Online Instrumental Variable Regression: Regret Analysis and Bandit Feedback

Riccardo Della Vecchia, Debabrota Basu

Online Instrumental Variable Regression: Regret Analysis and Bandit Feedback

Riccardo Della Vecchia
Équipe Scool, Univ. Lille, Inria, CNRS, Centrale Lille, UMR 9189- CRIStAL, F-59000 Lille, France

Debabrota Basu
Équipe Scool, Univ. Lille, Inria, CNRS, Centrale Lille, UMR 9189- CRIStAL, F-59000 Lille, France

Abstract

The independence of noise and covariates is a standard assumption in online linear regression with unbounded noise and linear bandit literature. This assumption and the following analysis are invalid in the case of endogeneity, i.e., when the noise and covariates are correlated. In this paper, we study the *online setting of Instrumental Variable (IV) regression*, which is widely used in economics to identify the underlying model from an endogenous dataset. Specifically, we upper bound the identification and oracle regrets of the popular Two-Stage Least Squares (2SLS) approach to IV regression but in the online setting. Our analysis shows that Online 2SLS (O2SLS) achieves $\mathcal{O}(d^2 \log^2 T)$ identification and $\mathcal{O}(\gamma\sqrt{dT \log T})$ oracle regret after T interactions, where d is the dimension of covariates and γ is the bias due to endogeneity. Then, we leverage O2SLS as an oracle to design OFUL-IV, a linear bandit algorithm. OFUL-IV can tackle endogeneity and achieves $\mathcal{O}(d\sqrt{T} \log T)$ regret. For datasets with endogeneity, we experimentally show the efficiency of OFUL-IV in terms of estimation error and regret.

Contents

1. Introduction

Online regression is a founding component of online learning (Kivinen et al., 2004), sequential testing (Kazerouni and Wein, 2021), contextual bandits (Foster and Rakhlin, 2020), and reinforcement learning (Ouhamma et al., 2022). Specially, online linear regression is widely used and analysed to design efficient algorithms with theoretical guarantees (Greene, 2003; Abbasi-Yadkori et al., 2011b; Hazan and Koren, 2012). In linear regression, the *outcome* (or output variable) $Y \in \mathbb{R}$, and the *input features* (or covariates, or treatments) $\mathbf{X} \in \mathbb{R}^d$ are related by a structural equation:

$$Y = \boldsymbol{\beta}^T \mathbf{X} + \eta,$$

where $\boldsymbol{\beta}$ is the *true parameter* and η is the observational noise with variance σ^2. *The goal is to estimate $\boldsymbol{\beta}$ from an observational dataset.* Two common assumptions in the analysis of linear regression are (i) bounded observations and covariates (Vovk, 1997; Bartlett et al., 2015; Gaillard et al., 2019), and (ii) *exogeneity*, i.e. independence of the noise η and the input features X ($\mathbb{E}[\eta|X] = 0$) (Abbasi-Yadkori et al., 2011b; Ouhamma et al., 2021). Under exogeneity, researchers have studied scenarios, where the observational noise is unbounded and has only bounded variance σ^2. Ouhamma et al. (2021) show that the unbounded stochastic setting asks for different technical analysis than the bounded adversarial setting popular in online regression literature. Additionally, in real-life, exogeneity is often violated, and we encounter *endogeneity*, i.e. dependence between noise and covariates ($\mathbb{E}[\eta|X] \neq 0$) (Greene, 2003; Angrist et al., 1996). Endogeneity arises due to omitted explanatory variables, measurement errors, dependence of the output and the covariates on unobserved confounding variables etc. (Wald, 1940; Mogstad et al., 2021; Zhu et al., 2022). In this paper, *we analyse online linear regression that aims to estimate $\boldsymbol{\beta}$ accurately from endogeneous observational data, where the noise is stochastic and unbounded.*

Instrumental Variable Regression. Historically, *Instrumental Variables* (IVs) are introduced to identify and quantify the causal effects of endogenous covariates (Newey and Powell, 2003). IVs are widely used in economics (Wright, 1928; Mogstad et al., 2021), causal inference (Rubin, 1974; Hernan and Robins, 2020; Harris et al., 2022), bio-statistics and epidemiology (Burgess et al., 2017).

Example 1.1. *Carneiro et al. (2011); Mogstad et al. (2021) aim to estimate the number of returning students to college using the National Longitudinal Survey of Youth data. The return depends on multiple covariates X, such as whether the individual attended college, her AFQT scores, her family*

income, her family conditions (mother's years of education, number of siblings etc.). Often the family conditions have unobserved confounding effects on the college attendance and scores. This endogenous nature of data leads to bias in traditional linear regression estimates, such as Ordinary Least Squares (OLS). To mitigate this issue, Carneiro et al. (2011); Mogstad et al. (2021) leverage two IVs (Z): average log income in the youth's county of residence at age 17, and the presence of a four-year college in the youth's county of residence at age 14[1]*. The logic is that a youth might find going to college more attractive when labour market opportunities are weaker and a college is nearby. Using these two IVs, the youth's attendance to college is estimated. Then, in the next stage this estimate of college attendance is used with family conditions to predict the return of the youth to college. This two stage regression approach with IVs produces a more accurate estimate of youths' return to the college than OLS models assuming exogeneity.*

This approach to conduct two stages of linear regression using instrumental variables is called *Two Stage Least Squares Regression* (2SLS) (Angrist and Imbens, 1995; Angrist et al., 1996). 2SLS has become the standard tool in economics, social sciences, and statistics to study the effect of treatments on outcomes involving endogeneity. Recently, in machine learning, researchers have extended traditional 2SLS techniques to nonlinear structures, non-compliant instruments, and corrupted observations using deep learning (Liu et al., 2020; Xu et al., 2020, 2021), graphical models (Stirn and Jebara, 2018), and kernel regression (Zhu et al., 2022), respectively. The existing analysis of 2SLS are asymptotic, i.e. what can be learned if we have access to infinite number of samples in an offline setting (Singh et al., 2020; Liu et al., 2020; Nareklishvili et al., 2022). In applications, this analysis is vacuous as one has access to only finite samples. Additionally, in practice, it is natural to acquire the data sequentially as treatments are chosen on-the-go, and then to learn the structural equation from the sequential data (Venkatraman et al., 2016). This setting motivates us to analyse the online extension of 2SLS, referred as O2SLS.

Additionally, in an interactive setting, if a policy maker aims to build more schools at some of the lower income areas as a form of intervention, she observes only the changes corresponding to it. This is referred as bandit feedback in online learning literature and studied under the linear bandit formulation (Abbasi-Yadkori et al., 2011a). This motivates us to further extend O2SLS to *linear bandits, where bandit feedback and endogeneity occur simultaneously.*

In this paper, we investigate these two questions:

1. *What is the upper bound on the loss in performance for deploying parameters estimated by* O2SLS *instead of the true parameters $\boldsymbol{\beta}$? How estimating the true parameters $\boldsymbol{\beta}$ influence different performance metrics under endogeneity?*

2. *Can we design efficient algorithms for linear bandits with endogeneity by using* O2SLS*?*

Our Contributions. Our investigation has led to

1. *A Non-asymptotic Analysis of* O2SLS*:* First, we identify three notions of regret: *identification regret*, *oracle regret*, and *population regret*. Though all of them are of same order under endeogeneity, we show that the relations are more nuanced under endogeneity and unbounded noise. We focus specifically on the identification regret, i.e. the sum of differences between the estimated parameters $\{\boldsymbol{\beta}_t\}_{t=1}^T$, and the true parameter $\boldsymbol{\beta}$, and *oracle regret*, i.e. the sum of differences between the losses incurred by the estimated parameters $\{\boldsymbol{\beta}_t\}_{t=1}^T$, and the true parameter $\boldsymbol{\beta}$. In Section 4, we theoretically show that O2SLS achieve $\mathcal{O}(d^2 \log^2 T)$ identification regret and $\mathcal{O}(d^2 \log^2 T + \gamma\sqrt{dT \log T})$ oracle regret after receiving T samples from the observational data. Identification regret of O2SLS is $d \log T$ multiplicative factor higher than regret of online linear regression under exogeneity, and oracle regret is $\mathcal{O}(\gamma\sqrt{dT \log T})$ additive factor higher. These are the costs that O2SLS pay for tackling endogeneity in two stages. In our knowledge, *we are the first to propose a non-asymptotic regret analysis of* O2SLS *with stochastic and unbounded noise.*

1. One can argue whether these are either sufficient or weak IVs. For simplicity, we assume sufficiency here, i.e. the IVs can decouple the unobserved confounding.

2. OFUL-IV *for Linear Bandits with Endogeneity:* In Section 5, we study the linear bandit problem with endogeneity. *We design an extension of OFUL algorithm used for linear bandit with exogeneity, namely* OFUL-IV, to tackle this problem. OFUL-IV uses O2SLS to estimate the parameters, and corresponding confidence bounds on $\boldsymbol{\beta}$ to balance exploration–exploitation. We show that OFUL-IV achieve $\mathcal{O}(d\sqrt{T}\log T)$ regret after T interactions. We experimentally show that OFUL-IV incur lower regret than OFUL under endogeneity (end of Section 5).

2. Related Work

Online Regression without Endogeneity. Our analysis of O2SLS extends the tools and techniques of online linear regression without endogeneity. Analysis of online linear regression began with (Foster, 1991; Littlestone et al., 1991). Vovk (1997, 2001) show that forward and ridge regressions achieve $\mathcal{O}(dY_{\max}^2 \log T)$ for outcomes with bound $Y_{\max}$. Bartlett et al. (2015) generalise the analysis further by considering the features known in hindsight. Gaillard et al. (2019) improve the analysis further to propose an optimal algorithm and a lower bound. *These works perform an adversarial analysis with bounded outcomes, covariates, and observational noise, while we focus on the stochastic setting.* Ouhamma et al. (2021) study the stochastic setting with bounded input features and unbounded noise. But they need to assume independence of noise and input features. *In this paper, we analyse online 2*SLS *under endogeneity and unbounded (stochastic) noise.* We do not assume to know the bound on the outcome and derive high probability bounds for any bounded sequence of features.

Linear Bandits without Endogeneity. Linear bandits generalise the setting of online linear regression under bandit feedback (Abbasi-Yadkori et al., 2011a, 2012; Foster and Rakhlin, 2020). To be specific, in bandit feedback, the algorithm observes only the outcomes for the input features that it has chosen to draw during an interaction. Popular algorithm design techniques, such as optimism-in-the-face-of-uncertainty and Thompson sampling, are extended to propose OFUL (Abbasi-Yadkori et al., 2012) and LinTS (Abeille and Lazaric, 2017), respectively. OFUL and LinTS algorithms demonstrate $\mathcal{O}(d\sqrt{T}\log T)$ and $\mathcal{O}(d^{1.5}\sqrt{T}\log T)$ regret guarantees under exogeneity assumption. *Here, we use* O2SLS *as a regression oracle to develop* OFUL-IV *for linear bandits with endogeneity. We prove that* OFUL-IV *achieves* $\mathcal{O}(d\sqrt{T}\log T)$ *regret.*

Instrument-armed Bandits. Kallus (2018) is the first to study endogeneity, and instrumental variables in stochastic bandit setting. Stirn and Jebara (2018) propose a Thompson sampling-type algorithm for stochastic bandits, where endogeneity arises due to non-compliant actions. But both Kallus (2018) and Stirn and Jebara (2018) study only the finite-armed bandit setting where arms are independent of each other. In this paper, *we study the stochastic linear bandit setting with endogeneity, which requires different techniques for analysis and algorithm design.*

3. Preliminaries: Instrumental Variables & Offline Two-stage Least Squares (2SLS)

We are given an observational dataset $\{\boldsymbol{x}_i, y_i\}_{i=1}^n$ consisting of n pairs of input features and outcomes, such that $y_i \in \mathbb{R}$ and $\boldsymbol{x}_i \in \mathbb{R}^d$.[2] These inputs and outcomes are stochastically generated using a linear model

$$y_i = \boldsymbol{\beta}^\top \boldsymbol{x}_i + \eta_i, \qquad \text{(Second stage)}$$

where $\boldsymbol{\beta} \in \mathbb{R}^d$ is the *unknown true parameter vector* of the linear model, and $\eta_i \sim \mathcal{N}(0, \sigma_\eta^2)$ is the unobserved error term representing all causes of y_i other than $\boldsymbol{x}_i$. It is assumed that the error terms η_i are independently and identically distributed, and have bounded variance σ^2. The parameter

2. Matrices and vectors are represented with bold capital and bold small letters, e.g. $\mathbf{A}$ and $\mathbf{a}$, respectively.

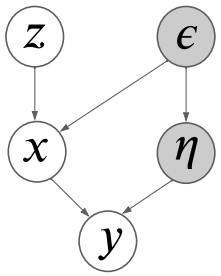


Figure 1: The DAG for 2SLS. The unobserved noises are $\boldsymbol{\epsilon}$ and η (in grey), while $\boldsymbol{z}, \boldsymbol{x}, y$ are observed quantities.

vector $\boldsymbol{\beta}$ quantifies the causal effect on y_i due to a unit change in a component of $\boldsymbol{x}_i$, while retaining other causes of y_i constant. The goal of linear regression is to estimate $\boldsymbol{\beta}$ by *minimising the square loss over the dataset* (Brier, 1950), i.e. $\hat{\boldsymbol{\beta}} \triangleq \operatorname{argmin}_{\boldsymbol{\beta}'} \sum_{i=1}(y_i - \boldsymbol{\beta}'^{\top} \boldsymbol{x}_i)^2$.

The obtained solution is called the Ordinary Least Square (OLS) estimate of $\boldsymbol{\beta}$ (Wasserman, 2004), and used as a corner stone of online regression (Gaillard et al., 2019) and linear bandit algorithms (Foster and Rakhlin, 2020). Specifically, if the input feature matrix $\mathbf{X}_n \in \mathbb{R}^{n \times d}$ is defined as $[\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n]^{\top}$, the outcome vector is $\boldsymbol{y}_n \triangleq [y_1, \ldots, y_n]^{\top}$, and the noise vector is $\boldsymbol{\eta}_n \triangleq [\eta_1, \ldots, \eta_n]^{\top}$, the OLS estimator is expressed as

$$\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}} \triangleq (\mathbf{X}_n^{\top}\mathbf{X}_n)^{-1}\mathbf{X}_n^{\top}\boldsymbol{y}_n = \boldsymbol{\beta} + (\mathbf{X}_n^{\top}\mathbf{X}_n)^{-1}\mathbf{X}_n^{\top}\boldsymbol{\eta}_n$$

If $\mathbf{X}_n$ *and* $\boldsymbol{\eta}_n$ *are independent*, the second term has zero expected value conditioned on $\mathbf{X}_n$. Hence, the OLS estimator is asymptotically unbiased, i.e. $\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}} \to \infty$ as $n \to \infty$.

In practice, the input features $\boldsymbol{x}$ and the noise $\boldsymbol{\eta}$ are often correlated (Greene, 2003, Chapter 8). As in Figure 1, this dependence, called endogeneity, is modelled with *a confounding unobserved random variable* $\boldsymbol{\epsilon}$. To compute an unbiased estimate of $\boldsymbol{\beta}$ under endogeneity, a popular technique is to introduce the Instrumental Variables (IVs) $\boldsymbol{z}$ (Angrist et al., 1996; Newey and Powell, 2003). IVs are chosen such that they are highly correlated with endogenous components of $\boldsymbol{x}$ (relevance condition) but are independent of the noise η (exogeneity condition for $\boldsymbol{z}$).

This leads to the Two-stage Least Squares (2SLS) approach to IV regression (Angrist and Imbens, 1995; Angrist et al., 1996). Here, we further assume that IVs, i.e. $\mathbf{Z}_n \triangleq [\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n]^{\top}$, cause linear effects on the endogenous covariates. Specifically, for the just-identified IVs,

$$\mathbf{X}_n = \mathbf{Z}_n\boldsymbol{\Theta} + \mathbf{E}_n, \qquad \text{(First stage)}$$

where $\boldsymbol{\Theta} \in \mathbb{R}^{d \times d}$ is an unknown first-stage parameter matrix and $\mathbf{E}_n \triangleq [\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n]^{\top}$ is the unobserved noise matrix leading to confounding in the second stage. This is a "classic" *multiple regression*, where the covariates $\boldsymbol{z}$ are independent of the noise terms $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma_{\boldsymbol{\epsilon}}^2 \mathbb{I}_d)$ (Wasserman, 2004, Ch. 13). Thus, the first-stage is amenable to OLS regression. This formulation leads us to the 2SLS estimator:

$$\widehat{\boldsymbol{\beta}}_{\mathrm{2SLS}} = \left(\mathbf{Z}_n^{\top}\mathbf{X}_n\right)^{-1}\mathbf{Z}_n^{\top}\boldsymbol{y}_n. \qquad \text{(2SLS)}$$

As long as $\mathbb{E}[\boldsymbol{z}_i\eta_i] = 0$ in the true model, we observe that

$$\widehat{\boldsymbol{\beta}}_{\mathrm{2SLS}} = \left(\mathbf{Z}_n^{\top}\mathbf{X}_n\right)^{-1}\mathbf{Z}_n^{\top}\mathbf{X}_n\boldsymbol{\beta} + \left(\mathbf{Z}_n^{\top}\mathbf{X}_n\right)^{-1}\mathbf{Z}_n^{\top}\boldsymbol{\eta}_n \xrightarrow{p} \boldsymbol{\beta},$$

as $n \to \infty$. This works because IV solves for the unique parameter that satisfies $\frac{1}{n}\mathbf{Z}_n^{\top}\eta \xrightarrow{p} 0$. Since $\boldsymbol{x}$ and $\boldsymbol{\eta}$ are correlated, 2SLS estimator is not unbiased in finite-time.

Assumption 3.1. *The assumptions for conducting 2*SLS *with just-identified IVs are (Greene, 2003):*

1. ***Well behaved data.*** *For every $n \in \mathbb{N}$, the matrices $\mathbf{Z}_n^\top \mathbf{Z}_n$ and $\mathbf{Z}_n^\top \mathbf{X}_n$ are full rank, and thus invertible.*
2. ***Endogeneity of $\boldsymbol{x}$.*** *The second stage input features $\boldsymbol{x}$ and noise η are not independent: $\boldsymbol{x} \not\perp\!\!\!\perp \eta$.*
3. ***Exogeneity of $\boldsymbol{z}$.*** *The IV random variables are independent of the noise in the second stage: $\boldsymbol{z} \perp\!\!\!\perp \eta$.*
4. ***Relevance Condition.*** *The variables $\boldsymbol{z}$ and $\boldsymbol{x}$ are correlated: $\boldsymbol{z} \not\perp\!\!\!\perp \boldsymbol{x}$. This implies that there exists $\mathfrak{r} > 0$:*
$$\left\| n \left(\mathbf{Z}_n^\top \mathbf{X}_n \right)^{-1} \right\|_2 \leq \frac{1}{\mathfrak{r}}. \tag{1}$$

4. Online Two-Stage Least Squares Regression

In this section, we describe the problem setting and schematic of *Online Two-Stage Least Squares Regression*, in brief O2SLS. Following that, we first define two notions of regret: identification and oracle, aimed at estimating the true parameter and producing accurate predictions. We provide a theoretical analysis of O2SLS and upper bound the two types of regret (Section 4.2).

O2SLS. In the online setting of IV regression, the data $(\boldsymbol{x}_1, \boldsymbol{z}_1, y_1), \ldots, (\boldsymbol{x}_T, \boldsymbol{z}_T, y_T), \ldots$ arrives in a stream. Following the 2SLS model (Figure 1), data is generated from

$$\begin{cases} \boldsymbol{x}_t = \boldsymbol{\Theta}^\top \boldsymbol{z}_t + \boldsymbol{\epsilon}_t \\ y_t = \boldsymbol{\beta}^\top \boldsymbol{x}_t + \eta_t, \end{cases} \tag{2}$$

such that $\boldsymbol{x}_t \not\perp\!\!\!\perp \eta_t$ and $\boldsymbol{z}_t \perp\!\!\!\perp \eta_t$ for all $t \in \mathbb{N}$. At each step t, the online IV regression algorithm is served with a new input feature $\boldsymbol{x}_t$ and an IV $\boldsymbol{z}_t$, and it aims to predict an outcome $\hat{y}_t \triangleq \boldsymbol{\beta}_t^\top \boldsymbol{x}_t \in \mathbb{R}$. Here, $\boldsymbol{\beta}_t$ is the estimate of the parameter at step t computed using current $(\boldsymbol{x}_t, \boldsymbol{z}_t)$ and the data $\{(\boldsymbol{x}_i, \boldsymbol{z}_i, y_i)\}_{i=1}$ observed so far. Following the prediction, Nature reveals the true outcome y_t. Quality of the prediction is evaluated using a square loss $\ell_t(\boldsymbol{\beta}_t) \triangleq (\widehat{y}_t - y_t)^2$ (Foster, 1991). The online protocol is the following.

At each round $t = 1, 2, \ldots, T$
1. $\boldsymbol{z}_t$ is sampled i.i.d. from an unknown distribution
2. $\boldsymbol{x}_t$ is sampled according to Equation (2) given $\boldsymbol{z}_t$
3. we compute an estimate $\boldsymbol{\beta}_\bullet$ and make a prediction $\widehat{y}_t = \boldsymbol{\beta}_\bullet^\top \boldsymbol{x}_t$
4. we observe the true y_t following Equation (2)
5. we incur in a loss $(y_t - \widehat{y}_t)^2 = (y_t - \boldsymbol{\beta}_\bullet^\top \boldsymbol{x}_t)^2$

In order to address this problem, we propose an online form of the 2SLS estimator. Thus, modifying Equation (2SLS), we obtain the O2SLS estimator that is computed for the prediction at time t, using information up to time $t-1$:

$$\boldsymbol{\beta}_{t-1} \triangleq \left(\sum_{s=1}^{t-1} \boldsymbol{z}_s \boldsymbol{x}_s^\top \right)^{-1} \sum_{s=1}^{t-1} \boldsymbol{z}_s y_s \tag{O2SLS}$$

We use the O2SLS estimator at step $t-1$ for the prediction $\widehat{y}_t = \boldsymbol{\beta}_{t-1}^\top \boldsymbol{x}_t$. We elaborate O2SLS in Algorithm 1.

Remark 4.1. *We could use $\boldsymbol{x}_t$ and $\boldsymbol{z}_t$ that we observe before committing to the estimate $\boldsymbol{\beta}_t$, and use it to predict $\widehat{y}_t$ (Vovk, 2001). Since we cannot use y_t for this estimate, we have to modify*

Algorithm 1 O2SLS

for $t = 1, 2, \ldots, T$ **do**
  Observe $\boldsymbol{z}_t$, $\boldsymbol{x}_t$
  Compute $\boldsymbol{\beta}_{t-1}$ according to Equation (O2SLS)
  Predict $\widehat{y}_t = \boldsymbol{\beta}_{t-1}^\top \boldsymbol{x}_t$
  Observe y_t and compute loss $\ell_t\left(\boldsymbol{\beta}_{t-1}\right)$
end for

2SLS to incorporate this additional knowledge. We skip this modification and use $\boldsymbol{\beta}_{t-1}$ to predict. Previously, Venkatraman et al. (2016) studied O2SLS for system identification but provided only an asymptotic analysis.

4.1 Defining Regrets: Identification and Oracle

To analyse the online regression algorithms, it is essential to define proper performance metrics, specifically regrets. Typically, regret quantifies what an online (or sequential) algorithm cannot achieve as it does not have access to the whole dataset rather observes it step by step. Here, we discuss and define different regrets that we leverage in our analysis of O2SLS.

In econometrics and bio-statistics, where 2SLS is popularly used the focus is accurate identification of the underlying structural model $\boldsymbol{\beta}$. Identifying $\boldsymbol{\beta}$ leads to understanding of the underlying economic or biological causal relations and their dynamics. In ML, Venkatraman et al. (2016) applied O2SLS for online linear system identification. Thus, given a sequence of estimators $\{\boldsymbol{\beta}_t\}_{t=1}^T$ and a sequence of covariates $\{\boldsymbol{x}_t\}_{t=1}^T$, the cost of identifying the true parameter $\boldsymbol{\beta}$ can be quantified by

$$\widetilde{R}_T(\boldsymbol{\beta}) \triangleq \sum_{t=1}^{T} (\boldsymbol{x}_t^\top \boldsymbol{\beta}_{t-1} - \boldsymbol{x}_t^\top \boldsymbol{\beta})^2. \tag{3}$$

We refer to $\widetilde{R}_T(\boldsymbol{\beta})$ as *identification regret* over horizon T. In the just identified setting that we are considering, the identification regret is equivalent to the regret of counterfactual prediction (Eqn. 5, Hartford et al. (2016)). Counterfactual predictions are important to study the causal questions: what would have changed in the outcome if Treatment a is used instead of treatment b. One of the modern applications of IVs are to facilitate such counterfactual predictions (Hartford et al., 2016; Bennett et al., 2019; Zhu et al., 2022).

Alternatively, one might be interested in evaluating and improving the quality of prediction obtained using an estimator $\{\boldsymbol{\beta}_t\}_{t=1}^T$ with respect to an underlying oracle (or expert), which is typically the case in statistical learning theory and forecasting (Foster, 1991; Cesa-Bianchi and Lugosi, 2006). If the oracle has access to the true parameters $\boldsymbol{\beta}$, the cost in terms of prediction that the estimators pay with respect to the oracle is $\bar{r}_t \triangleq \ell_t\left(\boldsymbol{\beta}_t\right) - \ell_t\left(\boldsymbol{\beta}\right)$. Thus, the regret in terms of the quality of prediction is defined as

$$\overline{R}_T(\boldsymbol{\beta}) \triangleq \sum_{t=1}^{T} (y_t - \boldsymbol{x}_t^\top \boldsymbol{\beta}_{t-1})^2 - \sum_{t=1}^{T} (y_t - \boldsymbol{x}_t^\top \boldsymbol{\beta})^2. \tag{4}$$

We refer to $\overline{R}_T(\boldsymbol{\beta})$ as the *oracle regret*. This regret is studied for stochastic analysis of online regression (Ouhamma et al., 2021) and is also useful for analysing bandit algorithms (Foster and Rakhlin, 2020).

As O2SLS is interesting for learning causal structures, we focus on the identification regret. On the other hand, to compare with the existing results in online linear regression, we also analyse the oracle regret of O2SLS. Though we know that they are of similar order (w.r.t. T) in the exogenous setting, we show that they differ significantly for O2SLS under endogeneity.

Remark 4.2. *In online learning theory focused on Empirical Risk Minimisation (ERM), another type of regret is considered where the oracle has access to the best offline estimator* $\boldsymbol{\beta}_T \triangleq \operatorname{argmin}_{\boldsymbol{\beta}} \sum_{t=1}^{T}(y_t - \boldsymbol{x}_t^\top \boldsymbol{\beta})^2$ *given the observations over* T *steps Cesa-Bianchi and Lugosi (2006). Thus, the new formulation of regret becomes*

$$R_T = \sum_{t=1}^{T}(y_t - \boldsymbol{x}_t^\top \boldsymbol{\beta}_{t-1})^2 - \min_{\boldsymbol{\beta}} \sum_{t=1}^{T}(y_t - \boldsymbol{x}_t^\top \boldsymbol{\beta})^2. \tag{5}$$

We refer to it as the population regret. Under exoegeneity, Ouhamma et al. (2021) shows that oracle regret and population regret differs by $o(\log^2 T)$*. We show that under endeogeneity their expected values differ by* $\Omega(T)$*. Thus, we avoid studying this notion of regret in this paper. More details are in Appendix E.*

4.2 Theoretical Analysis

Confidence Interval of $\boldsymbol{\beta}_t$. The central result in our analysis is concentration of the O2SLS estimates $\boldsymbol{\beta}_t$ around $\boldsymbol{\beta}$.

