Private Funding of Free Data: A Theoretical Framework

Rachel Harris Soloveichik (Bureau of Economic Analysis)

This paper develops a theoretical framework in which business buy both data services and other production inputs. Because data is only partially excludable, data sellers who charge high prices are vulnerable to piracy and may not be able to fund the fixed cost of creating data. And even if data can be sold at a high price, social welfare and productivity are still lower than they would be in a world in which data is freely shared (Coyle 2022). This paper demonstrates a potential solution to the theoretical framework which both avoids the problem of piracy and maximizes social welfare. If data is complimentary to a particular asset, then free data increases the market value of that particular asset. When the expected value increase is large enough, a private entity which owns a particular asset can increase their wealth by funding the fixed cost of data creation and then distributing their data for free.

The paper then argues that privately funded free data is a large investment category. To start out, the paper solves the theoretical framework to demonstrate that private funding of free data is a dominant solution for a broad range of plausible parameters. Next, the paper presents two empirical case studies of privately funded free data: individual credit files (Soloveichik 2023a) and tax data (Soloveichik 2023b). By itself, those two case studies account for $2 trillion of data creation in 2019. The paper then uses occupational task data from the Department of Labor (National Center for O*net Development 2022) and household task data from the American Time Use Survey (Bureau of Labor Statistics 2023) to calculate the ratio of overall data creation to the data creation in those two case studies. Based on those ratios, the paper calculates that total private creation of free data was $10 trillion in 2019. This $10 trillion of data creation is split evenly between $5 trillion of business investment and $5 trillion of household production.

JEL Codes: D12, E01, and G14