Capitalizing Data: Case Studies of Tax Forms and Individual Credit Reports

Authors Rachel Soloveichik, U.S. Bureau of Economic Analysis

Contact rachel.soloveichik@bea.gov

Date June 2023

Abstract Early papers on capitalizing data focused on complex digital data that are stored on supercomputers and managed by highly skilled computer scientists (Statistics Canada 2019) (Eurostat 2020) (Coyle 2022) (Calderon and Rassier 2022) (Mitchell et al. 2022). This paper studies two very different types of data: tax forms and individual credit reports. Both types of data are simple text records that can be stored on any computer or even on paper (Brenton 1964) and managed by workers with only a high school degree (Bureau of Labor Statistics 2022a) (Weedmark 2021). Despite their simplicity, these two data types are expensive to create. This paper estimates that tax forms had a creation cost of $0.4 trillion in 2017 and individual credit reports had a creation cost of $0.6 trillion in 2017.

Historical growth changes noticeably when tax forms and individual credit reports are tracked as long-lived intangible assets. Tax form creation rose rapidly between 1929 and 1950 due to the introduction of Social Security and widespread federal income taxes. As a result, including tax forms increases real gross domestic product (GDP) growth between 1929 and 1950 by 0.12 percentage point per year. In contrast, real credit report creation plummeted in 1970 due to the passage of the Fair Credit Reporting Act (FCRA). As a result, including individual credit reports decreases real GDP growth in 1970 by 1.6 percentage point. However, neither tax forms nor individual credit reports have much impact on recent GDP growth.

Keywords Capitalizing data, data, tax forms, individual credit reports, national accounts, data creation, long-lived intangible assets, gross domestic product, GDP

JEL Code D14, E01, and G14

Bureau of Economic Analysis
U.S. DEPARTMENT OF COMMERCE

Introduction

The United States economy depends on data. To start, data are integral inputs to many purchased services. Banks check credit scores before approving loans, insurers check risk factors and claims history before offering coverage or setting premiums, doctors check medical history before making diagnoses or prescribing treatments, and colleges check standardized test scores and grades before admitting students or recommending classes. In addition, employers use job references to evaluate workers and governments use tax forms to determine eligibility for benefits. Finally, households use reviews and online profiles to determine which businesses to buy from and which individuals to socialize with.

This paper studies two types of financial data: tax forms and individual credit reports. Tax forms record income and deductions, while individual credit reports record loans and payments. The creation of tax forms is mandatory, while the creation of individual credit reports is voluntary. Tax forms are created by virtually all entities, while individual credit reports are only created by lenders and borrowers. Even though these two types of data are separate, they are often used together in production. For example, a mortgage broker might use both tax forms and credit reports, along with other information like property titles or property appraisals, to see whether a loan applicant is likely to pay their mortgage on time.

This paper estimates that Americans spent 28 billion hours creating tax forms and individual credit reports in 2017. Millions of workers specialize in accounting, bookkeeping, loan origination, and bill collection (Bureau of Labor Statistics 2022b). In addition, the U.S. Department of Labor's task data shows that many generalist workers spend a portion of their time on tasks related to financial data (National Center for O*Net Development 2022). Furthermore, consumers spend significant amounts of time worrying about finances (Sergeyev et al. 2023) and often suffer from stress related to bills (Dackehag et al. 2019). Combining all these costs together, this paper estimates that tax forms had a creation cost of $0.4 trillion in 2017 and individual credit reports had a creation cost of $0.6 trillion in 2017.

The large cost of creating financial data is balanced out by significant benefits associated with data. Most obviously, tax forms are required by law and an entity that does not file tax forms may have its assets seized and an individual who refuses to file tax forms may go to prison (Gordon 1949). Tax forms are also helpful for businesses seeking banking services (Wu 2021) and individuals seeking vehicle loans (Chan et al. 2021). Individual credit reports are not required by law, but individuals without a credit report are rarely granted loans and are often denied jobs as well. Previous authors have shown that exogenous improvements to an individual's credit score decrease their interest rates (Jansen et al. 2022) and increase their earnings (Herkenhoff et al. 2021) (Dobbie et al. 2020) (Bos et al. 2018). In total, this paper estimates that tax forms and individual credit reports yielded a service flow of $1 trillion in 2017.

In recent years, national accountants have discussed including both purchased and own-account data in GDP as intangible capital assets (Rassier et al. 2019) (Eurostat 2020). This inclusion is a broadening of the current recommendation to only include purchased data in GDP (United Nations Statistics 2008, sec. 10.112-10.114). This paper includes government data and business data (both digital and non-digital) in GDP using a recently developed national accounting framework for data investment (Eurostat 2020). In general, tracking an expenditure as a capital investment raises measured GDP because both private capital investment and government capital depreciation are components of final output. This paper also presents revisions to personal consumption expenditures (PCE) and labor compensation that are associated with consumer data. Finally, household produced data are included in household production using the framework outlined in the U.S. Bureau of Economic Analysis' (BEA's) existing household production satellite account (Bridgman et al. 2022). Based on those frameworks, this paper presents calculations that tracking these two types of data as intangible assets increases measured GDP in 2017 by $0.7 trillion and measured household production in 2017 by $0.3 trillion.

This paper is divided into six sections. Section 1 describes the system in which these two types of data are created, distributed, and used. Section 2 uses official government publications and other sources to value the tax forms created each year. Section 3 uses existing economic research to value the individual credit reports created each year. Section 4 presents recalculations of nominal GDP and nominal household production when these two types of data are included as intangible capital assets. Section 5 presents price indexes for data and then uses those price indexes to recalculate overall GDP prices, overall household production prices, real GDP, and real household production. Finally, Section 6 presents calculations of private business productivity when both tax forms and individual credit reports are included in the production accounts.

Section 1: Tax Form and Individual Credit Report System

There are four main entities in the tax form and individual credit report system. First, there is a data creator that records financial data and submits it to a data platform. Second, there is a data platform which organizes, audits, and distributes data to a narrow list of specific users for a narrow list of permitted reasons. Third, there are the data owners who distribute their financial data to whomever they please for whatever reason they please. Finally, there is the data user. This section discusses each of the four main entities briefly.

To be clear, the two types of financial data studied in this paper are only a subset of all financial data. Large businesses have their own credit reports that are distinct from the individual credit reports of their owners (Sylla 2002). Banks keep some customer data in-house and use those data to retain profitable customers (Marquez 2002), to trade on inside information (Haselmann et al. 2022), or to pool with other banks on a reciprocal basis (DeNicola 2020). Conversely, many entities maintain internal records of supplier payments, worker payroll, and other useful financial data. Finally, hackers sometimes obtain data illegally and resell those data on the dark web (Manzi 2022). These other types of financial data, as well as nonfinancial data, may be studied in future research.

Data Creators

Tax forms are created by many different entities: employers record wages paid and health insurance status; banks and other financial institutions record investment income and mortgage interest; corporations and unincorporated businesses record revenue and expenses associated with their businesses; retailers record sales tax that has been collected on behalf of governments; governments record property tax payments; and households record personal income, deductions, and credits on individual tax forms. Later sections of this paper discuss how the national accounting treatment of tax form data depends on the sector which creates data, the sector which funds the data creation, and the sector which owns data. This section recognizes all tax forms as data and tracks total data creation.

Businesses often produce tax forms jointly with other data. For example, a human resource department might use the same employee time sheets to calculate taxes, payroll, leave balances, and seniority. In addition, large businesses often use the same underlying data to calculate both income that is measured using tax law and income that is measured using generally accepted accounting principles (Federal Accounting Standards Advisory Board 2022). Hence, it can be difficult to isolate the time which is spent on tax forms alone. This paper sidesteps that difficulty by using official government estimates of tax form creation time (Executive Office of the President 1981–2020). These official government estimates are assumed to adjust time for joint production. The next section presents calculations of tax form creation costs which multiply the official time spent on tax forms with an estimate of the cost per hour.

To be clear, the tax forms studied in this paper often contain inaccurate information. Businesses sometimes underreport profits to minimize their tax obligations (BEA 2022). In addition, both untrained individuals and sophisticated computers sometimes misreport information by accident (IRS 2022a). This inaccuracy can bias measured economic activity, and so BEA's published statistics adjust reported income for underreporting and misreporting (BEA 2022). However, this inaccuracy does not necessarily reduce the cost of creating tax forms. It is often more difficult to tell a plausible lie than the truth (Zuckerman et al. 1981). For example, a dishonest taxpayer might spend time faking receipts to support their overreported expenses. In practice, this paper does not try to measure the hypothetical cost of creating perfectly accurate tax forms. Instead, this paper uses official government estimates to measure the actual cost of creating the imperfectly accurate tax forms that do exist.