Lemma 4.1 (Confidence Ellipsoid for the Second-stage Parameters)**.** *Let us define the design matrix to be* $\mathbf{G}_{\boldsymbol{z},t} = \mathbf{Z}_t^\top \mathbf{Z}_t + \lambda \mathbb{I}_d$ *for some* $\lambda > 0$*. Then, for* σ_η*-sub-Gaussian first stage noise* η_t*, the true parameter* $\boldsymbol{\beta}$ *belongs to the set*

$$\mathcal{E}_t = \left\{ \boldsymbol{\beta} \in \mathbb{R}^d : \|\boldsymbol{\beta}_t - \boldsymbol{\beta}\|_{\widehat{\mathbf{H}}_t} \le \sqrt{\mathfrak{b}_t(\delta)} \right\}, \tag{6}$$

with probability at least $1-\delta \in (0,1)$*, for all* $t \ge 0$*. Here,* $\mathfrak{b}_t(\delta) \triangleq \frac{d\sigma_\eta^2}{4}\log\left(\frac{1+tL_z^2/\lambda}{\delta}\right)$*,* $\widehat{\mathbf{H}}_t \triangleq \widehat{\boldsymbol{\Theta}}_t^\top \mathbf{G}_{\boldsymbol{z},t}\widehat{\boldsymbol{\Theta}}_t$*, and* $\widehat{\boldsymbol{\Theta}}_t$ *is the estimate of the first-stage parameter at time* t *(Appendix J).*

Lemma 4.1 extends the well-known elliptical lemma for OLS and Ridge estimators under exoegeneity to the O2SLS estimator under endogeneity. It shows that the size of the confidence intervals induced by O2SLS estimate at time T is $\mathcal{O}(\sqrt{d\log T})$, which is of the same order as that of the exogenous elliptical lemma (Abbasi-Yadkori et al., 2011a).

Identification Regret Bound. Now, we state the identification regret upper bound of O2SLS and a brief proof sketch.

Theorem 4.1 (Identification Regret of O2SLS)**.** *If Assumption 3.1 holds true, then for bounded IVs* $\|\boldsymbol{z}\|^2 \le L_z^2$*, if* η_t *is the* σ_η*-sub-Gaussian second stage noise and* $\boldsymbol{\epsilon}_t$ *is the component-wise* $\sigma_{\boldsymbol{\epsilon}}$*-sub-Gaussian first stage noise, the regret of* O2SLS *at step* $T > 1$ *satisfies*

$$\widetilde{R}_T \le \underbrace{\mathfrak{b}_{T-1}(\delta)}_{\substack{\textit{Estimation}\\ \mathcal{O}(d\log T)}} \underbrace{\left((C_1^2 + dC_2^2)f(T) + C_4\right)}_{\substack{\textit{Second Stage Feature norm}\\ \mathcal{O}(d\log T)}}$$

with probability at least $1-\delta \in (0,1)$*. Here,* $\mathfrak{b}_{T-1}(\delta)$ *is the confidence bound of* O2SLS *estimate around* $\boldsymbol{\beta}$ *(Lemma 4.1) and* $f(T) \triangleq \left(\frac{C_3'}{\lambda} + \frac{\log(T)+1}{\lambda_{\min}(\boldsymbol{\Sigma})/2}\right)$*.* C_1*,* C_2*,* C_3'*,* C_4 *are* d *and* T*-independent positive constants (Appendix G), and* $\lambda_{\min}(\boldsymbol{\Sigma})$ *is the minimum eigenvalue of the true covariance matrix of IVs, i.e.* $\boldsymbol{\Sigma} \triangleq \mathbb{E}[\boldsymbol{z}\boldsymbol{z}^\top]$*.*

Proof Sketch. For brevity, we define $\Delta\boldsymbol{\beta}_{t-1} \triangleq \boldsymbol{\beta}_{t-1} - \boldsymbol{\beta}$. By applying Cauchy-Schwarz inequality in Eq. (3), we decouple the effects of parameter estimates and feature norms

$$\sum_{t=1}^{T}\left(\Delta\boldsymbol{\beta}_{t-1}^\top \boldsymbol{x}_t\right)^2 \le \sum_{t=1}^{T}\left\|\Delta\boldsymbol{\beta}_{t-1}\right\|_{\widehat{\mathbf{H}}_{t-1}}^2 \|\boldsymbol{x}_t\|_{\widehat{\mathbf{H}}_{t-1}^{-1}}^2$$

Now, we bound this term by (a) using the confidence bound to control the concentration of $\boldsymbol{\beta}_t$ around $\boldsymbol{\beta}$, and (b) by bounding the sum of feature norms.

Step a: Confidence Intervals of $\boldsymbol{\beta}_t$. We directly use Lemma 4.1 to bound $\left\|\Delta\boldsymbol{\beta}_{t-1}\right\|^2_{\widehat{\mathbf{H}}_{t-1}}$ by $\mathfrak{b}_{t-1}$.

Step b: Bounding the Second Stage Features. Now, we need to bound the sum of the feature norms. We use Lemma C.3 to obtain $\sum_{t=1}^{T}\|\boldsymbol{x}_t\|^2_{\widehat{\mathbf{H}}^{-1}_{t-1}} \leq \left(C_1^2 + dC_2^2\right) f(T) + C_4$. The idea is to substitute $\boldsymbol{x}_t$ with (First stage) equation. This leads to two terms $\sum_{t=1}^{T}\|\boldsymbol{\Theta}^\top \boldsymbol{z}_t\|^2_{\widehat{\mathbf{H}}^{-1}_{t-1}}$ and $\sum_{t=1}^{T}\|\boldsymbol{\epsilon}_t\|^2_{\widehat{\mathbf{H}}^{-1}_{t-1}}$. Then, we bound the first term by $C_1^2 f(T)$ using boundedness of the first-stage features and the concentration property of the minimum eigenvalue of the design matrix of the first stage, i.e. $\left\|\mathbf{G}^{-1}_{\boldsymbol{z},t-1}\right\|_2 \triangleq \left\|\left(\sum_{s=1}^{t-1}\boldsymbol{z}_s^T\boldsymbol{z}_s\right)^{-1}\right\|_2$. The concentration of the minimum eigenvalue leads to the term $f(T) \triangleq \left(\frac{C_3'}{\lambda} + \frac{\log(T)+1}{\lambda_{\min}(\boldsymbol{\Sigma})/2}\right)$.

Then, we bound $\sum_{t=1}^{T}\|\boldsymbol{\epsilon}_t\|^2_{\widehat{\mathbf{H}}^{-1}_{t-1}}$ using component-wise sub-Gaussianity of the first stage noise. This leads to a bound $dC_2^2 f(T) + C_4$ with probability $1-\delta$.

Final step. Since $\mathfrak{b}_{t-1}$ is non-decreasing in t, we conclude that $\sum_{t=1}^{T}\left(\Delta\boldsymbol{\beta}_{t-1}^\top \boldsymbol{x}_t\right)^2$ is upper bounded by $\mathfrak{b}_{T-1}(\delta)((C_1^2 + dC_2^2)f(T) + C_4)$. Thus, *we conclude that the identification regret of* O2SLS *is $\mathcal{O}(d^2\log^2 T)$ for bounded IVs and unbounded noises.*

Remark 4.3. *Theorem 4.1 entails a regret $\widetilde{R}_T = \mathcal{O}\left(d^2\log^2(T)\right)$, where d is dimension of IV. This regret bound is $d\log T$ more than the regret of online ridge regression, i.e. $\mathcal{O}(d\log T)$ (Gaillard et al., 2019). This is due to the fact that we perform d linear regressions in the first-stage and using the predictions of first stage for the second-stage regression. These two regression steps in cascade induce the proposed regret bound.*

Oracle Regret Bound. Now, we provide a proof sketch of the oracle regret. Further details are in Appendix G.

Theorem 4.2 (Oracle Regret of O2SLS)**.** *Under the same hypothesis of Theorem 4.1, the Oracle Regret of* O2SLS *at step $T > 1$ satisfies*

$$\overline{R}_T \leq \underbrace{\widetilde{R}_T}_{\substack{\text{Identif.}\\ \text{Regret}\\ \mathcal{O}(d^2\log^2 T)}} + \underbrace{\sqrt{\mathfrak{b}_{T-1}(\delta)}}_{\substack{\text{Estimation}\\ \mathcal{O}(\sqrt{d\log T})}}\left(\underbrace{\sigma_\eta C_1\sqrt{f(T)\log\left(\frac{\log T}{\delta}\right)}}_{\substack{\text{First Stage Feature Norm}\\ \mathcal{O}(\sqrt{\log T})}} \underbrace{+C_5\sqrt{2df(T)} + \sqrt{d}C_6}_{\substack{\text{Correlated noise}\\ \text{Concentration Term}\\ \mathcal{O}(\sqrt{d\log T})}} + \underbrace{\gamma L_{\widehat{\boldsymbol{\Theta}}^{-1}}\left(\frac{C_3}{\sqrt{\lambda}} + \frac{2\sqrt{2T}}{\sqrt{\lambda_{\min}(\boldsymbol{\Sigma})}}\right)}_{\substack{\text{Correlated noise}\\ \text{Bias Term}\\ \mathcal{O}(\gamma\sqrt{T})}}\right)$$

with probability at least $1-\delta \in (0,1)$. We define $\gamma \triangleq \|\boldsymbol{\gamma}\|_2 = \|\mathbb{E}[\eta_s\boldsymbol{\epsilon}_s]\|_2$. C_1, C_2, C_3', and C_4 are the d and T-independent positive constants (as in Thm. 4.1 and App. G). Constants $C_5 \triangleq 8e^2\left(\sigma_\eta^2+\sigma_{\boldsymbol{\epsilon}}^2\right)L_{\widehat{\boldsymbol{\Theta}}^{-1}}\sqrt{\log(2/\delta)}$, $C_6 \triangleq C_5\sqrt{\max\left\{\frac{1}{\lambda}, \frac{2}{\lambda_{\min}(\boldsymbol{\Sigma})}\right\}\log(2/\delta)}$, and $f(T) \triangleq \left(\frac{C_3'}{\lambda} + \frac{\log(T)+1}{\lambda_{\min}(\boldsymbol{\Sigma})/2}\right)$.

Proof Sketch. Using Equation (4) and Equation (2), we decompose the regret at step T as

$$\overline{R}_T = \underbrace{\sum_{t=1}^{T}\left(\Delta\boldsymbol{\beta}_{t-1}^\top\boldsymbol{x}_t\right)^2}_{(\bullet 1\bullet)} + 2\underbrace{\sum_{t=1}^{T}\eta_t\Delta\boldsymbol{\beta}_{t-1}^\top\Theta^\top\boldsymbol{z}_t}_{(\bullet 2\bullet)} + 2\underbrace{\sum_{t=1}^{T}\eta_t\Delta\boldsymbol{\beta}_{t-1}^\top\boldsymbol{\epsilon}_t}_{(\bullet 3\bullet)}.$$

The proof proceeds by bounding each of these three terms.

Term 1: Second Stage Regression Error. We observe that Term (•1•) is same as $\widetilde{R}_T$. By Theorem 4.1, we know $\widetilde{R}_T = \mathcal{O}(d^2 \log^2 T)$.

Term 2: Coupling of First-stage Data and Second-stage Parameter Estimation. Now, we bound the second term using concentration inequalities of martingales. First, we observe that $w_t \triangleq \left(\boldsymbol{\beta}_{t-1} - \boldsymbol{\beta}\right)^\top \Theta^\top \boldsymbol{z}_s$ is a martingale with respect to the filtration

$$\mathcal{F}_{t-1} = \sigma\left(\boldsymbol{z}_1, \boldsymbol{\epsilon}_1, \eta_1, \ldots, \boldsymbol{z}_{t-1}, \boldsymbol{\epsilon}_{t-1}, \eta_{t-1}, \boldsymbol{z}_t\right).$$

We also note that w_t is $\mathcal{F}_{t-1}-$measurable since $\boldsymbol{\beta}_{t-1}$ and $\boldsymbol{z}_t$ are too. By concentration property of scalar-valued martingale concentration (Theorem I.2), we get that with probability $1-\delta$

$$\begin{aligned}(\bullet 2\bullet) &\leq \left|\sum_{t=1}^{T} \eta_t w_t\right| \\ &\leq \sqrt{2\left(1+\sigma_\eta^2 \sum_{t=1}^{T} w_t^2\right) \log\left(\frac{\sqrt{1+\sigma_\eta^2 \sum_{t=1}^{T} w_t^2}}{\delta}\right)}.\end{aligned}$$

Now, we focus on bounding the quantity appearing under square root. By applying Cauchy-Schwarz inequality and a reasoning similar to bounding Term (•1•), we get $\sum_{t=1}^{T} w_t^2 \leq \mathfrak{b}_{T-1}(\delta) C_1^2 f(T)$. Hence, we conclude that Term (•2•) is $\mathcal{O}(\sqrt{d}\log T)$ ignoring the log log terms.

Term 3: Coupling of First- and Second-stage Noises. Finally, we bound Term (•3•) containing the correlation between the first- and second-stage noise. This term is referred as the self-fulfilling bias (Li et al., 2021). We bound this term by splitting it into two.

$$\sum_{t=1}^{T} \eta_t \Delta\boldsymbol{\beta}_{t-1}^\top \boldsymbol{\epsilon}_t = \underbrace{\sum_{t=1}^{T} \Delta\boldsymbol{\beta}_{t-1}^T \left(\boldsymbol{\epsilon}_t n_t - \boldsymbol{\gamma}\right)}_{\text{Martingale Concentration Term}} + \underbrace{\sum_{t=1}^{T} \Delta\boldsymbol{\beta}_{t-1}^\top \boldsymbol{\gamma}}_{\text{Bias Term}}$$

Here, $\boldsymbol{\gamma} \triangleq \mathbb{E}[\eta_s \boldsymbol{\epsilon}_s]$. This leads to the first term, which is a summation of martingale difference sequence and can be bounded using concentration inequalities in Lemma F.2. The technical challenge is to derive the sub-exponential parameters induced by $\boldsymbol{\epsilon}_t \eta_t$ in the martingale difference, since the individual terms are products of two dependent random variables $\boldsymbol{\epsilon}_t$ and η_t. By applying Bernstein's inequality on the martingale difference and $\left\|\Delta\boldsymbol{\beta}_{t-1}\right\|_{\widehat{\mathbf{H}}_{t-1}} \leq \sqrt{\mathfrak{b}_{T-1}(\delta)}$ with probability $1-\delta$, we obtain

$$\underbrace{\sqrt{\mathfrak{b}_{T-1}(\delta)}}_{\substack{\text{Estimation} \\ \mathcal{O}(\sqrt{d\log T})}} \underbrace{C_5\sqrt{2d\ f(T)} + C_6}_{\substack{\text{Correlated noise} \\ \text{Concentration Term} \\ \mathcal{O}(\sqrt{d\log T})}}$$

The Bias Term is the one where the correlation $\boldsymbol{\gamma}$ appears explicitly. We bound this term (Lemma F.3) by bounding the sum of the square root of the smallest eigenvalues of the first stage covariates design matrix $\sum_{t=1}^{T} \sqrt{\left\|\mathbf{G}_{\boldsymbol{z},t-1}^{-1}\right\|_2}$. We reuse the upper bound on the individual terms (Lemma B.2), where we show that the minimum eigenvalue of the first stage design matrix grows $\Omega(t)$. Thus, we get that $\sqrt{\lambda_{\max}\left(\mathbf{G}_{\boldsymbol{z},t}^{-1}\right)}$ is $\mathcal{O}(\frac{1}{\sqrt{t}})$. This leads to the following bound on the Bias Term

$$\underbrace{\sqrt{\mathfrak{b}_{T-1}(\delta)}}_{\substack{\text{Estimation} \\ \mathcal{O}(\sqrt{d\log T})}} \left(\underbrace{\gamma L_{\widehat{\boldsymbol{\Theta}}^{-1}}\left(\frac{C_3}{\sqrt{\lambda}} + 2\frac{\sqrt{2T}}{\sqrt{\lambda_{\min}(\boldsymbol{\Sigma})}}\right)}_{\substack{\text{Correlated noise - Bias Term} \\ \mathcal{O}(\|\boldsymbol{\gamma}\|_2\sqrt{T})}}\right)$$

Thus, we conclude the proof and get that the oracle regret of O2SLS is $\mathcal{O}(\gamma\sqrt{dT\log T} + d^2\log^2 T)$.

Remark 4.4. *Under exogeneity and unbounded stochastic noise, the oracle regret of online linear regression is $\mathcal{O}(d^2\log^2 T)$ (Ouhamma et al., 2021). Under endogeneity and unbounded stochastic noise,* O2SLS *incurs an extra $\mathcal{O}(\gamma\sqrt{dT\log T})$ factor in the oracle regret. This term appears due to the correlation between the second and the first-stage noises, and it is proportional to the degree of correlation between the noises in these two stages. Thus, the bias introduced by the correlation of noises acts as the dominant term. In 2*SLS *literature, this is referred as the self-fulfilling bias (Li et al., 2021). When the noises are independent, i.e. $\gamma = 0$, we retrieve an oracle regret of the same order as that of the exogenous case.*

5. Linear Bandits with Endogeneity: OFUL-IV

We formulate stochastic Linear Bandits with Endogeneity (LBE) with a two-stage linear model of data generation (Eqn. (2)). Then, we propose an index-based optimistic algorithm, OFUL-IV. Our analysis shows that OFUL-IV achieves $\mathcal{O}(d\sqrt{T}\log T)$ regret. Our experimental results show that OFUL-IV achieve lower regret and more accurately estimates than OFUL (Abbasi-Yadkori et al., 2011a).

In bandit setting, we observe $\boldsymbol{x}_t$ and y_t depending on arm (or intervention) $A_t \in \mathcal{A}_t$ drawn at time $t \in \{1, \ldots, T\}$.

$$\boldsymbol{x}_t = \boldsymbol{\Theta}^\top \boldsymbol{z}_{t,A_t} + \boldsymbol{\epsilon}_t \qquad \text{(LBE-first)}$$

$$y_t = \boldsymbol{\beta}^\top \boldsymbol{x}_t + \eta_t \qquad \text{(LBE-second)}$$

Here, y_t is the reward at round t. Each arm a corresponds to *a vector of IVs* $\boldsymbol{z}_{t,a} \in \mathcal{Z}_t \subset \mathbb{R}^d$, and a vector of endogenous variables $\boldsymbol{x}_{t,a} \in \mathcal{X}_t \subset \mathbb{R}^d$ (generated as per (LBE-first)). Here, $\mathcal{X}_t$ and $\mathcal{Z}_t$ are sets of IVs and endogenous variables corresponding to $\mathcal{A}_t$. Similar to regression setting, we have two sources of unobserved noises: $\boldsymbol{\epsilon}_t$ ($\sigma_{\boldsymbol{\epsilon}}^2\mathbb{I}$-sub-Gaussian) are i.i.d. vector error terms at round t which is independent of $\boldsymbol{z}$, and η_t (σ_η^2-sub-Gaussian), representing all causes of y_t other than $\boldsymbol{x}_t$. True parameters $\boldsymbol{\beta} \in \mathbb{R}^d$ and $\boldsymbol{\Theta} \in \mathbb{R}^{d\times d}$ are unknown to the agents. This is an extension of the classical stochastic linear bandit (Lattimore and Szepesvári, 2020, Ch. 19). Now, we state the protocol of LBE.

At each round $t = 1, 2, \ldots, T$, the agent
1. Observes a sample $\boldsymbol{z}_{t,a} \in \mathcal{Z}_t$ and $\boldsymbol{x}_{t,a} \in \boldsymbol{x}_t$ of contexts for all $a \in \mathcal{A}_t$
2. Chooses an arm $A_t \in \mathcal{A}_t$
3. Obtains the reward y_t computed from (LBE-first)
4. Updates the parameter estimates $\widehat{\boldsymbol{\Theta}}_t$ and $\boldsymbol{\beta}_t$

OFUL-IV: Algorithm Design. If the agent had full information in hindsight, she could infer the best arm (or intervention) in $\mathcal{A}_t$ as

$$a_t^* = \underset{a\in\mathcal{A}_t}{\operatorname{argmax}}\, \mathbb{E}[\boldsymbol{x}_{t,a}^\top \boldsymbol{\beta}]$$

We denote the corresponding variables as $\boldsymbol{z}_t^*$ and $\boldsymbol{x}_t^*$. Thus, choosing a^* can be shown as choosing $\boldsymbol{z}_t^*$ and $\boldsymbol{x}_t^*$. But the agent does not know them and aims to select $\{a_t\}_{t=1}^T$ leading to minimum regret (Eqn. (4)). Now, we extend the OFUL algorithm minimising regret in linear bandits with exogeneity (Abbasi-Yadkori et al., 2011a). The core idea is that the algorithm maintains a confidence set $\mathcal{B}_{t-1} \subseteq \mathbb{R}^d$ around the parameter $\boldsymbol{\beta}$, which is computed only using the observed data. Then, the

Algorithm 2 OFUL-IV

Input: Initialization parameters $\boldsymbol{\beta}_0, \widehat{\boldsymbol{\Theta}}_0, \mathfrak{b}'_0$
for $t = 1, 2, \dots, T$ **do**
  Observe $\boldsymbol{z}_{t,a} \in \mathcal{Z}_t$ and $\boldsymbol{x}_{t,a} \in \boldsymbol{x}_t$ for $a \in \mathcal{A}_t$
  Compute $\boldsymbol{\beta}_{t-1}$ according to Equation (O2SLS)
  Choose action A_t that solves Equation (7)
  Update $\boldsymbol{\beta}_t \leftarrow \boldsymbol{\beta}_{t-1}, \widehat{\boldsymbol{\Theta}}_t \leftarrow \widehat{\boldsymbol{\Theta}}_t, \mathfrak{b}'_t \leftarrow \mathfrak{b}'_t$
end for

algorithm chooses an optimistic estimate of $\widetilde{\boldsymbol{\beta}}_{t-1}$ from that confidence set:

$$\widetilde{\boldsymbol{\beta}}_{t-1} = \underset{\boldsymbol{\beta}' \in \mathcal{B}_{t-1}}{\operatorname{argmax}} \left(\max_{\boldsymbol{x} \in \mathcal{X}_t} \boldsymbol{x}^\top \boldsymbol{\beta}' \right)$$

Then, she chooses the action leading to $\boldsymbol{x}_t = \operatorname{argmax}_{\boldsymbol{x} \in \mathcal{X}_t} \boldsymbol{x}^\top \widetilde{\boldsymbol{\beta}}_t$, which maximizes the reward according to the estimate $\widetilde{\boldsymbol{\beta}}_t$. In brief, the algorithm chooses the pair $(\boldsymbol{x}_t, \widetilde{\boldsymbol{\beta}}_{t-1}) = \underset{(\boldsymbol{x}, \boldsymbol{\beta}') \in \mathcal{X}_t \times \mathcal{B}_{t-1}}{\operatorname{argmax}} \boldsymbol{x}^\top \boldsymbol{\beta}'$.

In order to tackle endogeneity, we choose to use the O2SLS estimate $\boldsymbol{\beta}_{t-1}$ computed using data observed till $t-1$. Then, we build an ellipsoid $\mathcal{B}_{t-1}$ around it, such that

$$\mathcal{B}_{t-1} \triangleq \left\{ \boldsymbol{\beta} \in \mathbb{R}^d : \|\boldsymbol{\beta}_t - \boldsymbol{\beta}\|_{\widehat{\mathbf{H}}_t} \leq \sqrt{\mathfrak{b}'_t(\delta)} \right\}$$

and

$$\mathfrak{b}'_t(\delta) \triangleq 2\sigma_\eta^2 \log \left(\frac{\det\left(\mathbf{G}_{\boldsymbol{z},t}\right)^{1/2} \lambda^{-d/2}}{\delta} \right).$$

Given this confidence interval, we optimistically choose the arm

$$A_t = \underset{a \in \mathcal{A}_t}{\operatorname{argmax}} \left\langle \boldsymbol{x}_{t,a}, \boldsymbol{\beta}_{t-1} \right\rangle + \sqrt{\mathfrak{b}'_{t-1}(\delta)} \, \|\boldsymbol{x}_{t,a}\|_{\widehat{\mathbf{H}}_{t-1}^{-1}}. \tag{7}$$

This arm selection index together with the O2SLS estimator yielding $\boldsymbol{\beta}_{t-1}$ construct the OFUL-IV (Algorithm 2).

Theorem 5.1. *Under the same assumptions and notations of Theorem 4.1 and Theorem 4.2, Algorithm 2 incurs a regret*

$$R_T \leq 2\sqrt{T} \underbrace{\sqrt{\mathfrak{b}_{T-1}(\delta)}}_{\substack{\textit{Estimation} \\ \mathcal{O}(\sqrt{d \log T})}} \underbrace{\sqrt{(C_1^2 + dC_2^2) f(T) + C_4}}_{\substack{\textit{Second Stage Feature norm} \\ \mathcal{O}(\sqrt{d \log T})}}$$

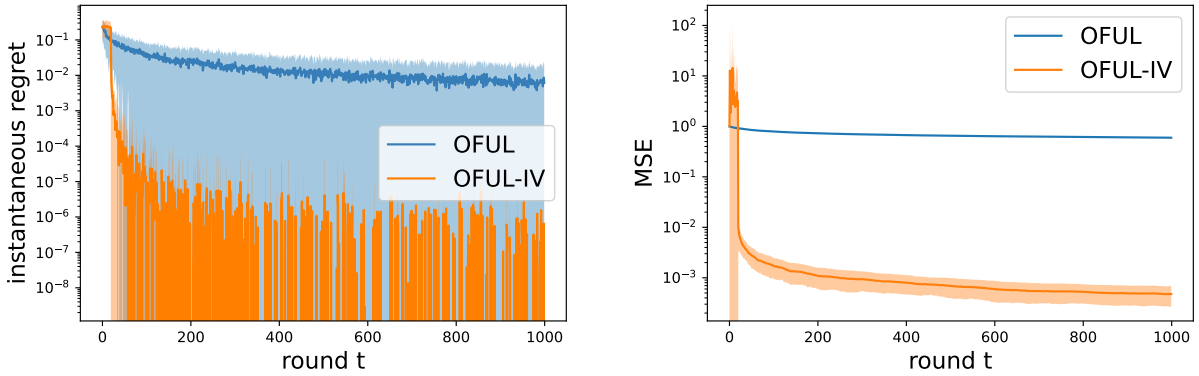
with probability $1 - \delta$ *and for horizon* $T > 1$.

Proof Sketch. **Step 1: Optimism.** We observe that $R_T = \sum_{t=1}^T \boldsymbol{\beta}^\top \boldsymbol{x}_* - \boldsymbol{\beta}^\top \boldsymbol{x}_t \triangleq \sum_{t=1}^T r_t$. Since $(\boldsymbol{x}_t, \widetilde{\boldsymbol{\beta}}_{t-1})$ is optimistic in $\mathcal{X}_t \times \mathcal{B}_t$, and $\boldsymbol{\beta} \in \mathcal{B}_t$, we obtain $r_t \leq (\widetilde{\boldsymbol{\beta}}_{t-1} - \boldsymbol{\beta})^\top \boldsymbol{x}_t$.

Step 2: Decomposition. Now, we decompose regret as

$$(\widetilde{\boldsymbol{\beta}}_{t-1} - \boldsymbol{\beta})^\top \boldsymbol{x}_t = (\widetilde{\boldsymbol{\beta}}_{t-1} - \boldsymbol{\beta}_{t-1})^\top \boldsymbol{x}_t + (\boldsymbol{\beta}_{t-1} - \boldsymbol{\beta})^\top \boldsymbol{x}_t.$$

The first term depends on tightness of confidence interval, while the second depends on accuracy of the estimate $\boldsymbol{\beta}_{t-1}$.



(a) Instantaneous regret (b) MSE w.r.t. $\boldsymbol{\beta}$

Figure 2: We compare instantaneous regrets (left) of OFUL and OFUL-IV in a linear bandit setting. We show the MSE between the parameters estimated by the two algorithms and the true parameter $\boldsymbol{\beta}$ (right). OFUL-IV incurs lower instantaneous regret and MSE.

