Mean	Std. Dev.	Min	Max
tenure	2,231	5.97785	5.510331	0	25.91667

```
. count if tenure==0
51
```

```
. describe
Contains data from C:\Program Files\stata14\ado\base\n\slsw88.dta
  obs:      2,246      NLSW, 1988 extract
  vars:      17      1 May 2014 22:52
  size:     60,642      (_dta has notes)
```

variable name	storage type	display format	value label	variable label
idcode	int	%8.0g		NLS id
age	byte	%8.0g		age in current year
race	byte	%8.0g	racelbl	race
married	byte	%8.0g	marlbl	married
never_married	byte	%8.0g		never married
grade	byte	%8.0g		current grade completed
collgrad	byte	%16.0g	gradlbl	college graduate
south	byte	%8.0g		lives in south
smsa	byte	%9.0g	smsalbl	lives in SMSA
c_city	byte	%8.0g		lives in central city
industry	byte	%23.0g	indlbl	industry
occupation	byte	%22.0g	occlbl	occupation
union	byte	%8.0g	unionlbl	union worker
wage	float	%9.0g		hourly wage
hours	byte	%8.0g		usual hours worked
t1l_exp	float	%9.0g		total work experience
tenure	float	%9.0g		job tenure (years)

Sorted by: idcode

We consider a model where the outcome, the treatment, and the controls are defined as follows:

- outcome y : wage
- treatment w : tenure
- controls x : age, race, married, collgrad, south, occupation

Furthermore, we consider two IVs to use in the IV estimation (when assuming endogenous treatment):

- instrument for w : c_city
- instrument for t : t1l_exp

The goodness of these instruments is just assumed; it is neither discussed nor tested because this is just an instructional example.

Before estimation, we generate the binary treatment variable, called `treatment`:

```
. capture drop treatment
. generate treatment=0 if tenure==0
(2,195 missing values generated)
. replace treatment=1 if tenure >0 & tenure !=.
(2,180 real changes made)
. tabulate treatment, mis
```

treatment	Freq.	Percent	Cum.
0	51	2.27	2.27
1	2,180	97.06	99.33
.	15	0.67	100.00
Total	2,246	100.00	

We then generate the continuous treatment (dose), which we call `tenure2`:

```
. capture drop tenure2
. quietly summarize tenure, detail
. generate tenure2=(tenure-0)/(r(max)-0)*100
(15 missing values generated)
. summarize tenure2
```

Variable	Obs	Mean	Std. Dev.	Min	Max
tenure2	2,231	23.06566	21.26173	0	100

We now have all the ingredients to apply `ctreatreg` to this example. Before we fit the `ct-ols` model (by assuming unconfoundedness) and then the `ct-iv` model (by assuming treatment endogeneity), we put variables into proper global macros:

```
. global xvars age i.race i.married i.collgrad i.south i.occupation
. global xvarh age married
```


► Example

Applying ctreatreg using ct-ols (unconfoundedness):

```
. xi: ctreatreg wage treatment $xvars, graphdrf delta(10) hetero($xvarh)
> model(ct-ols) ct(tenure2) ci(1) m(3) s(10)
```

(output omitted)

Source	SS	df	MS			
Model	14597.0383	24	608.209929	Number of obs	=	2,222
Residual	59355.7045	2,197	27.0167067	F(24, 2197)	=	22.51
				Prob > F	=	0.0000
				R-squared	=	0.1974
				Adj R-squared	=	0.1886
Total	73952.7428	2,221	33.2970477	Root MSE	=	5.1978

wage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
treatment	-.2258611	.7723743	-0.29	0.770	-1.740521	1.288799
age	.1767878	.227949	0.78	0.438	-.2702303	.6238058
_lrace_2	-.1996851	.2791678	-0.72	0.475	-.7471455	.3477754
_lrace_3	.1440398	1.032971	0.14	0.889	-1.881661	2.169741
_lmarried_1	1.715242	1.586304	1.08	0.280	-1.395571	4.826055
_lcollgrad_1	2.699153	.3225719	8.37	0.000	2.066575	3.33173
_lsouth_1	-1.066165	.2348634	-4.54	0.000	-1.526743	-.6055878
_loccupatio_2	.7195711	.4397573	1.64	0.102	-.1428125	1.581955
_loccupatio_3	-2.390577	.3711658	-6.44	0.000	-3.11845	-1.662705
_loccupatio_4	-.8981176	.6056345	-1.48	0.138	-2.085794	.2895586
_loccupatio_5	-2.447399	.7883316	-3.10	0.002	-3.993352	-.9014455
_loccupatio_6	-3.49792	.4760362	-7.35	0.000	-4.431448	-2.564392
_loccupatio_7	-5.015922	1.067796	-4.70	0.000	-7.109918	-2.921927
_loccupatio_8	-4.163116	.4537492	-9.17	0.000	-5.052938	-3.273293
_loccupatio_9	-4.049776	5.216821	-0.78	0.438	-14.28019	6.180642
_loccupatio_10	-5.074733	1.777695	-2.85	0.004	-8.560871	-1.588594
_loccupatio_11	-2.754774	1.344621	-2.05	0.041	-5.391636	-.1179119
_loccupatio_12	-2.218878	3.697292	-0.60	0.548	-9.469431	5.031675
_loccupatio_13	-3.025867	.5067538	-5.97	0.000	-4.019633	-2.0321
_ws_age	-.2638586	.2308734	-1.14	0.253	-.7166115	.1888944
_ws_married	-2.31864	1.600491	-1.45	0.148	-5.457274	.8199942
Tw_1	.126682	.0389813	3.25	0.001	.050238	.2031261
Tw_2	-.0026322	.0012225	-2.15	0.031	-.0050296	-.0002347
Tw_3	.0000203	.0000104	1.95	0.051	-9.27e-08	.0000407
_cons	1.878267	9.239643	0.20	0.839	-16.24108	19.99762

