Valid Causal Inference with (Some) Invalid Instruments

Jason Hartford, Victor Veitch, Dhanya Sridhar, Kevin Leyton-Brown
Causal inference with unobserved confounding
Causal inference with unobserved confounding

BMI: 50
\[ P(\text{Heart Disease} | \text{do}(\text{BMI} = 50)) \]

BMI: 23
\[ P(\text{Heart Disease} | \text{do}(\text{BMI}=23)) \]
Causal inference with unobserved confounding

BMI → Heart disease
Causal inference with unobserved confounding
Causal inference with unobserved confounding

BMI: 23

BMI: 50

Diet
Exercise
etc...

BMI

Heart disease
Causal inference with unobserved confounding

Diet → BMI → Heart disease
Exercise → BMI → Heart disease
etc... → BMI → Heart disease
Causal inference with unobserved confounding

Diet
Exercise
e tc...

BMI

Heart disease
Causal inference with unobserved confounding

Diet → BMI → Heart disease
Exercise → BMI → Heart disease
etc... → BMI → Heart disease
Causal inference with unobserved confounding

Instrument → BMI

Diet → BMI
Exercise → BMI
etc... → BMI

BMI → Heart disease
Causal inference with unobserved confounding

- Instrument
- BMI
- Heart disease
- Diet
- Exercise
- etc...

Outcome
Causal inference with unobserved confounding

Diagram:
- Instrument
- BMI
- Heart disease
- Treatment
- Outcome

Factors:
- Diet
- Exercise
- etc...
Causal inference with unobserved confounding
Deep IV: A Flexible Approach for Counterfactual Prediction

Jason Hartford¹  Greg Lewis²  Kevin Leyton-Brown¹  Matt Taddy²

Abstract

Counterfactual prediction requires understanding causal relationships between so-called treatment and outcome variables. This paper provides a recipe for augmenting deep learning methods to accurately characterize such relationships in the presence of instrument variables (IVs)—sources of treatment randomization that are conditionally independent from the outcomes. Our IV specification resolves into two prediction tasks that can be solved with deep neural nets: a first-stage network for treatment prediction and a second-stage network whose loss function involves integration over the conditional treatment distribution. This Deep IV framework allows us to take advantage of data to optimize the prices it charges its customers: in this case, price is the treatment variable and the customer’s decision about whether to buy a ticket is the outcome. There are two ways that a naïve analysis could lead to incorrect counterfactual predictions. First, imagine that price varies in the training data because the airline gradually increases prices as a plane fills. Around holidays, more people want to fly and hence planes become fuller leading to higher prices. So, in our training set we observe examples with high prices and high sales. A direct ML approach might incorrectly predict that if the airline were to increase prices at other times in the year they would also observe increased sales, whereas the true relationship between price and sales is surely negative. Typically we can observe holidays, and include them in the model, so that we can correct for their effects. This case
Causal inference with unobserved confounding

Deep IV: A Flexible Approach for Counterfactual Prediction

Jason Hartford 1 Greg Lewis 2 Kevin Leyton-Brown 1 Matt Taddy 2

Abstract

Counterfactual prediction requires understanding causal relationships between so-called treatment and outcome variables. This paper provides a recipe for augmenting deep learning methods to accurately characterize such relationships in the presence of instrument variables (IVs)—sources of treatment randomization that are conditionally independent from the outcomes. Our IV specification resolves into two prediction tasks that can be solved with deep neural nets: a first-stage network for treatment prediction and a second-stage network whose loss function involves integration over the conditional treatment distribution. This Deep IV framework allows us to take advantage of the rich empirical data to optimize the prices it charges. In this case, price is the treatment variable.