Step 3: Confidence Bound. Now, we can decouple the impact of parameter and observed data in both the terms using $\left\|\widetilde{\boldsymbol{\beta}}_{t-1}-\boldsymbol{\beta}_{t-1}\right\|_{\widehat{\mathbf{H}}_{t-1}}\|\boldsymbol{x}_t\|_{\widehat{\mathbf{H}}_{t-1}^{-1}}$ and $\left\|\boldsymbol{\beta}_{t-1}-\boldsymbol{\beta}\right\|_{\widehat{\mathbf{H}}_{t-1}}\|\boldsymbol{x}_t\|_{\widehat{\mathbf{H}}_{t-1}^{-1}}$, respectively. By construction of the optimistic confidence interval and concentrations bound of Lemma 4.1, both $\left\|\widetilde{\boldsymbol{\beta}}_{t-1}-\boldsymbol{\beta}_{t-1}\right\|_{\widehat{\mathbf{H}}_{t-1}}$ and $\left\|\boldsymbol{\beta}_{t-1}-\boldsymbol{\beta}\right\|_{\widehat{\mathbf{H}}_{t-1}}$ are bounded by $\sqrt{\mathfrak{b}'_{t-1}(\delta)}$. By determinant-trace inequality (Lemma A.1), we get that $\mathfrak{b}'_{T-1}(\delta)\leq\frac{d\sigma_\eta^2}{4}\log\left(\frac{1+TL_z^2/\lambda}{\delta}\right)$.

Final Step. Since the regret $R_T\leq\sqrt{T\sum_{t=1}^T r_t^2}$, we obtain

$$R_T\leq\sigma_\eta\sqrt{dT\log\left(\frac{1+TL_z^2/\lambda}{\delta}\right)\left(\sum_{t=1}^T\|\boldsymbol{x}_t\|^2_{\widehat{\mathbf{H}}_{t-1}^{-1}}\right)}.$$

Now, we bound the sum of the first-stage feature norms $\sum_{t=1}^T\|\boldsymbol{x}_t\|^2_{\widehat{\mathbf{H}}_{t-1}^{-1}}$ by $(C_1^2+dC_2^2)f(T)+C_4$ (Lemma C.3), which is $\mathcal{O}(d\log T)$. A detailed proof is in Appendix H.

Thus, we conclude that *regret of* OFUL-IV *is* $\mathcal{O}(d\sqrt{T}\log T)$. OFUL-IV achieves regret of similar order under endogeneity as OFUL achieves under exoegeneity.

Experimental Analysis. Now, we compare performance of OFUL-IV and OFUL (Abbasi-Yadkori et al., 2011a) for LBE setting. OFUL builds a confidence ellipsoid centered at $\boldsymbol{\beta}_{\mathsf{Ridge},t}$ to concentrate around $\boldsymbol{\beta}$, while OFUL-IV uses O2SLS to build an accurate estimate. We deploy the experiments in Python3 on a single Intel(R) Core(TM) i7-8665U CPU@1.90GHz. For each algorithm, we report the mean and standard deviation of instantaneous regret and mean square error $\|\boldsymbol{\beta}_t-\boldsymbol{\beta}\|^2$ over 100 runs. We run the algorithms with the same regularisation parameters equal to 10^{-3}. We denote the normal distribution with mean μ and standard deviation σ as $\mathcal{N}(\mu,\sigma)$, with $\mathcal{N}_n$ we indicate its multivariate extension to n dimensions. For each experiment, we sample the true parameters in our model once according to $\boldsymbol{\beta}\sim\mathcal{N}_{50}(\vec{1}0,\mathbb{I}_{50})$ and $\boldsymbol{\Theta}_{i,j}\sim\mathcal{N}(0,1)$ for each component. Then we sample at each time t the vectors $\boldsymbol{z}_{t,a}\sim\mathcal{N}_{50}(\vec{0},\mathbb{I}_{50})$, $\boldsymbol{\epsilon}_{t,a}\sim\mathcal{N}_{50}(\vec{0},\mathbb{I}_{50})$, and the scalar noise $\eta_{t,a}=\frac{1}{13}\left(\widetilde{\eta}_{t,a}+\sum_{i=1}^{12}\boldsymbol{\epsilon}_{t,a,i}\right)$ where $\widetilde{\eta}_{t,a}\sim\mathcal{N}(0,1)$.

The estimates obtained by OFUL-IV achieves 3-order less error than those of OFUL (Fig. 2b). Thus, OFUL-IV leads to lower regret than OFUL for linear bandits with endogeneity (Fig. 2a). Further experimental details and results of regression are deferred to Appendix K.1.

6. Conclusions and Future Works

In this paper, we study online IV regression, specifically the online 2SLS algorithm, for unbounded noise and endogenous data. We analyse the finite-time identification and oracle regrets of O2SLS. We observe that O2SLS incurs $\mathcal{O}(d^2 \log^2 T)$ identification regret, which is $d \log T$ higher than that of online linear regression under exogeneity. In contrast, O2SLS achieves $\mathcal{O}(\|\boldsymbol{\gamma}\|_2 \sqrt{dT \log T})$ oracle regret as the correlation between the noises in the two-stages dominate the identification regret. But these two are of the same order in the exogenous setting. Following that, we study stochastic linear bandits with endogeneity. We propose OFUL-IV that uses O2SLS to estimate the model parameters. We show that OFUL-IV achieves $\mathcal{O}(d\sqrt{T} \log T)$ regret. We experimentally show that OFUL-IV yields more accurate estimates of the true parameter and thus, lower regret.

For simplicity, we consider the just-identified IVs. In future, we will like to extend our algorithms and analysis to weakly or over-identified IVs (Greene, 2003). Additionally, O2SLS and OFUL-IV work if the IVs are already specified. There has been significant work to identify IVs in offline setting (Newey and Powell, 2003; Chen et al., 2020). Still, it is an open question how optimally IVs can be identified online, while O2SLS is performed simultaneously.

References

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. *Advances in neural information processing systems*, 24, 2011a.

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Online least squares estimation with self-normalized processes: An application to bandit problems. *arXiv preprint arXiv:1102.2670*, 2011b.

Yasin Abbasi-Yadkori, David Pal, and Csaba Szepesvari. Online-to-confidence-set conversions and application to sparse stochastic bandits. In *Artificial Intelligence and Statistics*, pages 1–9. PMLR, 2012.

Marc Abeille and Alessandro Lazaric. Linear thompson sampling revisited. In *AISTATS*, 2017.

Joshua D Angrist and Guido W Imbens. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American statistical Association*, 90 (430):431–442, 1995.

Joshua D Angrist, Guido W Imbens, and Donald B Rubin. Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, 91(434):444–455, 1996.

Peter L Bartlett, Wouter M Koolen, Alan Malek, Eiji Takimoto, and Manfred K Warmuth. Minimax fixed-design linear regression. In *Conference on Learning Theory*, pages 226–239. PMLR, 2015.

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

Glenn W. Brier. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78:1–3, 1950.

Stephen Burgess, Dylan S Small, and Simon G Thompson. A review of instrumental variable estimators for mendelian randomization. *Statistical methods in medical research*, 26(5):2333–2355, 2017.

Pedro Carneiro, James J Heckman, and Edward J Vytlacil. Estimating marginal returns to education. *American Economic Review*, 101(6):2754–81, 2011.

Nicolo Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge university press, 2006.

Jiafeng Chen, Daniel L Chen, and Greg Lewis. Mostly harmless machine learning: learning optimal instruments in linear iv models. *arXiv preprint arXiv:2011.06158*, 2020.

Dean P Foster. Prediction in the worst case. *The Annals of Statistics*, pages 1084–1090, 1991.

Dylan Foster and Alexander Rakhlin. Beyond UCB: Optimal and efficient contextual bandits with regression oracles. In *International Conference on Machine Learning*, pages 3199–3210. PMLR, 2020.

David A Freedman. On tail probabilities for martingales. *the Annals of Probability*, pages 100–118, 1975.

Pierre Gaillard, Sébastien Gerchinovitz, Malo Huard, and Gilles Stoltz. Uniform regret bounds over $\mathbb{R}^d$ for the sequential linear regression problem with the square loss. In *Algorithmic Learning Theory*, pages 404–432. PMLR, 2019.

William H Greene. *Econometric analysis*. Pearson Education India, 2003.

Keegan Harris, Dung Daniel T Ngo, Logan Stapleton, Hoda Heidari, and Steven Wu. Strategic instrumental variable regression: Recovering causal relationships from strategic responses. In *International Conference on Machine Learning*, pages 8502–8522. PMLR, 2022.

Jason Hartford, Greg Lewis, Kevin Leyton-Brown, and Matt Taddy. Counterfactual prediction with deep instrumental variables networks. *arXiv preprint arXiv:1612.09596*, 2016.

Elad Hazan and Tomer Koren. Linear regression with limited observation. In *29th International Conference on Machine Learning, ICML 2012*, pages 807–814, 2012.

MA Hernan and J Robins. *Causal Inference: What if*. Boca Raton: Chapman & Hill/CRC, 2020.

Nathan Kallus. Instrument-armed bandits. In *Algorithmic Learning Theory*, pages 529–546. PMLR, 2018.

Abbas Kazerouni and Lawrence M Wein. Best arm identification in generalized linear bandits. *Operations Research Letters*, 49(3):365–371, 2021.

Jyrki Kivinen, Alexander J Smola, and Robert C Williamson. Online learning with kernels. *IEEE transactions on signal processing*, 52(8):2165–2176, 2004.

Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

Jin Li, Ye Luo, and Xiaowei Zhang. Self-fulfilling bandits: Dynamic selection in algorithmic decision-making. *arXiv preprint arXiv:2108.12547*, 2021.

Nicholas Littlestone, Philip M Long, and Manfred K Warmuth. On-line learning of linear functions. In *Proceedings of the twenty-third annual ACM symposium on Theory of computing*, pages 465–475, 1991.

Ruiqi Liu, Zuofeng Shang, and Guang Cheng. On deep instrumental variables estimate. *arXiv preprint arXiv:2004.14954*, 2020.

Magne Mogstad, Alexander Torgovitsky, and Christopher R Walters. The causal interpretation of two-stage least squares with multiple instrumental variables. *American Economic Review*, 111 (11):3663–98, 2021.

Maria Nareklishvili, Nicholas Polson, and Vadim Sokolov. Deep partial least squares for iv regression. *arXiv preprint arXiv:2207.02612*, 2022.

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

Francesco Orabona. A modern introduction to online learning. *arXiv preprint arXiv:1912.13213*, 2019.

Reda Ouhamma, Odalric Maillard, and Vianney Perchet. Stochastic online linear regression: the forward algorithm to replace ridge. *arXiv preprint arXiv:2111.01602*, 2021.

Reda Ouhamma, Debabrota Basu, and Odalric-Ambrym Maillard. Bilinear exponential family of mdps: Frequentist regret bound with tractable exploration and planning. *arXiv preprint arXiv:2210.02087*, 2022.

Donald B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688, 1974.

Amandeep Singh, Kartik Hosanagar, and Amit Gandhi. Machine learning instrument variables for causal inference. In *Proceedings of the 21st ACM Conference on Economics and Computation*, pages 835–836, 2020.

Andrew Stirn and Tony Jebara. Thompson sampling for noncompliant bandits. *arXiv preprint arXiv:1812.00856*, 2018.

Arun Venkatraman, Wen Sun, Martial Hebert, J Bagnell, and Byron Boots. Online instrumental variable regression with applications to online linear system identification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30, 2016.

Volodya Vovk. Competitive on-line linear regression. *Advances in Neural Information Processing Systems*, 10, 1997.

Volodya Vovk. Competitive on-line statistics. *International Statistical Review*, 69(2):213–248, 2001.

Martin J Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press, 2019.

Abraham Wald. The fitting of straight lines if both variables are subject to error. *The annals of mathematical statistics*, 11(3):284–300, 1940.

Larry Wasserman. *All of statistics: a concise course in statistical inference*, volume 26. Springer, 2004.

Philip G Wright. *Tariff on animal and vegetable oils*. Macmillan Company, New York, 1928.

Liyuan Xu, Yutian Chen, Siddarth Srinivasan, Nando de Freitas, Arnaud Doucet, and Arthur Gretton. Learning deep features in instrumental variable regression. *arXiv preprint arXiv:2010.07154*, 2020.

Liyuan Xu, Heishiro Kanagawa, and Arthur Gretton. Deep proxy causal learning and its application to confounded bandit policy evaluation. *arXiv preprint arXiv:2106.03907*, 2021.

Yuchen Zhu, Limor Gultchin, Arthur Gretton, Matt J Kusner, and Ricardo Silva. Causal inference with treatment measurement error: a nonparametric instrumental variable approach. In *Uncertainty in Artificial Intelligence*, pages 2414–2424. PMLR, 2022.

Appendix

A. Useful Results

Notations for scalars, vectors, matrices: We indicate in bold vectors and matrices, e.g. the vector and matrix (matrices are also usually capitalized) $\boldsymbol{v} \in \mathbb{R}^d, \mathbf{A} \in \mathbb{R}^{d\times d}$; while scalars do not use the bold notation, e.g. the scalar $s \in \mathbb{R}$. We indicate the determinant of matrix $\mathbf{A}$ with $\det(\mathbf{A})$ and its trace with $\mathrm{Tr}(\mathbf{A})$. For a $x \in \mathbb{R}_{\geq 0}$, we indicate the function that takes as input x, and gives as output the least integer greater than or equal to x as $\lceil x \rceil$ (ceiling function). We indicate the identity matrix of dimension d with $\mathbb{I}_d$.

A.1 Random Variables

Random variables follow the previous convention if they are scalar, vectors, or random matrix variables. We adopt the following convention when we talk about sub-Gaussians and sub-exponential random variables.

Definition A.1 (Sub-Gaussian r.v.)**.** *A random variable X with mean $\mu = \mathbb{E}[X]$ is sub-Gaussian if there is a positive number σ such that*

$$\mathbb{E}\left[e^{\lambda(X-\mu)}\right] \leq e^{\sigma^2\lambda^2/2} \quad \text{for all } \lambda \in \mathbb{R}.$$

Definition A.2 (Sub-exponential r.v.)**.** *A random variable X with mean $\mu = \mathbb{E}[X]$ is sub-exponential if there are non-negative parameters (ν, α) such that*

$$\mathbb{E}\left[e^{\lambda(X-\mu)}\right] \leq e^{\nu^2\lambda^2/2} \quad \text{for all } |\lambda| < \frac{1}{\alpha}$$

A.2 Norms of Vectors and Matrices

Definition A.3 (ℓ_p-norms)**.** *For a vector $\boldsymbol{v} \in \mathbb{R}^n$, we express its ℓ_p-norm as $\|\boldsymbol{v}\|_p$ for $p \geq 0$. A special case is the Euclidean ℓ_2-norm denoted as $\|\cdot\|_2$, which is induced by classical scalar product on $\mathbb{R}^n$ denoted by $\langle\cdot,\cdot\rangle$.*

Given a rectangular matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$ with $n \geq m$, we write its ordered singular values as

$$\sigma_{\max}(\mathbf{A}) = \sigma_1(\mathbf{A}) \geq \sigma_2(\mathbf{A}) \geq \cdots \geq \sigma_m(\mathbf{A}) = \sigma_{\min}(\mathbf{A}) \geq 0$$

The minimum and maximum singular values have the variational characterization

$$\sigma_{\max}(\mathbf{A}) = \max_{\boldsymbol{v}\in\mathbb{S}^{m-1}} \|\mathbf{A}\boldsymbol{v}\|_2 \quad \text{and} \quad \sigma_{\min}(\mathbf{A}) = \min_{\boldsymbol{v}\in\mathbb{S}^{m-1}} \|\mathbf{A}\boldsymbol{v}\|_2,$$

where $\mathbb{S}^{d-1} \triangleq \left\{\boldsymbol{v} \in \mathbb{R}^d \mid \|\boldsymbol{v}\|_2 = 1\right\}$ is the Euclidean unit sphere in $\mathbb{R}^d$.

Definition A.4 (ℓ_2-operator norm)**.** *The spectral or ℓ_2-operator norm of $\mathbf{A}$ is defined as*

$$|||\mathbf{A}|||_2 \triangleq \sigma_{\max}(\mathbf{A}) \,. \tag{8}$$

Since covariance matrices are symmetric, we also focus on the set of symmetric matrices in $\mathbb{R}^d$, denoted $\mathcal{S}^{d\times d} = \left\{\mathbf{Q} \in \mathbb{R}^{d\times d} \mid \mathbf{Q} = \mathbf{Q}^{\mathrm{T}}\right\}$, as well as the subset of positive semidefinite matrices given by

$$\mathcal{S}_+^{d\times d} \triangleq \left\{\mathbf{Q} \in \mathcal{S}^{d\times d} \mid \mathbf{Q} \succeq 0\right\}.$$

From standard linear algebra, we recall the facts that any matrix $\mathbf{Q} \in \mathcal{S}^{d\times d}$ is diagonalizable via a unitary transformation, and we use $\lambda(\mathbf{Q}) \in \mathbb{R}^d$ to denote its vector of eigenvalues, ordered as

$$\lambda_{\max}(\mathbf{Q}) = \lambda_1(\mathbf{Q}) \geq \lambda_2(\mathbf{Q}) \geq \cdots \geq \lambda_d(\mathbf{Q}) = \lambda_{\min}(\mathbf{Q}).$$

Note that a matrix $\mathbf{Q}$ is positive semidefinite-written $\mathbf{Q} \succeq 0$ for short-if and only if $\lambda_{\min}(\mathbf{Q}) \geq 0$.

Remark A.1 (Rayleigh-Ritz variational characterization of eigenvalues)**.** *We remind also the Rayleigh-Ritz variational characterization of the minimum and maximum eigenvalues-namely*

$$\lambda_{\max}(\mathbf{Q}) = \max_{\boldsymbol{v}\in\mathbb{S}^{d-1}} \boldsymbol{v}^\top \mathbf{Q}\boldsymbol{v} \quad \textit{and} \quad \lambda_{\min}(\mathbf{Q}) = \min_{\boldsymbol{v}\in\mathbb{S}^{d-1}} \boldsymbol{v}^\top \mathbf{Q}\boldsymbol{v}.$$

Remark A.2. *For any symmetric matrix $\mathbf{Q}$, the ℓ_2-operator norm can be written as*

$$|||\mathbf{Q}|||_2 = \max\left\{\lambda_{\max}(\mathbf{Q}), |\lambda_{\min}(\mathbf{Q})|\right\},$$

by virtue of which it inherits the variational representation $|||\mathbf{Q}|||_2 = \max_{\boldsymbol{v}\in\mathbb{S}^{d-1}} \left|\boldsymbol{v}^\top \mathbf{Q}\boldsymbol{v}\right|$.

Corollary A.1. *Given a rectangular matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$ with $n \geq m$, suppose that we define the m dimensional symmetric matrix $\mathbf{R} = \mathbf{A}^{\mathrm{T}}\mathbf{A}$. We then have the relationship*

$$\lambda_j(\mathbf{R}) = (\sigma_j(\mathbf{A}))^2 \quad \textit{for } j = 1, \ldots, m$$

We now introduce norms that are induced by positive semi-definite matrices in the following way.

Definition A.5. *For any vector $y \in \mathbb{R}^n$ and matrix $\mathbf{A} \in \mathcal{S}_+^{n\times n}$, let us define the norm $\|\boldsymbol{y}\|_{\mathbf{A}} \triangleq \sqrt{\boldsymbol{y}^T\mathbf{A}\boldsymbol{y}} = \sqrt{\langle \boldsymbol{y}, \mathbf{A}\boldsymbol{y}\rangle}$.*

Throughout the paper we will need often to bound matrix induced norms using ℓ_2-norms for operators, the following results shows how this can be done easily for generic matrices. We specialize this result as we need in the text.

Proposition A.1. *Take $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n\times n}$, with $\mathbf{B} \succeq 0$ positive semi-definite and $\boldsymbol{x} \in \mathbb{R}^n$*

$$\|\mathbf{A}\boldsymbol{x}\|_{\mathbf{B}}^2 = \|\boldsymbol{x}\|_{\mathbf{A}^\top\mathbf{B}\mathbf{A}}^2 \leq |||\mathbf{B}|||_2 |||\mathbf{A}|||_2^2 \|\boldsymbol{x}\|_2^2 \tag{9}$$

Proof. The equality holds since we can rewrite

$$\|\mathbf{A}\boldsymbol{x}\|_{\mathbf{B}}^2 = \langle \mathbf{A}\boldsymbol{x}, \mathbf{B}\mathbf{A}\boldsymbol{x}\rangle = \langle \boldsymbol{x}, \mathbf{A}^\top\mathbf{B}\mathbf{A}\boldsymbol{x}\rangle = \|\boldsymbol{x}\|_{\mathbf{A}^\top\mathbf{B}\mathbf{A}}^2.$$

The inequality follows by the definition of ℓ_2-norms, where we further substitute $\boldsymbol{y} = \mathbf{A}\boldsymbol{x}$, to get

$$\langle \mathbf{A}\boldsymbol{x}, \mathbf{B}\mathbf{A}\boldsymbol{x}\rangle = \frac{\langle \mathbf{A}\boldsymbol{x}, \mathbf{B}\mathbf{A}\boldsymbol{x}\rangle}{\|\mathbf{A}\boldsymbol{x}\|_2^2} \frac{\|\mathbf{A}\boldsymbol{x}\|_2^2}{\|\boldsymbol{x}\|_2^2} \|\boldsymbol{x}\|_2^2 = \frac{\langle \boldsymbol{y}, \mathbf{B}\boldsymbol{y}\rangle}{\|\boldsymbol{y}\|_2^2} \frac{\|\mathbf{A}\boldsymbol{x}\|_2^2}{\|\boldsymbol{x}\|_2^2} \|\boldsymbol{x}\|_2^2 \leq |||\mathbf{B}|||_2 |||\mathbf{A}|||_2^2 \|\boldsymbol{x}\|_2^2.$$

We note that the inequality holds trivially for x in the null space of $\mathbf{A}$, therefore, in the previous case, we can safely divide by $\|\mathbf{A}\boldsymbol{x}\|_2$ and $\|\boldsymbol{x}\|_2$. □

A.3 Technical Lemmas

Lemma A.1 (Determinant-Trace Inequality)**.** *Suppose* $\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_t \in \mathbb{R}^d$ *and for any* $1 \leq s \leq t$, $\|\boldsymbol{z}_s\|_2 \leq L_z$. *Let* $\mathbf{G}_{\boldsymbol{z},t} = \lambda \mathbb{I} + \sum_{s=1}^{t} \boldsymbol{z}_s \boldsymbol{z}_s^\top$ *for some* $\lambda > 0$. *Then,*

$$\det\left(\mathbf{G}_{\boldsymbol{z},t}\right) \leq \left(\lambda + tL_z^2/d\right)^d \tag{10}$$

Proof. Let $\alpha_1, \alpha_2, \ldots, \alpha_d$ be the eigenvalues of $\mathbf{G}_{\boldsymbol{z},t}$. Since $\mathbf{G}_{\boldsymbol{z},t}$ is positive definite, its eigenvalues are positive. Also, note that $\det(\mathbf{G}_{\boldsymbol{z},t}) = \prod_{s=1}^{d} \alpha_s$ and $\operatorname{Tr}\left(\mathbf{G}_{\boldsymbol{z},t}\right) = \sum_{s=1}^{d} \alpha_s$. By inequality of arithmetic and geometric means,

$$\sqrt[d]{\alpha_1 \alpha_2 \cdots \alpha_d} \leq \frac{\alpha_1 + \alpha_2 + \cdots + \alpha_d}{d}.$$

Therefore, $\det\left(\mathbf{G}_{\boldsymbol{z},t}\right) \leq \left(\operatorname{Tr}\left(\mathbf{G}_{\boldsymbol{z},t}\right)/d\right)^d$.

Now, it remains to upper bound the trace:

$$\operatorname{Tr}\left(\mathbf{G}_{\boldsymbol{z},t}\right) = \operatorname{Tr}(\lambda \mathbb{I}_d) + \sum_{s=1}^{t} \operatorname{Tr}\left(\boldsymbol{z}_s \boldsymbol{z}_s^\top\right) = d\lambda + \sum_{s=1}^{t} \|\boldsymbol{z}_s\|_2^2 \leq d\lambda + tL_z^2$$

and the lemma follows. □

B. Concentration of The Minimum Eigenvalue of The Design Matrix

The aim of the section is to find a concentration result for the minimum eigenvalue of the design matrix, which, in turn, gives us a concentration of the ℓ_2-norm of the inverse of the design matrix $\left\|\mathbf{G}_{\boldsymbol{z},t}^{-1}\right\|_2$.

We start by staging two know results that we use in order to derive Lemma B.2. Lemma B.1 is a direct corollary of Weyl's theorem for eigenvalues (see for example Exercise 6.1 in (Wainwright, 2019)).

Lemma B.1. *For two symmetric matrices* $\mathbf{A}$ *and* $\mathbf{B}$

$$|\lambda_{\min}(\mathbf{A}) - \lambda_{\min}(\mathbf{B})| \leq \|\mathbf{A} - \mathbf{B}\|_2. \tag{11}$$

The following, is a classical concentration result for the covariance matrix using the ℓ_2-norm for matrices, for a proof of this result we refer the reader to Corollary 6.20 in (Wainwright, 2019).

Theorem B.1 (Estimation of covariance matrices)**.** *Let* $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_t$ *be i.i.d. zero-mean random vectors with covariance* $\boldsymbol{\Sigma}$ *such that* $\|\boldsymbol{z}_s\|_2 \leq L_{\boldsymbol{z}}$ *almost surely. Then for all* $\delta > 0$*, the sample covariance matrix* $\widehat{\boldsymbol{\Sigma}}_t = \frac{1}{t}\sum_{s=1}^{t} \boldsymbol{z}_s \boldsymbol{z}_s^\top$ *satisfies*

$$\mathbb{P}\left[\left\|\widehat{\boldsymbol{\Sigma}}_t - \boldsymbol{\Sigma}\right\|_2 \geq \delta\right] \leq 2d \exp\left(-\frac{t\delta^2}{(2L_{\boldsymbol{z}}^2\|\boldsymbol{\Sigma}\|_2 + \delta)}\right),$$

this means that with probability at least $1 - \delta$

$$\left\|\widehat{\boldsymbol{\Sigma}}_t - \boldsymbol{\Sigma}\right\|_2 \leq \frac{4L_{\boldsymbol{z}}^2}{t}\log\left(\frac{2d}{\delta}\right) + 2\sqrt{\frac{2L_{\boldsymbol{z}}^2}{t}\log\left(\frac{2d}{\delta}\right)\|\boldsymbol{\Sigma}\|_2}.$$

Now, we use this concentration bound together with the bound on the difference of the minimum eigenvalues of two symmetric matrices in order to bound the maximum eigenvalue of the inverse of the design matrix.

Lemma B.2 (Well-behavedness of First-stage Design Matrix)**.** *Let* $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_t$ *be i.i.d. zero-mean random vectors with covariance* $\boldsymbol{\Sigma}$ *such that* $\|\boldsymbol{z}_s\|_2 \leq L_{\boldsymbol{z}}$ *almost surely. We denote the regularized design matrix as* $\mathbf{G}_{\boldsymbol{z},t} = \lambda \mathbb{I}_d + \sum_{s=1}^{t} \boldsymbol{z}_s \boldsymbol{z}_s^\top$*. For all* $\delta > 0$ *and regularization parameter* $\lambda > 0$*, we observe that*

$$\left\|\mathbf{G}_{\boldsymbol{z},t}^{-1}\right\|_2 = \lambda_{\max}\left(\mathbf{G}_{\boldsymbol{z},t}^{-1}\right) \leq \begin{cases} \frac{1}{\lambda} \text{ if } t \leq C_3 \\ \frac{2}{t\,\lambda_{\min}(\boldsymbol{\Sigma})} \text{ if } t > C_3 \end{cases}.$$

Here, $C_3 > 0$ *is a constant defined by Equation* (13) *and* $\lambda_{\min}(\boldsymbol{\Sigma})$ *is the minimum eigenvalue of the true covariance matrix of* $\boldsymbol{z}$*, i.e.* $\boldsymbol{\Sigma} \triangleq \mathbb{E}[\boldsymbol{z}\boldsymbol{z}^\top]$.