(output omitted)

Results show a not very good R -squared with a negative and nonsignificant ATE, equal to around -0.22 . It means that, on average, over all values taken by job tenure, the effect of tenure on wage is negative. However, the plot of the DRF (figure 1) shows that the relationship is weakly increasing and quite precisely estimated for lower values of the dose; it is more strongly increasing but less precisely estimated for higher levels of the dose.

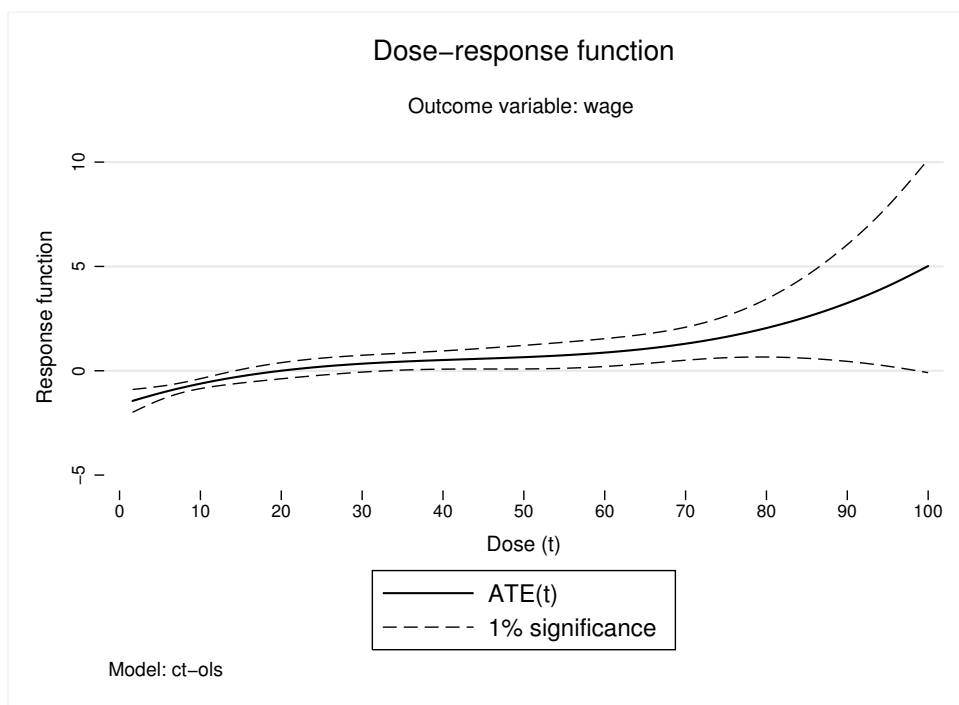


Figure 1. DRF of job tenure on wage; exogenous treatment case

By default, `ctreatreg` also plots the derivative of the DRF along with its confidence interval, as illustrated in figure 2. We can see that the derivative of the DRF is a parabola because the DRF is a cubic function. The minimum of the derivative is found around a dose of 50, where the DRF correctly exhibits a flex point.