Individual credit reports are also created by many different entities: banks and other financial institutions report detailed information on loans and payments, sellers of goods and services report negative information like bills sent to collections, and governments report negative information like bankruptcy petitions. This paper does not attempt to measure the inputs which are used to produce individual credit reports. Instead, it uses previous research to estimate the service flow from credit report data. The standard capital formulas assume that the initial creation cost of a capital asset equals the net present value of its expected future services (Jorgenson 1963). Accordingly, estimates of the tax form creation which are based on initial creation cost can be compared to estimates of the individual credit report creation cost which are based on the net present value of future service flows.

Borrowers know that individual credit reports are bundled with other financial services (Walczak and Borkan 2016). To start out, loan applicants explicitly consent to credit report creation. In addition, popular financial advisors have told Americans for more than a century that using financial services helps build credit (Carpenter 1910) (Orman 2019). Consistent with popular financial advice, loan applicants who are just barely able to borrow money have higher future credit scores than otherwise similar loan applicants who are just barely unable to borrow money (Di Maggio et al. 2022) and individuals who keep a secured credit card open have higher future credit scores than individuals who close their card (Santucci 2016). Building credit is an especially well-advertised bundle component for secured credit cards (Levy et al. 2016) and credit builder loans (Burke et al. 2019). Based on these facts, individual credit reports are clearly not an externality under the official guidelines for national accounting (United Nations 2008, sec. 3.92). This paper tracks credit reports as purchased data creation services that are bundled together with other financial services. Tracking data creation does not change total output for banks or other financial service providers. Instead, financial service output is split into two components: data creation and other financial services.

Individual credit reports are generally accurate. The credit bureau industry claims that less than one percent of individuals have errors significant enough to change their credit score by a material amount and 95 percent of individuals who disputed an error were satisfied with the outcome of their dispute (Turner et al. 2011). Even a critical article by Consumer Reports acknowledged that only eleven percent of individuals had account information errors (Fox 2021). At first glance, this high level of accuracy appears inconsistent with claims by credit repair agencies to “fix” credit scores. However, many supposed credit repair agencies do not actually change credit scores at all (Federal Trade Commission 2012); and even the large credit score increases provided by legitimate credit repair agencies (Huntley 2019) are mostly illusory. Credit repair agencies typically increase credit scores by disputing negative information and thereby temporarily removing that information from the credit score formula. But potential lenders can still see the negative information and still deny loan applications as a result (Ulzheimer 2022). In other words, credit repair agencies game the credit score formula without actually improving the underlying data. In the rare case that credit repair agencies provide long-term value, they do so by identifying genuine errors in the reports and demanding that those errors be corrected (Whiting 2022). Hence, reports are more accurate due to the involvement of credit repair agencies.

Data Platforms

Data platforms have four data jobs. First, they organize the scattered data supplied by separate data creators into a combined file. Second, this combined file is then audited for problems that must be resolved. Third, they store the combined file for future reference. Finally, they distribute the combined file to any entity with a valid request. In addition to these four data jobs, data platforms also perform a few other tasks: government tax agencies collect payments as necessary and credit bureaus sell analysis services for an additional fee. These additional tasks may be important—but they are not directly related to the financial data system. Accordingly, this paper focuses on the four data jobs.

Data platforms are small relative to the value of the data they hold. The Internal Revenue Service (IRS) spends less than $0.50 cents for every $100 that it collects (IRS 2023a). Similarly, the credit bureau industry (NAICS 56145) earned less than $4 billion of revenue in 2017 from individual rating services. This small size is possibly due to the fact that data which are supplied to data platforms are already formatted by data creators (IRS 2023a) (Datalinx 2022). In addition, data platforms rarely handle data problems on their own. Instead, tax authorities notify data creators about problems and request that the creators resolve the data problems themselves (Taxpayer Advocate Service 2022). Similarly, credit bureaus pass along disputes to the data creator and request that the creators resolve those disputes themselves (Newman and Frontino 2015). Data platforms also have little control over the data that they hold. The distribution of tax forms is heavily regulated by law and taxpayers are guaranteed privacy and confidentiality (IRS 2023b). Credit bureaus are required by law (Newman and Frontino 2015) and credit bureau policy (Hartley 2019) to report accurate data which are based on actual transactions to any entity with a valid request (Redford 1941). Intuitively, data platforms are only intermediaries between data creators, data owners, and data users with no economic ownership or actual control over data. Because data platforms are so small, even a substantial change to their output has little impact on this paper’s measures of data creation or data usage.

Data Owners

Businesses and individuals can generally distribute their data to any entity they choose. In some cases, the business or individual distributes the data directly. For example, potential tenants might give a landlord printed copies of their tax forms (Baker 2022) and credit reports (Farivar 2021) as part of a rental application. In other cases, the business or individual authorizes a platform to distribute data to a specified user. For example, potential tenants might sign a paper authorizing a landlord to request tax forms from the IRS and credit reports from a credit bureau. Direct distribution and authorized distribution have similar economic impacts, so this paper does not distinguish between them. Because businesses and individuals have so much control over their data, this paper treats them as the primary owner of their data.

Governments also distribute data to specific entities for specific uses. Examples of permitted tax form distribution include administration of unemployment insurance (Dunlap 1951), economic measurement (BEA 2022), and criminal investigation (IRS 2017). Examples of permitted individual credit report distribution include child support collection and court cases (DeNicola 2020). Because governments have some control over data, this paper treats them as a partial owner of data.

Partial ownership by multiple parties likely yields higher social welfare than either complete secrecy or full sharing. Data are nonrival and complementary to other data, and so their contribution to economic output is maximized when they are shared broadly (Coyle 2022). At the same time, businesses and individuals value privacy and prefer limited distribution of data (Acquisti et al. 2016). Each business and each individual has their own desired level of privacy, and it would be difficult for the government to set a data distribution level which works for everyone. Governments resolve these contradictory preferences by distributing a minimum amount of data for purposes that are believed to be important enough to override privacy preferences. Businesses and individuals who desire to distribute more data than that minimum amount can then decide to distribute their data to any entity they choose.

“Free” distribution of financial data also likely yields higher social welfare than data sales. Broad sharing of data requires marginal prices for data that may be too low to cover the cost of data creation (Coyle and Diepeveen 2022). In this respect, data are similar to other products whose creation involves high fixed costs and low marginal costs. The most studied methods to finance these products are either provision by governments (Samuelson 1954) or regulated monopolies (Posner 1968). But recent papers have shown that individual ownership is an alternative method that can increase welfare (Jones and Tonetti 2020). In particular, an individual can finance the creation of their own financial data, give those data for free to any entity with a valid reason for its request, and then benefit when that entity uses those data to provide services or employment to the data owner.

Data Users

Governments use tax forms constantly. Tax agencies use tax forms to determine how much tax each individual or corporation is required to pay and social insurance programs use tax forms to allocate benefits. Statistical agencies and policymakers also use tax form data to measure the economy (BEA 2022) and the Federal Reserve Bank of New York uses a sample of individual credit reports to measure household debt and other important outcomes (Lee and Klaauw 2010). This paper does not study exactly how each government agency uses data in its operations. Instead, it simply accepts that financial data are important production inputs. This paper assumes that government programs take data from businesses and individuals without giving anything of value in return. This assumption is equivalent to treating data given to governments as taxes in-kind.

Private businesses use financial data to make useful predictions (Farboodi and Veldkamp 2022). Data can predict outcomes either by directly influencing individual behavior (Sergeyev et al. 2023) or by merely being correlated with difficult to observe variables. Banks and other lenders use tax forms and individual credit reports to predict late payment or loan default (Martin 2022) (Chatterjee et al. 2020), universities and other educational institutions use tax forms to identify students who are eligible for financial aid (Dynarski et al. 2013), property insurers use credit reports to predict the likelihood of insurance claims (Federal Trade Commission 2007), and hiring managers use tax forms to determine previous salary history (Feldmann 2017) and use credit reports to predict future job performance (Kiviat 2019). As with governments, this paper does not study exactly how each business uses data in its operations. Instead, it simply accepts that financial data are important production inputs.