Proof. First, we aim to find a lower bound for the smallest eigenvalue of the design, matrix where we set the regularization parameter λ to zero. We denote the 'non-regularized' design matrix as $\mathbf{G}_{\boldsymbol{z},t}^{\lambda=0}$. For $t \geq 1$, we observe that $\mathbf{G}_{\boldsymbol{z},t}^{\lambda=0}/t \triangleq \widehat{\boldsymbol{\Sigma}}_t$. Thus, by applying Equation (11), we obtain

$$\left|\lambda_{\min}\left(\mathbf{G}_{\boldsymbol{z},t}^{\lambda=0}/t\right) - \lambda_{\min}(\boldsymbol{\Sigma})\right| \leq \frac{4L_{\boldsymbol{z}}^2}{t}\log\left(\frac{2d}{\delta}\right) + 2\sqrt{\frac{2L_{\boldsymbol{z}}^2}{t}\log\left(\frac{2d}{\delta}\right)\|\boldsymbol{\Sigma}\|_2}.$$

Further substituting $A \triangleq 2L_{\boldsymbol{z}}^2 \log\left(\frac{2d}{\delta}\right)$ leads to the following lower bound for the minimum eigenvalue

$$\lambda_{\min}\left(\mathbf{G}_{\boldsymbol{z},t}^{\lambda=0}\right) \geq \max\left\{0, t\left(\lambda_{\min}(\boldsymbol{\Sigma}) - 2A/t - 2\sqrt{A\,\lambda_{\max}(\boldsymbol{\Sigma})/t}\right)\right\}.$$

Here, $\lambda_{\max}(\mathbf{\Sigma})$ and $\lambda_{\min}(\mathbf{\Sigma})$ is the maximum and minimum eigenvalues of the true covariance matrix of $\boldsymbol{z}$, i.e. $\mathbf{\Sigma} \triangleq \mathbb{E}[\boldsymbol{z}\boldsymbol{z}^\top]$. By well-behavedness assumption of the IV, both of them are positive and bounded reals.

Now, from the variational definition of the minimum eigenvalues, we have

$$\lambda_{\min}(\mathbf{G}_{\boldsymbol{z},t}) \geq \lambda_{\min}\left(\mathbf{G}_{\boldsymbol{z},t}^{\lambda=0}\right) + \lambda,$$

which implies that $\lambda_{\min}(\mathbf{G}_{\boldsymbol{z},t}) \geq \lambda$ for all $t \geq 0$, with equality for $t = 0$. Thus, we have

$$\lambda_{\min}(\mathbf{G}_{\boldsymbol{z},t}) \geq \max\left\{\lambda, \lambda + t\left(\lambda_{\min}(\mathbf{\Sigma}) - 2A/t - 2\sqrt{A\,\lambda_{\max}(\mathbf{\Sigma})/t}\right)\right\}. \tag{12}$$

Let us consider the second term inside maximum of Equation (12), and we split it in the following way

$$\lambda + t\,\lambda_{\min}(\mathbf{\Sigma}) - 2A - 2\sqrt{tA\,\lambda_{\max}(\mathbf{\Sigma})} = t\,\underbrace{\frac{\lambda_{\min}(\mathbf{\Sigma})}{2}}_{\text{Term (A)}} + \underbrace{\left(\frac{t}{2}\,\lambda_{\min}(\mathbf{\Sigma}) - 2\sqrt{tA\,\lambda_{\max}(\mathbf{\Sigma})} + \lambda - 2A\right)}_{\text{Term (B)}}$$

Now we study for which values Term (B) is non-negative. The corresponding second order polynomial equation is obtained substituting $u = \sqrt{t}$, and reads

$$u^2\,\lambda_{\min}(\mathbf{\Sigma}) - 4u\sqrt{A\,\lambda_{\max}(\mathbf{\Sigma})} + 2(\lambda - 2A) = 0\,,$$

which has two solutions given by

$$u_\pm = \frac{2\sqrt{A\,\lambda_{\max}(\mathbf{\Sigma})} \pm \sqrt{4A\,\lambda_{\max}(\mathbf{\Sigma}) + 2(2A - \lambda)\,\lambda_{\min}(\mathbf{\Sigma})}}{\lambda_{\min}(\mathbf{\Sigma})}\,.$$

In particular for $t > \lceil u_+ \rceil$, Term (A) ≥ 0, and Equation (12) reads

$$\lambda_{\min}(\mathbf{G}_{\boldsymbol{z},t}) \geq \max\{\lambda, t\,\lambda_{\min}(\mathbf{\Sigma})/2\}\,.$$

Therefore, for $t > \lceil 2\lambda/\,\lambda_{\min}(\mathbf{\Sigma})\rceil$ and $t > \lceil u_+ \rceil$ we have that $\lambda_{\min}(\mathbf{G}_{\boldsymbol{z},t}) \geq t\,\lambda_{\min}(\mathbf{\Sigma})/2$.

Putting the results together, we conclude that

$$\lambda_{\min}(\mathbf{G}_{\boldsymbol{z},t}) \geq t\,\lambda_{\min}(\mathbf{\Sigma})/2 \quad \text{for} \quad t > C_3 \triangleq \max\{\lceil 2\lambda/\,\lambda_{\min}(\mathbf{\Sigma})\rceil, \lceil u_+ \rceil\}\,, \tag{13}$$

while for $t \leq C_3$, we retain the trivial lower bound of the minimum eigenvalue, i.e. λ.

In summary, we have

$$\lambda_{\min}(\mathbf{G}_{\boldsymbol{z},t}) \geq \begin{cases} \lambda \text{ if } t \leq C_3 \\ t\,\lambda_{\min}(\mathbf{\Sigma})/2 \text{ if } t > C_3 \end{cases} \iff \lambda_{\max}\big(\mathbf{G}_{\boldsymbol{z},t}^{-1}\big) \leq \begin{cases} \frac{1}{\lambda} \text{ if } t \leq C_3 \\ \frac{2}{t\,\lambda_{\min}(\mathbf{\Sigma})} \text{ if } t > C_3 \end{cases}.$$

□

C. Technical Lemmas for the Endogeneous Setting

In this section, we present some useful Lemmas that we used in the proofs of the regret bounds of O2SLS and OFUL-IV.

Remark C.1. *In Appendix J, we describe that the first stage regression in* O2SLS *can be expressed as running d independent ridge regressions for each column of $\boldsymbol{\Theta}$ (Equation (27)). Since the standard analysis of each of the ridge regressions assume independent and sub-Gaussian noise added in the linear model (cf. Theorem J.1; (Ouhamma et al., 2021)), we assume that each component of the first stage noise, i.e. $\boldsymbol{\epsilon}_{t,i}$, corresponding to the i-th ridge regression is* sub-Gauss$(\sigma_{\boldsymbol{\epsilon}})$. *Thus, we obtain that $\mathbb{E}\|\boldsymbol{\epsilon}_t\|_2^2 \leq d\sigma_{\boldsymbol{\epsilon}}^2$. We use this result throughout this section.*

Lemma C.1 (Bounding the First-stage Parameters)**.** *Given the relevance condition in Assumption 3.1 and a regularization parameter $\lambda > 0$, we have the following upper bound for the inverse of the estimated parameter (Equation (27)) in the first-stage regression:*

$$\left\|\widehat{\boldsymbol{\Theta}}_t^{-1}\right\|_2 \leq \frac{\lambda + L_{\boldsymbol{z}}^2}{\mathfrak{r}} \triangleq L_{\widehat{\boldsymbol{\Theta}}^{-1}} \tag{14}$$

Proof. By the sub-multiplicativity of the matrix norms we have

$$\left\|\widehat{\boldsymbol{\Theta}}_t^{-1}\right\|_2 = \left\|\left(\mathbf{Z}_t^\top \mathbf{X}_t\right)^{-1}\left(\mathbf{Z}_t^\top \mathbf{Z}_t + \lambda \mathbb{I}\right)\right\|_2 \leq \left\|\left(\mathbf{Z}_t^\top \mathbf{X}_t\right)^{-1}\right\|_2 \left\|\mathbf{Z}_t^\top \mathbf{Z}_t + \lambda \mathbb{I}\right\|_2$$

Then, by the sub-additivity of the norms and the variational definition of the biggest eigenvalue

$$\begin{aligned}
\left\|\mathbf{Z}_t^\top \mathbf{Z}_t + \lambda \mathbb{I}\right\|_2 &\leq \left\|\mathbf{Z}_t^\top \mathbf{Z}_t\right\|_2 + \|\lambda \mathbb{I}\|_2 \leq \max_{\boldsymbol{v} \in \mathbb{S}^{d-1}} \left\langle \boldsymbol{v}, \sum_{s=1}^t \boldsymbol{z}_s \boldsymbol{z}_s^\top \boldsymbol{v} \right\rangle + \lambda \\
&= \max_{\boldsymbol{v} \in \mathbb{S}^{d-1}} \sum_{s=1}^t \langle \boldsymbol{v}, \boldsymbol{z}_s \rangle^2 + \lambda \leq \sum_{s=1}^t \|\boldsymbol{z}_s\|_2^2 + \lambda \\
&\leq t L_{\boldsymbol{z}}^2 + \lambda
\end{aligned}$$

Then, we note that the quantity $\left\|\left(\mathbf{Z}_t^\top \mathbf{X}_t\right)^{-1}\right\|_2 \leq \frac{1}{t\mathfrak{r}}$ by the definition of relevance, which implies

$$\left\|\widehat{\boldsymbol{\Theta}}_t^{-1}\right\|_2 \leq \frac{t L_{\boldsymbol{z}}^2 + \lambda}{t\mathfrak{r}} = \frac{L_{\boldsymbol{z}}^2 + \lambda/t}{\mathfrak{r}} \leq \frac{L_{\boldsymbol{z}}^2 + \lambda}{\mathfrak{r}} \tag{15}$$

□

Lemma C.2 (Bounding the Impact of First-stage Noise)**.** *For first stage noises that is component-wise sub-Gaussian($\sigma_{\boldsymbol{\epsilon}}$), and first-stage parameter estimates satisfying $\left\|\widehat{\boldsymbol{\Theta}}_t^{-1}\right\|_2 \leq L_{\widehat{\boldsymbol{\Theta}}^{-1}}$ (Lemma C.1), we have that*

$$\sum_{t=1}^T \|\boldsymbol{\epsilon}_t\|^2_{\widehat{\boldsymbol{\Theta}}_{t-1}^{-1} \mathbf{G}_{\boldsymbol{z},t-1}^{-1} \widehat{\boldsymbol{\Theta}}_{t-1}^{-\top}} \leq \underbrace{d\sigma_{\boldsymbol{\epsilon}}^2 L^2_{\widehat{\boldsymbol{\Theta}}^{-1}} \left(\frac{C_3 + 1}{\lambda} + 2\frac{\log(T) + 1}{\lambda_{\min}(\boldsymbol{\Sigma})}\right) + C_4}_{\mathcal{O}(d \log T)}$$

with probability at least $1 - \delta$ and with C_4 a constant defined in Equation (17).

Proof. The proof follows using chain of inequalities from Proposition A.1 and Lemma C.1,

$$\begin{aligned}\sum_{t=1}^{T}\|\boldsymbol{\epsilon}_t\|^2_{\widehat{\boldsymbol{\Theta}}_{t-1}^{-1}\mathbf{G}_{\boldsymbol{z},t-1}^{-1}\widehat{\boldsymbol{\Theta}}_{t-1}^{-\top}} &\le \sum_{t=1}^{T}\left\|\mathbf{G}_{\boldsymbol{z},t-1}^{-1}\right\|_2\left\|\widehat{\boldsymbol{\Theta}}_{t-1}^{-1}\right\|_2^2\|\boldsymbol{\epsilon}_t\|_2^2\\ &\le L^2_{\widehat{\boldsymbol{\Theta}}^{-1}}\sum_{t=1}^{T}\left\|\mathbf{G}_{\boldsymbol{z},t-1}^{-1}\right\|_2\left(\left(\|\boldsymbol{\epsilon}_t\|_2^2-\mathbb{E}\,\|\boldsymbol{\epsilon}_t\|_2^2\right)+\mathbb{E}\,\|\boldsymbol{\epsilon}_t\|_2^2\right)\\ &\le \underbrace{L^2_{\widehat{\boldsymbol{\Theta}}^{-1}}\sum_{t=1}^{T}\left\|\mathbf{G}_{\boldsymbol{z},t-1}^{-1}\right\|_2\left(\|\boldsymbol{\epsilon}_t\|_2^2-\mathbb{E}\,\|\boldsymbol{\epsilon}_t\|_2^2\right)}_{\text{I}}+\underbrace{L^2_{\widehat{\boldsymbol{\Theta}}^{-1}}\sum_{t=1}^{T}\left\|\mathbf{G}_{\boldsymbol{z},t-1}^{-1}\right\|_2\mathbb{E}\,\|\boldsymbol{\epsilon}_t\|_2^2}_{\text{II}}\end{aligned}$$

Term I: We now assume $\|\boldsymbol{\epsilon}_t\|_2^2-\mathbb{E}\,\|\boldsymbol{\epsilon}_t\|_2^2\sim\text{sub-exp}(\nu,\alpha)$ where the correct values of ν,α can simply be taken from the results on the square of sub-Gaussians and gives that $\nu\triangleq d4\sqrt{2}\sigma_{\boldsymbol{\epsilon}}^2$ and $\alpha\triangleq 4\sigma_{\boldsymbol{\epsilon}}^2$.

Now, given an $X\sim\text{sub-exp}(\nu,\alpha)$, we have that its rescaling by a constant c is distributed according to $\frac{X}{c}\sim\text{sub-exp}\left(\frac{v}{c},\frac{c}{\alpha}\right)$ which follows by sustituting $\lambda\to\lambda/c$ into the definitions

$$\mathbb{E}\left[e^{\lambda X/c}\right]\le e^{\lambda^2\nu^2/2c^2},\qquad\forall|\lambda/c|\le\frac{1}{\alpha}\Leftrightarrow|\lambda|\le\frac{c}{\alpha}$$

Using this, we can rescale by the factor t in the following

$$\mathbb{P}\left[\sum_{t=m}^{n}\frac{\|\boldsymbol{\epsilon}_t\|_2^2-\mathbb{E}\,\|\boldsymbol{\epsilon}_t\|_2^2}{t}\ge\mu\right]\le\mathbb{E}\left[e^{\lambda\sum_{t=m}^{n}\frac{\|\boldsymbol{\epsilon}_t\|^2-\mathbb{E}\|\epsilon_t\|^2}{t}}\right]e^{-\lambda\mu}\le e^{\sum_{t=m}^{n}\lambda^2\nu^2/2t^2-\lambda\mu}\le e^{\frac{\lambda^2\nu^2}{2}\left(\frac{1}{m-1}-\frac{1}{n}\right)-\lambda\mu}$$

which holds $\forall|\lambda|\le\frac{m}{\alpha}$ thanks to the following series of inequalities $\sum_m^n\frac{1}{t^2}\le\int_{m-1}^n 1/t^2dt=(-\frac{1}{t}|_{m-1}^n=-\frac{1}{n}+\frac{1}{m-1}$. This proves that

$$\sum_{t=C_3}^{T}\frac{\|\varepsilon_t\|_2^2-\mathbb{E}\,\|\varepsilon_t\|_2^2}{t}\sim\text{sub-exp}\left(\nu\sqrt{\frac{1}{C_3-1}-\frac{1}{T}},\frac{C_3}{\alpha}\right)\tag{16}$$

We bound the following summation using Lemma B.2:

$$\begin{aligned}&\sum_{t=1}^{T}\left\|\mathbf{G}_{\boldsymbol{z},t-1}^{-1}\right\|_2\left(\|\boldsymbol{\epsilon}_t\|_2^2-\mathbb{E}\,\|\boldsymbol{\epsilon}_t\|_2^2\right)\\ &\quad=\sum_{t=0}^{C_3}\lambda_{\max}\left(\mathbf{G}_{\boldsymbol{z},t}^{-1}\right)\left(\|\boldsymbol{\epsilon}_{t+1}\|_2^2-\mathbb{E}\,\|\boldsymbol{\epsilon}_{t+1}\|_2^2\right)+\sum_{t=C_3+1}^{T-1}\lambda_{\max}\left(\mathbf{G}_{\boldsymbol{z},t}^{-1}\right)\left(\|\boldsymbol{\epsilon}_{t+1}\|_2^2-\mathbb{E}\,\|\boldsymbol{\epsilon}_{t+1}\|_2^2\right)\\ &\quad\le\frac{1}{\lambda}\sum_{t=0}^{C_3}\left(\|\boldsymbol{\epsilon}_{t+1}\|_2^2-\mathbb{E}\,\|\boldsymbol{\epsilon}_{t+1}\|_2^2\right)+\frac{2}{\lambda_{\min}(\boldsymbol{\Sigma})}\sum_{t=C_3+1}^{T-1}\frac{\|\boldsymbol{\epsilon}_{t+1}\|_2^2-\mathbb{E}\,\|\boldsymbol{\epsilon}_{t+1}\|_2^2}{t}\\ &\quad\le\frac{C_3+1}{\lambda}\left(d4\sqrt{2}\sigma_{\boldsymbol{\epsilon}}^2\sqrt{2\log(1/\delta)}+\frac{1}{2\sigma_{\boldsymbol{\epsilon}}^2}\log(1/\delta)\right)\\ &\qquad\qquad+\frac{2}{\lambda_{\min}(\boldsymbol{\Sigma})}\left(\sqrt{2\nu^2\left(\frac{1}{C_3}-\frac{1}{T}\right)\log(1/\delta)}+\frac{2C_3}{\alpha}\log(1/\delta)\right)\end{aligned}$$

Since term I can be upperbounded by a constant $\mathcal{O}(1)$ we just name this constant C_4 where we also substitute back the definitions $\nu=d4\sqrt{2}\sigma_{\boldsymbol{\epsilon}}^2$ and $\alpha=4\sigma_{\boldsymbol{\epsilon}}^2$:

$$C_4\triangleq L^2_{\widehat{\boldsymbol{\Theta}}^{-1}}\frac{C_3+1}{\lambda}\left(d4\sqrt{2}\sigma_{\boldsymbol{\epsilon}}^2\sqrt{2\log(1/\delta)}+\frac{1}{2\sigma_{\boldsymbol{\epsilon}}^2}\log(1/\delta)\right)$$

$$+ \frac{2L^2_{\widehat{\boldsymbol{\Theta}}^{-1}}}{\lambda_{\min}(\boldsymbol{\Sigma})}\left(d4\sqrt{2}\sigma^2_{\boldsymbol{\epsilon}}\sqrt{\frac{2}{C_3}\log(1/\delta)} + \frac{2C_3}{4\sigma^2_{\boldsymbol{\epsilon}}}\log(1/\delta)\right) \tag{17}$$

Term II: The proof follows using the high probability bound that we introduce in Lemma B.2, plus the estimate $\sum_{k=1}^{n}\frac{1}{k} < \log(n)+1$:

$$\begin{aligned} d\sigma^2_{\boldsymbol{\epsilon}}L^2_{\widehat{\boldsymbol{\Theta}}^{-1}}\sum_{t=1}^{T}\lambda_{\max}\left(\mathbf{G}^{-1}_{\boldsymbol{z},t-1}\right) &\leq dL^2_{\boldsymbol{\epsilon}}L^2_{\widehat{\boldsymbol{\Theta}}^{-1}}\left(\sum_{t=0}^{C_3}\frac{1}{\lambda} + \sum_{t=C_3+1}^{T-1}\frac{2}{\lambda_{\min}(\boldsymbol{\Sigma})t}\right) \\ &\leq dL^2_{\boldsymbol{\epsilon}}L^2_{\widehat{\boldsymbol{\Theta}}^{-1}}\left(\frac{C_3+1}{\lambda} + 2\frac{\log(T)+1}{\lambda_{\min}(\boldsymbol{\Sigma})}\right). \end{aligned}$$

Therefore, putting **Term I** and **Term II** together, we obtain

$$\sum_{t=1}^{T}\|\boldsymbol{\epsilon}_t\|^2_{\widehat{\boldsymbol{\Theta}}_{t-1}^{-1}\mathbf{G}^{-1}_{\boldsymbol{z},t-1}\widehat{\boldsymbol{\Theta}}^{-\top}_{t-1}} \leq \underbrace{d\sigma^2_{\boldsymbol{\epsilon}}L^2_{\widehat{\boldsymbol{\Theta}}^{-1}}\left(\frac{C_3+1}{\lambda} + 2\frac{\log(T)+1}{\lambda_{\min}(\boldsymbol{\Sigma})}\right) + C_4}_{\mathcal{O}(d\log T)}$$

□

Lemma C.3 (Bounding the Sum of Feature Norms)**.** *Under the same conditions of Lemma C.2 plus first-stage parameters with bounded ℓ_2-norm $\|\boldsymbol{\Theta}\|_2 \leq L_{\boldsymbol{\Theta}}$ and bounded IVs $\|\boldsymbol{z}\|^2 \leq L_z^2$ we have that*

$$\sum_{t=1}^{T}\|\boldsymbol{x}_t\|^2_{\widehat{\boldsymbol{\Theta}}_{t-1}^{-1}\mathbf{G}^{-1}_{\boldsymbol{z},t-1}\widehat{\boldsymbol{\Theta}}^{-\top}_{t-1}} \leq \underbrace{L^2_{\widehat{\boldsymbol{\Theta}}^{-1}}\left(L^2_{\boldsymbol{\Theta}}L^2_{\boldsymbol{z}} + d\sigma^2_{\boldsymbol{\epsilon}}\right)\left(\frac{C_3+1}{\lambda} + 2\frac{\log(T)+1}{\lambda_{\min}(\boldsymbol{\Sigma})}\right) + C_4}_{\mathcal{O}(d\log T)}$$

Proof. We start by substituting the First Stage equations inside the norm and using a Triangle Inequality

$$\begin{aligned} \sum_{t=1}^{T}\|\boldsymbol{x}_t\|^2_{\widehat{\boldsymbol{\Theta}}_{t-1}^{-1}\mathbf{G}^{-1}_{\boldsymbol{z},t-1}\widehat{\boldsymbol{\Theta}}^{-\top}_{t-1}} &= \sum_{t=1}^{T}\left\|\boldsymbol{\Theta}^\top\boldsymbol{z}_t + \boldsymbol{\epsilon}_t\right\|^2_{\widehat{\boldsymbol{\Theta}}_{t-1}^{-1}\mathbf{G}^{-1}_{\boldsymbol{z},t-1}\widehat{\boldsymbol{\Theta}}^{-\top}_{t-1}} \\ &\leq \sum_{t=1}^{T}\left\|\boldsymbol{\Theta}^\top\boldsymbol{z}_t\right\|^2_{\widehat{\boldsymbol{\Theta}}_{t-1}^{-1}\mathbf{G}^{-1}_{\boldsymbol{z},t-1}\widehat{\boldsymbol{\Theta}}^{-\top}_{t-1}} + \sum_{t=1}^{T}\|\boldsymbol{\epsilon}_t\|^2_{\widehat{\boldsymbol{\Theta}}_{t-1}^{-1}\mathbf{G}^{-1}_{\boldsymbol{z},t-1}\widehat{\boldsymbol{\Theta}}^{-\top}_{t-1}} \\ &\leq L^2_{\widehat{\boldsymbol{\Theta}}^{-1}}\left(L^2_{\boldsymbol{\Theta}}L^2_{\boldsymbol{z}} + d\sigma^2_{\boldsymbol{\epsilon}}\right)\left(\frac{C_3+1}{\lambda} + 2\frac{\log(T)+1}{\lambda_{\min}(\boldsymbol{\Sigma})}\right) + C_4 \end{aligned}$$

where in the last inequality we used the result of Lemma C.2 for the second term and the following chain of inequality for the following term

$$\begin{aligned} \sum_{t=1}^{T}\left\|\boldsymbol{\Theta}^\top\boldsymbol{z}_t\right\|^2_{\widehat{\boldsymbol{\Theta}}_{t-1}^{-1}\mathbf{G}^{-1}_{\boldsymbol{z},t-1}\widehat{\boldsymbol{\Theta}}^{-\top}_{t-1}} &\leq L^2_{\boldsymbol{\Theta}}L^2_{\boldsymbol{z}}L^2_{\widehat{\boldsymbol{\Theta}}^{-1}}\sum_{t=1}^{T}\lambda_{\max}\left(\mathbf{G}^{-1}_{\boldsymbol{z},t-1}\right) \\ &\leq L^2_{\boldsymbol{\Theta}}L^2_{\boldsymbol{z}}L^2_{\widehat{\boldsymbol{\Theta}}^{-1}}\left(\sum_{t=0}^{C_3}\frac{1}{\lambda} + \sum_{t=C_3+1}^{T-1}\frac{2}{\lambda_{\min}(\boldsymbol{\Sigma})t}\right) \\ &\leq L^2_{\boldsymbol{\Theta}}L^2_{\boldsymbol{z}}L^2_{\widehat{\boldsymbol{\Theta}}^{-1}}\left(\frac{C_3+1}{\lambda} + 2\frac{\log(T)+1}{\lambda_{\min}(\boldsymbol{\Sigma})}\right). \end{aligned}$$

□

D. Elliptical Lemma for the Endogeneous Setting

Lemma D.1 (Confidence Ellipsoid for the Second-stage Parameters)**.** *Let us define the design matrix to be* $\mathbf{G}_{\boldsymbol{z},t} = \mathbf{Z}_t^\top \mathbf{Z}_t + \lambda \mathbb{I}_d$ *for some* $\lambda > 0$ *and* $\widehat{\mathbf{H}}_t \triangleq \widehat{\boldsymbol{\Theta}}_t^\top \mathbf{G}_{\boldsymbol{z},t} \widehat{\boldsymbol{\Theta}}_t$*. Then, for* σ_η*-sub-Gaussian second stage noise* η_t*, the true parameter* $\boldsymbol{\beta}$ *belongs to the set*

$$\mathcal{E}_t = \left\{ \boldsymbol{\beta} \in \mathbb{R}^d : \|\boldsymbol{\beta}_t - \boldsymbol{\beta}\|_{\widehat{\mathbf{H}}_t} \leq \sqrt{\mathfrak{b}_t(\delta)} \right\},$$

with probability at least $1 - \delta \in (0,1)$*, for all* $t \geq 0$*. Here,* $\mathfrak{b}_t(\delta) \triangleq \frac{d\sigma_\eta^2}{4} \log\left(\frac{1 + tL_z^2/\lambda}{\delta}\right)$.

