Nonparametric Jackknife Instrumental Variable Estimation and Confounding Robust Surrogate Indices

Aurélien Bibaut, Nathan Kallus, Apoorva Lal

June 25, 2024

The non-parametric instrumental variable (NPIV) problem is, given N observations, $O_i = (A_i, S_i, Y_i)$ $i = 1, \ldots, N$, on instrumental variables (IVs) A, interventions $S \in \mathbb{R}^p$, and outcomes $Y \in \mathbb{R}$, to find a function $h : \mathbb{R}^p \to \mathbb{R}$ satisfying

$$\mathbb{E}[Y - h(S) \mid A] = 0. \tag{1}$$

This can be motivated by a structural (*i.e.*, causal) model $Y = h^\star(S) + \epsilon$, where S and ϵ can be endogenous due, *e.g.*, to the presence of common confounders (so that $h^\star(S) \neq \mathbb{E}[Y \mid S]$), but A and ϵ are exogenous so that Eq. (1) holds for $h^\star$. A variety of work studies the estimation of $h^\star$ and inference on functionals thereof under nonparametric restrictions on $h^\star$ as we receive additional observations N from a fixed $O = (A, S, Y)$ distribution [Ai and Chen, 2003, 2007, 2012, Bennett and Kallus, 2023, Bennett et al., 2019, 2023a,b,b, Chen and Pouzo, 2009, 2012, Darolles et al., 2011, Dikkala et al., 2020, Hartford et al., 2017, Kremer et al., 2022, 2023, Newey and Powell, 2003, Santos, 2011, Severini and Tripathi, 2006, 2012, Singh et al., 2019, Zhang et al., 2023].

In this paper, we tackle NPIV in the challenging many-weak-IV setting, where $A \in \{1, \ldots, K\}$ is discrete and we only see so many of each value, namely for each $a \in [K]$, we have n i.i.d. observations of $(Y, S) \mid A = a$, forming $N = nK$ observations in total. Then, we can have only $n \ll N$ observations for each value $A = a$, which is challenging as it limits our ability to consistently estimate the moment $\mathbb{E}[Y - h(S) \mid A = a]$ and thus to empirically check the fitness of a candidate h function, which is the general approach of many NPIV estimators above.

In the linear setting where $h^\star(S) = S^\top \beta^\star$, Eq. (1) reduces to solving $\mathbb{E}[Y \mid A] = \mathbb{E}[S \mid A]^\top \beta$ for β. This motivates the two-stage least squares (2SLS) approach of estimating β by ordinary least squares (OLS) of Y on the "first-stage" OLS prediction of S given A (for discrete A this is simply the sample means of S for each A value). However, when $n \ll N$, even as $N \to \infty$ this can incur non-vanishing bias because the first-stage regression may not converge at all [Angrist et al., 1999, Bibaut et al., 2024, Peysakhovich and Eckles, 2018]. JIVE [Angrist et al., 1999] addresses this by regressing Y on a prediction of S given A based on OLS using all the data *except* the datapoint on which we make the prediction. This renders the errors from the first stage uncorrelated with the second stage so they average out to zero so that we regain consistency [Chao et al., 2012].

The NPIV analog, which we tackle, is, however, unresolved. It is also rather nuanced because when S is continuous but A is discrete, a nonparametric $h^\star$ is generally not uniquely identified by Eq. (1), which involves just K moments but a general function h. Nonetheless, certain *linear functionals* of $h^\star$, meaning $\theta_0 = \mathbb{E}[h^\star(S)\alpha(S)]$ for some α, may still be uniquely identified, meaning $\theta_0 = \mathbb{E}[h(S)\alpha(S)]$ for *any* h satisfying Eq. (1).

This problem setting is of particular interest in digital experimentation, where the rapid pace of innovation means we have many (K) historical randomized experiments (with serial numbers A), which can be used to instrument for the effect ($h^\star$) of short-term surrogate observations (S) on long-term outcomes (Y) even in the presence of unobserved confounding between the two, but where each experiment has a certain sample size (n). If we know this effect, we can construct a surrogate index $h^\star(S)$ such that average treatment effects (ATEs) on Y are the same as those on $h^\star(S)$. Moreover, the ATEs on $h^\star(S)$ is a linear functional thereof. Then, for novel experiments predicting long-term ATEs before observing long-term outcomes can be phrased as inference on a linear functional of a solution to Eq. (1).

In this paper we also tackle the question how to reliably do this inference in the presence of underidentified and nonparametric $h^\star$, which is another significant challenge, besides solving Eq. (1) in the many-weak-IV setting. We furthermore extend the simple instrumentation identification to account for the possibility that short-term surrogate observations do not fully mediate the treatment effects on long-term outcomes (that is, there is exclusion violation).