This paper assumes that business which use financial data implicitly purchase them from their owner. These implicit purchases are often part of a bundled transaction. For example, a department store might offer special coupons to customers who open a store credit card, an employer might offer all job applicants a regular position but also offer those job applicants who consent to have their credit checked an opportunity to apply for highly paid management positions, or an equipment supplier might offer discounts on leases to small business owners who personally cosign for their leased equipment. This paper is focused on tracking data services, and so adjusts all bundled transactions to explicitly track both the value of data given to firms and the implicit payments that firms give in return for those data. For customer data, this paper adjusts the value of purchased products upward and business intermediate inputs upward to include the value of data provided by customers. In the example given earlier, the department store's sales would be adjusted upward to include value of the special coupon given to customers who open a store credit card and the department store's intermediate inputs would be adjusted upward to include the value of data provided by customers who open a store credit card. These two adjustments precisely cancel out so that tracking customer data does not change measured business value added. Similarly, this paper shifts the value of data provided by workers from labor compensation to business intermediate inputs without changing measured gross operating surplus. In the example given earlier, the value of the opportunity to apply for management positions is considered an implicit payment for the data provided by job applicants. This implicit payment is shifted from labor compensation to a payment for data provided by job applicants. Finally, this paper shifts the value of data provided by self-employed individuals from general proprietors' income to capital services from data without changing either measured business value added or measured gross operating surplus. In the example given earlier, the discount on the business equipment lease is considered an implicit payment for the data provided by the small business owner. This implicit payment is shifted from general proprietors' income to a payment for the value of data provided by the small business owner. To be clear, all of these treatments only apply to data usage. The treatment of implicit payments for data creation services by banks and other businesses was described in an earlier section.

This paper does not study business risk taking, employee benefits, homeownership, or other indirect behavior changes that may be impacted by financial data. It is very common for individuals to consider their future tax liabilities when planning an activity. Forward-looking individuals may also consider the expected damage to their credit reports when planning a financial risk. Conversely, governments may structure their tax laws or businesses may structure their payment system to influence future behavior by individuals. The combined impact of these efforts is theoretically ambiguous and difficult to solve without a complete model. For now, this paper only focuses on the direct impact of financial data.

Section 2: Valuing Tax Form Creation

Time Spent on Federal Tax Forms

This paper's primary source for the time spent on federal tax forms is the official government publication "The Information Collection Budget of the United States" (Executive Office of the President 1989–2020). Since 1989, this publication has estimated the full time required by each government form, starting with the initial record keeping and finishing when a form is submitted (Little 1988). Between 1977 and 1988, this publication excluded time spent on initial record keeping and therefore reported a much lower level of time spent on government forms. Each edition of "The Information Collection

Budget of the United States" is careful to distinguish between changes in time estimates due to actual changes in required tax forms[1] and changes in time estimates due to changes in the model used to estimate the time associated with each form.[2] This paper combines the estimated impact of actual changes to required tax forms each year to calculate a hypothetical index of the time that would be spent on federal tax forms if the number of people impacted by each tax form and the technology used to file taxes were both held fixed.

This hypothetical index is adjusted to reflect actual time spent on tax forms. Precise information on the number of people impacted by each tax form was not be located, so the total number of forms filed serves as a proxy for the tax burden of all required forms. Similarly, precise information on the role of software was not located. Instead, this paper presents estimates that are based on a sudden methodology change. In 2010, the U.S. Department of the Treasury decreased its estimate of the time associated with 1040 forms from 3.7 billion hours to 2.4 billion hours (Executive Office of the President 2011). This paper assumes this decrease represents a sudden recognition of the gradual growth in tax software usage between the previous benchmark year of 1989 and the new benchmark year of 2009. This paper's estimates use published IRS data on the share of tax returns that are filed electronically to interpolate and extrapolate the role of software over the period 1989 to 2020. Before 1989, very few tax returns were filed electronically and therefore software is assumed to have had little impact.[3] Official publications studying tax creation time before 1977 were not located so the number of federal tax forms filed each year is used as an extrapolator. This time series is available from IRS reports back to 1917,[4] but the exact forms studied and the fiscal year covered are not always consistent. In order to avoid a trend break, this paper calculates growth rates for each year from the IRS report for that year and then chains those growth rates to calculate an index of time spent on federal tax forms from 1929 onwards.

At the time this paper was written, official publications studying tax form time after 2017 had not been published. For now, the number of federal tax forms and the share of tax forms filed electronically are

---

1. Table 1 of "The Information Collection Budget of the United States: 2017" reports a 13 percent rise in tax time due to "discretionary agency action." The text clarifies that this increase is due to a new methodology. Accordingly, this paper reclassifies that 13 percent rise as a methodology revision.

2. The Executive Office of the President does not measure time burdens itself. Instead, their publication is a summary of raw data collected by the Office of Management and Budget under the Paperwork Reduction Acts of 1980 and 1995. Curious readers may browse the raw burden imposed by individual forms online.

3. It may be true that calculators and mainframe computers were introduced before 1989. However, analog equivalents like adding machines were common technologies from the late 1800s (Eschner 2017).

4. Information on the number of non-income tax returns is not available before 1950. This paper uses BEA's estimates of federal non-income tax revenue as an extrapolator for the period 1929 to 1949.

used as extrapolators. By construction, these extrapolators are smooth and do not reflect laws that took effect after 2017. Because recent growth is hard to measure, this paper's discussion focuses on the long-term growth impact of including tax forms.

Time Burden for State and Local Government Tax Forms

Official state publications comparable to "The Information Collection Budget of the United States" were not located. State governments each have their own tax forms and their own filing requirements. It would be a mammoth undertaking to measure all the time spent on each state's tax form. Instead, the estimates presented in this paper use industry literature to measure the time spent on state and local government tax forms. One study estimated that state tax form costs accounted for approximately 27 percent of total corporate tax forms in the 1990s (Gupta and Mills 2003). Another study estimated that state-required forms accounted for approximately 30 percent of small business time spent on government-required data in the 1970s (Office of the Chief Counsel for Advocacy 1979). This paper uses BEA's published statistics on the share of taxes[5] paid to the federal government to extrapolate that the federal government share of tax form creation time rose from 36 percent in 1929 to 71 percent in 2021.

In total, this paper presents calculations that Americans spent 11.5 billion hours filling out tax forms in 2017. The tax forms studied include forms that are: completed by ordinary employees filling out paperwork required by human resources, completed in-house by employees who specialize in tax paperwork, outsourced to specialty firms, completed by the self-employed, completed by homeowners, and even forms that are completed by households which are out of scope for GDP. The 11.5 billion hours of tax form time is time is not evenly distributed. Most individuals spend only a few hours per year on tax forms—but a few specialists spend almost their entire work time on tax forms.

This paper's estimate of 11.5 billion hours of tax form time is much higher than a recent paper's estimate of 3.8 billion hours devoted to regulatory paperwork (Trebbi and Zhang 2022).[6] Trebbi and Zhang 2022's much lower number is likely due to their paper's narrower focus. Their paper focuses on in-house time and therefore omits any paperwork that is outsourced to specialty firms.[7] In addition, their paper appears to focus on employees who specialize in regulatory paperwork and therefore omits

5. Mandatory contributions to government social insurance programs like Social Security are treated as taxes.

6. That paper reports that 1.34 percent of a firm's wage bill is devoted to regulatory paperwork. This paper multiplies that percent by the number of workers and the average hours per worker to get total paperwork time.

7. Page 5 of their paper claims that outsourced paperwork is negligible and therefore omitting it does not bias their results much, but the accounting, bookkeeping, and payroll industries employ more than 1 million workers. In addition, other industries like lawyers or financial advisors also provide some outsourced tax help.

time spent by ordinary employees on general forms like the W-4. Finally, their paper focuses on employees and therefore misses time spent by the self-employed, homeowners, and households.

Hourly Costs for Tax Form Time

This paper's primary source for the hourly costs of tax forms is the Quarterly Census of Employment and Wages (QCEW). The QCEW reports that employees in the tax preparation industry (North American Industrial Classification System (NAICS) 541213) received wages of $635 per week in 2017. The American Community Survey reports that employees in the broad accounting sector worked around 40 hours per week. Hence, tax preparation workers earned a little more than $15 per hour.[8] QCEW information on average weekly wages in the tax preparation industry is available back to 1990. One might think that wages in earlier years could be extrapolated using QCEW data on average weekly wages in the overall accounting industry (Standard Industrial Classification (SIC) 872 for 1988–1989 and SIC 893 before 1988). However, the broader accounting industry requires more skill than the tax preparation industry, and so it consequently pays much higher weekly wages. Instead, wages for lower-skilled workers who are employed at credit bureaus and collection services (SIC 732) are used as a proxy for tax preparation worker wages. Before 1975, this paper uses self-reported annual earnings in the Current Population Survey, self-reported annual earnings in the decennial Census of Population, and BEA's price index for PCE professional services as proxies. Between 1929 and 2020, estimated current-dollar tax preparation worker wages rose from $0.58 per hour to $18.72 per hour.

These hourly wage costs are adjusted to include non-wage costs. To start out, employees typically receive non-cash benefits like health insurance. More importantly, businesses have non-labor costs like office space and computers. This paper uses Economic Census data on business revenue and wages to calculate total non-wage costs for the years 1977, 1982, 1987, 1992, 1997, 2002, 2007, 2012, and 2017.[9] For other years, this paper uses the ratio of gross output to labor compensation in BEA's production accounts and the historical ratio of non-cash benefits to cash wages reported in National Income and Product Accounts (NIPA) tables 6.2A-D and 6.3A-D to interpolate and extrapolate non-wage costs. In total, this paper calculates that the current-dollar hourly cost of tax preparation rose from $1.55 per hour in 1929 to $42.31 per hour in 2020.