Deep Generalized Method of Moments for Instrumental Variable Analysis

Andrew Bennett
Cornell University
awb222@cornell.edu

Nathan Kallas
Cornell University
kallas@cornell.edu

Tobias Schnabel
Microsoft Research
tbs4@cornell.edu

Abstract

Instrumental variable analysis is a powerful tool for estimating causal effects when randomization or full control of confounders is not possible. The application of standard methods such as 2SLS, GMM, and more recent variants are significantly impeded when the causal effects are complex, the instruments are high-dimensional, and/or the treatment is high-dimensional. In this paper, we propose the DeepGMM algorithm based on a new variational reformula-
Causal inference with unobserved confounding

Deep IV: A Flexible Approach for Counterfactual Prediction

Kernel Instrumental Variable Regression

Deep Generalized Method of Moments for Instrumental Variable Analysis

Abstract

Instrumental variable (IV) regression is a strategy for learning causal relationships in observational data. If measurements of input $X$ and output $Y$ are confounded, the causal relationship can nonetheless be identified if an instrumental variable $Z$ is available that influences $X$ directly, but is conditionally independent of $Y$ given $X$ and the unmeasured confounder. The classic two-stage least squares algorithm (2SLS) simplifies the estimation problem by modeling all relationships as linear functions. We propose kernel instrumental variable regression (KIV), a nonparametric generalization of 2SLS, modeling relations among $X$, $Y$, and $Z$ as
Causal inference with unobserved confounding

Minimax Estimation of Conditional Moment Models

Nishanth Dikkala  
MIT  
nishanthd@csail.mit.edu

Greg Lewis  
Microsoft Research  
glewis@microsoft.com

Lester Mackey  
Microsoft Research  
lmackey@microsoft.com

Vasilis Syrgkanis  
Microsoft Research  
vasy@microsoft.com

Abstract

We develop an approach for estimating models described via conditional moment restrictions, with a prototypical application being non-parametric instrumental variable estimation. We introduce a minimax criterion function, under which the estimation problem can be thought of as solving a zero-sum game between a minimax minimizer and a maximizer who is optimizing over the hypothesis space of the target model and an adversary who is optimizing over the hypothesis space of the test function space. We analyze the statistical estimation rate of the resulting estimator for arbitrary hypothesis spaces, with respect to an appropriate analogue of the mean squared error metric, as well as linear function spaces.

Deep IV: A Flexible Framework for Inference with Unobserved Confounding

Rahul Kallus  
Cornell University  
kallus@cornell.edu

Tobias Schnabel  
Microsoft Research  
tbs4@cornell.edu

Abstract

Deep IV is a powerful tool for estimating causal effects when unobserved confounders is not possible. The application of deep learning to causal inference is relatively recent and, more recent variants are significantly more complex. In this paper, we propose the DeepGMM method based on a new variational reformula-
Causal inference with unobserved confounding

Learning Deep Features in Instrumental Variable Regression

Liyuan Xu
Gatsby Unit
liyuan.jo.19@ucl.ac.uk

Yutian Chen
DeepMind
yutianc@google.com

Nando de Freitas
DeepMind
nandodefreitas@google.com

Arnaud Doucet
DeepMind
arnauddoucet@google.com

Siddharth Srinivasan
University of Washington
sidsrini@cs.washington.edu

Arthur Gretton
Gatsby Unit
arthur.gretton@gmail.com

November 3, 2020

Abstract

Instrumental variable (IV) regression is a standard strategy for learning causal relationships between confounded treatment and outcome variables from observational data by utilizing an instrumental variable, which affects the outcome from the instrument to the treatment; and stage 2 performs linear regression conditioned on the instrument. We propose a novel method, deep feature instrumental variable regression (DFIV), which utilizes deep neural networks to construct the relationship between the instrumental and the treatment. The algorithm alternates between stage 1 and stage 2, using a neural network trained to represent the nonparametric instrumental variable. This algorithm is shown to be more robust to violations of the assumption of a linear instrumental variable, and performs significantly better than state-of-the-art methods in both simulated and empirical settings.
Instrumental variable assumptions

Gene → BMI → Heart Disease

Lifestyle
Instrumental variable assumptions

Gene → BMI

BMI → Heart Disease

Gene → Relevance

Lifestyle

BMI
Instrumental variable assumptions

Gene → BMI

BMI → Heart Disease

Lifestyle

Gene → Relevance → BMI

Exclusion
Instrumental variable assumptions

- Gene
- BMI
- Heart Disease
- Unconfounded instrument
- Lifestyle
- Relevance
- Exclusion
Example: Mendelian randomization

Gene k

Gene 1
- Valid

Gene 2
- Valid

Gene k
- Invalid

BMI

Lifestyle

Heart Disease

Multiple candidate ‘instruments’, only some valid

See Davey Smith & Hemani 2014 for a review
One way to be unbiased

\[ E[\hat{g}_{z_1}(t)] = g(t) \]