In this paper, we develop both a novel estimator for $h^\star$ in the nonparametric many-weak-IV setting and methods for debiased inference on surrogate-predicted ATEs.

Taken together, our methods and results provide new ways to conduct long-term causal inference in challenging, but practically very relevant, settings.

References

C. Ai and X. Chen. Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica*, 71(6):1795–1843, 2003.

C. Ai and X. Chen. Estimation of possibly misspecified semiparametric conditional moment restriction models with different conditioning variables. *Journal of Econometrics*, 141(1): 5–43, 2007.

C. Ai and X. Chen. The semiparametric efficiency bound for models of sequential moment restrictions containing unknown functions. *Journal of Econometrics*, 170(2):442–457, 2012.

J. D. Angrist, G. W. Imbens, and A. B. Krueger. Jackknife instrumental variables estimation. *Journal of Applied Econometrics*, 14(1):57–67, 1999.

A. Bennett and N. Kallus. The variational method of moments. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 85(3):810–841, 2023.

A. Bennett, N. Kallus, and T. Schnabel. Deep generalized method of moments for instrumental variable analysis. *Advances in neural information processing systems*, 32, 2019.

A. Bennett, N. Kallus, X. Mao, W. Newey, V. Syrgkanis, and M. Uehara. Inference on strongly identified functionals of weakly identified functions, 2023a.

A. Bennett, N. Kallus, X. Mao, W. Newey, V. Syrgkanis, and M. Uehara. Source condition double robust inference on functionals of inverse problems. *arXiv preprint arXiv:2307.13793*, 2023b.

A. Bibaut, W. Chou, S. Ejdemyr, and N. Kallus. Learning the covariance of treatment effects across many weak experiments. *arXiv preprint arXiv:2402.17637*, 2024.

J. C. Chao, N. R. Swanson, J. A. Hausman, W. K. Newey, and T. Woutersen. Asymptotic distribution of jive in a heteroskedastic iv regression with many instruments. *Econometric Theory*, 28(1):42–86, 2012.

X. Chen and D. Pouzo. Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals. *Journal of Econometrics*, 152(1):46–60, 2009.

X. Chen and D. Pouzo. Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals. *Econometrica*, 80(1):277–321, 2012.

S. Darolles, Y. Fan, J.-P. Florens, and E. Renault. Nonparametric instrumental regression. *Econometrica*, 79(5):1541–1565, 2011.

N. Dikkala, G. Lewis, L. Mackey, and V. Syrgkanis. Minimax estimation of conditional moment models. *Advances in Neural Information Processing Systems*, 33:12248–12262, 2020.

J. Hartford, G. Lewis, K. Leyton-Brown, and M. Taddy. Deep iv: A flexible approach for counterfactual prediction. In *International Conference on Machine Learning*, pages 1414–1423. PMLR, 2017.

H. Kremer, J.-J. Zhu, K. Muandet, and B. Schölkopf. Functional generalized empirical likelihood estimation for conditional moment restrictions. In *International Conference on Machine Learning*, pages 11665–11682. PMLR, 2022.

H. Kremer, Y. Nemmour, B. Schölkopf, and J.-J. Zhu. Estimation beyond data reweighting: Kernel method of moments. In *International Conference on Machine Learning*, pages 17745–17783. PMLR, 2023.

W. K. Newey and J. L. Powell. Instrumental variable estimation of nonparametric models. *Econometrica*, 71(5):1565–1578, 2003.

A. Peysakhovich and D. Eckles. Learning causal effects from many randomized experiments using regularized instrumental variables. In *Proceedings of the 2018 World Wide Web Conference*, pages 699–707, 2018.

A. Santos. Instrumental variable methods for recovering continuous linear functionals. *Journal of Econometrics*, 161(2):129–146, 2011.

T. A. Severini and G. Tripathi. Some identification issues in nonparametric linear models with endogenous regressors. *Econometric Theory*, 22(2):258–278, 2006.

T. A. Severini and G. Tripathi. Efficiency bounds for estimating linear functionals of nonparametric regression models with endogenous regressors. *Journal of Econometrics*, 170(2):491–498, 2012.

R. Singh, M. Sahani, and A. Gretton. Kernel instrumental variable regression. *Advances in Neural Information Processing Systems*, 32, 2019.

R. Zhang, M. Imaizumi, B. Schölkopf, and K. Muandet. Instrumental variable regression via kernel maximum moment loss. *Journal of Causal Inference*, 11(1):20220073, 2023.