8. The actual calculations smooth hourly wages over three years to minimize volatility.

9. These non-wage costs are adjusted to exclude earnings for self-employed tax preparation workers. In particular, each establishment is assumed to have one self-employed worker whose earnings are 2,000 times hourly wages.

A cost of $42.31 per hour in 2020 might seem high, but it is actually much lower than the average hourly fee[10] charged by accountants (National Society of Accountants 2020). Similarly, this paper's estimated cost of $3.91 per hour in 1964 is lower than the hourly fee that an early paper used to estimate tax form creation costs (Attiat and Ott 1969). Consistent with the high hourly cost of tax preparation, individuals often choose to file simple tax forms even though they could reduce their tax liability by filing complex tax forms. One academic paper used this preference for simpler forms to calculate a disutility cost of $39 to $131 per hour for untrained individuals in the benchmark year of 2016 (Benzarti 2020). This high disutility cost of creating tax forms matches a survey in which the majority of Americans have a negative feeling about doing their taxes and a quarter of Americans hate doing taxes (Pew Research Center 2013). Hence, this paper's estimates of hourly tax preparation costs are actually conservative.

Tax form creation costs are calculated by multiplying time with hourly cost. This paper estimated earlier that Americans spent 11.5 billion hours on tax forms in 2017 and that the cost of tax creation was $35.66 per hour. Hence, the value of tax forms created was $0.4 trillion (0.0115*35.66) in 2017. Similar calculations are done for each year from 1929 to 2020.

Section 3: Valuing Individual Credit Reports

Economists generally study individual credit reports using longitudinal panels that follow a sample of individuals over time. These longitudinal panels are often called consumer credit panels—but they include home mortgages (Lee and Klaauw 2010) and small business loans (Buchak 2019) that BEA classifies as part of the business sector. Recent papers have used the FCRA's mandated deletion of bankruptcy information to measure the impact of individual credit reports on specific outcomes. This mandated deletion occurs on a specific date that is not connected with changes to an individual's life circumstances, abilities, or other data. Therefore, regression discontinuities can isolate the impact of credit score changes from other factors. Each of the papers that are used in this section studies a different population that has a different local average treatment effect from mandated deletion. For example, deleting bankruptcy information might substantially increase the credit score of someone who has a recent record of reliable bill payment but only slightly increase the credit score of someone who has a recent record of many unpaid bills. This paper assumes a linear relationship between credit scores and the full data contained in an individual credit report. Based on that assumption, this paper adjusts for population differences by dividing each specific outcome with the credit score change associated

10. Hourly fees may be higher because they exclude time spent learning about forms and other non-billable time.

with mandated deletion of bankruptcy information in that particular population. This paper then combines the adjusted specific outcomes to estimate the total outcomes associated with a one-point increase to credit scores and a total service flow from the average credit report.

Lower Interest Rates

Lower interest rates are the most obvious benefit associated with individual credit reports. A study of the auto loan sector estimates that credit scores increased by 17 points and interest rates decreased from 9 percent to 8.8 percent following the removal of a bankruptcy flag (Jansen et al. 2022). In other words, an individual who increases their credit score by one point decreases interest payments by 0.13 percent [(1-8.8/9)/17]. Information on the interest rate paid for other loans was not reported in that paper, so this paper assumes that the same relative decrease applies to all loan types. Consistent with that assumption, another study found that removal of a chapter 13 bankruptcy flag increased the share of individuals with a mortgage from 41 percent to 43 percent (Dobbie et al. 2020). Based on the Economic Census and BEA's housing data, Americans paid $173 billion in consumer interest and $334 billion in mortgage interest in 2017. This paper also estimates that Americans paid approximately $200 billion in implicit interest in 2017 on services that are provided in advance and billed later.[11] Spread among the 245 million Americans with credit reports,[12] this equals $706 of consumer interest, $1,363 of mortgage interest, and $816 of implicit interest per person. Holding total loan balances fixed, an individual who improves their credit score by one point saves $1 (706*0.13%) of consumer loan interest, $2 (1363*0.13%) of mortgage interest, and $1 ($816*0.13%) of implicit interest each year.

Lower Insurance Premiums

Insurance companies routinely quote higher premiums to customers with bad credit (Gusner 2021). To the best of my knowledge, no economics paper has yet studied the impact of an exogenous change in credit scores on premium rates. One study, however, estimated that individuals with credit scores in the first decile have claims 44 percent higher than individuals with credit scores in the 3rd decile (Federal Trade Commission 2007). The Federal Reserve also reports that credit scores ranged from under 500 to over 800 with a 10th percentile score around 550 and a 25th percentile score around 650 (Goodman et al. 2021). Accordingly, each credit score point reduces claims by approximately 0.3 percent [(0.44/1.44)/100]. This paper assumes that the relative impact of credit scores on premium rates is the

---

11. The $200 billion is very approximate and is loosely based on bad debt losses reported in the IRS's Statistics of Corporate Income. Results with that component of credit score benefits excluded are available upon request.

12. These 245 million Americans include individuals whose credit reports do not contain enough data to calculate a credit score but exclude the 18 million American adults without any credit report at all.

same as the relative impact of credit scores on claims. This assumption is reasonable in a competitive insurance market where premiums track expected costs closely and expected costs are highly correlated with expected claims.[13] The Economic Census reports that Americans paid a total of $319 billion in consumer insurance premiums and $109 billion in homeowner insurance premiums in 2017. Spread among the 245 million Americans with credit reports, this works out to $1,302 in consumer insurance premiums and $445 in homeowner insurance premiums per person. Hence, an individual who increases their credit score by one point saves $4 (1,302*0.3%) on consumer insurance premiums and $1 (445*0.3%) on homeowners insurance premiums.

Higher Employee Compensation

Better employment opportunities are one of the greatest benefits associated with high credit scores. In a survey, 14 percent of individuals with poor credit reported missing out on a job opportunity due to their poor credit (Moon 2020). One recent paper estimated that wage earnings increased by 0.9 percent for individuals who benefited from the mandated deletion of a bankruptcy flag (Herkenhoff et al. 2021). These numbers are consistent with a previous study of Swedish pawnshop users that found a 3 percentage point increase in employment for individuals who benefited from the mandated deletion of negative data older than 3 years (Bos et al. 2018) and another study of U.S. checking account users that found spending on consumption increased after banks changed their accounting system in a way that decreased the number of overdrafts recorded (Di Maggio et al. 2020). Even a paper that claims to have found a minimal change to earnings actually found a statistically significant increase for individuals who benefited from the mandated deletion of a chapter 13 bankruptcy flag (Dobbie et al. 2020). This paper's estimate of the wage earnings increase associated with individual credit reports is based on the Dobbie et al. paper because it is the only paper that reports standard credit scores.[14] In particular, that paper reports that an exogenous credit score increase of one point increases employee compensation by 0.035 percentage point (0.2 percentage point employment increase from a 5.71 point score increase).

Laws restricting the usage of credit checks by employers reduce the impact of credit scores on employment. Previous research has shown that these laws increase the job-finding rate for individuals with likely credit problems (Friedberg et al. 2021) (Ballance et al. 2020) but harm minorities (Bartik and Nelson 2019) and individuals living in areas with low average credit scores (Cortes et al. 2022). This paper uses the state laws reported in table OA1 of Friedberg et al. 2021 and expert judgment about the

13. Payment for successful claims is the largest cost for insurance companies but they also have administrative costs related to claims. For example, an insurer might hire a lawyer to fight a claim it believes is spurious.

14. Herkenhoff et al. 2021 reports a proprietary credit score which is not comparable to the standard scores.

impact of local government laws to estimate that the share of Americans who could have their credit checked during a job application fell from 100 percent in 2007 to 77 percent in 2016. Accordingly, this paper assumes that the wage earnings increase associated with a one point credit score increase fell from 0.035 percentage point in 2007 to 0.027 (0.035*0.77) percentage point in 2016 and afterwards. BEA reports that total wages and employee benefits were $12 trillion in 2017. Spread across 263 million American adults, this is $39,544 of wages and employee benefits per person. Hence, a one point credit score increase raises employee compensation by $11 (39,544*0.00027).