Proof. We can rewrite

$$\boldsymbol{\beta}_t - \boldsymbol{\beta} = \left(\mathbf{Z}_t^\top \mathbf{X}_t\right)^{-1} \mathbf{Z}_t^\top \boldsymbol{\eta}_t = \widehat{\boldsymbol{\Theta}}_t^{-1} \left(\mathbf{Z}_t^\top \mathbf{Z}_t + \lambda \mathbb{I}_d\right)^{-1} \mathbf{Z}_t^\top \boldsymbol{\eta}_t = \widehat{\boldsymbol{\Theta}}_t^{-1} \mathbf{G}_{\boldsymbol{z},t}^{-1} \mathbf{Z}_t^\top \boldsymbol{\eta}_t.$$

Take $\boldsymbol{x} \in \mathbb{R}^d$ and by the Cauchy–Schwarz inequality we have

$$\begin{aligned}
\overbrace{\boldsymbol{x}^\top \boldsymbol{\beta}_t - \boldsymbol{x}^\top \boldsymbol{\beta}}^{\text{(I)}} &= \boldsymbol{x}^\top \widehat{\boldsymbol{\Theta}}_t^{-1} \mathbf{G}_{\boldsymbol{z},t}^{-1} \mathbf{Z}_t^\top \boldsymbol{\eta}_t = \left\langle \widehat{\boldsymbol{\Theta}}_t^{-\top} \boldsymbol{x}, \mathbf{G}_{\boldsymbol{z},t}^{-1} \mathbf{Z}_t^\top \boldsymbol{\eta}_t \right\rangle \\
&\leq \left\| \widehat{\boldsymbol{\Theta}}_t^{-\top} \boldsymbol{x} \right\|_{\mathbf{G}_{\boldsymbol{z},t}^{-1}} \left\| \mathbf{G}_{\boldsymbol{z},t}^{-1} \mathbf{Z}_t^\top \boldsymbol{\eta}_t \right\|_{\mathbf{G}_{\boldsymbol{z},t}} = \underbrace{\left\| \widehat{\boldsymbol{\Theta}}_t^{-\top} \boldsymbol{x} \right\|_{\mathbf{G}_{\boldsymbol{z},t}^{-1}}}_{\text{(II)}} \underbrace{\left\| \mathbf{Z}_t^\top \boldsymbol{\eta}_t \right\|_{\mathbf{G}_{\boldsymbol{z},t}^{-1}}}_{\text{(III)}}.
\end{aligned}$$

The choice we will make for $\boldsymbol{x}$ is the following

$$\boldsymbol{x} \triangleq \widehat{\boldsymbol{\Theta}}_t^\top \mathbf{G}_{\boldsymbol{z},t} \widehat{\boldsymbol{\Theta}}_t (\boldsymbol{\beta}_t - \boldsymbol{\beta}) = \widehat{\mathbf{H}}_t (\boldsymbol{\beta}_t - \boldsymbol{\beta}), \tag{18}$$

which leads to the following rewriting for the three previous terms.

Term (I): From a simple substitution and the definition of a norm induced by a matrix we have

$$\begin{aligned}
\boldsymbol{x}^\top \boldsymbol{\beta}_t - \boldsymbol{x}^\top \boldsymbol{\beta} &= \langle \boldsymbol{x}, \boldsymbol{\beta}_t - \boldsymbol{\beta} \rangle = \left\langle \widehat{\boldsymbol{\Theta}}_t^\top \mathbf{G}_{\boldsymbol{z},t} \widehat{\boldsymbol{\Theta}}_t (\boldsymbol{\beta}_t - \boldsymbol{\beta}), \boldsymbol{\beta}_t - \boldsymbol{\beta} \right\rangle \\
&= \|\boldsymbol{\beta}_t - \boldsymbol{\beta}\|^2_{\widehat{\boldsymbol{\Theta}}_t^\top \mathbf{G}_{\boldsymbol{z},t} \widehat{\boldsymbol{\Theta}}_t}.
\end{aligned}$$

Term (II): First, we rewrite the following term

$$\begin{aligned}
\left\| \widehat{\boldsymbol{\Theta}}_t^{-\top} \boldsymbol{x} \right\|_{\mathbf{G}_{\boldsymbol{z},t}^{-1}} &= \left\langle \widehat{\boldsymbol{\Theta}}_t^{-\top} \boldsymbol{x}, \widehat{\boldsymbol{\Theta}}_t^{-\top} \boldsymbol{x} \right\rangle_{\mathbf{G}_{\boldsymbol{z},t}^{-1}} = \boldsymbol{x}^\top \widehat{\boldsymbol{\Theta}}_t^{-1} \mathbf{G}_{\boldsymbol{z},t}^{-1} \widehat{\boldsymbol{\Theta}}_t^{-\top} \boldsymbol{x} = \|\boldsymbol{x}\|_{\widehat{\boldsymbol{\Theta}}_t^{-1} \mathbf{G}_{\boldsymbol{z},t}^{-1} \widehat{\boldsymbol{\Theta}}_t^{-\top}} \\
&= \|\boldsymbol{x}\|_{\left(\widehat{\boldsymbol{\Theta}}_t^\top \mathbf{G}_{\boldsymbol{z},t} \widehat{\boldsymbol{\Theta}}_t\right)^{-1}},
\end{aligned}$$

and, once again, we substitute the definition of $\boldsymbol{x}$ in Equation (18):

$$\begin{aligned}
\|\boldsymbol{x}\|_{\left(\widehat{\boldsymbol{\Theta}}_t^\top \mathbf{G}_{\boldsymbol{z},t} \widehat{\boldsymbol{\Theta}}_t\right)^{-1}} &= \|\widehat{\boldsymbol{\Theta}}_t^\top \mathbf{G}_{\boldsymbol{z},t} \widehat{\boldsymbol{\Theta}}_t (\boldsymbol{\beta}_t - \boldsymbol{\beta})\|_{\left(\widehat{\boldsymbol{\Theta}}_t^\top \mathbf{G}_{\boldsymbol{z},t} \widehat{\boldsymbol{\Theta}}_t\right)^{-1}} \\
&= \|\boldsymbol{\beta}_t - \boldsymbol{\beta}\|_{\widehat{\boldsymbol{\Theta}}_t^\top \mathbf{G}_{\boldsymbol{z},t} \widehat{\boldsymbol{\Theta}}_t}.
\end{aligned}$$

Terms (III): We bound the last term using Theorem I.1 for the first inequality, and Lemma A.1 in the second inequality:

$$\begin{aligned}
\left\| \mathbf{Z}_t^\top \boldsymbol{\eta}_t \right\|_{\mathbf{G}_{\boldsymbol{z},t}^{-1}} &= \left\| \sum_{s=1}^t \eta_s \boldsymbol{z}_s \right\|_{\mathbf{G}_{\boldsymbol{z},t}^{-1}} \leq \sqrt{2(\sigma_\eta/2)^2 \log\left(\frac{\det\left(\mathbf{G}_{\boldsymbol{z},t}\right)^{1/2} \lambda^{-d/2}}{\delta}\right)} \\
&\leq \sqrt{\frac{d\sigma_\eta^2}{4} \log\left(\frac{1 + tL_z^2/\lambda}{\delta}\right)}.
\end{aligned}$$

Finally, from our initial decomposition, dividing on both sides by $\|\boldsymbol{\beta}_t - \boldsymbol{\beta}\|_{\widehat{\boldsymbol{\Theta}}_t^\top \mathbf{G}_{\boldsymbol{z},t} \widehat{\boldsymbol{\Theta}}_t}$, we get

$$\|\boldsymbol{\beta}_t - \boldsymbol{\beta}\|_{\widehat{\boldsymbol{\Theta}}_t^\top \mathbf{G}_{\boldsymbol{z},t} \widehat{\boldsymbol{\Theta}}_t} \leq \sqrt{\frac{d\sigma_\eta^2}{4} \log\left(\frac{1 + tL_z^2/\lambda}{\delta}\right)}.$$

□

Remark D.1. *We note that the ellipsoid bound has the following order in d and t while neglecting the constants:*

$$\mathfrak{b}_t(\delta) = \mathcal{O}\left(d \log(t)\right)$$

E. A Detailed Discussion on Different Definitions of Regret

In econometrics, the focus has historically been given to the correct identification of the estimator $\boldsymbol{\beta}$. This is because the community has been mainly interested in discovering causal relations and assigning causal meaning to the parameters in the regression. Instead, the primary concern of the statistics and statistical learning community has been arguably on generalization. Interestingly, this tension between these two efforts has not been analyzed until recently since the two communities have worked in two different settings, the endogenous and the exogenous ones, where this conflict is not apparent. In fact, under exogeneity the problem of identifying correctly the true parameter $\boldsymbol{\beta}$ is solved at the same time as the one of having good generalization since the OLS estimator is a consistent estimator. With the exogeneity hypothesis is possible to perform the two tasks of identification and generalization at the same time. This is not true if we introduce the more realistic assumption of endogeneity. In this case, the Minimum Mean Squared Error Estimator (MMSEE) becomes the following,

$$\boldsymbol{\beta}^{\mathsf{MMSEE}} = \mathbb{E}\left[\boldsymbol{x}\boldsymbol{x}^\top\right]^{-1}\mathbb{E}[\boldsymbol{x}y] = \boldsymbol{\beta} + \mathbb{E}\left[\boldsymbol{x}\boldsymbol{x}^\top\right]^{-1}\mathbb{E}[\boldsymbol{x}\eta].$$

This follows by solving for $\boldsymbol{\beta}^*$ in $\partial_{\beta^*}\mathbb{E}_{\boldsymbol{x}y}\left[(y-\boldsymbol{x}\boldsymbol{\beta}^*)^2\right] = 0$, and by substituting for the definition of y in $\mathbb{E}[\boldsymbol{x}y] = \mathbb{E}\left[\boldsymbol{x}\boldsymbol{x}^\top\right]\boldsymbol{\beta}^*$ (provided that we can invert the covariance matrix $\mathbb{E}\left[\boldsymbol{x}\boldsymbol{x}^\top\right]$). Under exogeneity $\mathbb{E}[\boldsymbol{x}\eta] = 0$ and the MMSEE conicedes with $\boldsymbol{\beta}$. Instead, in the more realistic case of endogenous noise $\mathbb{E}[\boldsymbol{x}\eta] \neq 0$ and the MMSEE estimator is biased with respect to the oracle $\boldsymbol{\beta}$. This reveals that if we want to perform well at prediction time, we don't actually want an estimator for the oracle $\boldsymbol{\beta}$, but we want to use an estimator of the MMSE, which can be done in by taking the Empirical Risk Minimizer (ERM). This would lead us to the OLS estimator for regression but without the nice properties of unbiasedness that it has in the exogenous setting. If we don't account for this endogeneity, we end up with a biased estimate that generalizes well but has no causal meaning.

On the contrary, we are instead interested in finding notions of *regret* that preserve the causal interpretation of the estimates, and this naturally leads to extending the *instrumental variables* analysis to the online setting. This motivates the inspection of different notions of regret, which differ from the *population regret* for a time-dependent estimator $\boldsymbol{\beta}_{t-1}$ defined as $R_T(\boldsymbol{\beta}_T) \triangleq \sum_{t=1}^T (y_t - \boldsymbol{x}_t^\top\boldsymbol{\beta}_{t-1})^2 - \min_{\boldsymbol{\beta}}\sum_{t=1}^T(y_t - \boldsymbol{x}_t^\top\boldsymbol{\beta})^2$ where we indicated with $\boldsymbol{\beta}_T = \arg\min_{\boldsymbol{\beta}}\sum_{t=1}^T(y_t-\boldsymbol{x}_t^\top\boldsymbol{\beta})^2$. This is studied in the online learning literature but under the exogeneity assumptions. Therefore, we introduce two alternative regrets that measure the performance of an estimator $\boldsymbol{\beta}_{t-1}$ compared to the oracle. They are the *oracle regret* $\overline{R}_T(\boldsymbol{\beta})$ and the *identification regret* $\widetilde{R}_T(\boldsymbol{\beta})$ below

$$\overline{R}_T(\boldsymbol{\beta}) = \sum_{t=1}^T (y_t - \boldsymbol{x}_t^\top\boldsymbol{\beta}_{t-1})^2 - \sum_{t=1}^T (y_t - \boldsymbol{x}_t^\top\boldsymbol{\beta})^2 \quad \text{and} \quad \widetilde{R}_T(\boldsymbol{\beta}) = \sum_{t=1}^T (\boldsymbol{x}_t^\top\boldsymbol{\beta}_{t-1} - \boldsymbol{x}_t^\top\boldsymbol{\beta})^2.$$

Theorem E.1 shows that an estimator that performs well in terms of *oracle regret* is instead a bad choice for the *population regret*, and the other way around.

Theorem E.1. *For any estimator, the quantity* $\Delta_T \triangleq R_T(\boldsymbol{\beta}_T) - \overline{R}_T(\boldsymbol{\beta})$ *is lower bound in expectation by* $\mathbb{E}[\Delta_T] = \Omega(T)$

Proof Sketch. Solving for $\boldsymbol{\beta}_T$ leads to the OLS estimator for data up to time T: $\boldsymbol{\beta}_T = \boldsymbol{\beta} + \left(\sum_{t=1}^T \boldsymbol{x}_t^\top\boldsymbol{x}_t\right)^{-1}\sum_{t=1}^T \boldsymbol{x}_t^\top\eta_t$ provided that we can invert the design matrix in the previous expression. Using this expression we ca rewrite Δ_T using $\Delta\boldsymbol{\beta}_T \triangleq \boldsymbol{\beta} - \boldsymbol{\beta}_T$ and $G_T \triangleq \sum_{t=1}^T \boldsymbol{x}_t^\top\boldsymbol{x}_t$ as follows

$$\Delta_T = \sum_{t=1}^T \left(\boldsymbol{x}_t^\top\Delta\boldsymbol{\beta}_T\right)^2 + 2\sum_{t=1}^T \eta_t\boldsymbol{x}_t^\top\Delta\boldsymbol{\beta}_T = 3\|\Delta\boldsymbol{\beta}_T\|_{G_T}^2.$$

Thanks to Lemma B.2, we know that the minimum eigenvalue of G_T is $\Omega(T)$ which implies $\|\Delta\boldsymbol{\beta}_T\|_{G_T}^2 \geq \|\Delta\boldsymbol{\beta}_T\|_2^2 \lambda_{\min}(G_T) \gtrsim \|\Delta\boldsymbol{\beta}_T\|_2^2 T$. Furthermore, we can bound away from zero in expectation $\|\Delta\boldsymbol{\beta}_T\|$ by using Cramér–Rao bound on each component of the estimator $\boldsymbol{\beta}_T$, for which we have $\mathbb{E}\left[(\boldsymbol{\beta}_{T,i} - \boldsymbol{\beta}_i)^2\right] \geq \frac{[1+b'(\boldsymbol{\beta}_i)]^2}{I(\boldsymbol{\beta}_i)} + b(\boldsymbol{\beta}_i)^2$ where $b(\boldsymbol{\beta}_i) = \mathbb{E}[\boldsymbol{\beta}_{T,i}] - \boldsymbol{\beta}_i$ is the bias of the estimator and $I(\boldsymbol{\beta}_i)$ is the Fisher Information evaluated at $\boldsymbol{\beta}_i$. We know that $\boldsymbol{\beta}_T$ is a biased estimator of $\boldsymbol{\beta}$ in the endogenous setting, therefore the bias is strictly positive for at least one component, and this concludes the proof.

F. Lemmas on Correlation between First and Second Stages

In the first part of this section, we derive a concentration result for the quantity

$$S_t \triangleq \sum_{s=1}^{t} \Delta\boldsymbol{\beta}_{s-1}^{\top} \left(\boldsymbol{\epsilon}_s \eta_s - \boldsymbol{\gamma}\right),$$

which we call the Martingale Concentration Term in the proof of Theorem 4.2. We can prove that S_t is a martingale adapted to the filtration $\mathcal{F}_t \triangleq \sigma\left(\boldsymbol{\epsilon}_{s:t}, \eta_{1:t}, z_{1:t}\right)$ (or equivalently a sum of martingale difference sequence), by proving that: (a) $\mathbb{E}\left[|S_t|\right] \leq \infty$ and (b) $\mathbb{E}\left[S_{t+1} \mid \mathcal{F}_t\right] = S_t$. The first condition is immediate and the second can be easily verified since:

$$\mathbb{E}\left[S_{t+1} \mid \mathcal{F}_t\right] = \mathbb{E}\left[\Delta\boldsymbol{\beta}_t^{\top}\left(\boldsymbol{\epsilon}_{t_1}\eta_{t+1} - \boldsymbol{\gamma}\right) + \sum_{s=1}^{t} \Delta\boldsymbol{\beta}_{s-1}^{\top}\left(\boldsymbol{\epsilon}_s\eta_s - \boldsymbol{\gamma}\right) \middle| \mathcal{F}_t\right] = 0 + \sum_{s=1}^{t} \Delta\boldsymbol{\beta}_{s-1}^{\top}\left(\boldsymbol{\epsilon}_s\eta_s - \boldsymbol{\gamma}\right) = S_t.$$

Then, the idea is to apply the following theorem on concentration bounds for martingale difference sequences for the first term in eq. (24). In our case the martingale difference sequence is $\{(\Delta\boldsymbol{\beta}_{s-1}^{\top}\left(\boldsymbol{\epsilon}_s\eta_s - \boldsymbol{\gamma}\right), \mathcal{F}_s)\}_{s=1}^{\infty}$.

Theorem F.1 (Concentration Bounds for Martingale Difference Sequences Wainwright (2019))**.** *Let $\{(D_k, \mathcal{F}_k)\}_{k=1}^{\infty}$ be a martingale difference sequence, and suppose that $\mathbb{E}\left[e^{\lambda D_k} \mid \mathcal{F}_{k-1}\right] \leq e^{\lambda^2\nu_k^2/2}$ almost surely for any $|\lambda| < 1/\alpha_k$. Then the following hold:*

1. *The sum $\sum_{k=1}^{n} D_k$ is sub-exponential with parameters $\left(\sqrt{\sum_{k=1}^{n}\nu_k^2}, \alpha_*\right)$, where $\alpha_* := \max_{k=1,\ldots,n} \alpha_k$.*
2. *The sum satisfies the concentration inequality*

$$\mathbb{P}\left[\left|\sum_{k=1}^{n} D_k\right| \geq t\right] \leq \begin{cases} 2e^{-\frac{t^2}{2\sum_{k=1}^{n}\nu_k^2}} & \text{if } 0 \leq t \leq \frac{\sum_{k=1}^{n}\nu_k^2}{\alpha_*}, \\ 2e^{-\frac{1}{2\alpha_*}} & \text{if } t > \frac{\sum_{k=1}^{n}\nu_k^2}{\alpha_*}. \end{cases}$$

To apply the previous theorem, we derive in the following the sub-exponentiality parameters ν_s^*, α_s^* for the martingale difference $D_s \triangleq \Delta\boldsymbol{\beta}_{s-1}^{\top}\left(\boldsymbol{\epsilon}_s\eta_s - \boldsymbol{\gamma}\right)$ such that $\mathbb{E}\left[e^{\lambda\Delta\boldsymbol{\beta}_{s-1}^{\top}(\boldsymbol{\epsilon}_s\eta_s - \boldsymbol{\gamma})} \middle| \mathcal{F}_{s-1}\right] \leq e^{\lambda^2\nu_s^{*2}/2}$ a.s. $\forall|\lambda| < 1/\alpha_s^*$ and then we apply the previous theorem. The bound is derived in the following lemma.

Lemma F.1 (Square and product of non-independent sub-Gaussian random variables)**.** *Given two non-independent random variables $X \sim$sub-Gauss(σ_X) and $Y \sim$sub-Gauss(σ_Y), we can prove the two following things:*

1. *X^2 is sub-exp$\left(4\sqrt{2}\sigma^2, 4\sigma^2\right)$;*
2. *the recentered random variable XY is sub-exp$\left(4\sqrt{2}\left(\sigma_X^2 + \sigma_Y^2\right), 2\left(\sigma_X^2 + \sigma_Y^2\right)\right)$.*

Proof. We prove the two statements in order and we use the first result to prove the second for the case of non independent random variables, which leads to different constants with respect to the result for independent random variables.

1. We start bounding the rescaled p-th power of X

$$\begin{aligned} \mathbb{E}\left[|X/\sigma\sqrt{2}|^p\right] &= \int_0^{\infty} \mathbb{P}\left\{|X/\sigma\sqrt{2}|^p \geq u\right\} du && \text{(integral identity for positive r.v.)} \\ &= \frac{1}{\sqrt{2}\sigma}\int_0^{\infty} \mathbb{P}\left\{|X| \geq t\sqrt{2}\sigma\right\} pt^{p-1}dt && \text{(change of variable } u = t^p\text{)} \\ &\leq \int_0^{\infty} 2e^{-t^2} pt^{p-1}dt && \text{(by } \sigma\text{-sub-Gaussianity)} \end{aligned}$$

$$= p\boldsymbol{\gamma}(p/2) \qquad \text{(set } t^2 = s \text{ and use definition of Gamma function))}$$

By multiplying the previous inequality on both sides by the constant $(\sqrt{2}\sigma)^p$ we obtain $\mathbb{E}\left[|X|^p\right] \leq p2^{\frac{p}{2}}\sigma^p\Gamma(p/2)$.
Now, let $Y = X^2$ and $\mu_Y = \mathbb{E}[Y]$. By power series expansion and since $\Gamma(r) = (r-1)!$ for an integer r, we have:

$$\mathbb{E}\left[e^{\lambda(Y-\mu_Y)}\right] = 1 + \lambda\mathbb{E}\left[Y-\mu_Y\right] + \sum_{r=2}^{\infty}\frac{\lambda^r\mathbb{E}\left[(Y-\mu_Y)^r\right]}{r!} \leq 1 + \sum_{r=2}^{\infty}\frac{\lambda^r\mathbb{E}\left[|X|^{2r}\right]}{r!}$$
$$\leq 1 + \sum_{r=2}^{\infty}\frac{\lambda^r 2r 2^r\sigma^{2r}\Gamma(r)}{r!} = 1 + \sum_{r=2}^{\infty}\lambda^r 2^{r+1}\sigma^{2r} = 1 + \frac{8\lambda^2\sigma^4}{1-2\lambda\sigma^2}$$

By making $|\lambda| \leq 1/\left(4\sigma^2\right)$, we have $1/\left(1-2\lambda\sigma^2\right) \leq 2$. Finally, since $(\forall\alpha)1+\alpha \leq e^{\alpha}$, we have that a sub-Gaussian variable X with parameter σ is sub-exponential with parameters $(4\sqrt{2}\sigma^2, 4\sigma^2)$, in fact we have:

$$\mathbb{E}\left[e^{\lambda\left(X^2-\mathbb{E}[X^2]\right)}\right] \leq e^{16\lambda^2\sigma^4} \quad \forall|\lambda| \leq 1/\left(4\sigma^2\right).$$

2. We notice that

$$XY - \mathbb{E}[XY] = \left(\frac{X-Y}{2}\right)^2 - \mathbb{E}\left[\left(\frac{X-Y}{2}\right)^2\right] - \left(\left(\frac{X+Y}{2}\right)^2 - \mathbb{E}\left[\left(\frac{X+Y}{2}\right)^2\right]\right) \triangleq Z_1 - Z_2$$

where we defined $Z_1 \triangleq \left(\frac{X-Y}{2}\right)^2 - \mathbb{E}\left[\left(\frac{X-Y}{2}\right)^2\right]$ and $Z_2 \triangleq \left(\frac{X+Y}{2}\right)^2 - \mathbb{E}\left[\left(\frac{X+Y}{2}\right)^2\right]$. We have to take into account the dependence, and we start with the sum/difference between X, Y

$$\mathbb{E}\left[e^{\lambda(X+Y)}\right] \leq \sqrt{\mathbb{E}\left[e^{2\lambda X}\right]}\sqrt{\mathbb{E}\left[e^{2\lambda Y}\right]} \leq \sqrt{e^{\frac{4\lambda^2\sigma_X^2}{2}}}\sqrt{e^{\frac{4\lambda^2\sigma_X^2}{2}}} = e^{\frac{\lambda^2}{2}\left[2\left(\sigma_X^2+\sigma_Y^2\right)\right]}$$

where we used Cauchy-Schwarz inequality and the sub-Gaussianity of X and Y. This proves for rescaled variables that both $X+Y$ and $X-Y$ are sub-Gauss$\left(\sqrt{2}\sqrt{\sigma_X^2+\sigma_Y^2}\right)$, therefore their rescaled versions $\frac{X+Y}{2}$ and $\frac{X-Y}{2}$ are sub-Gauss$\left(\frac{\sqrt{\sigma_X^2+\sigma_Y^2}}{\sqrt{2}}\right)$. At this point we use the result on square of sub-Gaussian random variables at point 1, to have that Z_1 and Z_2 are both sub-exp $\left(4\sqrt{2}\frac{\sigma_X^2+\sigma_Y^2}{2}, 4\frac{\sigma_X^2+\sigma_Y^2}{2}\right)$.
Again we use Cauchy-Schwarz due to the dependency between the random variables Z_1 and Z_2 in the first inequality in the next equation

$$\mathbb{E}\left[e^{\lambda(XY-E[XY])}\right] = \mathbb{E}\left[e^{\lambda(Z_1-Z_2)}\right] \leq \sqrt{\mathbb{E}\left[e^{2\lambda Z_1}\right]\mathbb{E}\left[e^{-2\lambda Z_2}\right]} \leq e^{+\frac{4\lambda^2}{2}8\left(\sigma_X^2+\sigma_Y^2\right)^2}$$

which holds for $\lambda \leq \frac{1}{2\left(\sigma_X^2+\sigma_Y^2\right)}$. This proves that the sub-exponential parameters are indeed $\nu \triangleq 4\sqrt{2}\left(\sigma_X^2+\sigma_Y^2\right)$ and $\alpha \triangleq 2\left(\sigma_X^2+\sigma_Y^2\right)$.

□

Lemma F.2 (Concentration of Correlated First and Second-stage Noise)**.** *For sub-Gaussian first- and second-stage noises with parameters σ_η and $\sigma_{\boldsymbol{\epsilon}}$, and first-stage parameter estimates satisfying $\left\|\widehat{\boldsymbol{\Theta}}_t^{-1}\right\|_2 \leq L_{\widehat{\boldsymbol{\Theta}}^{-1}}$ (Lemma C.1), we show that*

$$\left|\sum_{t=1}^{T}\Delta\boldsymbol{\beta}_{t-1}^{\top}\left(\boldsymbol{\epsilon}_t\eta_t-\boldsymbol{\gamma}\right)\right|\leq 8e^2\left(\sigma_\eta^2+\sigma_{\boldsymbol{\epsilon}}^2\right)\sqrt{d}L_{\widehat{\boldsymbol{\Theta}}^{-1}}\sqrt{\mathfrak{b}_{T-1}(\delta)}\left(\sqrt{2\log(2/\delta)\left(\frac{C_3+1}{\lambda}+2\frac{\log(T)+1}{\lambda_{\min}(\boldsymbol{\Sigma})}\right)}\right.$$
$$\left.+\sqrt{\max\left\{\frac{1}{\lambda},\frac{2}{\lambda_{\min}(\boldsymbol{\Sigma})}\right\}\log(2/\delta)}\right)$$

with probability at least $1-\delta\in[0,1)$.

Proof. The proof proceeds in steps and the main technical difficulty arises from the dependence between the random variables, which we tackle using Lemma F.1 plus some techniques derived from the equivalent characterizations of sub-Gaussians and sub-exponential random variables.

