Review of Economic Studies (2014) **81**, 608–650 doi: 10.1093/restud/rdt044
© The Author 2013. Published by Oxford University Press on behalf of The Review of Economic Studies Limited.
Advance access publication 24 November 2013

# Inference on Treatment Effects after Selection among High-Dimensional Controls†

ALEXANDRE BELLONI

*Duke University*

VICTOR CHERNOZHUKOV

*MIT*

and

CHRISTIAN HANSEN

*University of Chicago*

*First version received October* 2012; *final version accepted October* 2013 (*Eds.*)

We propose robust methods for inference about the effect of a treatment variable on a scalar outcome in the presence of very many regressors in a model with possibly non-Gaussian and heteroscedastic disturbances. We allow for the number of regressors to be larger than the sample size. To make informative inference feasible, we require the model to be approximately sparse; that is, we require that the effect of confounding factors can be controlled for up to a small approximation error by including a relatively small number of variables whose identities are unknown. The latter condition makes it possible to estimate the treatment effect by selecting approximately the right set of regressors. We develop a novel estimation and uniformly valid inference method for the treatment effect in this setting, called the "post-double-selection" method. The main attractive feature of our method is that it allows for imperfect selection of the controls and provides confidence intervals that are valid uniformly across a large class of models. In contrast, standard post-model selection estimators fail to provide uniform inference even in simple cases with a small, fixed number of controls. Thus, our method resolves the problem of uniform inference after model selection for a large, interesting class of models. We also present a generalization of our method to a fully heterogeneous model with a binary treatment variable. We illustrate the use of the developed methods with numerical simulations and an application that considers the effect of abortion on crime rates.

*Key words*: Treatment effects, Partially linear model, High-dimensional-sparse regression, Inference under imperfect model selection, Uniformly valid inference after model selection, Average treatment effects, Lasso, Orthogonality of estimating equations with respect to nuisance parameters.

*JEL Codes*: C01

## 1. INTRODUCTION

Many empirical analyses focus on estimating the structural, causal, or treatment effect of some variable on an outcome of interest. For example, we might be interested in estimating the causal

†This is a revision of a 2011 ArXiv/CEMMAP paper entitled "Estimation of Treatment Effects with High-Dimensional Controls".

- Now let's consider a related idea to Double ML

- Now let's consider a related idea to Double ML
- This is known as post double selection (Belloni et al., 2014)

- Now let's consider a related idea to Double ML

- This is known as post double selection (Belloni et al., 2014)

- It is a useful way to estimate treatment effects in linear models

- Now let's consider a related idea to Double ML
- This is known as post double selection (Belloni et al., 2014)
- It is a useful way to estimate treatment effects in linear models
- Same setup as Double ML, but here $g(\cdot)$ and $m(\cdot)$ are linear-in-parameters

- Now let's consider a related idea to Double ML
- This is known as post double selection (Belloni et al., 2014)
- It is a useful way to estimate treatment effects in linear models
- Same setup as Double ML, but here $g(\cdot)$ and $m(\cdot)$ are linear-in-parameters

$$
\begin{aligned}
Y &= D \cdot \theta + g(X) + U, & \mathbb{E}[U|X, D] = 0 \\
D &= m(X) + V, & \mathbb{E}[V|X] = 0
\end{aligned}
$$

PDS steps:

PDS steps:

1. Use LASSO to separately select $X$

PDS steps:

1. Use LASSO to separately select $X$
   - First on $Y = g(X) + \tilde{U}$

PDS steps:

1. Use LASSO to separately select $X$
   - First on $Y = g(X) + \tilde{U}$
   - Then on $D = m(X) + V$

PDS steps:

1. Use LASSO to separately select $X$
   - First on $Y = g(X) + \tilde{U}$
   - Then on $D = m(X) + V$
2. Regress $Y$ on $D$ and the *union* of the selected $X$'s from step 1

- Called "post double selection" because the final regression is on the set of $X$'s that have been doubly selected (first in the outcome equation, then in the selection equation)

- Called "post double selection" because the final regression is on the set of $X$'s that have been doubly selected (first in the outcome equation, then in the selection equation)

- Key idea is that we avoid regularization bias by only looking at the selection part of LASSO (not the shrinkage part)

## Usefulness of PDS

- For an example, let's re-evaluate Donohue and Levitt (2001)

## Usefulness of PDS

- For an example, let's re-evaluate Donohue and Levitt (2001)
- Their claim: legalizing abortion reduces crime

## Usefulness of PDS

- For an example, let's re-evaluate Donohue and Levitt (2001)
- Their claim: legalizing abortion reduces crime
    - Intuition: unwanted children are most likely to become criminals

## Usefulness of PDS

- For an example, let's re-evaluate Donohue and Levitt (2001)

- Their claim: legalizing abortion reduces crime
  - Intuition: unwanted children are most likely to become criminals

- Use a "two-way fixed effects" model on state-level panel data:

## Usefulness of PDS

- For an example, let's re-evaluate Donohue and Levitt (2001)
- Their claim: legalizing abortion reduces crime
  - Intuition: unwanted children are most likely to become criminals
- Use a "two-way fixed effects" model on state-level panel data:

$$y_{st} = \alpha a_{st} + \beta w_{st} + \delta_s + \gamma_t + \varepsilon_{st}$$

## Usefulness of PDS

- For an example, let's re-evaluate Donohue and Levitt (2001)
- Their claim: legalizing abortion reduces crime
  - Intuition: unwanted children are most likely to become criminals
- Use a "two-way fixed effects" model on state-level panel data:

$$y_{st} = \alpha a_{st} + \beta w_{st} + \delta_s + \gamma_t + \varepsilon_{st}$$

where $s$ is US state, $t$ is time, and $a_{st}$ is the abortion rate (15-25 years prior)

- $y_{st}$ are various measures of crime (property, violent, murder, ...)

- $y_{st}$ are various measures of crime (property, violent, murder, ...)
- $w_{st}$ are state-level controls (prisoners per capita, police per capita, ...)

## Re-evaluating Donohue and Levitt (2001)

- A potential issue with Donohue and Levitt (2001): specification of $w_{st}$

## Re-evaluating Donohue and Levitt (2001)

- A potential issue with Donohue and Levitt (2001): specification of $w_{st}$
- We might think we should include highly flexible forms of elements of $w_{st}$

## Re-evaluating Donohue and Levitt (2001)

- A potential issue with Donohue and Levitt (2001): specification of $w_{st}$
- We might think we should include highly flexible forms of elements of $w_{st}$
- Indeed, when Belloni et al. (2014) do this, the SE's get larger

Re-evaluating Donohue and Levitt (2001)

- A potential issue with Donohue and Levitt (2001): specification of $w_{st}$
- We might think we should include highly flexible forms of elements of $w_{st}$
- Indeed, when Belloni et al. (2014) do this, the SE's get larger
- All previous results are diminished in magnitude and have 2x-3x larger SE's

## Re-evaluating Donohue and Levitt (2001)

- A potential issue with Donohue and Levitt (2001): specification of $w_{st}$
- We might think we should include highly flexible forms of elements of $w_{st}$
- Indeed, when Belloni et al. (2014) do this, the SE's get larger
- All previous results are diminished in magnitude and have 2x-3x larger SE's
- The PDS approach is also useful for other regression designs such as DiD