Higher Self-employment Earnings

Finally, higher self-employment earnings are an important benefit associated with individual credit reports. A paper that used administrative earnings data from the period 1995 to 2008 estimated that individuals who benefited from the mandated deletion of a bankruptcy flag increased wage earnings by 0.9 percent and increased self-employment earnings by 7.5 percent (Herkenhoff et al. 2021). That paper did not investigate the reason for those higher self-employed earnings, but small business owners are often personally liable for their business's debts, and therefore an improvement to a business owner's individual credit is equivalent to an increase in the business's credit score. The industry literature suggests that businesses with better credit scores have better loan options (Cole 2014) and have an easier time hiring workers (Graham et al. 2022) and buying supplies (Motola 2020). Accordingly, it is not surprising that businesses earn higher profits after their owner's individual credit improves. To the best of my knowledge, no state has passed laws meaningfully restricting the ability of investors, workers, suppliers, or others to check the individual credit of business owners. Therefore, it seems likely that the relative impact of individual credit reports on self-employment earnings did not change between 1995 to 2017. This paper presents calculations that a one point credit score increase raises self-employment earnings by 0.29 percentage point (0.035*7.5/0.9). BEA reports that total proprietors' income was $1.5 trillion in 2017. Spread across 263 million American adults, this is $5,703 in self-employment income per person. Hence, an exogenous credit score increase of one point increase raises self-employment earnings by $16 (5703*0.0029).[15]

In total, this paper calculates that Americans receive $36 (2+1+1+4+1+11+16) per year from a one point improvement in their individual credit score. This $36 of benefits per point is likely a conservative estimate of the total benefits associated with a credit score increase. Individuals sometimes share their credit report with friends or family by cosigning loans or otherwise guaranteeing their obligations.

---

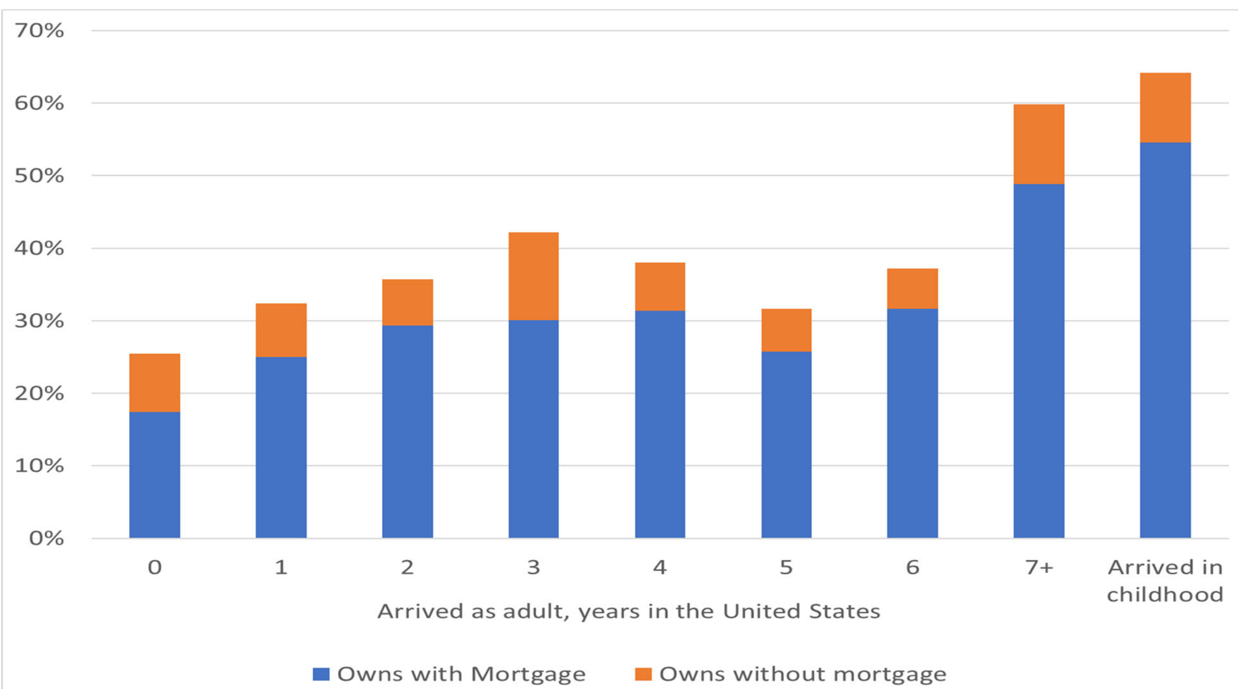
15. The administrative records used by Herkenhoff et al. 2021 report that self-employment earnings are only 6 percent of total earnings. In contrast, BEA's published data show that proprietors' income is 13 percent of total earnings. These different shares are likely related to BEA's adjustment for underreporting and misreporting.

Information on the value of good credit to friends and family was not located. To be conservative, this paper ignores those benefits and focuses on the direct benefits of good credit to the individual only.

Measuring the Average Services from Individual Credit Reports

The papers cited in the previous section focus on individuals who benefited from the deletion of serious negative information. To the best of my knowledge, no economics paper has measured the benefits of the average credit report relative to no credit report. Industry sources suggest that individuals without a valid American credit score are routinely treated as if they have poor credit (Daly 2021). This treatment is particularly striking for recent immigrants who may have good credit in their birth country (French 2020). This paper focuses on prime-aged immigrants who were born in developed countries to parents who already had United States citizenship. These immigrants have had United States citizenship all their lives, and so they have the same rights to work, vote, and travel as individuals born in the United States. Conversely, these immigrants have the same obligation to file United States tax forms, and so they can be assumed to have similar tax forms as individuals born in the United States. This paper uses the mortgage status of those immigrants, as reported in the American Community Survey,[16] to proxy for the treatment of Americans without any credit reports.

Figure 1: Mortgage Status of Immigrants By Time Since Arrival



16. The American Community Survey reports mortgage status by household. This paper assumes that a mortgage is held by the head and their spouse, but not other household members.

Figure 1 shows that individuals who have recently arrived in the United States are much less likely to have a mortgage. The results shown in figure 1 do not control for demographics or housing tenure. Immigrants who have spent more time in the United States tend to be older, and older people are known to be more likely to own a home. Therefore, controlling for demographics decreases the mortgage gap between brand-new immigrants and long-term immigrants from 31 percentage points to 26 percentage points. Furthermore, recent immigrants automatically have a shorter housing tenure because they must have moved to come to the United States. As a robustness check, this paper restricted its sample to individuals who report moving in the past year.[17] This further control slightly reduces the mortgage gap from 26 percentage points to 24 percentage points.

This paper converts the 24 percentage point mortgage difference into a credit score. Earlier research showed that removal of a chapter 13 bankruptcy code raised average credit scores by 5.71 points and the share of the population with a mortgage by 1.74 percentage points (Dobbie et al. 2020). Accordingly, a 24 percentage point increase in the share of the population with a mortgage is roughly equivalent to a 78 point credit score increase [5.71*24/1.74]. Information on the share of immigrants with a credit report was not located, so this paper assumes that immigrants have the same 93 percent (245/263) credit report ownership as the overall adult population. Hence, this paper calculates that people without a credit report are treated as if they had a credit score 84 points (78/0.93) worse than otherwise similar people. This paper calculates that the average American adult receives $2,784 (84*0.93*36) in services from their credit report each year. To be clear, this $2,784 in services is a net number that includes the negative services for individuals with bad credit,[18] and the zero services for individuals without any credit report. Across the population of 263 million adults, this sums up to $0.7 trillion of data services.

Translating Individual Credit Report Services into Individual Credit Report Creation

Information that directly measures the value of newly created reports was not located. In the absence of direct information, this paper uses the standard capital service formula (Jorgenson 1963) to estimate the ratio of new report creation costs to service flows for each year. When the real rate of return is fixed at 7 percent,[19] the 2017 ratio of new report creation costs to service flows is 81 percent. This estimated ratio is then multiplied by the previous section's estimate of $0.7 trillion of capital services to yield an estimate of $0.6 trillion of data creation in the benchmark year of 2017. This 81 percent ratio is sensitive

17. A substantial portion of American Community Survey respondents who report their year of immigration as the current year also report being in their current home for more than one year. This may reflect misreporting.

18. A score 84 points below average is slightly above the 25th percentile of the distribution (Goodman et al. 2021). Individuals can refuse to share credit reports, but this refusal to share implicitly reveals a very low credit score.

19. This paper's actual calculations use a 10 percent nominal rate of return and 3 percent expected price growth.

to the depreciation rate. A later section of this paper shows that the depreciation rate increased and the ratio of new report creation costs to service flows decreased after the Fair Credit Reporting Act of 1970 mandated that most negative information be deleted after 7 years.