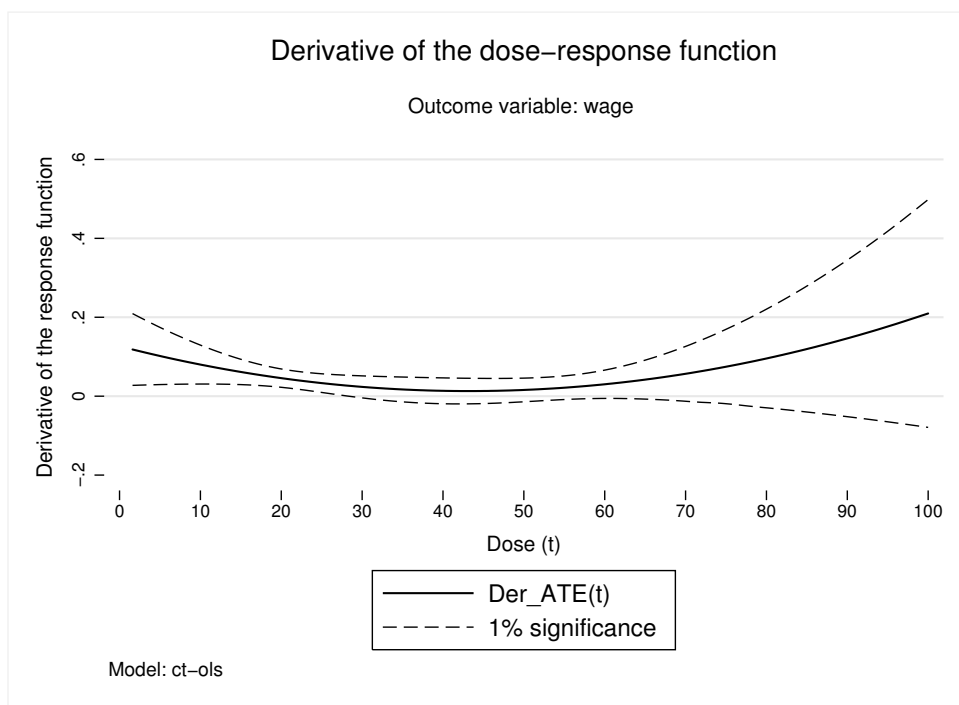


Figure 2. Derivative of the DRF of job tenure on wage; exogenous treatment case

Finally, we can estimate the DRF with bootstrapped standard errors by using the `ctreatreg` postestimation command `boot_drf` with 20 replications.

```
. quietly xi: ctreatreg wage treatment $xvars, graphdrf delta(10) hetero($xvarh)
> model(ct-ols) ct(tenure2) ci(1) m(3) s(25)
. boot_drf, rep(20)
(output omitted)
```

We obtain the plot displayed in figure 3, where we see that the bootstrapped and analytical standard errors show a very similar pattern.

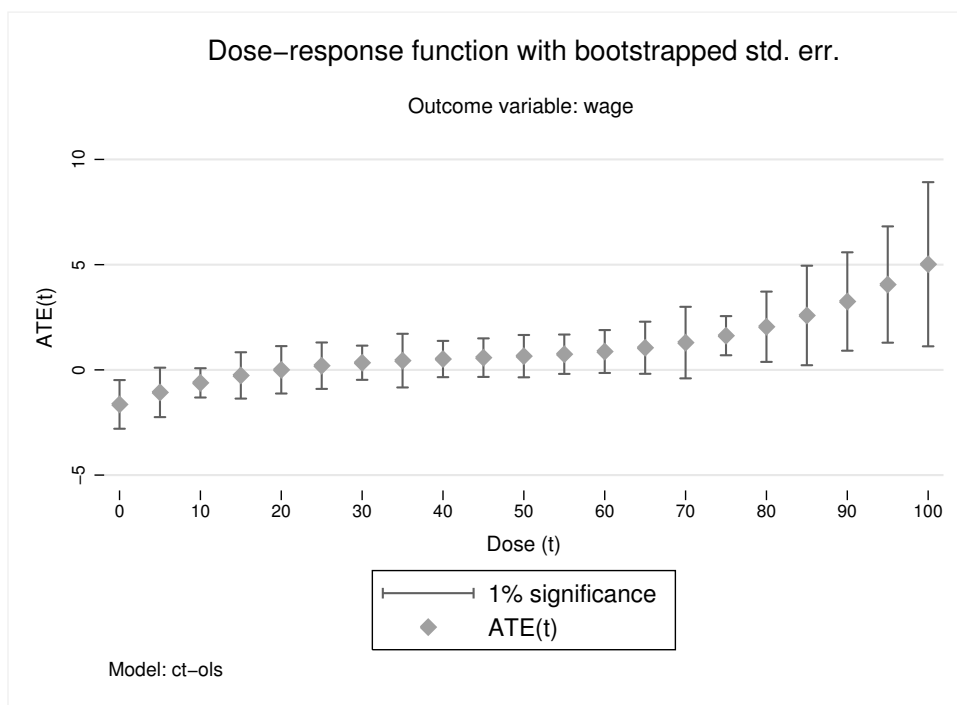


Figure 3. DRF of job tenure on wage; exogenous treatment case with bootstrapped standard errors

◀

► Example

Applying `ctreatreg` using `ct-iv` (treatment endogeneity):

```
. xi: ctreatreg wage treatment $xvars, graphdrf delta(10) hetero($xvarh)
> model(ct-iv) ct(tenure2) ci(1) m(3) s(10) estype(twostep) iv_t(ttl_exp)
> iv_w(c_city)
(output omitted)
```

```

Heckman selection model -- two-step estimates      Number of obs      =      2,222
(regression model with sample selection)          Censored obs        =         50
                                                  Uncensored obs      =      2,172
                                                  Wald chi2(19)       =      215.05
                                                  Prob > chi2         =      0.0000