[c.f. Hartwig et al. 2017]
One way to be unbiased

\[ E[\hat{g}_{z_2}(t)] = g(t) \]

[c.f. Hartwig et al. 2017]
One way to be unbiased

\[ E[\hat{g}_{z_3}(t)] \neq g(t) \]

[cf. Hartwig et al. 2017]
One way to be unbiased

\[ E[\hat{g}_{z_4}(t)] = g(t) \]

[c.f. Hartwig et al. 2017]
One way to be unbiased

\[ E[\hat{g}_{z_5}(t)] \neq g(t) \]

[\text{c.f. Hartwig et al. 2017}]
One way to be unbiased

\[ E[\hat{g}_{z_6}(t)] = g(t) \]

[c.f. Hartwig et al. 2017]
One way to be unbiased

\[ E[\hat{g}_{z_7}(t)] = g(t) \]

[c.f. Hartwig et al. 2017]
One way to be unbiased

\[ E[\hat{g}_{z_8}(t)] \neq g(t) \]

[c.f. Hartwig et al. 2017]
One way to be unbiased

\[ E[\hat{g}_{z_9}(t)] \neq g(t) \]

[c.f. Hartwig et al. 2017]
ModelIV algorithm in two steps:

Input: lower bound on the number of valid instruments, $V$

1. If you have $k$ instrumental variables, fit an *ensemble* of $k$ different instances of DeepIV / DeepGMM / Kernel IV / etc.

2. Output the ‘Venter mode’ of the ensemble (mean of $V$ closest estimates).
Model IV

Model IV algorithm in two steps:

Input: lower bound on the number of valid instruments, $V$

1. If you have $k$ instrumental variables, fit an ensemble of $k$ different instances of DeepIV / DeepGMM / Kernel IV / etc.

2. Output the ‘Venter mode’ of the ensemble (mean of $V$ closest estimates).
ModelIV

ModelIV algorithm in two steps:

Input: lower bound on the number of valid instruments, $V$

1. If you have $k$ instrumental variables, fit an ensemble of $k$ different instances of DeepIV / DeepGMM / Kernel IV / etc.

2. Output the ‘Venter mode’ of the ensemble (mean of $V$ closest estimates).
**ModelIV**

**ModelIV** algorithm in two steps:

Input: lower bound on the number of valid instruments, \( V \)

1. If you have \( k \) instrumental variables, fit an **ensemble** of \( k \) different instances of DeepIV / DeepGMM / Kernel IV / etc.

2. Output the ‘Venter mode’ of the ensemble (mean of \( V \) closest estimates).
Model IV
Model IV
Model IV

![Graph showing response to treatment with Model IV]
Model IV
Model IV
Model IV
Theorem: If each estimator is consistent and modal validity hold, ModelIV is a consistent estimator for the true effect $E[y \mid do(p), x]$.
Results summary

ModelIV…

- Performs well in finite sample simulations. Successfully removes most of the bias introduced by invalid instruments.

- Converges at the same rate as the underlying estimators, even on worst case distributions.

Theorem 2. For some test point \((t, x)\), let \(Z = \{\beta_1, \ldots, \beta_k\}\) be \(k\) estimates of the causal effect of \(t\) at \(x\). Assume,

[Bounded estimates] Each estimate is bounded by some constants, \([a_i, b_i]\]

[Convergent estimators] Each estimator converges in mean squared error at a rate \(n^{-r}\) (where \(r = \frac{1}{2}\) if the estimator achieves the parametric rate), and hence each estimator has finite variance, \(\text{Var}(\hat{\beta}_i) = \frac{\sigma^2}{n}\) for some \(\sigma_i\).

Then, if \(\sigma = \max_{i \in Y} \sigma_i\) there exists \(a, C\), such that

\[
E[(\text{ModeIV}(Z) - \beta)^2 - (\frac{1}{2} \sum_{i \in Y} \beta_i - \beta)^2] \leq 9kC\sigma n^{-r}.
\]
• Instrumental variable approaches allow you to estimate causal effects with unobserved confounding.

• ModelV is the first nonparametric procedure that is robust to invalid instruments & is a simple black box procedure.