This paper extrapolates nominal report creation from 1929 until 2020 using an index that is an equally weighted geometric average of collection costs and expected bad debt.[20] Collection costs are estimated using self-reported wage earnings for bill and account collectors and the ratio of non-wage compensation to wage earnings that was estimated earlier for tax preparation workers. Expected bad debt is estimated using the total balance of individual loans that are dischargeable in bankruptcy (home mortgages and small business loans, but not student loans). These loan data are taken from the Federal Reserve's website and the Historical Statistics of the United States (Department of Commerce 1975). To be clear, neither collection costs nor expected bad debt are perfect proxies for data creation costs. Some collection costs are processing costs that are not directly related to data on the borrower's creditworthiness. Furthermore, some bad debt is due to easy to observe shocks like borrower death and therefore does not yield any additional data on the borrower's creditworthiness. Nevertheless, both series contain components which are related to the creation costs of data on the borrower's creditworthiness. In the absence of specific information on those components, this paper uses the full value of these series as proxies for nominal data creation over time.

Section 4: Nominal GDP and Nominal Household Production

This section combines digital and non-digital tax forms into one intangible asset type and digital and non-digital credit reports into another intangible asset type. This choice is made for a combination of practical and theoretical reasons. Practically, reliable information splitting digital data and non-digital data was not located. Theoretically, the storage method has little impact on data production, data depreciation, or data usage. Accordingly, splitting digital data and non-digital data would add measurement complexity without changing anything theoretically important. Readers might worry that combining digital and non-digital data conflicts with the official guideline's focus on computerized data (United Nations 2008, sec. 10.109). However, national accountants studying the digital economy routinely include similar non-digital products in their analysis. For example, BEA's digital economy satellite account includes television programs throughout the entire time period studied (Highfill and

20. Bad debt is considered a transfer by official national accounting guidelines (United Nations 2008, sec 10.211) and is therefore not a cost for national accounting purposes. However, it is a cost for business accounting purposes.

Surfield 2022) even though the United States television remained analog for several years after the digital economy satellite account started (Cole 2009). Similarly, previous BEA research on specific data-related activities does not distinguish between activities that create digital data and activities that create non-digital data (Calderon and Rassier 2022). Hence, this paper's treatment of non-digital data does not conflict with national accounting practices even if it is inconsistent with the official guidelines.

Figure 2: Financial Data Creation Relative to Other Capital Categories

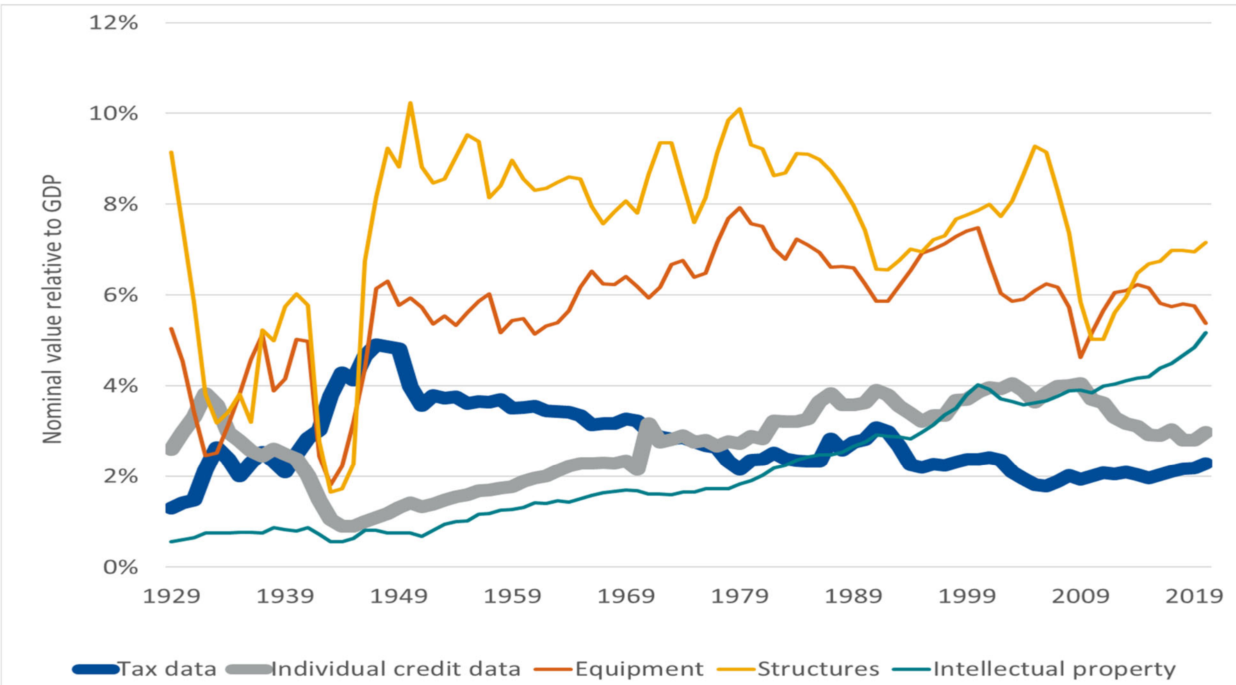


Figure 2 shows that the ratio of tax form creation plus individual credit report creation to nominal GDP has hovered around 5 percent from 1929 onwards. Interestingly, this ratio does not appear to be impacted by long-term economic trends like the introduction of computers or the decline of manufacturing. To remind readers, other examples of long-lived data include: business credit reports (Farboodi et al. 2022), individual arrest records, job references, social media profiles, insurance claims, etc. An unpublished working paper used back-of-the-envelope calculations to estimate that the total ratio of data creation to GDP may be as high as 30 percent (Soloveichik 2023).

Splitting Data Creation by the Funding Sector

The primary source used to split tax forms by sector is a report that lists the time required for each federal tax form (Taxpayer Advocate Service 2008). In most cases, the form's title and description make clear which sector is the creator. For example, employment tax forms (940 and 941 series) are clearly created by employers on behalf of their workers. However, such a simple allocation is not possible for individual income tax forms (1040 series). This paper uses the time required for each type of 1040 form (Frankel 2017) and expert judgment to infer the sector share for that particular form. This paper also uses expert judgment and imputation to estimate sector shares for state and local government tax forms. In 2008, this paper estimates that business tax forms accounted for 8.9 billion hours of data creation time, homeowner tax forms accounted for 1.0 billion hours of data creation time, and consumer tax forms accounted for the remaining 2.2 billion hours of data creation time. NIPA table 2.4.5U reports that consumers purchased $22.3 billion of tax preparation services in 2008. Section 2 of this paper calculated that the hourly cost of tax preparation was $24.41 in 2008. Hence, this paper estimates that consumers purchased 0.9 billion hours of data creation services (22.3/24.41) and household production accounted for the remaining 1.3 billion hours of data creation. This split is important because purchased services are included in GDP but household production is not.

Sources splitting individual credit reports by sector were not located. For now, this paper assumes that the household share of individual credit report creation equals the household share of individual tax form creation. This paper then uses expert judgment to split the residual credit report creation between data which are bundled together with purchased consumer services, data which are created or purchased by homeowners, and data which are created or purchased by businesses. Based on that expert judgment, this paper presents calculations that the $589 billion of individual credit report creation in 2017 is divided as follows: $57 billion of household data creation, $365 billion of purchased consumer data creation services, $24 billion of data created or funded by homeowners, and $122 billion of data created or funded by small businesses. This paper then uses the hourly cost of tax preparation to proxy for the hourly cost of individual credit report creation. Based on that proxy, this paper presents calculations that time spent on individual credit report creation in 2017 is divided as follows: 1.6 billion hours of data creation by households, 10.2 billion hours of purchased consumer data creation services, 0.7 billion hours of data creation by homeowners or funded by homeowners, and 4.0 billion hours of data creation by small businesses or funded by small businesses.

Total consumer time devoted to financial data creation was 2.9 billion (1.3+1.6) hours in 2008.[21] In comparison, the American Time Use Survey (ATUS) reports that Americans spent 2.6 billion hours on "financial management" in the same year and another 0.8 billion hours on "purchasing financial services" (Bureau of Labor Statistics 2023). Given the known time devoted to non-data financial

---

21. In addition, homeowners may also spend some non-work time on data creation.

activities like cash withdrawals, it seems likely that the ATUS tracks most consumer data creation time in household production time. This paper presents calculations of household production value which hold total household production time fixed but use a different labor cost for household data creation than for other household production.

Current Accounting for Data Creation by Sector

Private business expenditures on financial data do not currently impact measures of GDP that are based on industry value added. If financial data creation services are purchased from a specialist firm, they are included as output of the producing company and intermediate input for the purchasing company. This increase to value added for the producing company is precisely canceled out by the decrease to value added for the purchasing company so that sales of private business data creation services to businesses have no net impact on total value added (just as any other intermediate input). Own-account financial data created by private businesses are not included as either output or intermediate input and therefore do not directly impact any industry's value added in official BEA statistics. Similarly, private business financial data do not directly impact measures of GDP that are based on final demand.

Government and nonprofit expenditures on financial data creation do impact measures of GDP. For those sectors, BEA measures output (and implicitly value added) based on costs rather than market revenue. Wages for workers who create financial data are included in total labor costs and therefore implicitly included in measured value-added of the government or non-profit entity producing data. Furthermore, a government purchase of financial data services from a private business raises measured value-added for the private business without lowering measured value-added for the government.