1. We bound the p-th moment of the random variable $\mathbf{z}_i\triangleq(\boldsymbol{\epsilon}_{s,i}\eta_s-\boldsymbol{\gamma}_i)$ using the result from Lemma F.1 for the product of two non independent random variables. The random variable $\mathbf{z}_i$ is the centered product of η_s and $\boldsymbol{\epsilon}_{s,i}$ which are sub-Gaussians, therefore it is sub-exp (ν_i,α_i) with $\nu_i\triangleq 4\sqrt{2}\left(\sigma_{\boldsymbol{\epsilon}}^2+\sigma_\eta^2\right),\alpha_i\triangleq 2\left(\sigma_{\boldsymbol{\epsilon}}^2+\sigma_\eta^2\right)$. We can also take $K_5\triangleq\max\left\{\alpha_i,\frac{\nu_i}{\sqrt{2}}\right\}=4\left(\sigma_{\boldsymbol{\epsilon}}^2+\sigma_\eta^2\right)$ which implies $E\left[e^{\lambda\mathbf{z}_i}\right]\leq e^{\lambda^2K_5^2}\quad\forall|\lambda|\leq\frac{1}{K_5}$, and together with the inequality $|x|^p\leq p^p\left(e^x+e^{-x}\right)$ we obtain

$$\mathbb{E}\left[\left|\frac{\mathbf{z}_i}{K_5}\right|^p\right]\leq\mathbb{E}\left[p^p\left(\exp\left(\frac{\mathbf{z}_i}{K_5}\right)+\exp\left(-\frac{\mathbf{z}_i}{K_5}\right)\right)\right]\leq p^p2e^{K_5^2/K_5^2}\leq 2ep^p$$

The previous inequality directly implies an inequality on the p-th norm of the random variable $\boldsymbol{\epsilon}_{s,i}\eta_s-\boldsymbol{\gamma}_i$

$$\sqrt[p]{\mathbb{E}\left[|\boldsymbol{\epsilon}_{s,i}\eta_s-\boldsymbol{\gamma}_i|^p\right]}\leq\sqrt[p]{2e}K_5p\leq 2eK_5p=4e\left(\sigma_{\boldsymbol{\epsilon}}^2+\sigma_\eta^2\right)p\triangleq K_2p$$

2. We finally find the sub-exponential parameters for the scalar product $\Delta\boldsymbol{\beta}_{s-1}^T(\boldsymbol{\epsilon}_s\eta_s-\boldsymbol{\gamma})$. Using the sub-exponential characterization with L^p norm we have $\forall p\geq 1$

$$\left\|\Delta\boldsymbol{\beta}_{s-1}^T\left(\boldsymbol{\epsilon}_s\eta_s-\boldsymbol{\gamma}\right)\right\|_p\leq\sum_{i=1}^{d}\left|\Delta\boldsymbol{\beta}_{s-1,i}\right|\left\|\boldsymbol{\epsilon}_{s,i}\eta_s-\boldsymbol{\gamma}\right\|_p\leq\sum_{i=1}^{d}\left|\Delta\boldsymbol{\beta}_{s-1,i}\right|K_2p=K_sp \tag{19}$$

where $K_s\triangleq\sum_{i=1}^{d}\left|\Delta\boldsymbol{\beta}_{s-1,i}\right|K_2\triangleq 4e\left(\sigma_{\boldsymbol{\epsilon}}^2+\sigma_\eta^2\right)\sum_{i=1}^{d}\left|\Delta\boldsymbol{\beta}_{s-1,i}\right|$ We are ready to bound the moment-generating function and derive the

$$\begin{aligned}
\mathbb{E}\left[e^{\lambda\sum_i\Delta\boldsymbol{\beta}_{s-1,i}(\boldsymbol{\epsilon}_{s,i}\eta_s-\boldsymbol{\gamma}_i)}\mid\mathcal{F}_{s-1}\right]&=\mathbb{E}\left[\sum_{p=0}^{\infty}\frac{\lambda^p\left(\Delta\boldsymbol{\beta}^\top(\boldsymbol{\epsilon}_s\eta_s-\boldsymbol{\gamma})\right)^p}{p!}\right] && \text{(series expansion)}\\
&=\sum_{p=0}^{\infty}\frac{\lambda^p}{p!}\mathbb{E}\left[\left(\Delta\boldsymbol{\beta}^\top(\boldsymbol{\epsilon}_s\eta_s-\boldsymbol{\gamma})\right)^p\right] && \text{(linearity of expectation)}\\
&=1+\sum_{p=2}^{\infty}\frac{\lambda^p\mathbb{E}\left[\left(\Delta\boldsymbol{\beta}^\top(\boldsymbol{\epsilon}_s\eta_s-\boldsymbol{\gamma})\right)^p\right]}{p!} && \text{(1st moment = 0)}\\
&\leq 1+\sum_{p=2}^{\infty}\frac{\lambda^pK_s^pp^p}{p!} && \text{(Equation (19))}\\
&\leq 1+\sum_{p=2}^{\infty}\lambda^pK_s{}^pe^p && \text{(Stirling's approximation } p!\geq\tfrac{p^p}{e^p})
\end{aligned}$$

$$
\begin{aligned}
&= 1 + \frac{(\lambda K_s e)^2}{1 - \lambda K_s e} && \text{(expression valid } \forall |\lambda| \leq \tfrac{1}{K_s e}) \\
&\leq 1 + 2\,(\lambda K_s e)^2 && \text{(valid for } |\lambda| \leq \tfrac{1}{2K_s e}) \\
&\leq e^{2\lambda^2 K_s^2 e^2} && \text{(using } 1 + x \leq e^{x^2} \text{ for all } x \in \mathbb{R})
\end{aligned}
$$

Therefore we have that $\Delta\boldsymbol{\beta}_{s-1}^T(\boldsymbol{\epsilon}_s\eta_s - \boldsymbol{\gamma})$ is sub-exp$(2K_s e,\, 2K_s e)$. We finally substitute for K_s and we obtain

$$
\Delta\boldsymbol{\beta}_{s-1}^T(\boldsymbol{\epsilon}_s\eta_s - \boldsymbol{\gamma}) \sim \text{sub-exp}\left(8e^2\left(\sigma_{\boldsymbol{\epsilon}}^2 + \sigma_\eta^2\right)\left\|\Delta\boldsymbol{\beta}_{s-1}\right\|_1,\, 8e^2\left(\sigma_{\boldsymbol{\epsilon}}^2 + \sigma_\eta^2\right)\left\|\Delta\boldsymbol{\beta}_{s-1}\right\|_1\right)
$$

3. We are now ready to derive the concentration for the sum of the martingale difference sequence using Theorem F.1. From Theorem F.1 we have that $\left|\sum_{t=1}^T D_t\right| \leq \sqrt{2\log(2/\delta)\sum_t \nu_s^2} + 2\alpha^* \log(2/\delta)$ with probability bigger than $1-\delta$. With the following substitutions:

$$
D_t \to \Delta\boldsymbol{\beta}_{t-1}^\top\left(\boldsymbol{\epsilon}_t\eta_t - \boldsymbol{\gamma}\right), \quad \alpha_* \to 8e^2\left(\sigma_{\boldsymbol{\epsilon}}^2 + \sigma_\eta^2\right)\max_s\left\|\Delta\boldsymbol{\beta}_{s-1}\right\|_1, \quad \nu_s \to 8e^2\left(\sigma_{\boldsymbol{\epsilon}}^2 + \sigma_\eta^2\right)\left\|\Delta\boldsymbol{\beta}_{s-1}\right\|_1
$$

we obtain that with probability bigger than $1-\delta$

$$
\begin{aligned}
\left|\sum_{t=1}^T \Delta\boldsymbol{\beta}_{t-1}^\top\left(\boldsymbol{\epsilon}_t\eta_t - \boldsymbol{\gamma}\right)\right| \leq \sqrt{2\log\left(\frac{2}{\delta}\right)(8e^2)^2\left(\sigma_\eta^2 + \sigma_{\boldsymbol{\epsilon}}^2\right)^2\sum_{t=1}^T\left\|\Delta\boldsymbol{\beta}_{t-1}\right\|_1^2} \\
+ 8e^2\left(\sigma_\eta^2 + \sigma_{\boldsymbol{\epsilon}}^2\right)\max_t\left\|\Delta\boldsymbol{\beta}_{t-1}\right\|_1\log(2/\delta)
\end{aligned}
$$

4. We study the term $\left\|\Delta\boldsymbol{\beta}_{t-1}\right\|_1$, its sum and maximum over t, which we need to substitute in the previous concentration bound. We can bound $\left\|\Delta\boldsymbol{\beta}_{t-1}\right\|_1^2 \leq d\left\|\Delta\boldsymbol{\beta}_{t-1}\right\|_2^2$ and then use some matrix tricks in the following for the individual terms

$$
\begin{aligned}
\left\|\Delta\boldsymbol{\beta}_{t-1}\right\|_2^2 &= \left\|\widehat{\mathbf{H}}^{-1/2}\widehat{\mathbf{H}}^{1/2}\Delta\boldsymbol{\beta}_{t-1}\right\|_2^2 \\
&\leq \left\|\widehat{\mathbf{H}}_{t-1}^{-1/2}\right\|_2^2\left\|\Delta\boldsymbol{\beta}_{t-1}\right\|_{\widehat{\mathbf{H}}_{t-1}}^2 && \text{(Proposition A.1)} \\
&\leq \left\|\widehat{\mathbf{H}}_{t-1}^{-1}\right\|_2\mathfrak{b}_{t-1}(\delta) && \text{(Lemma D.1)} \\
&\leq L_{\widehat{\boldsymbol{\Theta}}^{-1}}^2\left\|\mathbf{G}_{\boldsymbol{z},t-1}^{-1}\right\|_2\mathfrak{b}_{T-1}(\delta) && \text{(again Proposition A.1)}
\end{aligned}
$$

When we take the sum over the rounds we have

$$
\begin{aligned}
\sum_{t=1}^T\left\|\Delta\boldsymbol{\beta}_{t-1}\right\|_1^2 \leq d\sum_{t=1}^T\left\|\Delta\boldsymbol{\beta}_{t-1}\right\|_2^2 &\leq dL_{\widehat{\boldsymbol{\Theta}}^{-1}}^2\mathfrak{b}_{T-1}(\delta)\sum_{t=1}^T\left\|\mathbf{G}_{\boldsymbol{z},t-1}^{-1}\right\|_2 && \text{(Proposition A.1)} \\
&\leq dL_{\widehat{\boldsymbol{\Theta}}^{-1}}^2\mathfrak{b}_{T-1}(\delta)\left(\frac{C_3+1}{\lambda} + 2\frac{\log(T)+1}{\lambda_{\min}(\boldsymbol{\Sigma})}\right) && \text{(from Lemma B.2)}
\end{aligned}
$$

For the maximum, we have instead

$$
\max_t\left\|\Delta\boldsymbol{\beta}_{t-1}\right\|_1 \leq \sqrt{d}L_{\widehat{\boldsymbol{\Theta}}^{-1}}\sqrt{\mathfrak{b}_{T-1}(\delta)\|\mathbf{G}_{\boldsymbol{z},t-1}^{-1}\|_2} \leq \sqrt{d}L_{\widehat{\boldsymbol{\Theta}}^{-1}}\sqrt{\mathfrak{b}_{T-1}(\delta)\max\left\{\frac{1}{\lambda}, \frac{2}{\lambda_{\min}(\boldsymbol{\Sigma})}\right\}}
$$

Finally, we can substitute back these expressions in the bound at the previous point. We have that with a probability bigger than $1-\delta$

$$
\left|\sum_{t=1}^T \Delta\boldsymbol{\beta}_{t-1}^\top\left(\boldsymbol{\epsilon}_t\eta_t - \boldsymbol{\gamma}\right)\right| \leq \sqrt{2\log(2/\delta)\,(8e^2)^2\left(\sigma_\eta^2 + \sigma_{\boldsymbol{\epsilon}}^2\right)^2 dL_{\widehat{\boldsymbol{\Theta}}^{-1}}^2\mathfrak{b}_{T-1}(\delta)\left(\frac{C_3+1}{\lambda} + 2\frac{\log(T)+1}{\lambda_{\min}(\boldsymbol{\Sigma})}\right)}
$$

$$+ \left(8e^2\right)\left(\sigma_\eta^2 + \sigma_\epsilon^2\right)\sqrt{d}L_{\widehat{\boldsymbol{\Theta}}^{-1}}\sqrt{\mathfrak{b}_{T-1}(\delta)\max\left\{\frac{1}{\lambda}, \frac{2}{\lambda_{\min}(\boldsymbol{\Sigma})}\right\}\log(2/\delta)}$$

□

Lemma F.3 (Bias of Correlated First and Second-stage Noise)**.** *Under the same hypothesis of previous lemmas we can bound the bias term in the following way*

$$\sum_{t=1}^{T}\Delta\boldsymbol{\beta}_{t-1}^{\top}\boldsymbol{\gamma} \leq \|\boldsymbol{\gamma}\|_2 L_{\widehat{\boldsymbol{\Theta}}^{-1}}\sqrt{\mathfrak{b}_T(\delta)}\left(\frac{C_3}{\sqrt{\lambda}} + 2\frac{\sqrt{2T}}{\sqrt{\lambda_{\min}(\boldsymbol{\Sigma})}}\right)$$

Proof. The bias term has a much simpler analyis

$$\begin{aligned}
\sum_{t=1}^{T}\Delta\boldsymbol{\beta}_{t-1}^{\top}\boldsymbol{\gamma} &\leq \sum_{t=1}^{T}\|\Delta\boldsymbol{\beta}_{t-1}\|_{\widehat{\mathbf{H}}_t}\|\boldsymbol{\gamma}\|_{\widehat{\mathbf{H}}_t^{-1}} && \text{(from Cauchy-Schwarz)}\\
&\leq \sqrt{\mathfrak{b}_{T-1}(\delta)}\sum_{t=1}^{T}\|\boldsymbol{\gamma}\|_{\widehat{\boldsymbol{\Theta}}_t^{-1}\mathbf{G}_{\boldsymbol{z},t}^{-1}\widehat{\boldsymbol{\Theta}}_t^{-\top}} && \text{(from Lemma D.1 and } \mathfrak{b}_{T-1} \text{ increasing)}\\
&\leq \sqrt{\mathfrak{b}_{T-1}(\delta)}\|\boldsymbol{\gamma}\|_2 L_{\widehat{\boldsymbol{\Theta}}^{-1}}\sum_{t=1}^{T}\sqrt{\left\|\mathbf{G}_{\boldsymbol{z},t-1}^{-1}\right\|_2} && \text{(from Proposition A.1)}\\
&\leq \sqrt{\mathfrak{b}_{T-1}(\delta)}\|\boldsymbol{\gamma}\|_2 L_{\widehat{\boldsymbol{\Theta}}^{-1}}\left(\sum_{t=1}^{C_3}\frac{1}{\sqrt{\lambda}} + \sum_{t=C_3+1}^{T-1}\sqrt{\frac{2}{\lambda_{\min}(\boldsymbol{\Sigma})t}}\right) && \text{(from Lemma B.2)}\\
&\leq \sqrt{\mathfrak{b}_{T-1}(\delta)}\|\boldsymbol{\gamma}\|_2 L_{\widehat{\boldsymbol{\Theta}}^{-1}}\left(\frac{C_3}{\sqrt{\lambda}} + 2\frac{\sqrt{2T}}{\sqrt{\lambda_{\min}(\boldsymbol{\Sigma})}}\right)
\end{aligned}$$

where the last inequality follows from $\sum_{k=1}^{n}\frac{1}{\sqrt{k}} = \sum_{k=1}^{n}\int_{k-1}^{k}\frac{dx}{\sqrt{k}} \leq \sum_{k=1}^{n}\int_{k-1}^{k}\frac{dx}{\sqrt{x}} = \int_0^n\frac{dx}{\sqrt{x}} = 2\sqrt{n}$.
□

G. Regret Analysis for IV Regression: O2SLS

In this section, we elaborate on the proofs and techniques to bound the regret of O2SLS.

Remark G.1. *In Appendix J, we describe that the first stage regression in O2SLS can be expressed as running d independent ridge regressions for each column of $\boldsymbol{\Theta}$ (Equation (27)). Since the standard analysis of each of the ridge regressions assume independent and sub-Gaussian noise added in the linear model (cf. Theorem J.1; (Ouhamma et al., 2021)), we assume that each component of the first stage noise, i.e. $\boldsymbol{\epsilon}_{t,i}$, corresponding to the i-th ridge regression is* sub-Gauss$(\sigma_{\boldsymbol{\epsilon}})$. *Thus, we obtain that $\mathbb{E}\|\boldsymbol{\epsilon}_t\|_2^2 \leq d\sigma_{\boldsymbol{\epsilon}}^2$. For the rest of the paper, we use this bound for the expected value of the norm of the first stage noise.*

Theorem G.1 (Identification Regret of O2SLS)**.** *If Assumption 3.1 holds true, then for a first stage noises that is a componentwise sub-Gaussian($\sigma_{\boldsymbol{\epsilon}}$) r.v., a second stage noise that is sub-Gaussian(σ_η), first-stage parameters with bounded ℓ_2-norm $\|\!|\boldsymbol{\Theta}|\!\|_2 \leq L_{\boldsymbol{\Theta}}$, bounded IVs $\|\boldsymbol{z}\|^2 \leq L_z^2$ and a fixed expect value $\mathbb{E}[\eta_s \boldsymbol{\epsilon}_s] \triangleq \boldsymbol{\gamma} \in \mathbb{R}^d$, the regret of O2SLS at step T is*

$$\widetilde{R}_T \leq \underbrace{\mathfrak{b}_{T-1}(\delta)}_{\substack{\textit{Estimation} \\ \mathcal{O}(d\log T)}} \left(\underbrace{L^2_{\widehat{\boldsymbol{\Theta}}^{-1}} \left(L^2_{\boldsymbol{\Theta}} L^2_{\boldsymbol{z}} + d\sigma^2_{\boldsymbol{\epsilon}} \right) \left(\frac{C_3+1}{\lambda} + 2\frac{\log(T)+1}{\lambda_{\min}(\boldsymbol{\Sigma})} \right) + C_4}_{\substack{\textit{Second Stage Feature norm} \\ \mathcal{O}(d\log T)}} \right) = \mathcal{O}\left(d^2 \log^2(T)\right)$$

with probability at least $1-\delta \in [0,1)$. Here, $\mathfrak{b}_{T-1}(\delta)$ is the confidence interval defined by Lemma 4.1, the estimates of the first-stage parameters are bounded according to Equation (14), i.e. $\|\!|\widehat{\boldsymbol{\Theta}}_t^{-1}|\!\|_2 \leq L_{\widehat{\boldsymbol{\Theta}}^{-1}}$, C_3 is defined in Lemma B.2, C_4 in Lemma C.2, $\lambda > 0$ is the regularization parameter of the first stage and $\lambda_{\min}(\boldsymbol{\Sigma})$ is the minimum eigenvalue of the true covariance matrix of IVs, i.e. $\boldsymbol{\Sigma} \triangleq \mathbb{E}[\boldsymbol{z}\boldsymbol{z}^\top]$.

Remark G.2. *In order to shorten the result and clarify the dimension d-dependence and T-dependence in Theorem 4.1, we further define $C_1 \triangleq L_{\widehat{\boldsymbol{\Theta}}^{-1}} L_{\boldsymbol{\Theta}} L_{\boldsymbol{z}}$, $C_2 \triangleq L_{\widehat{\boldsymbol{\Theta}}^{-1}} \sigma_{\boldsymbol{\epsilon}}$. We also define the constants C_3 and C_4 respectively in Equation (13) and Equation (17). We define $C_3' \triangleq C_3 + 1$ and $f(T) \triangleq \left(\frac{C_3'}{\lambda} + \frac{\log(T)+1}{\lambda_{\min}(\boldsymbol{\Sigma})/2} \right)$. Since these constants are d- and T-independent, and $\mathfrak{b}_{T-1}(\delta)$ is $\mathcal{O}\left(d\log(T)\right)$, we obtain that the Identificaiton Regret of O2SLS is $\mathcal{O}\left(d^2\log^2(T)\right)$.*

Proof of Theorem. We bound it (a) using the confidence bound to control the concentration of $\boldsymbol{\beta}_t$ around $\boldsymbol{\beta}$, and (b) by bounding the sum of feature norms according to the following decomposition.

Step 1: By applying Cauchy-Schwarz inequality, we first decouple the effect of parameter estimation and the feature norms

$$\left(\boldsymbol{\beta}_{t-1} - \boldsymbol{\beta}\right)^\top \boldsymbol{x}_t \leq \left\|\boldsymbol{\beta}_{t-1} - \boldsymbol{\beta}\right\|_{\widehat{\mathbf{H}}_t} \|\boldsymbol{x}_t\|_{\widehat{\mathbf{H}}_t^{-1}} \leq \sqrt{\mathfrak{b}_{t-1}(\delta)}\, \|\boldsymbol{x}_t\|_{\left(\widehat{\boldsymbol{\Theta}}_{t-1}^\top \mathbf{G}_{\boldsymbol{z},t-1} \widehat{\boldsymbol{\Theta}}_{t-1}\right)^{-1}} \tag{20}$$

where we used the definition of $\widehat{\mathbf{H}}_t \triangleq \widehat{\boldsymbol{\Theta}}_{t-1}^\top \mathbf{G}_{\boldsymbol{z},t-1} \widehat{\boldsymbol{\Theta}}_{t-1}$ and Lemma D.1. The last inequality holds with probability at least $1-\delta$. Since $\mathfrak{b}_t$ is monotonically increasing in t, by Equation (20),

$$\sum_{t=1}^{T} \left(\left(\boldsymbol{\beta}_{t-1} - \boldsymbol{\beta}\right)^\top \boldsymbol{x}_t \right)^2 \leq \mathfrak{b}_{T-1}(\delta) \sum_{t=1}^{T} \|\boldsymbol{x}_t\|^2_{\left(\widehat{\boldsymbol{\Theta}}_{t-1}^\top \mathbf{G}_{\boldsymbol{z},t-1} \widehat{\boldsymbol{\Theta}}_{t-1}\right)^{-1}}.$$

Step 2: Now, we need to bound the sum of the feature norms, for which we can directly use the result of Lemma C.3

$$\sum_{t=1}^{T} \|\boldsymbol{x}_t\|^2_{\widehat{\boldsymbol{\Theta}}_{t-1}^{-1}\mathbf{G}_{\boldsymbol{z},t-1}^{-1}\widehat{\boldsymbol{\Theta}}_{t-1}^{-\top}} \leq L^2_{\widehat{\boldsymbol{\Theta}}^{-1}}\left(L^2_{\boldsymbol{\Theta}}L^2_{\boldsymbol{z}} + d\sigma^2_{\boldsymbol{\epsilon}}\right)\left(\frac{C_3+1}{\lambda} + 2\frac{\log(T)+1}{\lambda_{\min}(\boldsymbol{\Sigma})}\right) + C_4 \tag{21}$$

Step 3: By combining the results of the previous steps and considering the definition of $\mathfrak{b}_{T-1}(\delta)$, we conclude that we can bound the Identification Regret as follows, and its orders is $\mathcal{O}\left(d^2\log^2(T)\right)$

$$\sum_{t=1}^{T}\left(\left(\boldsymbol{\beta}_{t-1}-\boldsymbol{\beta}\right)^\top \boldsymbol{x}_t\right)^2 \leq \underbrace{\mathfrak{b}_{T-1}(\delta)}_{\mathcal{O}(d\log T)}\left(\underbrace{L^2_{\widehat{\boldsymbol{\Theta}}^{-1}}\left(L^2_{\boldsymbol{\Theta}}L^2_{\boldsymbol{z}} + d\sigma^2_{\boldsymbol{\epsilon}}\right)\left(\frac{C_3+1}{\lambda} + 2\frac{\log(T)+1}{\lambda_{\min}(\boldsymbol{\Sigma})}\right) + C_4}_{\mathcal{O}(d\log T)}\right)$$

□

Theorem G.2 (Oracle Regret of O2SLS)**.** *If Assumption 3.1 hold true, then for first stage noises that is componentwise sub-Gaussian($\sigma_{\boldsymbol{\epsilon}}$) and second stage noise that is sub-Gaussian(σ_η), first-stage parameters with bounded ℓ_2-norm $\|\boldsymbol{\Theta}\|_2 \leq L_{\boldsymbol{\Theta}}$, bounded IVs $\|\boldsymbol{z}\|^2 \leq L_z^2$ and bounded expect value $\mathbb{E}[\eta_s\boldsymbol{\epsilon}_s] \triangleq \boldsymbol{\gamma}$, the regret of O2SLS at step T is*

$$\begin{aligned}
\overline{R}_T \leq\ & \underbrace{\mathfrak{b}_{T-1}(\delta)}_{\substack{\textit{Estimation}\\ \mathcal{O}(d\log T)}}\left(\underbrace{L^2_{\widehat{\boldsymbol{\Theta}}^{-1}}\left(L^2_{\boldsymbol{\Theta}}L^2_{\boldsymbol{z}} + d\sigma^2_{\boldsymbol{\epsilon}}\right)\left(\frac{C_3+1}{\lambda} + 2\frac{\log(T)+1}{\lambda_{\min}(\boldsymbol{\Sigma})} + C_4\right)}_{\substack{\textit{Second Stage Feature norm}\\ \mathcal{O}(d\log T)}}\right)\\
& + \underbrace{\sqrt{\mathfrak{b}_{T-1}(\delta)}}_{\substack{\textit{Estimation}\\ \mathcal{O}(\sqrt{d\log T})}}\left(\underbrace{\sigma_\eta L_{\widehat{\boldsymbol{\Theta}}^{-1}}L_{\boldsymbol{\Theta}}L_{\boldsymbol{z}}\sqrt{\left(\frac{C_3+1}{\lambda} + 2\frac{\log(T)+1}{\lambda_{\min}(\boldsymbol{\Sigma})}\right)}}_{\substack{\textit{First Stage Feature Norm}\\ \mathcal{O}(\sqrt{\log T})}}\right.\\
& \underbrace{+8e^2\left(\sigma_\eta^2+\sigma_{\boldsymbol{\epsilon}}^2\right)\sqrt{d}L_{\widehat{\boldsymbol{\Theta}}^{-1}}\left(\sqrt{2\log\left(\frac{2}{\delta}\right)\left(\frac{C_3+1}{\lambda} + 2\frac{\log(T)+1}{\lambda_{\min}(\boldsymbol{\Sigma})}\right)} + \sqrt{\max\left\{\frac{1}{\lambda}, \frac{2}{\lambda_{\min}(\boldsymbol{\Sigma})}\right\}}\log\frac{2}{\delta}\right)}_{\substack{\textit{Correlated noise}\\ \textit{Concentration Term}\\ \mathcal{O}(\sqrt{d\log T})}}\\
& \left. + \underbrace{\|\boldsymbol{\gamma}\|_2 L_{\widehat{\boldsymbol{\Theta}}^{-1}}\left(\frac{C_3}{\sqrt{\lambda}} + 2\frac{\sqrt{2T}}{\sqrt{\lambda_{\min}(\boldsymbol{\Sigma})}}\right)}_{\substack{\textit{Correlated noise}\\ \textit{Bias Term}\\ \mathcal{O}(\|\boldsymbol{\gamma}\|_2\sqrt{T})}}\right)\\
=\ & \mathcal{O}\left((d\log T)^2 + \|\boldsymbol{\gamma}\|_2\sqrt{dT\log T}\right)
\end{aligned}$$

with probability at least $1-\delta \in [0,1)$. Here, $\mathfrak{b}_{T-1}(\delta)$ is the confidence interval defined by Lemma 4.1, the estimates of the first-stage parameters are bounded according to Equation (14)*,*

i.e. $\|\widehat{\boldsymbol{\Theta}}_t^{-1}\|_2 \leq L_{\widehat{\boldsymbol{\Theta}}^{-1}}$, C_3 *is defined in Lemma B.2,* C_4 *in Lemma C.2,* $\lambda > 0$ *is the regularization parameter of the first stage and* $\lambda_{\min}(\boldsymbol{\Sigma})$ *is the minimum eigenvalue of the true covariance matrix of IVs, i.e.* $\boldsymbol{\Sigma} \triangleq \mathbb{E}[\boldsymbol{z}\boldsymbol{z}^\top]$.

Remark G.3. *In order to shorten the result and clarify the dimension* d*-dependence and* T*-dependence in Theorem 4.2, we further define* $C_1 \triangleq L_{\widehat{\boldsymbol{\Theta}}^{-1}} L_{\boldsymbol{\Theta}} L_{\boldsymbol{z}}$, $C_2 \triangleq L_{\widehat{\boldsymbol{\Theta}}^{-1}} \sigma_{\boldsymbol{\epsilon}}$. *We also define the constants* C_3 *and* C_4 *respectively in Equation* (13) *and Equation* (17). *Finally, the constants* $C_5 \triangleq 8e^2\left(\sigma_\eta^2 + \sigma_{\boldsymbol{\epsilon}}^2\right) L_{\widehat{\boldsymbol{\Theta}}^{-1}} \sqrt{\log(2/\delta)}$, $C_6 \triangleq C_5 \sqrt{\max\left\{\frac{1}{\lambda}, \frac{2}{\lambda_{\min}(\boldsymbol{\Sigma})}\right\}}\log(2/\delta)$, *and* $f(T) \triangleq \left(\frac{C_3'}{\lambda} + \frac{\log(T)+1}{\lambda_{\min}(\boldsymbol{\Sigma})/2}\right)$. *We also define* $\gamma \triangleq \|\boldsymbol{\gamma}\|_2 = \|\mathbb{E}[\eta_s \boldsymbol{\epsilon}_s]\|_2$. *Since these constants are* d*- and* T*-independent, and* $\mathfrak{b}_{T-1}(\delta)$ *is* $\mathcal{O}\left(d\log(T)\right)$, *we obtain that the Identificaiton Regret of* O2SLS *is* $\mathcal{O}\left(d^2\log^2(T)\right)$ *and the Oracle Regret is* $\mathcal{O}(d^2\log^2 T + \gamma\sqrt{dT\log T})$.