```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
tenure2						
age	.2754417	.5995455	0.46	0.646	-.8996459	1.450529
_Irace_2	4.883963	5.678379	0.86	0.390	-6.245454	16.01338
_Irace_3	-5.935405	10.02299	-0.59	0.554	-25.5801	13.70929
_Imarried_1	.9848613	3.053211	0.32	0.747	-4.999322	6.969044
_Icollgrad_1	-.7735144	2.671136	-0.29	0.772	-6.008844	4.461815
_Isouth_1	-2.230267	1.940938	-1.15	0.251	-6.034436	1.573903
_Ioccupatio_2	-1.854261	3.678048	-0.50	0.614	-9.063102	5.35458
_Ioccupatio_3	1.142482	3.494785	0.33	0.744	-5.707171	7.992135
_Ioccupatio_4	-3.492005	5.128079	-0.68	0.496	-13.54286	6.558845
_Ioccupatio_5	3.216274	6.502177	0.49	0.621	-9.527759	15.96031
_Ioccupatio_6	3.173788	5.332911	0.60	0.552	-7.278525	13.6261
_Ioccupatio_7	-4.144287	9.199841	-0.45	0.652	-22.17564	13.88707
_Ioccupatio_8	-.0788224	4.295322	-0.02	0.985	-8.497499	8.339855
_Ioccupatio_9	13.30952	43.12609	0.31	0.758	-71.21606	97.8351
_Ioccupatio_10	-3.30663	16.39857	-0.20	0.840	-35.44723	28.83397
_Ioccupatio_11	.15048	12.24041	0.01	0.990	-23.84028	24.14124
_Ioccupatio_12	2.790015	30.87441	0.09	0.928	-57.72271	63.30274
_Ioccupatio_13	7.52296	4.249082	1.77	0.077	-.8050881	15.85101
tll_exp	2.674497	.1994488	13.41	0.000	2.283585	3.06541
_cons	-20.20849	21.89844	-0.92	0.356	-63.12864	22.71166
treatment						
age	-.0411952	.0197378	-2.09	0.037	-.0798805	-.0025099
_Irace_2	-.3772631	.1481417	-2.55	0.011	-.6676154	-.0869108
_Irace_3	4.241509
_Imarried_1	-.1866555	.1368208	-1.36	0.172	-.4548194	.0815084
_Icollgrad_1	-.0239416	.1832938	-0.13	0.896	-.3831908	.3353076
_Isouth_1	-.0170888	.1291613	-0.13	0.895	-.2702404	.2360627
_Ioccupatio_2	.0794301	.2837233	0.28	0.780	-.4766574	.6355175
_Ioccupatio_3	-.1461987	.21451	-0.68	0.496	-.5666305	.2742331
_Ioccupatio_4	-.1138135	.3431414	-0.33	0.740	-.7863584	.5587314
_Ioccupatio_5	.001804	.4545683	0.00	0.997	-.8891336	.8927416
_Ioccupatio_6	-.2590639	.2508861	-1.03	0.302	-.7507916	.2326638
_Ioccupatio_7	-.1739169	.5023217	-0.35	0.729	-1.158449	.8106155
_Ioccupatio_8	-.1712341	.2485614	-0.69	0.491	-.6584054	.3159373
_Ioccupatio_9	3.903505
_Ioccupatio_10	4.389021
_Ioccupatio_11	4.386132
_Ioccupatio_12	4.21906
_Ioccupatio_13	.0566438	.3059878	0.19	0.853	-.5430813	.6563688
c_city	.0300973	.1400446	0.21	0.830	-.2443852	.3045798
_cons	3.981387	.8266223	4.82	0.000	2.361237	5.601537
mills						
lambda	-42.75893	116.9072	-0.37	0.715	-271.8928	186.375
rho						
sigma	-1.00000	42.758929				

(output omitted)

Instrumental variables (2SLS) regression

Source	SS	df	MS	Number of obs	=	2,222
Model	5979.39919	24	249.141633	F(24, 2197)	=	20.92
Residual	67973.3437	2,197	30.9391642	Prob > F	=	0.0000
				R-squared	=	0.0809
				Adj R-squared	=	0.0708
Total	73952.7428	2,221	33.2970477	Root MSE	=	5.5623

