CONTINUATION OF AUTHOR'S RESPONSE TO REVIEWER MD5A

6. Relevance condition: We have stated the assumptions in the main paper as in [5], Chapter 16 and [4]. We state the technical version of the relevance condition as a lower bound on the minimum eigenvalue of the covariance matrix between Z and X in our Appendix B.1, which is the finite time version of the definition of relevance stated in econometrics books, such as (Greene, 2012). We refer to Eq. (12) in page 17 of our paper for details. Now, with the extra pages of the camera ready version, we add this formal version in the main paper following Assumption 2.1.

7. Interpretation of regression coefficients: We tried to provide an interpretation of β in a linear equation when we stated: "the parameter vector β quantifies the causal effect on y_i due to a unit change in a component of x_i, while retaining other causes of y_i constant." (first column, page 4). Until that point, we have not discussed the endogeneous setting or 2SLS estimates.

8. Discussion on the conventional assumptions for instrumental variables: Our intention was to simplify the exposition in the main paper without technically stating all the assumptions, which we state technical versions for the interested reader in the Appendix. We now add the precise statements in the main body and also add the technical assumptions of relevance in Appendix B.1, Equation (12) to the main. Eq. (12) lower bounds the minimum amount of correlation between Z and X necessary to carry out the finite time analysis. Our Instrumental Variables assumptions follow more closely the exposition made in [4], P. 248, and [5], Chapter 16, which use the nomenclature of the CS literature and are stronger with respect to what is usually assumed in econometrics literature (Greene, 2012) because they want to to be "structural", i.e. model agnostic. For a comparison between the two literatures and formalisms, we refer to [6]. On the oter side, these definitions allow for nonparametric extensions of our present work (e.g. kerenls and deep networks). In fact, following the previous references from the CS literature, we introduce IV's through structural assumptions (which are generalized for stuctural models through the use of causal diagrams) and therefore without the need to specify a model (e.g. a linear one like ours). In essence, we adopt Pearl's definitions: 1. The equations of interest are "structural", not "regression". 2. The error term η stands for all the factors that affect Y when X is held constant. 3. The instrument Z should be independent of η. 4. The instrument Z should not affect Y when X is held constant (exclusion restriction). 5. The instrument Z should not be independent of X (relevance).

(The exclusion restriction, condition 4, is actually redundant. It follows from conditions 2 and 3.) These conditions do not rely on the specific functional form of the equations (e.g. where η can be non-additive). They are also applicable to a system of multiple equations in which X (and other factors) affect Y through several intermediate variables. Furthermore, when the form of the structural equations is unknown, an instrumental variable Z can still be defined through the equations $X = g(Z, \epsilon)$ and $Y = f(X, \eta)$ where f and g are two arbitrary functions through the previous properties [4]. In any case, our analysis only uses $\mathbb{E}[\eta|\mathbf{X}] \neq 0$ to define endogeneity, which we specify as early as page 2, column 1. To clarify this confusion, we now add a footnote on page 1: "In literature, there are two conventions about exogeneity. In the first case, the exogeneous setting is defined by the independence of the noise η and the covariates $\mathbf{X}$ [5]. In another convention, exogeneity is defined by $\mathbb{E}[\eta|\mathbf{X}] = 0$ (Greene, 2012). The second one is a slightly milder condition than the first one. Hereafter, in this paper, we use $\mathbb{E}[\eta|\mathbf{X}] = 0$ as the requirement for exogeneity."

9. Optimism and the exploration-exploitation dilemma: By "optimistically", it is common practice to refer to the optimism-in-face-of-uncertainty principle which is arguably the most used paradigm to design of algorithms dealing with exploration-exploitation dilemma. In the simple bandit strategy employed in UCB, "optimism" comes in the form of an upper confidence bound (hence the acronym UCB). In linear bandit case (as here), this means to choose an estimated β_t inside the $1 - \delta$ confidence ellipsoid

10. O2SLS and original contribution: In the paper in Sec. 3, page 5, second column, we write, "In ML, Venkatraman et al. [2016] applied O2SLS for online linear system identification." and at page 3 we write "Previously, Venkatraman et al. [2016] studied O2SLS for system identification but provided only asymptotic analysis.", furthermore at page 2, in "**Our Contributions.**" paragraph we state in point "1. A non-asymptotic analysis of O2SLS." that this is one of the original contributions, therefore differing from the cited paper. We said that the idea of binging 2SLS to an online learning setting was already established in the cited paper, but the algorithm's implementation differs, it is studied in a very different setting (adversarial), it has different regret guarantees (asymptotic, therefore much weaker), and finally, it is therefore compared experimentally with different non-stochastic algorithms like Follow-The-Regularized-Leader and Online Newton Step.

More in detail, it differs for example, in the following ways: 1. The analysis in [7] is done for the nonstochastic/adversarial case, which assumes boundedness (see discussion on this in point 4. in this rebuttal) and does not handle the stochastic unbounded outcomes and noise. These two settings are very different: the first is the worst-case regret (assuming boundedness), and the second studies the regret for typical behaviour assuming a general underlying stochastic process for data generation (see, for example, [1] for a book introduction to the two settings). 2. We add an L2 regularization term, and therefore, the two algorithms differ. We do this for experimental stability and to be able to analyse the result bounding the regret with finite time guarantees. 3. The results in [7] are of the form $\frac{1}{T}\mathrm{Regret}_T \to 0$ or $\mathbb{E}[\|\beta_t - \beta\|] \to 0$ which are clearly asymptotic and would be subsumed by our analysis if it is extended to stochastic unbounded noise.

Presently, we already have some lines of comparison with [7], but we will add these other points with the extra space in the updated paper.

We hope that we have answered your comments. We look forward to respond further if you have other queries.

Bibliography:

[1] Orabona, Francesco. "A modern introduction to online learning." (2019).

[2] Peixoto, Caio, Yuri Saporito, and Yuri Fonseca. "Nonparametric Instrumental Variable Regression through Stochastic Approximate Gradients." (2024).

[3] Hoi, Steven CH, et al. "Online learning: A comprehensive survey."(2021)

[4] J Pearl. "Causality: Models, reasoning and inference." 2000.

[5] MA Hernan and J Robins. "Causal Inference: What if." 2024.

[6] Imbens, Guido W. "Potential outcome and directed acyclic graph approaches to causality: Relevance for empirical practice in economics." (2020)

[7] Arun Venkatraman, Wen Sun, Martial Hebert, J Bagnell, and Byron Boots. "Online instrumental variable regression with applications to online linear system identification." 2016