Proof of Theorem. By Equation (4), defining $\Delta\boldsymbol{\beta}_{t-1} \triangleq \left(\boldsymbol{\beta}_{t-1} - \boldsymbol{\beta}\right)$, the instantaneous regret at step t is

$$\begin{aligned}
\overline{R}_t &\triangleq \ell_t\left(\boldsymbol{\beta}_{t-1}\right) - \ell_t\left(\boldsymbol{\beta}\right) = \left(y_t - \boldsymbol{\beta}_{t-1}^\top \boldsymbol{x}_t\right)^2 - \left(y_t - \boldsymbol{\beta}^\top \boldsymbol{x}_t\right)^2 \\
&= \left(\left(\boldsymbol{\beta}_{t-1} - \boldsymbol{\beta}\right)^\top \boldsymbol{x}_t - \eta_t\right)^2 - \eta_t^2 = \left(\left(\boldsymbol{\beta}_{t-1} - \boldsymbol{\beta}\right)^\top \boldsymbol{x}_t\right)^2 + 2\eta_t\left(\boldsymbol{\beta}_{t-1} - \boldsymbol{\beta}\right)^\top \boldsymbol{x}_t \\
&= \left(\Delta\boldsymbol{\beta}_{t-1}^\top \boldsymbol{x}_t\right)^2 + 2\eta_t \Delta\boldsymbol{\beta}_{t-1}^\top \boldsymbol{x}_t
\end{aligned}$$

Since $\boldsymbol{x}_t = \boldsymbol{\Theta}^\top \boldsymbol{z}_t + \boldsymbol{\epsilon}_t$ by (First stage), the second term can be rewritten substituting as

$$2\eta_t \Delta\boldsymbol{\beta}_{t-1}^\top \boldsymbol{x}_t = 2\eta_t \Delta\boldsymbol{\beta}_{t-1}^\top \boldsymbol{\Theta}^\top \boldsymbol{z}_t + 2\eta_t \Delta\boldsymbol{\beta}_{t-1}^\top \boldsymbol{\epsilon}_t$$

Therefore, the cumulative regret or regret $\overline{R}_T$ by horizon T is

$$\sum_{t=1}^T \overline{R}_t = \underbrace{\sum_{t=1}^T \left(\Delta\boldsymbol{\beta}_{t-1}^\top \boldsymbol{x}_t\right)^2}_{(\bullet 1 \bullet)} + 2\underbrace{\sum_{t=1}^T \eta_t \Delta\boldsymbol{\beta}_{t-1}^\top \boldsymbol{\Theta}^\top \boldsymbol{z}_t}_{(\bullet 2 \bullet)} + 2\underbrace{\sum_{t=1}^T \eta_t \Delta\boldsymbol{\beta}_{t-1}^\top \boldsymbol{\epsilon}_t}_{(\bullet 3 \bullet)} \tag{22}$$

The proof proceeds by bounding each of the three terms individually.

Term 1: Second-stage Regression Error. The first term $(\bullet 1 \bullet)$ quantifies the error introduced by the second stage regression. This is exactly equal to the Identification Regret that we bounded in Theorem G.1. Therefore we know that it is bounded as

$$(\bullet 1 \bullet) \leq \underbrace{\mathfrak{b}_{T-1}(\delta)}_{\mathcal{O}(d\log T)} \underbrace{L_{\widehat{\boldsymbol{\Theta}}^{-1}}^2 \left(L_{\boldsymbol{\Theta}}^2 L_{\boldsymbol{z}}^2 + d\sigma_{\boldsymbol{\epsilon}}^2\right)\left(\frac{C_3+1}{\lambda} + 2\frac{\log(T)+1}{\lambda_{\min}(\boldsymbol{\Sigma})} + C_4\right)}_{\mathcal{O}(d\log T)} = \mathcal{O}\left(d^2\log^2(T)\right)$$

Term 2: Coupling of First-stage Data and Second-stage Parameter Estimation. Now, we bound $(\bullet 2 \bullet)$ using martingale inequalities similar to the ones used for the confidence intervals to derive a uniform high probability bound.

Step 1: Following Theorem I.2, we define

$$w_s \triangleq \left(\boldsymbol{\beta}_{s-1} - \boldsymbol{\beta}\right)^\top \boldsymbol{\Theta}^\top \boldsymbol{z}_s \quad \text{and} \quad \mathcal{F}_{t-1} \triangleq \sigma\left(\boldsymbol{z}_1, \boldsymbol{\epsilon}_1, \eta_1, \ldots, \boldsymbol{z}_{t-1}, \boldsymbol{\epsilon}_{t-1}, \eta_{t-1}, \boldsymbol{z}_t\right).$$

It is immediate to verify that the hypothesis are satisfied, since w_t is $\mathcal{F}_{t-1}$-measurable as $\boldsymbol{\beta}_{t-1}$ and $\boldsymbol{z}_t$ are too. Bearing in mind this substitution we have

$$\left|\sum_{t=1}^{T}\eta_t w_t\right| \leq \sqrt{2\left(1+\sigma_\eta^2\sum_{t=1}^{T}w_t^2\right)\log\left(\frac{\sqrt{1+\sigma_\eta^2\sum_{t=1}^{T}w_t^2}}{\delta}\right)} \tag{23}$$

with probability at least $1-\delta$.

Step 2: Thus, we proceed like for the first term in the *Step 2* for **Term 1**:

$$\begin{aligned}
\sum_{t=1}^{T}\left\langle \boldsymbol{\beta}_{t-1}-\boldsymbol{\beta}, \boldsymbol{\Theta}^\top \boldsymbol{z}_t\right\rangle^2 &\leq \mathfrak{b}_{T-1}(\delta)\sum_{t=1}^{T}\left\|\boldsymbol{\Theta}^\top \boldsymbol{z}_t\right\|^2_{\left(\widehat{\boldsymbol{\Theta}}_{t-1}^\top \mathbf{G}_{\boldsymbol{z},T-1}\widehat{\boldsymbol{\Theta}}_{t-1}\right)^{-1}} \\
&\qquad\qquad \text{(Cauchy-Schwarz and Lemma D.1)} \\
&\leq \mathfrak{b}_{T-1}(\delta)\sum_{t=1}^{T}\left\|\mathbf{G}_{\boldsymbol{z},t-1}^{-1}\right\|_2\left\|\widehat{\boldsymbol{\Theta}}_{t-1}^{-1}\right\|_2^2\left\|\boldsymbol{\Theta}^\top \boldsymbol{z}_t\right\|_2^2 \qquad \text{(Proposition A.1)} \\
&\leq \mathfrak{b}_{T-1}(\delta)L^2_{\widehat{\boldsymbol{\Theta}}^{-1}}\sum_{t=1}^{T}\left\|\mathbf{G}_{\boldsymbol{z},t-1}^{-1}\right\|_2\|\boldsymbol{\Theta}\|_2^2\|\boldsymbol{z}_t\|_2^2 \\
&\qquad\qquad \text{(Proposition A.1 and definition of } L^2_{\widehat{\boldsymbol{\Theta}}^{-1}}) \\
&\leq \mathfrak{b}_{T-1}(\delta)L^2_{\widehat{\boldsymbol{\Theta}}^{-1}}L^2_{\boldsymbol{\Theta}}L^2_{\boldsymbol{z}}\sum_{t=1}^{T}\left\|\mathbf{G}_{\boldsymbol{z},t-1}^{-1}\right\|_2 \\
&\qquad\qquad \text{(boundedness of } \boldsymbol{z}_t \text{ and definition of } L_{\boldsymbol{\Theta}}) \\
&= \underbrace{\mathfrak{b}_{T-1}(\delta)}_{\mathcal{O}(d\log T)}\underbrace{L^2_{\widehat{\boldsymbol{\Theta}}^{-1}}L^2_{\boldsymbol{\Theta}}L^2_{\boldsymbol{z}}\left(\frac{C_3+1}{\lambda}+2\frac{\log(T)+1}{\lambda_{\min}(\boldsymbol{\Sigma})}\right)}_{\mathcal{O}(\log T)} \qquad \text{(from Lemma B.2)} \\
&= \mathcal{O}\left(d(\log T)^2\right)
\end{aligned}$$

where in the first inequality we also used the fact that $\mathfrak{b}_{t-1}(\delta)$ is monotonically increasing in t, to take the radii outside the summation.

Step 3: Thus, substituting inside Equation (23), the order of (•2•) is negligible with respect to term (•1•)

$$\begin{aligned}
2\sum_{t=1}^{T}\eta_t(\boldsymbol{\beta}_{t-1}-\boldsymbol{\beta})^\top\boldsymbol{\Theta}^\top\boldsymbol{z}_t &\leq \underbrace{\sqrt{2\left(1+\mathfrak{b}_{T-1}(\delta)\sigma_\eta^2L^2_{\widehat{\boldsymbol{\Theta}}^{-1}}L^2_{\boldsymbol{\Theta}}L^2_{\boldsymbol{z}}\left(\frac{C_3+1}{\lambda}+2\frac{\log(T)+1}{\lambda_{\min}(\boldsymbol{\Sigma})}\right)\right)}}_{\mathcal{O}(\sqrt{d}\log T)} \\
&\qquad \underbrace{\sqrt{\log\left(\sqrt{1+\mathfrak{b}_{T-1}(\delta)\sigma_\eta^2L^2_{\widehat{\boldsymbol{\Theta}}^{-1}}L^2_{\boldsymbol{\Theta}}L^2_{\boldsymbol{z}}\left(\frac{C_3+1}{\lambda}+2\frac{\log(T)+1}{\lambda_{\min}(\boldsymbol{\Sigma})}\right)}\Big/\delta\right)}}_{\mathcal{O}\left(\sqrt{\log(\sqrt{d}\log T)}\right)} \\
&= \mathcal{O}\left(\sqrt{d}\log T\sqrt{\log(\sqrt{d}\log T)}\right)
\end{aligned}$$

with probability at least $1-\delta$.

Term 3: Coupling of First- and Second-stage Noises. We bound the term (•3•) containing the self-fulfilling bias, i.e. the correlation between the first- and second-stage noise, by splitting it into two contributions.

$$\begin{aligned}\sum_{t=1}^{T}\eta_t(\underbrace{\boldsymbol{\beta}_{t-1}-\boldsymbol{\beta}}_{\Delta\boldsymbol{\beta}_{t-1}})^\top\boldsymbol{\epsilon}_t &= \sum_{t=1}^{T}\Delta\boldsymbol{\beta}_{t-1}^\top(\boldsymbol{\epsilon}_t\eta_t)+\sum_{t=1}^{T}\Delta\boldsymbol{\beta}_{t-1}^\top\mathbb{E}_t[\boldsymbol{\epsilon}_t\eta_t]-\sum_{t=1}^{T}\Delta\boldsymbol{\beta}_{t-1}^\top\mathbb{E}_t[\boldsymbol{\epsilon}_t\eta_t]\\ &= \sum_{t=1}^{T}\Delta\boldsymbol{\beta}_{t-1}^\top\boldsymbol{\epsilon}_t\eta_t-\sum_{t=1}^{T}\Delta\boldsymbol{\beta}_{t-1}^\top\boldsymbol{\gamma}+\sum_{t=1}^{T}\Delta\boldsymbol{\beta}_{t-1}^\top\boldsymbol{\gamma}\\ &= \underbrace{\sum_{t=1}^{T}\Delta\boldsymbol{\beta}_{t-1}^T(\boldsymbol{\epsilon}_t n_t-\boldsymbol{\gamma})}_{\substack{\text{Martingale Concentration Term}\\ \mathcal{O}(\log(T))}}+\underbrace{\sum_{t=1}^{T}\Delta\boldsymbol{\beta}_{t-1}^\top\boldsymbol{\gamma}}_{\substack{\text{Bias Term}\\ \mathcal{O}(\sqrt{T})}}\end{aligned}\tag{24}$$

Now we can use Lemma F.2 and Lemma F.3 to conclude that term (•3•) is bounded by the following quantity

$$\underbrace{8e^2\left(\sigma_\eta^2+\sigma_{\boldsymbol{\epsilon}}^2\right)L_{\widehat{\boldsymbol{\Theta}}^{-1}}\sqrt{d\mathfrak{b}_{T-1}(\delta)}\left(\sqrt{2\log(\frac{2}{\delta})\left(\frac{C_3+1}{\lambda}+2\frac{\log(T)+1}{\lambda_{\min}(\boldsymbol{\Sigma})}\right)}+\sqrt{\max\left\{\frac{1}{\lambda},\frac{2}{\lambda_{\min}(\boldsymbol{\Sigma})}\right\}\log\frac{2}{\delta}}\right)}_{d\log(T)}$$
$$+\underbrace{\sqrt{\mathfrak{b}_{T-1}(\delta)}\|\boldsymbol{\gamma}\|_2 L_{\widehat{\boldsymbol{\Theta}}^{-1}}\left(\frac{C_3}{\sqrt{\lambda}}+2\frac{\sqrt{2T}}{\sqrt{\lambda_{\min}(\boldsymbol{\Sigma})}}\right)}_{\mathcal{O}(\|\boldsymbol{\gamma}\|_2\sqrt{dT\log T})}$$

□

H. Regret Analysis for IV Linear Bandits: OFUL-IV

Theorem H.1. *Under the same assumptions as that of Theorem 4.2, Algorithm 2 incurs a regret*

$$R_T = \sum_{t=1}^{T} r_t \leq 2\sqrt{T}\sqrt{\mathfrak{b}_{T-1}(\delta)}\sqrt{L^2_{\widehat{\boldsymbol{\Theta}}^{-1}}\left(L^2_{\boldsymbol{\Theta}}L^2_{\boldsymbol{z}} + d\sigma^2_{\boldsymbol{\epsilon}}\right)\left(\frac{C_3+1}{\lambda} + 2\frac{\log(T)+1}{\lambda_{\min}(\boldsymbol{\Sigma})}\right) + C_4}$$

with probability $1-\delta$*, horizon* $T > 1$ *and* C_4 *is defined in eq.* (17)*.*

Proof. The instantaneous regret r_t reads

$$\begin{aligned}
\langle \boldsymbol{\beta}, \boldsymbol{x}_* \rangle - \langle \boldsymbol{\beta}, \boldsymbol{x}_t \rangle &\leq \left\langle \widetilde{\boldsymbol{\beta}}_{t-1}, \boldsymbol{x}_t \right\rangle - \langle \boldsymbol{\beta}, \boldsymbol{x}_t \rangle && \text{(since } (\boldsymbol{x}_t, \widetilde{\boldsymbol{\beta}}_{t-1}) \text{ is optimistic inside } \mathcal{X}_t \times \mathcal{B}_t) \\
&= \langle \boldsymbol{\beta}_{t-1} - \boldsymbol{\beta}, \boldsymbol{x}_t \rangle + \left\langle \widetilde{\boldsymbol{\beta}}_{t-1} - \boldsymbol{\beta}_{t-1}, \boldsymbol{x}_t \right\rangle && \text{(summing and subtracting } \boldsymbol{\beta}_{t-1}) \\
&\leq \left\|\boldsymbol{\beta}_{t-1} - \boldsymbol{\beta}\right\|_{\widehat{\boldsymbol{\Theta}}_{t-1}^\top \mathbf{G}_{\boldsymbol{z},t-1} \widehat{\boldsymbol{\Theta}}_{t-1}} \left\|\boldsymbol{x}_t\right\|_{\left(\widehat{\boldsymbol{\Theta}}_{t-1}^\top \mathbf{G}_{\boldsymbol{z},t-1} \widehat{\boldsymbol{\Theta}}_{t-1}\right)^{-1}} \\
&\qquad + \left\|\widetilde{\boldsymbol{\beta}}_{t-1} - \boldsymbol{\beta}_{t-1}\right\|_{\widehat{\boldsymbol{\Theta}}_{t-1}^\top \mathbf{G}_{\boldsymbol{z},t-1} \widehat{\boldsymbol{\Theta}}_{t-1}} \left\|\boldsymbol{x}_t\right\|_{\left(\widehat{\boldsymbol{\Theta}}_{t-1}^\top \mathbf{G}_{\boldsymbol{z},t-1} \widehat{\boldsymbol{\Theta}}_{t-1}\right)^{-1}} \\
&&& \text{(Cauchy-schwarz ineq.)} \\
&\leq 2\sqrt{\mathfrak{b}_{t-1}(\delta)} \left\|\boldsymbol{x}_t\right\|_{\left(\widehat{\boldsymbol{\Theta}}_{t-1}^\top \mathbf{G}_{\boldsymbol{z},t-1} \widehat{\boldsymbol{\Theta}}_{t-1}\right)^{-1}} && \text{(Lemma D.1)}
\end{aligned}$$

The last inequality uses the concentration of $\boldsymbol{\beta}_t$ around the true value $\boldsymbol{\beta}$, and the fact that we choose $\widetilde{\boldsymbol{\beta}}_{t-1}$ inside $\mathcal{B}_{t-1}$. In both cases, the two norms are bounded by the radius of the ellipsoid, i.e. $\sqrt{\mathfrak{b}_{t-1}(\delta)}$. Since we already know in this case how to concentrate the sum of the features norms $\sum_{t=1}^{T} \|\boldsymbol{x}_t\|^2_{\widehat{\boldsymbol{\Theta}}_{t-1}^{-1} \mathbf{G}_{\boldsymbol{z},t-1}^{-1} \widehat{\boldsymbol{\Theta}}_{t-1}^{-\top}}$ from Lemma C.3, we bound the cumulative regret using Cauchy-Schwarz inequality in the first inequality below, and we substitute the bound on the instantaneous regret that we just obtained:

$$\begin{aligned}
R_T &\leq \sqrt{T\sum_{t=1}^{T} r_t^2} \leq 2\sqrt{T\sum_{t=1}^{T} \mathfrak{b}_{t-1}(\delta) \left\|\boldsymbol{x}_t\right\|^2_{\widehat{\boldsymbol{\Theta}}_{t-1}^{-1} \mathbf{G}_{\boldsymbol{z},t-1}^{-1} \widehat{\boldsymbol{\Theta}}_{t-1}^{-\top}}} \\
&\leq 2\sqrt{T}\sqrt{\mathfrak{b}_{T-1}(\delta)}\sqrt{\sum_{t=1}^{T} \left\|\boldsymbol{x}_t\right\|^2_{\widehat{\boldsymbol{\Theta}}_{t-1}^{-1} \mathbf{G}_{\boldsymbol{z},t-1}^{-1} \widehat{\boldsymbol{\Theta}}_{t-1}^{-\top}}}
\end{aligned} \tag{25}$$

where in the second inequality, we used the fact that the radius $\mathfrak{b}_{t-1}(\delta)$ is monotonically increasing in t. Now, we can use Lemma C.3 to bound the sum of feature norms, and putting all together, we obtain the following bound:

$$R_T \leq 2\sqrt{T}\underbrace{\sqrt{\mathfrak{b}_{T-1}(\delta)}}_{\mathcal{O}(\sqrt{d\log T})}\underbrace{\sqrt{L^2_{\widehat{\boldsymbol{\Theta}}^{-1}}\left(L^2_{\boldsymbol{\Theta}}L^2_{\boldsymbol{z}} + d\sigma^2_{\boldsymbol{\epsilon}}\right)\left(\frac{C_3+1}{\lambda} + 2\frac{\log(T)+1}{\lambda_{\min}(\boldsymbol{\Sigma})}\right) + C_4}}_{\mathcal{O}(\sqrt{d\log T})} = \mathcal{O}\left(d\sqrt{T}\log T\right).$$

□

I. Concentration of Scalar and Vector-valued Martingales

We look for deviations of the vector martingales $\sum_{s=1}^{t} \eta_s \boldsymbol{z}_s$ and the scalar valued martingale

$$\sum_{t=1}^{T} \eta_t \left(\boldsymbol{\beta}_{t-1} - \boldsymbol{\beta}\right)^\top \boldsymbol{\Theta}^\top \boldsymbol{z}_t$$

from their expected values. These results are required in the proof of Theorem G.2. The first martingale is vector-valued while the second is scalar-valued. For the vector-valued martingale, we want to bound its deviations when its values are weighted by the inverse of its design matrix $\mathbf{G}_{\boldsymbol{z},t}^{-1}$ like it appears in $\|\boldsymbol{s}_t\|^2_{\mathbf{G}_{\boldsymbol{z},t}^{-1}}$. The design matrix $\mathbf{G}_{\boldsymbol{z},t}^{-1}$ is itself derived from the martingale. Hence, it is called the 'self-normalized bound'. The following theorems were introduced in (Abbasi-Yadkori et al., 2011a, 2012) for the two cases. We state and prove them here for completeness. We leverage the fact that the first and second stage noises are sub-Gaussian random variables.

Lemma I.1. *Let $\boldsymbol{\mu} \in \mathbb{R}^d$ be arbitrary and consider for any $t \geq 0$*

$$m_t^{\boldsymbol{\mu}} \triangleq \prod_{s=1}^{t} \exp\left(\frac{\eta_s \langle \boldsymbol{\mu}, \boldsymbol{z}_s \rangle}{\sigma_2} - \frac{1}{2} \langle \boldsymbol{\mu}, \boldsymbol{z}_s \rangle^2\right).$$

Let τ be a stopping time with respect to the filtration $\{\mathcal{F}_t\}_{t=0}^{\infty}$. Then, $m_\tau^{\boldsymbol{\mu}}$ is a.s. well-defined and $\mathbb{E}\left[m_\tau^{\boldsymbol{\mu}}\right] \leq 1$.

Proof. We claim that $\left\{m_t^{\boldsymbol{\mu}}\right\}_{t=0}^{\infty}$ is a supermartingale. Let

$$d_t^{\boldsymbol{\mu}} \triangleq \exp\left(\frac{\eta_t \langle \boldsymbol{\mu}, \boldsymbol{z}_t \rangle}{\sigma_2} - \frac{1}{2} \langle \boldsymbol{\mu}, \boldsymbol{z}_t \rangle^2\right).$$

Observe that by conditional R-sub-Gaussianity of η_t we have $\mathbb{E}\left[d_t^{\boldsymbol{\mu}} \mid \mathcal{F}_{t-1}\right] \leq 1$. Clearly, $d_t^{\boldsymbol{\mu}}$ is $\mathcal{F}_t$-measurable, as is $m_t^{\boldsymbol{\mu}}$. Further,

$$\mathbb{E}\left[m_t^{\boldsymbol{\mu}} \mid \mathcal{F}_{t-1}\right] = \mathbb{E}\left[m_1^{\boldsymbol{\mu}} \cdots d_{t-1}^{\boldsymbol{\mu}} d_t^{\boldsymbol{\mu}} \mid \mathcal{F}_{t-1}\right] = d_1^{\boldsymbol{\mu}} \cdots d_{t-1}^{\boldsymbol{\mu}} \mathbb{E}\left[d_t^{\boldsymbol{\mu}} \mid \mathcal{F}_{t-1}\right] \leq m_{t-1}^{\boldsymbol{\mu}},$$

showing that $\left\{m_t^{\boldsymbol{\mu}}\right\}_{t=0}^{\infty}$ is indeed a supermartingale and in fact $\mathbb{E}\left[m_t^{\boldsymbol{\mu}}\right] \leq 1$.

Now, we argue that $m_\tau^{\boldsymbol{\mu}}$ is well-defined. By the convergence theorem for nonnegative supermartingales, $M_\infty^{\boldsymbol{\mu}} = \lim_{t\to\infty} m_t^{\boldsymbol{\mu}}$ is almost surely well-defined. Hence, $m_\tau^{\boldsymbol{\mu}}$ is indeed well-defined independently of whether $\tau < \infty$ holds or not. Next, we show that $\mathbb{E}\left[m_\tau^{\boldsymbol{\mu}}\right] \leq 1$. For this let $Q_t^{\boldsymbol{\mu}} = M_{\min\{\tau,t\}}^{\boldsymbol{\mu}}$ be a stopped version of $\left(m_t^{\boldsymbol{\mu}}\right)_t$. By Fatou's Lemma, $\mathbb{E}\left[m_\tau^{\boldsymbol{\mu}}\right] = \mathbb{E}\left[\liminf_{t\to\infty} Q_t^{\boldsymbol{\mu}}\right] \leq \liminf_{t\to\infty} \mathbb{E}\left[Q_t^{\boldsymbol{\mu}}\right] \leq 1$, showing that $\mathbb{E}\left[m_\tau^{\boldsymbol{\mu}}\right] \leq 1$ indeed holds. □

Next lemma uses the "method of mixtures" technique, (Lattimore and Szepesvári, 2020) Chapter 20.

Lemma I.2. *Let $\{\mathcal{F}_t\}_{t=0}^{\infty}$ be a filtration. Let τ be a stopping time with respect to the filtration $\{\mathcal{F}_t\}_{t=0}^{\infty}$. Then, for any $\delta > 0$, with probability $1 - \delta$*

$$\|\boldsymbol{s}_\tau\|^2_{\mathbf{G}_{\boldsymbol{z},\tau}^{-1}} \leq 2\sigma_2^2 \log\left(\frac{\det\left(\mathbf{G}_{\boldsymbol{z},\tau}\right)^{1/2} \lambda^{-d/2}}{\delta}\right).$$

Proof. We decompose $\mathbf{G}_{\boldsymbol{z},t}$ according to the following notation in order to ease the notation

$$\mathbf{G}_{\boldsymbol{z},t} \triangleq \lambda \mathbb{I}_d + \sum_{s=1}^{t} \boldsymbol{z}_s \boldsymbol{z}_s^\top = \mathbf{V} + \mathbf{V}_t$$

where $\mathbf{V}$ and $\mathbf{V}_t$ are defined by $\mathbf{V} \triangleq \lambda \mathbb{I}_d$ and $\mathbf{V}_t \triangleq \sum_{s=1}^{t} \boldsymbol{z}_s \boldsymbol{z}_s^\top$. We can rewrite for example $m_t^{\boldsymbol{\mu}}$ as follows $m_t^{\boldsymbol{\mu}} = \exp\left(\frac{\langle \boldsymbol{\mu}, \boldsymbol{s}_t \rangle}{\sigma_2} - \frac{1}{2}\|\boldsymbol{\mu}\|_{\mathbf{V}_t}^2\right)$.