wage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
treatment	-5.935915	51.67792	-0.11	0.909	-107.2786	95.40679
_ws_age	1.446693	3.919612	0.37	0.712	-6.23984	9.133225
_ws_married	-17.08296	26.39882	-0.65	0.518	-68.85222	34.6863
Tw_1	.3203173	.1779275	1.80	0.072	-.0286064	.669241
Tw_2	-.005373	.0051574	-1.04	0.298	-.0154869	.0047409
Tw_3	.0000309	.0000396	0.78	0.435	-.0000467	.0001085
age	-1.516218	3.919952	-0.39	0.699	-9.203418	6.170981
_Irace_2	-.5161937	.9440787	-0.55	0.585	-2.367574	1.335186
_Irace_3	.5907224	1.478986	0.40	0.690	-2.309635	3.491079
_Imarried_1	16.15121	25.5332	0.63	0.527	-33.92053	66.22296
_Icollgrad_1	2.457484	.4774821	5.15	0.000	1.521121	3.393848
_Isouth_1	-1.00309	.2542077	-3.95	0.000	-1.501602	-.5045771
_Ioccupatio_2	.6295055	.5350773	1.18	0.240	-.4198048	1.678816
_Ioccupatio_3	-2.1374	.5723671	-3.73	0.000	-3.259837	-1.014963
_Ioccupatio_4	-.0207587	.7577837	-0.03	0.978	-1.506806	1.465289
_Ioccupatio_5	-2.534627	.8942808	-2.83	0.005	-4.288352	-.7809032
_Ioccupatio_6	-3.507123	1.332124	-2.63	0.009	-6.119476	-.8947696
_Ioccupatio_7	-3.528173	1.252746	-2.82	0.005	-5.984864	-1.071482
_Ioccupatio_8	-3.675224	.9923129	-3.70	0.000	-5.621193	-1.729254
_Ioccupatio_9	-5.300002	5.703895	-0.93	0.353	-16.48559	5.88559
_Ioccupatio_10	-3.290802	2.179257	-1.51	0.131	-7.564422	.9828187
_Ioccupatio_11	-2.040387	1.638178	-1.25	0.213	-5.252926	1.172152
_Ioccupatio_12	-2.706283	4.019555	-0.67	0.501	-10.58881	5.176242
_Ioccupatio_13	-3.178334	.5584581	-5.69	0.000	-4.273495	-2.083173
_cons	64.34971	187.35	0.34	0.731	-303.052	431.7515

Instrumented:	treatment _ws_age _ws_married Tw_1 Tw_2 Tw_3
Instruments:	age _Irace_2 _Irace_3 _Imarried_1 _Icollgrad_1 _Isouth_1 _Ioccupatio_2 _Ioccupatio_3 _Ioccupatio_4 _Ioccupatio_5 _Ioccupatio_6 _Ioccupatio_7 _Ioccupatio_8 _Ioccupatio_9 _Ioccupatio_10 _Ioccupatio_11 _Ioccupatio_12 _Ioccupatio_13 probw _ps_age _ps_married T_hatp_1 T_hatp_2 T_hatp_3

(output omitted)

We see that the ATE becomes even more negative (-5.93) but is still insignificant. However, the DRF (figure 4) sets out a pattern similar to the previous model, with a slight cubic form having a flex point around a dose of 60 this time.

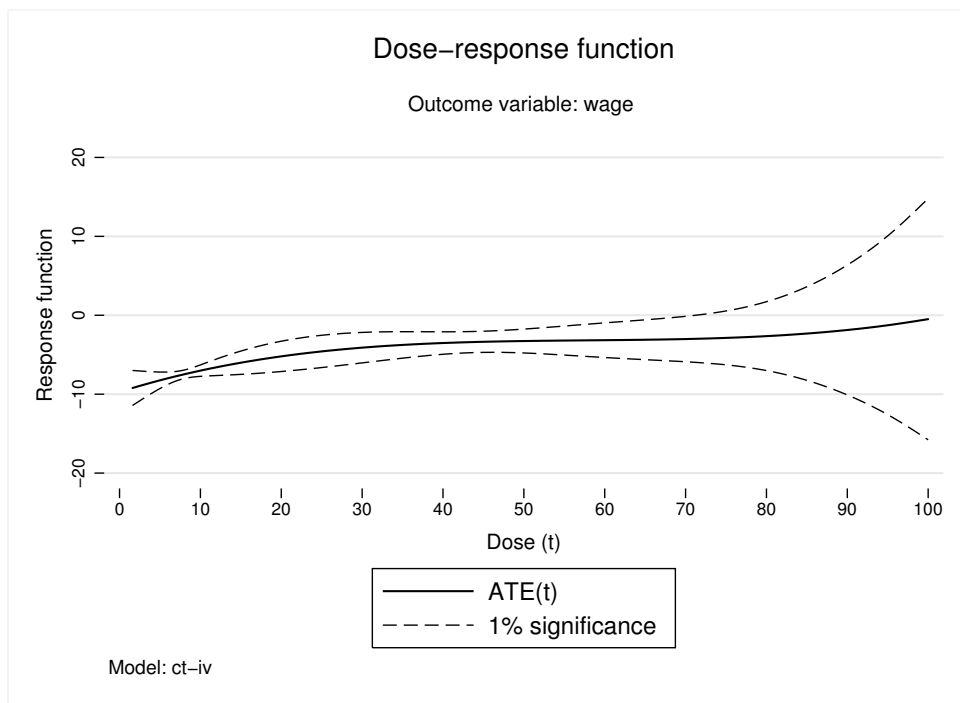


Figure 4. DRF of job tenure on wage; endogenous treatment case

Of course, such results have to be taken just as illustrative, because we have no idea about the quality of the instruments used, in particular about their exogeneity.