Consumer purchases of financial data creation services are already included in measured GDP. Accounting firms explicitly sell financial data creation services to consumers, and financial firms often implicitly bundle financial data creation services together with their primary financial services. As a result, consumer purchases of financial data creation services are currently included in PCE. Household produced data are not included in measured GDP.

Proposed Accounting for Data Creation and Data Ownership by Sector

This paper proposes to track all data as intangible capital assets. This treatment raises measured business output by the newly included data investment and raises measured government output by the newly included depreciation on data capital. Consumer purchases of data are already included in final output as a component of PCE. In order to avoid double-counting, this paper shifts those purchases from PCE services to PCE consumer durables without changing nominal GDP at all. To remind readers, owner-occupied housing is treated as a business in GDP (United Nations Statistics Division 2008, sec. 6.37), and therefore housing-related financial data are capital assets of the real estate sector rather than consumer durables. Of course, the capital services associated with owner-occupied housing contribute to imputed rent and are therefore indirectly included in PCE. Finally, both household creation of data and the services from consumer data are included in household production but not in GDP.

Data are often owned by a different sector than the one which funded their creation. This paper tracks transfers of financial data from one sector to another as a capital transfer. On the one hand, consumers often fund the creation of data that are used by businesses. For example, a young adult might build a good credit score by using a consumer credit card responsibly and later use the good credit score to take out a small business loan. This is treated as a negative purchase of consumer durables and a positive business investment. On the other hand, businesses often fund the creation of data that are used by consumers. For example, a middle-aged adult might use their small business credit card to improve their credit score and later use that improved credit score to buy a recreational vehicle for their retirement. This is treated as a positive purchase of consumer durables and a negative business investment. These two revisions precisely cancel out so that there is no net impact on GDP. However, data that businesses give to their workers are considered a component of labor compensation and data that businesses give to their owners are considered a component of capital compensation. Therefore, both measured business output and measured GDP rise by the value of the newly included data. For example, consumers often use a W-2 form that is supplied by their employer to apply for personal loans or college financial aid. Finally, data that businesses give to the government are considered in-kind taxes and therefore measured business output rises by the value of the newly included data and measured government output rises by the value of the newly included data depreciation.

Government services from financial data are assumed to equal depreciation (BEA 2022). The normal lifespan for tax forms is 3 years (IRS 2022b) and the normal lifespan for credit reports after FCRA is 7 years. This paper treats tax forms as a one-hoss shay with a lifespan of 3 years and individual credit reports as a one-hoss shay with a lifespan of 7 years (Diewert 2004). To the best of my knowledge, no economics paper has estimated depreciation for individual credit reports before the FCRA was passed in 1970. One recent paper on credit scores calibrated a model where individuals have a 2.5 percent annual probability of death and a 1.1-1.3 percent annual probability of changing their credit risk type (Chatterjee et al. 2020).[22] This low annual probability of changing type is consistent with psychology research showing that credit scores are correlated with personality traits like conscientiousness (Jones

---

22. Other drafts of Chatterjee et al. 2020 calibrate change probabilities and therefore yield different depreciation.

and Kramer 2022) and that such relative personality traits are stable throughout adulthood (Soldz and Vaillant 1999).[23] If these calibrated parameters apply to a pre-FCRA world, then depreciation would have been 3.7 percent (2.5+1.1*.5+1.3*0.5) per year. This calibrated depreciation rate is very close to the 4 percent depreciation rate for data assumed in a recent paper (Statistics Canada 2019). Based on those depreciation rates, this paper estimates that governments receive services equal to 49 percent of their tax form stock each year, 24 percent of their individual credit report stock for each year after 1970, and 4 percent of their individual credit report stock for each year before 1970. Businesses and consumers receive services equal to government services plus a 7 percent real rate of return that is based on the average return for housing and stocks over the past 150 years (Jorda et al. 2019).

Revisions to Nominal GDP Associated with Financial Data

Figure 3: Revisions to Nominal GDP from Capitalizing Tax Forms

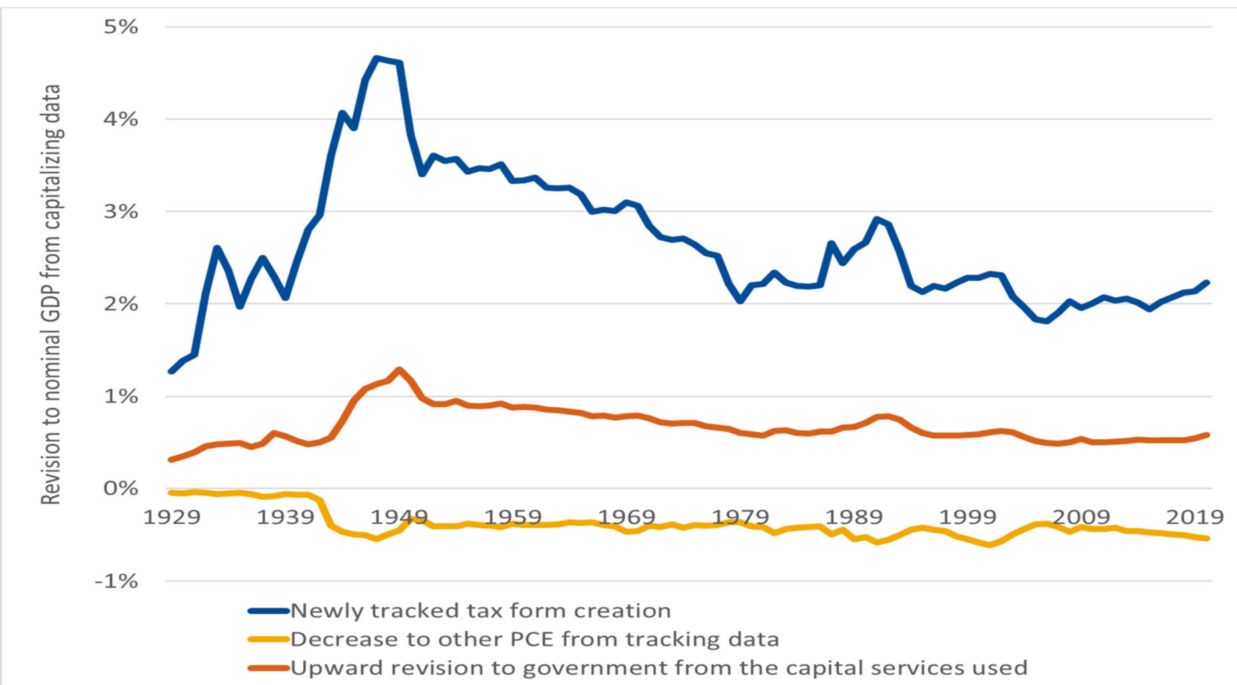
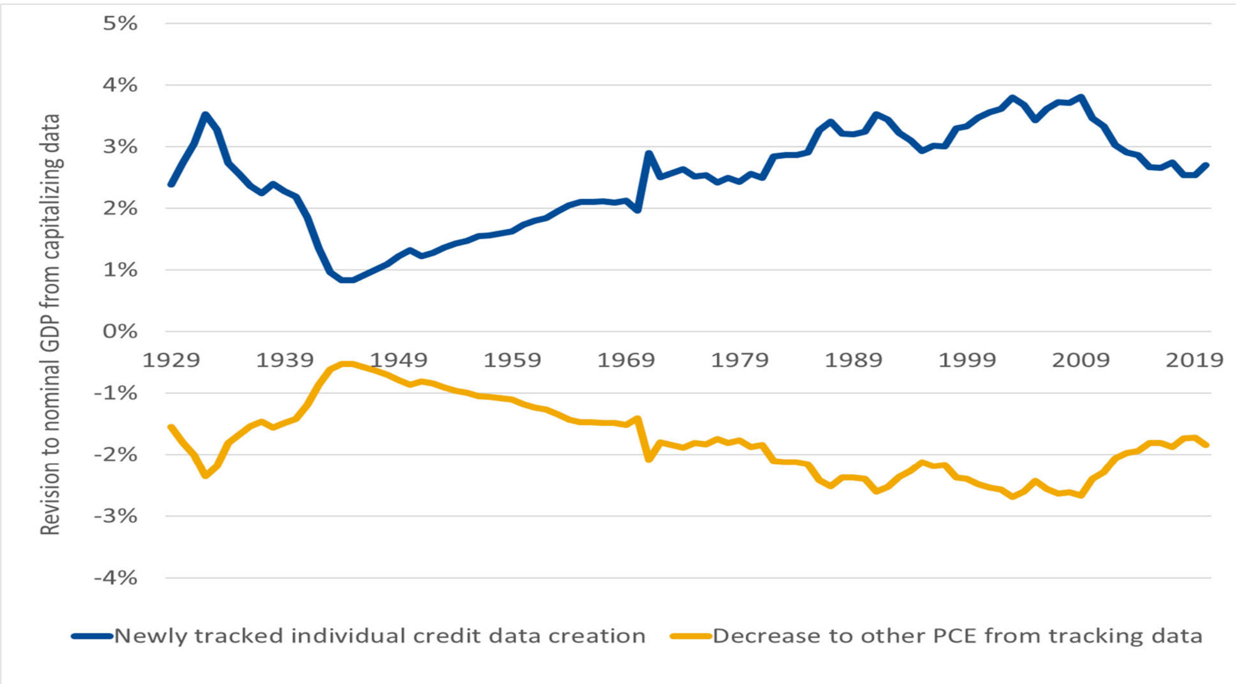