Let $\boldsymbol{\mu}$ be a Gaussian random variable which is independent of all the other random variables and whose covariance is $\mathbf{V}^{-1}$. Define

$$m_t \triangleq \mathbb{E}\left[m_t^{\boldsymbol{\mu}} \mid \mathcal{F}_\infty\right],$$

where $\mathcal{F}_\infty$ is the tail σ-algebra of the filtration i.e. the σ-algebra generated by the union of the all events in the filtration. Clearly, we still have $\mathbb{E}[m_\tau] = \mathbb{E}\left[\mathbb{E}\left[m_\tau^{\boldsymbol{\mu}} \mid \boldsymbol{\mu}\right]\right] \leq 1$. Let us calculate m_t. Let f denote the density of $\boldsymbol{\mu}$ and for a positive definite matrix $\mathbf{P}$ let $c(\mathbf{P}) = \sqrt{(2\pi)^d / \det(\mathbf{P})} = \int \exp\left(-\frac{1}{2}\boldsymbol{x}^\top \mathbf{P} \boldsymbol{x}\right) d\boldsymbol{x}$. Then,

$$\begin{aligned}
m_t &= \int_{\mathbb{R}^d} \exp\left(\langle \boldsymbol{\mu}, \boldsymbol{s}_t \rangle - \frac{1}{2}\|\boldsymbol{\mu}\|_{\mathbf{V}_t}^2\right) f(\boldsymbol{\mu}) d\boldsymbol{\mu} \\
&= \int_{\mathbb{R}^d} \exp\left(-\frac{1}{2}\left\|\boldsymbol{\mu} - \mathbf{V}_t^{-1}\boldsymbol{s}_t\right\|_{\mathbf{V}_t}^2 + \frac{1}{2}\left\|\boldsymbol{s}_t\right\|_{\mathbf{V}_t^{-1}}^2\right) f(\boldsymbol{\mu}) d\boldsymbol{\mu} \\
&= \frac{1}{c(\mathbf{V})} \exp\left(\frac{1}{2}\left\|\boldsymbol{s}_t\right\|_{\mathbf{V}_t^{-1}}^2\right) \int_{\mathbb{R}^d} \exp\left(-\frac{1}{2}\left\{\left\|\boldsymbol{\mu} - \mathbf{V}_t^{-1}\boldsymbol{s}_t\right\|_{\mathbf{V}_t}^2 + \|\boldsymbol{\mu}\|_{\mathbf{V}}^2\right\}\right) d\boldsymbol{\mu}
\end{aligned}$$

Elementary calculation shows that if $\mathbf{P}$ is positive semi-definite and $\mathbf{Q}$ is positive definite

$$\|\boldsymbol{x} - a\|_{\mathbf{P}}^2 + \|x\|_{\mathbf{Q}}^2 = \left\|\boldsymbol{x} - (\mathbf{P}+\mathbf{Q})^{-1}\mathbf{P}a\right\|_{\mathbf{P}+\mathbf{Q}}^2 + \|a\|_{\mathbf{P}}^2 - \|\mathbf{P}a\|_{(\mathbf{P}+\mathbf{Q})^{-1}}^2.$$

Therefore,

$$\begin{aligned}
\left\|\boldsymbol{\mu} - \mathbf{V}_t^{-1}\boldsymbol{s}_t\right\|_{\mathbf{V}_t}^2 + \|\boldsymbol{\mu}\|_{\mathbf{V}}^2 &= \left\|\boldsymbol{\mu} - (\mathbf{V}+\mathbf{V}_t)^{-1}\boldsymbol{s}_t\right\|_{\mathbf{V}+\mathbf{V}_t}^2 + \left\|\mathbf{V}_t^{-1}\boldsymbol{s}_t\right\|_{\mathbf{V}_t}^2 - \|\boldsymbol{s}_t\|_{(\mathbf{V}+\mathbf{V}_t)^{-1}}^2 \\
&= \left\|\boldsymbol{\mu} - (\mathbf{V}+\mathbf{V}_t)^{-1}\boldsymbol{s}_t\right\|_{\mathbf{V}+\mathbf{V}_t}^2 + \|\boldsymbol{s}_t\|_{\mathbf{V}_t^{-1}}^2 - \|\boldsymbol{s}_t\|_{(\mathbf{V}+\mathbf{V}_t)^{-1}}^2,
\end{aligned}$$

which gives

$$\begin{aligned}
m_t &= \frac{1}{c(\mathbf{V})} \exp\left(\frac{1}{2}\left\|\boldsymbol{s}_t\right\|_{(\mathbf{V}+\mathbf{V}_t)^{-1}}^2\right) \int_{\mathbb{R}^d} \exp\left(-\frac{1}{2}\left\|\boldsymbol{\mu} - (\mathbf{V}+\mathbf{V}_t)^{-1}\boldsymbol{s}_t\right\|_{\mathbf{V}+\mathbf{V}_t}^2\right) d\boldsymbol{\mu} \\
&= \frac{c(\mathbf{V}+\mathbf{V}_t)}{c(\mathbf{V})} \exp\left(\frac{1}{2}\left\|\boldsymbol{s}_t\right\|_{(\mathbf{V}+\mathbf{V}_t)^{-1}}^2\right) = \left(\frac{\det(\mathbf{V})}{\det(\mathbf{V}+\mathbf{V}_t)}\right)^{1/2} \exp\left(\frac{1}{2}\left\|\boldsymbol{s}_t\right\|_{(\mathbf{V}+\mathbf{V}_t)^{-1}}^2\right) \\
&= \left(\frac{\det(\mathbf{V})}{\det(\mathbf{V}+\mathbf{V}_t)}\right)^{1/2} \exp\left(\frac{1}{2}\left\|\boldsymbol{s}_t\right\|_{(\mathbf{V}+\mathbf{V}_t)^{-1}}^2\right)
\end{aligned}$$

Now, from $\mathbb{E}[m_\tau] \leq 1$, we obtain

$$\begin{aligned}
\mathbb{P}\left[\|\boldsymbol{s}_\tau\|_{(\mathbf{V}+\mathbf{V}_\tau)^{-1}}^2 > 2\log\left(\frac{\det(\mathbf{V}+\mathbf{V}_\tau)^{1/2}}{\delta \det(\mathbf{V})^{1/2}}\right)\right] &= \mathbb{P}\left[\frac{\exp\left(\frac{1}{2}\|\boldsymbol{s}_\tau\|_{(\mathbf{V}+\mathbf{V}_\tau)^{-1}}^2\right)}{\delta^{-1}\left(\det(\mathbf{V}+\mathbf{V}_\tau)/\det(\mathbf{V})\right)^{\frac{1}{2}}} > 1\right] \\
&\leq \mathbb{E}\left[\frac{\exp\left(\frac{1}{2}\|\boldsymbol{s}_\tau\|_{(\mathbf{V}+\mathbf{V}_\tau)^{-1}}^2\right)}{\delta^{-1}\left(\det(\mathbf{V}+\mathbf{V}_\tau)/\det(\mathbf{V})\right)^{\frac{1}{2}}}\right] \\
&= \mathbb{E}[m_\tau]\,\delta \leq \delta
\end{aligned}$$

and substituting back the definition of $\mathbf{G}_{\boldsymbol{z},t}$ gives

$$\mathbb{P}\left[\|\boldsymbol{s}_\tau\|^2_{\mathbf{G}^{-1}_{\boldsymbol{z},\tau}} > 2\log\left(\frac{\det\left(\mathbf{G}_{\boldsymbol{z},\tau}\right)^{1/2}}{\delta\lambda^{d/2}}\right)\right] \leq \delta.$$

□

Theorem I.1 (Self-Normalized Bound for Vector-Valued Martingales)**.** *Let $\{\mathcal{F}_t\}_{t=0}^{\infty}$ be a filtration. Let $\{\eta_t\}_{t=1}^{\infty}$ be a real-valued stochastic process such that η_t is $\mathcal{F}_t$-measurable and η_t is conditionally σ_2-sub-Gaussian for some $\sigma_2 \geq 0$ i.e. $\forall\lambda \in \mathbb{R}$ holds*

$$\mathbb{E}\left[e^{\lambda\eta_t} \mid \mathcal{F}_{t-1}\right] \leq \exp\left(\frac{\lambda^2\sigma_2^2}{2}\right).$$

Let $\{\boldsymbol{z}_t\}_{t=1}^{\infty}$ be an $\mathbb{R}^d$-valued stochastic process such that $\boldsymbol{z}_t$ is $\mathcal{F}_{t-1}$-measurable. For any $t \geq 0$, define $\boldsymbol{s}_t = \sum_{s=1}^{t}\eta_s\boldsymbol{z}_s$. Then, for any $\delta > 0$, with probability at least $1-\delta$, for all $t \geq 0$,

$$\|\boldsymbol{s}_t\|^2_{\mathbf{G}^{-1}_{\boldsymbol{z},t}} \leq 2\sigma_2^2\log\left(\frac{\det\left(\mathbf{G}_{\boldsymbol{z},t}\right)^{1/2}\lambda^{-d/2}}{\delta}\right)$$

Proof. We will use a stopping time construction, which goes back at least to (Freedman, 1975). Define the bad event

$$B_t(\delta) = \left\{\omega \in \Omega : \|\boldsymbol{s}_t\|^2_{\mathbf{G}^{-1}_{\boldsymbol{z},t}} > 2\sigma_2^2\log\left(\frac{\det\left(\mathbf{G}_{\boldsymbol{z},t}\right)^{1/2}\det(\mathbf{V})^{-1/2}}{\delta}\right)\right\}$$

We are interested in bounding the probability that $\bigcup_{t\geq 0}B_t(\delta)$ happens. Define $\tau(\omega) = \min\{t \geq 0 : \omega \in B_t(\delta)\}$, with the convention that $\min\emptyset = \infty$. Then, τ is a stopping time. Further, $\bigcup_{t\geq 0}B_t(\delta) = \{\omega : \tau(\omega) < \infty\}$ Thus,

$$\begin{aligned}\mathbb{P}\left[\bigcup_{t\geq 0}B_t(\delta)\right] &= \mathbb{P}[\tau < \infty] = \mathbb{P}\left[\|\boldsymbol{s}_\tau\|^2_{\mathbf{G}^{-1}_{\boldsymbol{z},\tau}} > 2\sigma_2^2\log\left(\frac{\det\left(\mathbf{G}_{\boldsymbol{z},\tau}\right)^{1/2}\det(\mathbf{V})^{-1/2}}{\delta}\right), \tau < \infty\right] \\ &\leq \mathbb{P}\left[\|\boldsymbol{s}_\tau\|^2_{\mathbf{G}^{-1}_{\boldsymbol{z},\tau}} > 2\sigma_2^2\log\left(\frac{\det\left(\mathbf{G}_{\boldsymbol{z},\tau}\right)^{1/2}\det(\mathbf{V})^{-1/2}}{\delta}\right)\right] \leq \delta.\end{aligned}$$

□

Lemma I.3. *Let $(\mathcal{F}_t)_{t\geq 0}$ be a filtration such that w_t is $\mathcal{F}_{t-1}$ measurable and η_t is $\mathcal{F}_t$ measurable and is conditionally σ_2-sub-Gaussian. Let τ be a stopping time w.r.t. to this filtration i.e. the event $\{\tau \leq t\}$ belongs to $\mathcal{F}_t$. The following sequence of random variables is a martingale with respect to $\mathcal{F}_t$: $s_t = \sum_{s=1}^{t}\eta_s w_s$. Furthermore, for any $\delta > 0, \sigma_2 > 0$, with probability at least $1-\delta$:*

$$|s_\tau| \leq \sigma_2\sqrt{2\left(1/\sigma_2^2 + \sum_{t=1}^{\tau}w_t^2\right)\log\left(\frac{\sqrt{1+\sigma_2^2\sum_{t=1}^{\tau}w_t^2}}{\delta}\right)}$$

Proof. The fact that it is a martingale follows from the conditional sub-Gaussianity. Then for $\lambda \in \mathbb{R}^d, t > 0$ we define

$$m_t^\lambda = \exp\left(\frac{\lambda\, s_t}{\sigma_2} - \frac{\lambda^2}{2}\sum_{s=1}^{t}w_s^2\right)$$

$$d_t^\lambda = \frac{\lambda \eta_t w_t}{\sigma_2} - \frac{\lambda^2}{2} w_t^2$$

Since τ be a stopping time with respect to the filtration $\{\mathcal{F}_t\}_{t=0}^\infty$ we can show that m_τ^λ is well-defined almost surely and $\mathbb{E}\left[m_\tau^\lambda\right] \leq 1$. We start by proving that $\left\{m_t^\lambda\right\}_{t=0}^\infty$ is a supermartingale. Clearly, d_t^λ is $\mathcal{F}_t$-measurable, as is m_t^λ. Further,

$$\mathbb{E}\left[m_t^\lambda \mid \mathcal{F}_{t-1}\right] = \mathbb{E}\left[m_1^\lambda \cdots d_{t-1}^\lambda d_t^\lambda \mid \mathcal{F}_{t-1}\right] = d_1^\lambda \cdots d_{t-1}^\lambda \mathbb{E}\left[d_t^\lambda \mid \mathcal{F}_{t-1}\right] \leq m_{t-1}^\lambda,$$

showing that $\left\{m_t^\lambda\right\}_{t=0}^\infty$ is indeed a supermartingale. Next we show that m_τ^λ is always well-defined and $\mathbb{E}\left[m_\tau^\lambda\right] \leq 1$. First define $\widetilde{M} = m_\tau^\lambda$ and note that $\widetilde{M}(\omega) = M_{\tau(\omega)}^\lambda(\omega)$. Thus, when $\tau(\omega) = \infty$, we need to argue about $M_\infty^\lambda(\omega)$. By the convergence theorem for nonnegative supermartingales, $\lim_{t\to\infty} m_t^\lambda(\omega)$ is well-defined, which means m_τ^λ is well-defined, independently of whether $\tau < \infty$ holds or not. Now let $Q_t^\lambda = M_{\min\{\tau,t\}}^\lambda$ be a stopped version of m_t^λ. We proceed by using Fatou's Lemma to show that $\mathbb{E}\left[m_\tau^\lambda\right] = \mathbb{E}\left[\liminf_{t\to\infty} Q_t^\lambda\right] \leq \liminf_{t\to\infty} \mathbb{E}\left[Q_t^\lambda\right] \leq 1$.

Let $\Lambda \sim \mathbb{N}\left(0, \sigma_2^2\right)$ be a Gaussian random variable and define $m_t = \mathbb{E}\left[m_t^\Lambda \mid F^\infty\right]$. Clearly, we still have $\mathbb{E}\left[m_t\right] = \mathbb{E}\left[\mathbb{E}\left[m_t^\Lambda \mid \Lambda\right]\right] \leq 1$. Let us calculate m_t. We will need the density λ which is $f(\lambda) = \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\lambda^2/2\sigma_2^2}$. Now, it is easy to write m_t explicitly

$$\begin{aligned} m_t = \mathbb{E}\left[m_t^\Lambda \mid \mathcal{F}_\infty\right] &= \int_{-\infty}^{\infty} m_t^\lambda f(\lambda) d\lambda = \sqrt{\frac{1}{2\pi\sigma_2^2}} \int_{\infty}^{\infty} \exp\left(\frac{\lambda s_t}{\sigma_2} - \frac{\lambda^2}{2} \sum_{s=1}^t w_s^2\right) e^{-\lambda^2/2\sigma_2^2} d\lambda \\ &= \exp\left(\frac{s_t^2}{2\sigma_2^2\left(1/\sigma_2^2 + \sum_{s=1}^t w_s^2\right)}\right) \sqrt{\frac{1}{1+\sigma_2^2 \sum_{s=1}^t w_s^2}} \end{aligned}$$

where we have used that $\int_{-\infty}^{\infty} \exp\left(a\lambda - b\lambda^2\right) = \exp\left(a^2/(4b)\right)\sqrt{\pi/b}$.

To finish the proof, we use Markov's inequality and the fact that $\mathbb{E}\left[m_\tau\right] \leq 1$:

$$\begin{aligned} &\mathbb{P}\left[|s_\tau| \geq R\sqrt{2\left(1/\sigma_2^2 + \sum_{t=1}^\tau w_t^2\right)\log\left(\frac{\sqrt{1+\sigma_2^2\sum_{t=1}^\tau w_t^2}}{\delta}\right)}\right] \\ &= \mathbb{P}\left[\frac{\left(\sum_{t=1}^\tau \eta_t w_t\right)^2}{2\sigma_2^2\left(1/\sigma_2^2 + \sum_{t=1}^\tau w_t^2\right)} \geq \log\left(\frac{\sqrt{1+\sigma_2^2\sum_{t=1}^\tau w_t^2}}{\delta}\right)\right] \\ &= \mathbb{P}\left[\exp\left(\frac{s_\tau^2}{2\sigma_2^2\left(1/\sigma_2^2 + \sum_{t=1}^\tau w_t^2\right)}\right) \geq \frac{\sqrt{1+\sigma_2^2\sum_{t=1}^\tau w_t^2}}{\delta}\right] \\ &= \mathbb{P}\left[m_\tau \geq \frac{1}{\delta}\right] \leq \frac{\mathbb{E}\left[m_\tau\right]}{1/\delta} \leq \delta \end{aligned}$$

□

Theorem I.2 (Self-normalized Bound for Scalar Valued Martingales)**.** *Under the same assumptions as the previous theorem, for any $\delta > 0$, with probability at least $1-\delta$, for all $t \geq 0$,*

$$\left|\sum_{s=1}^t \eta_t w_t\right| \leq \sigma_2 \sqrt{2\left(1/\sigma_2^2 + \sum_{s=1}^t w_s\right)\log\left(\frac{\sqrt{1+\sigma_2^2\sum_{s=1}^t w_s^2}}{\delta}\right)} \tag{26}$$

Proof. Define the "bad" event

$$B_t(\delta) = \left\{ \omega \in \Omega : \frac{\left(\sum_{s=1}^t \eta_s w_s\right)^2}{1/\sigma_2^2 + \sum_{t=1}^\tau w_t^2} > 2\sigma_2^2 \ln\left(\frac{\sqrt{1+\sigma_2^2 \sum_{s=1}^t w_s^2}}{\delta}\right) \right\}$$

We are interested in bounding the probability that $\bigcup_{t\geq 0} B_t(\delta)$ happens. Define $\tau(\omega) = \min\{t \geq 0 : \omega \in B_t(\delta)\}$, with the convention that $\min \emptyset = \infty$. Then, τ is a stopping time. Further, $\bigcup_{t\geq 0} B_t(\delta) = \{\omega : \tau(\omega) < \infty\}$ Thus, by the previous theorem it holds that

$$\begin{aligned}
\mathbb{P}\left[\bigcup_{t\geq 0} B_t(\delta)\right] &= \mathbb{P}[\tau < \infty] = \mathbb{P}\left[\frac{\left(\sum_{s=1}^t \eta_s w_s\right)^2}{1/\sigma_2^2 + \sum_{t=1}^\tau w_t^2} > 2\sigma_2^2 \ln\left(\frac{\sqrt{1+\sigma_2^2 \sum_{s=1}^t w_s^2}}{\delta}\right) \text{ and } \tau < \infty\right] \\
&= \mathbb{P}\left[\frac{\left(\sum_{s=1}^t \eta_s w_s\right)^2}{1/\sigma_2^2 + \sum_{t=1}^\tau w_t^2} > 2\sigma_2^2 \ln\left(\frac{\sqrt{1+\sigma_2^2 \sum_{s=1}^t w_s^2}}{\delta}\right)\right] \leq \delta
\end{aligned}$$

□

J. Parameter Estimation and Concentration in First Stage

There are many ways of addressing the regression problem in the first stage and they fundamentally reduce to a choice for the regularizer in the regression. If we do not introduce such regularizer, we are left with a system of multiple regressions that can be solved with standard OLS (estimator). Another choice is to introduce a Frobenius norm regularizer. We introduce the parameter $\lambda > 0$ and a regularization term $\lambda \|\|\boldsymbol{\Theta}\|\|_F^2$, which is used to penalize the model complexity. By choosing the Frobenius norm, the system of equations decouples again but each with a regularizer term. Thus, we end up with d independent linear equations that we try to fit separately. More interesting settings could try to solve the optimization problem jointly by a regularizer that couples the equations (e.g. (Wainwright, 2019) provides concentration results for such settings). This will be interesting to investigate in future works.

We indicate with $\underline{\boldsymbol{\theta}}_j$ the j-th column of the matrix $\boldsymbol{\Theta}$, then

$$\begin{aligned}
\widehat{\boldsymbol{\Theta}}_t &\in \operatorname*{argmin}_{\boldsymbol{\Theta}} \sum_{s=1}^{t} \left\| \boldsymbol{x}_s^\top - \boldsymbol{z}_s^\top \boldsymbol{\Theta} \right\|_2^2 + \lambda \|\|\boldsymbol{\Theta}\|\|_F^2 \\
&= \operatorname*{argmin}_{\boldsymbol{\Theta}} \sum_{s=1}^{t} \left(\sum_{j=1}^{d} \left(\boldsymbol{x}_{s,j} - \boldsymbol{z}_s^\top \underline{\boldsymbol{\theta}}_j \right)^2 + \lambda \sum_{j=1}^{d} \left\| \underline{\boldsymbol{\theta}}_j \right\|_2^2 \right) \\
&= \operatorname*{argmin}_{\{\underline{\boldsymbol{\theta}}_j\}_{j=1}^d} \sum_{j=1}^{d} \left(\sum_{s=1}^{t} \left(\boldsymbol{x}_{s,j} - \boldsymbol{z}_s^\top \underline{\boldsymbol{\theta}}_j \right)^2 + \lambda \left\| \underline{\boldsymbol{\theta}}_j \right\|_2^2 \right)
\end{aligned}$$

Clearly, we can compute separately the columns $\widehat{\boldsymbol{\theta}}_{t,j}$ of $\widehat{\boldsymbol{\Theta}}_t$ as

$$\widehat{\boldsymbol{\theta}}_{t,j} \in \operatorname*{argmin}_{\underline{\boldsymbol{\theta}}_j} \sum_{s=1}^{t} \left(\boldsymbol{x}_{s,j} - \boldsymbol{z}_s^\top \underline{\boldsymbol{\theta}}_j \right)^2 + \lambda \left\| \underline{\boldsymbol{\theta}}_j \right\|_2^2 \tag{27}$$

with solution

$$\widehat{\boldsymbol{\theta}}_{t,j} = \left(\mathbf{Z}_t^\top \mathbf{Z}_t + \lambda \mathbb{I}_d \right)^{-1} \mathbf{Z}_t^\top \boldsymbol{x}_{t,j}$$

where $\boldsymbol{x}_{t,j}$ is a vector with components $x_{1,j}, \ldots, x_{t,j}$. The solution to the independent quadratic optimization problems is in matrix notation equal to

$$\widehat{\boldsymbol{\Theta}}_t = \left(\mathbf{Z}_t^\top \mathbf{Z}_t + \lambda \mathbb{I}_d \right)^{-1} \mathbf{Z}_t^\top \mathbf{X}_t$$

From the decomposition of the problem into multiple independent regressions, we understand that it is enough to concentrate the individual columns of $\widehat{\boldsymbol{\Theta}}_t$ around the ones of $\boldsymbol{\Theta}$ and then to use a union bound to put things together.

Theorem J.1 (Confidence Ellipsoid for Columns in First Stage)**.** *Define $\boldsymbol{x}_t = \boldsymbol{z}_t^\top \underline{\boldsymbol{\theta}}_j + \epsilon_{t,j}$ with $\epsilon_{t,j}$ is $\sigma_{\boldsymbol{\epsilon}}$-sub-Gaussian and assume that $\left\| \underline{\boldsymbol{\theta}}_j \right\|_2 \leq S$. Then, for any $\delta > 0$, with probability at least $1 - \delta$, for all $t \geq 0$, $\underline{\boldsymbol{\theta}}_j$ lies in the set*

$$\mathcal{E}_t = \left\{ \underline{\boldsymbol{\theta}}_j \in \mathbb{R}^d : \left\| \widehat{\underline{\boldsymbol{\theta}}}_{t,j} - \underline{\boldsymbol{\theta}}_j \right\|_{\mathbf{G}_{\boldsymbol{z},t}} \leq \sigma_{\boldsymbol{\epsilon}} \sqrt{2 \log \left(\frac{\det \left(\mathbf{G}_{\boldsymbol{z},t} \right)^{1/2} \det(\lambda \mathbb{I}_d)^{-1/2}}{\delta} \right)} + \lambda^{1/2} S \right\}$$

Furthermore, if for all $t \geq 1, \|\boldsymbol{z}_t\|_2 \leq L_z$ then with probability at least $1 - \delta$, for all $t \geq 0$,

$$\left\| \widehat{\underline{\boldsymbol{\theta}}}_{t,j} - \underline{\boldsymbol{\theta}}_j \right\|_{\mathbf{G}_{\boldsymbol{z},t}} \leq \sigma_{\boldsymbol{\epsilon}} \sqrt{d \log \left(\frac{1 + t L_z^2 / \lambda d}{\delta} \right)} + \lambda^{1/2} S$$

Corollary J.1 (Confidence Ellipsoid for First Stage)**.** *Under the conditions of the previous theorem, for any $\delta > 0$, with probability at least $1 - \delta$, for all $t \geq 0$*

$$\left\|\widehat{\boldsymbol{\Theta}}_t - \boldsymbol{\Theta}\right\|_F^2 = \sum_{j=1}^{d} \left\|\widehat{\underline{\boldsymbol{\theta}}}_{t,j} - \underline{\boldsymbol{\theta}}_j\right\|_2^2 \leq d\left(\sigma_{\boldsymbol{\epsilon}}\sqrt{d\log\left(\frac{1 + tL_z^2/\lambda d}{\delta}\right)} + \lambda^{1/2}S\right)$$

K. Experiments

K.1 Experimental Analysis of O2SLS

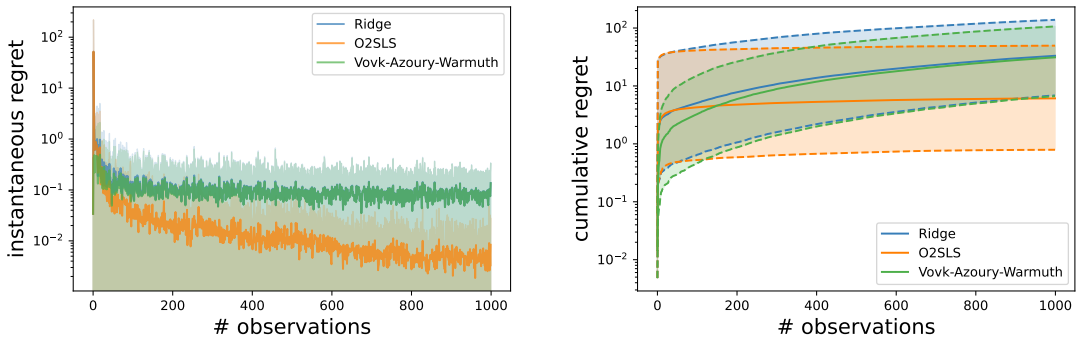
Now, we aim to compare the empirical performance of O2SLS estimator with the online ridge regression (Ridge) and the Vovk-Azoury-Warmuth Forecaster (VAWf) (Orabona, 2019). The Ridge estimator is given by

$$\boldsymbol{\beta}_t^{\mathsf{Ridge}} = \left(\mathbf{X}_t^\top \mathbf{X}_t + \lambda_{\mathsf{Ridge}} \mathbb{I}_d\right)^{-1} \mathbf{X}_t^T \boldsymbol{y}_t,$$

while VAWf reads

$$\boldsymbol{\beta}_t^{\mathsf{VAWf}} = \left(\mathbf{X}_{t+1}^\top \mathbf{X}_{t+1} + \lambda_{\mathsf{VAWf}} \mathbb{I}_d\right)^{-1} \mathbf{X}_t^T \boldsymbol{y}_t.$$

We choose the regularisation parameters to be all equal to 10^{-3}. We deploy the experiments in Python3 on a single Intel(R) Core(TM) i7-8665U CPU@1.90GHz. An experiment consisting of 100 runs of an algorithm takes approximately ten minutes. We denote the normal distribution with mean μ and standard deviation σ as $\mathcal{N}(\mu, \sigma)$, with $\mathcal{N}_n$ we indicate its multivariate extension to n dimensions. We consider the true parameter of the model to be generated as $\boldsymbol{\beta} \sim \mathcal{N}_{50}(\vec{1}0, \mathbb{I}_{50})$ and $\boldsymbol{\Theta}_{i,j} \sim \mathcal{N}(0,1)$ for each of its entry. At each time step we sample the vectors $\boldsymbol{z}_t \sim \mathcal{N}_{50}(\vec{0}, \mathbb{I}_{50})$, and $\boldsymbol{\epsilon}_t \sim \mathcal{N}_{50}(\vec{0}, \mathbb{I}_{50})$. The endogenous noise in the second stage is $\eta_t = (\widetilde{\eta}_t + \sum_{i=1}^{12} \boldsymbol{\epsilon}_{t,i})/13$, where $\widetilde{\eta}_t \sim \mathcal{N}(0,1)$ is a r.v. independent from all the others. From Figure 3, we observe that in the presence of endogeneity, Ridge and VAWf have very similar performances and are sensibly worse than O2SLS according to the instantaneous *identification regret*.



(a) Instantaneous regret (b) Cumulative regret

Figure 3: Performance of O2SLS for online linear regression compared with online Ridge and VAWf in terms of *identification regret*. O2SLS incurs lower regret compared to the other two. We average our curves over 100 samples and the shaded area indicates one standard deviation. The $y-$axis is in logarithmic scale.