◀

6 A Monte Carlo experiment for testing `ctreatreg`'s reliability

This section provides a Monte Carlo experiment to check whether `ctreatreg` complies with the consistency of the DRF estimates and to assess its correctness from a computational point of view. The first step is that of defining a data-generating process (DGP) as follows:

$$\left\{ \begin{array}{l} w = 1(50 + 60x_1 - 30x_2 + 25z_w + a > 0) \\ y_0 = 0.1 + 0.2x_1 + 0.3x_2 + e \\ y_1 = 0.3 + 0.6x_1 + 0.3x_2 + h(t) + e \\ t = 0.4x_1 + 0.5x_2 + 0.1z_t + u \\ h(t) = 0.888t - 0.023t^2 + 0.00017t^3 \end{array} \right.$$

We have assumed, for simplifying the model, that $e_1 = e_0 = e$ and that

$$\begin{cases} x_1 & \sim U(0, 1) \times 100 \\ x_2 & \sim U(0, 1) \times 100 \\ z_w & \sim N(15, 1) \\ z_t & \sim N(100, 1) \end{cases}$$

We are interested in comparing the performance of **ctreatreg** when i) the error terms have a joint normal distribution and treatment exogeneity is assumed; ii) the error terms have a normal joint distribution and treatment endogeneity is assumed; iii) the error terms have a nonnormal joint distribution and treatment endogeneity is assumed.

In the first and second cases, we assume the following:

$$\begin{aligned} (a, u) &\sim N(0, \mathbf{\Omega}) \\ \mathbf{\Omega} &= \begin{pmatrix} \sigma_a^2 & \sigma_{a,u} \\ \sigma_{a,u} & \sigma_u^2 \end{pmatrix} = \begin{pmatrix} \sigma_a^2 & \rho_{a,u}\sigma_a\sigma_u \\ & \sigma_u^2 \end{pmatrix} \\ \sigma_a^2 &= 0.65, \quad \sigma_u^2 = 1, \quad \rho_{a,u} = 0.7 \end{aligned}$$

We then suppose that the correlation between a and e_0 can be either equal or different from 0. In the latter case, w is endogenous. Therefore, we assume the following DGP:⁸

$$\begin{aligned} e &= \eta + \gamma a + v \\ v &\sim N(0, 1) \\ \gamma &= \sqrt{\rho^2/(1 - \rho^2)} \\ \rho &= \text{corr}(e, a) \\ \eta &= 0.0001 \end{aligned}$$

When $\rho = 0$, the model **ct-ols** would be the appropriate one; otherwise, the model **ct-iv** should be used.

By z_w and z_t , we indicate the IVs for w and t , directly correlated with w and t , respectively, but (directly) uncorrelated with y_1 and y_0 . Given these assumptions, the DGP is completed by the potential outcome equation $y_i = y_{0i} + w_i(y_{1i} - y_{0i})$, generating the observable outcome (or response) y .

The DGP is simulated 200 times using a sample size of 100, 500, 1,000, 3,000, and 10,000 to evaluate both finite- and large-sample properties of the estimation proposed by **ctreatreg**. For each simulation, we get a different data matrix $(x_1, x_2, y, w, t, z_w, z_t)$ on which we apply the two models (**ct-ols** and **ct-iv**) implemented by **ctreatreg**.

8. The coefficient γ is equal to $\{\rho^2/(1 - \rho^2)\}^{-1/2}$, where $\rho = \text{corr}(e_0, a)$. To get this result, put $x = e$ and $y = a$. We know that $\text{corr}(x, y) = \text{cov}(x, y)/\text{sd}(x)\text{sd}(y)$. We can see that, while $\text{var}(y) = 1$ by assumption, $\text{var}(x) = \gamma^2 + 1$. Moreover, $\text{cov}(x, y) = \text{cov}(\eta + \gamma a + v, a) = \text{cov}(\eta + \gamma a, a) + \text{cov}(v, a) = \text{cov}(\eta + \gamma a, a) = \text{cov}(\gamma a, a) = \gamma \text{cov}(a, a) = \gamma \text{var}(a) = \gamma$. Thus, $\rho = \gamma/(\gamma^2 + 1)^{-1/2}$, which implies that $\gamma = \{\rho^2/(1 - \rho^2)\}^{-1/2}$.

6.1 Case 1: Exogeneity

We start by assuming $\rho = 0$, that is, zero correlation between the error term of the outcome equation (e) and the error term of the selection equation (a). Under this assumption, w is exogenous. Moreover, we assume a strong correlation between the selection and the dose equation, as implied by a correlation between a and u equal to 0.7.

Results are set out in table 1. The value of ATE obtained by the `ct-ols` estimator is really close to the true ATE (9.22), and the confidence interval at 5% significance for this estimator strictly contains that value. But also, the percentage bias of `ct-iv` is very low (0.86%) and comparable with `ct-ols` (0.81%), sufficient to imply that the 5% significance contains the true ATE even in this case.

Table 1. Mean test of ATE from Monte Carlo results using `ctreatreg`; exogenous selection is assumed.

	Mean	Standard error	[95% confidence interval]	
ATE (true value)	9.22	-	-	-
ATE— <code>ct-ols</code>	9.21	0.01	9.19	9.22
ATE— <code>ct-iv</code>	9.20	0.01	9.19	9.22
% bias of OLS	0.81	0.04	0.73	0.90
% bias of IV	0.86	0.04	0.77	0.94

Note: $\rho = 0$, 10,000 observations, 200 simulations.

These results confirm what was expected, thus showing that `ct-ols` behaves correctly. Thus, when assuming exogeneity, an analyst may reliably use `ctreatreg` with the option `ct-ols`.

6.2 Case 2: Endogeneity

If we assume that $\rho = 0.7$, that is, a high positive correlation between the error term of the outcome equation (e) and the error term of the selection equation (a), then w becomes endogenous. For the sake of comparison, we still assume the same strong correlation between the selection and the dose equation (0.7).

Table 2 shows that results are again coherent with the theoretical predictions. Indeed, the percentage bias of model `ct-ols` is rather high and equal to around 18%, whereas the bias of `ct-iv` is around 1%. Furthermore, and more importantly, the 95% mean test confidence interval for `ct-iv` contains the true ATE.

Table 2. Mean test of ATE from Monte Carlo results using `ctreatreg`; endogenous selection is assumed

	Mean	Standard error	[95% confidence interval]	
ATE (true value)	9.22	-	-	-
ATE— <code>ct-ols</code>	7.53	0.01	7.51	7.55
ATE— <code>ct-iv</code>	9.22	0.01	9.20	9.24
% bias of OLS	18.26	0.11	18.05	18.48
% bias of IV	1.28	0.07	1.15	1.41

Note: $\rho = 0.7$, 10,000 observations, 200 simulations.

As expected, this implies that `ct-iv` is a consistent estimator in the presence of selection endogeneity, thus leading to a reliable estimation of the true value of ATE.

These results confirm the reliability of the model and of `ctreatreg` under both selection exogeneity or selection endogeneity.

Finally, figure 5 plots the DRF along with the 95% interval confidence lines for both models. This is done with the `graphdrf` option of `ctreatreg`. Results clearly confirm our expectations.

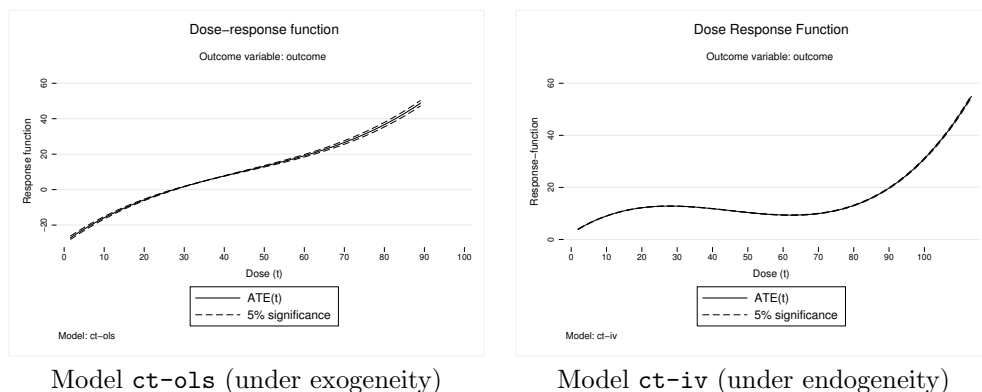


Figure 5. Graphical representation of the DRF using `ctreatreg` options `ct-ols` and `ct-iv` under exogeneity and endogeneity, respectively

6.3 Case 3: Endogeneity and joint nonnormal error terms

In this case, we assume that $\rho = 0.7$ and

$$\begin{aligned}(a, u) &\sim F(0, \mathbf{\Omega}) \\ \mathbf{\Omega} &= \begin{pmatrix} \sigma_a^2 & \sigma_{a,u} \\ \sigma_{a,u} & \sigma_u^2 \end{pmatrix} = \begin{pmatrix} \sigma_a^2 & \rho_{a,u}\sigma_a\sigma_u \\ \rho_{a,u}\sigma_a\sigma_u & \sigma_u^2 \end{pmatrix} \\ \sigma_a^2 &= 1, \quad \sigma_u^2 = 1, \quad \rho_{a,u} = 0.7 \\ a &\sim U(5, 10), \quad u \sim \chi^2(5)\end{aligned}$$

where $F(\cdot)$ is a nonnormal joint distribution with covariance matrix $\mathbf{\Omega}$ and mean 0, and a and u are, respectively, marginally uniform within $(5, 10)$ and marginally chi-squared with 5 degrees of freedom.

The model is completed by assuming that the error term e of the y equation is also nonnormally distributed,

$$e = \eta + \gamma a + (1 - v) \quad (10)$$

$$v \sim \chi^2(1) \quad (11)$$

thus having a quite strong chi-squared asymmetric distribution.

The results in table 3 illustrate the percentage bias of the **ct-iv** estimator in the normal and the nonnormal setting using 50 simulations and increasing sample size. The IV approach is inconsistent in finite samples also assuming the normal setting, and this is confirmed by the 12% bias appearing in the first column. By increasing the sample size, however, the bias tends to disappear, becoming lower than 1% for a size of 10,000.

Table 3. Percentage bias of ATE from a Monte Carlo simulation using the estimation option **ct-iv** when treatment is endogenous and errors are jointly nonnormally distributed

	Sample size				
	100	500	1,000	3,000	10,000
ct-iv (normal setting)	12.03	4.36	3.61	1.79	0.77
ct-iv (nonnormal setting)	22.46	9.95	6.87	5.12	2.18

Note: $\rho = 0.7$, 50 simulations, true value of ATE = 8.755.

When we use the **ct-iv** option in a nonnormal setting, however, results stress a higher percentage bias, which is decreasing but not disappearing with a larger sample size. When the sample size is 10,000, the bias is 2.18%, which is unexpectedly not dramatically high. We can conclude—at least in this stylized DGP—that **ctreatreg** also seems rather robust in the case of departures from the joint normality of the error

terms of the first step of the Heckman procedure and in the nonnormality of the error of the outcome equation.

Although drawing on the nonnormal joint distribution of the errors, these results still consider smooth distribution functions of the errors. In the case of nonsmooth distribution functions (as in cases characterized by large probability mass in some specific points of the support of the continuous treatment t), previous results might be questionable. Such a case, however, is beyond the purpose of the present model, because we are interested in modeling a continuous treatment setting. When treatment continuity is poor and some points have large probability mass, then a multiple-treatment approach may be more suitable (Cattaneo 2010; Cattaneo, Drukker, and Holland 2013).

7 Conclusion

In this article, I presented `ctreatreg`, a command for estimating DRFs through a regression approach where i) treatment is continuous, ii) individuals may react heterogeneously to observable confounders, and iii) the selection into treatment may be endogenous.

Two estimation procedures are contemplated by this command: one based on OLS under CMI, and one based on IV when selection endogeneity is assumed.

We saw an application to real data, for testing the impact of job tenure on wages.

To test the reliability of the formulas and of their associated Stata implementation, we performed a Monte Carlo experiment. The results showed that the models' formulas are fairly reliable because estimates comply with the expected results. The proposed estimators also appear to be quite robust when departures from the errors' joint normality are allowed, although this may rely on styled DGP assumptions.

8 References

- Adorno, V., C. Bernini, and G. Pellegrini. 2007. The impact of capital subsidies: New estimations under continuous treatment. *Giornale degli Economisti e Annali di Economia* 66: 67–92.
- Angrist, J. D., and J.-S. Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.
- Bia, M., C. A. Flores, A. Flores-Lagunes, and A. Mattei. 2014. A Stata package for the application of semiparametric estimators of dose–response functions. *Stata Journal* 14: 580–604.
- Bia, M., and A. Mattei. 2008. A Stata package for the estimation of the dose–response function through adjustment for the generalized propensity score. *Stata Journal* 8: 354–373.

- Cattaneo, M. D. 2010. Efficient semiparametric estimation of multi-valued treatment effects under ignorability. *Journal of Econometrics* 155: 138–154.
- Cattaneo, M. D., D. M. Drukker, and A. D. Holland. 2013. Estimation of multivalued treatment effects under conditional independence. *Stata Journal* 13: 407–450.
- Cerulli, G. 2014. *ivtreatreg*: A command for fitting binary treatment models with heterogeneous response to treatment and unobservable selection. *Stata Journal* 14: 453–480.
- Guardabascio, B., and M. Ventura. 2014. Estimating the dose-response function through a generalized linear model approach. *Stata Journal* 14: 141–158.
- Heckman, J. J. 1979. Sample selection bias as a specification error. *Econometrica* 47: 153–161.
- Hirano, K., and G. W. Imbens. 2004. The propensity score with continuous treatments. In *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, ed. A. Gelman and X.-L. Meng, 73–84. Chichester, UK: Wiley.
- Imai, K., and D. A. van Dyk. 2004. Causal inference with general treatment regimes: Generalizing the propensity score. *Journal of the American Statistical Association* 99: 854–866.
- Imbens, G. W., and J. D. Angrist. 1994. Identification and estimation of local average treatment effects. *Econometrica* 62: 467–475.
- Robertson, C., P. Boyle, C. C. Hsieh, G. J. MacFarlane, and P. Maisonneuve. 1994. Some statistical considerations in the analysis of case-control studies when the exposure variables are continuous measurements. *Epidemiology* 5: 164–170.
- Royston, P., and W. Sauerbrei. 2008. *Multivariable Model-building: A Pragmatic Approach to Regression Analysis Based on Fractional Polynomials for Modelling Continuous Variables*. Chichester, UK: Wiley.
- Rubin, D. B. 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 66: 688–701.
- Wooldridge, J. M. 1997. On two stage least squares estimation of the average treatment effect in a random coefficient model. *Economics Letters* 56: 129–133.
- . 2003. Further results on instrumental variables estimation of average treatment effects in the correlated random coefficient model. *Economics Letters* 79: 185–191.
- . 2010. *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. Cambridge, MA: MIT Press.

About the authors

Giovanni Cerulli is a researcher at CNR-IRCrES, National Research Council of Italy, Institute for Research on Sustainable Economic 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 main research interests are 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.