Figure 3 shows that including tax forms increases the level of GDP in World War II much more than it increases the level of GDP in 1929. As a result, measured GDP growth between 1929 and 1945 increases noticeably when tax forms are included in GDP. This growth revision is associated with the introduction of Social Security (DeWitt 2010) and the expansion of personal income taxes in World War II (Thorndike 2022). After World War II, the U.S. Department of the Treasury consolidated the separate forms

23. Average personality often changes over a person's lifecycle. For example, teenagers are more impulsive than middle-aged adults. Data users generally know (approximate) age and can adjust for lifecycle effects.

required for Social Security and for personal income taxes into combined forms and therefore slightly reduced the number of forms required for each employee (Dunlap 1951). The growth of the modern social insurance state has been extensively documented by previous authors (Fishback 2020). Accordingly, the increasing value of tax forms shown in figure 3 is not novel to this paper. However, this paper's focus on the impact of including tax forms as data in GDP is new to the literature.

Figure 4: Revisions to Nominal GDP from Capitalizing Individual Credit Reports



In contrast, figure 4 shows that capitalizing individual credit reports has little net impact on the level of GDP. It is true that the ratio of individual credit report creation to nominal GDP ranges from a minimum of 1 percent during World War II to a maximum of 4 percent during the Great Depression and the Great Recession. However, this variation in measured data creation is mostly canceled out by variation in other PCE. Intuitively, including data mostly shifts spending from one final output category to another without changing the total. Government usage of individual credit reports is minimal, and therefore the line showing government revisions is omitted from figure 4 and future figures on individual credit reports.

Revisions to Nominal Household Production Associated with Financial Data

Figure 5: Revisions to Household Production from Tax Forms

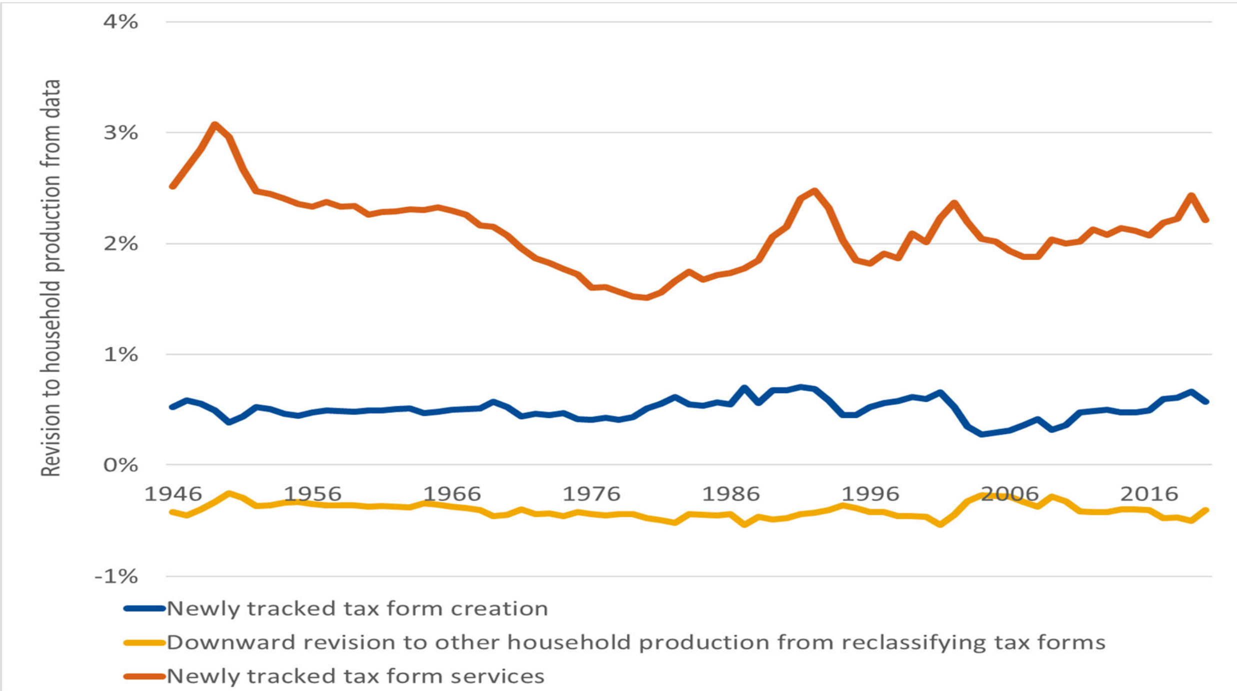
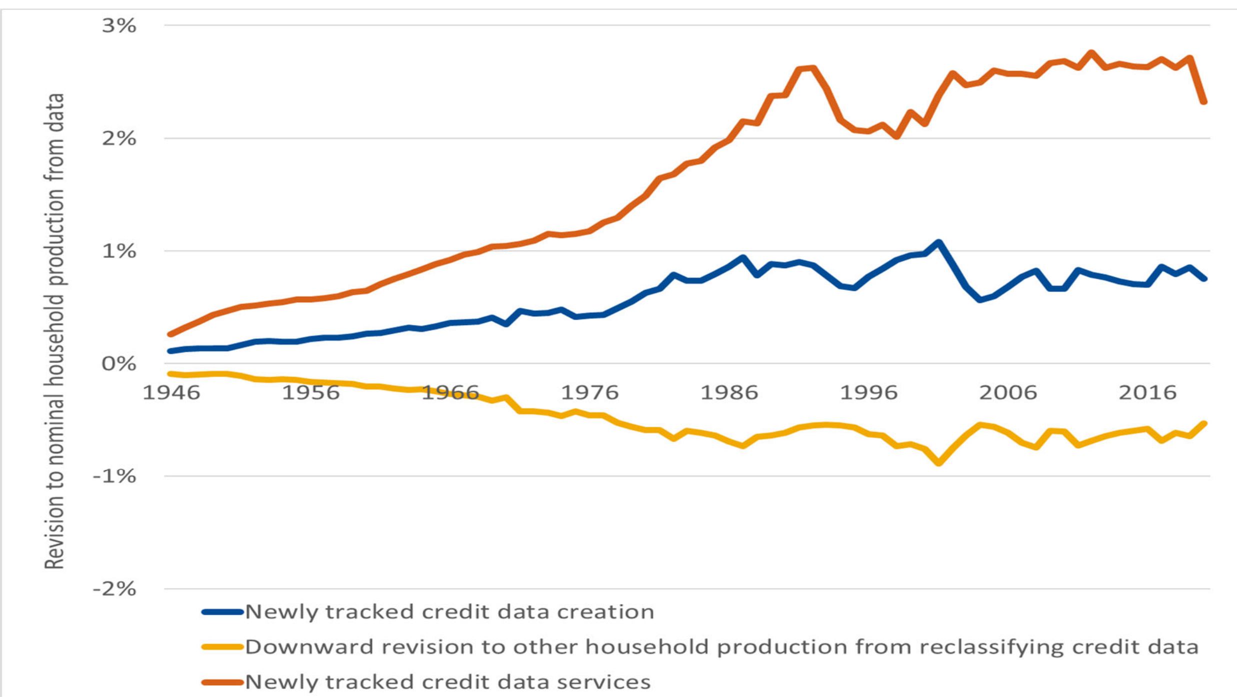


Figure 6: Revisions to Nominal Household Production from Including Individual Credit Reports



Figures 5 and 6 show that including household data creation increases the level of household production by a larger percentage than the level increase to GDP shown in figures 3 and 4. To start out, BEA's household production satellite account includes services from consumer durables in its measure of output (Bridgman et al. 2022). Accordingly, both financial data created in the household and financial data purchased from businesses contribute to household output when they yield services to the household. In addition, this paper presents estimates of total household production which use a specialist wage to value data creation time that is higher than the generalist wage used to value other household production time. As a result, the value of newly tracked data creation is larger than the downward revision to other household production. The net impact of all three revisions is an increase in household production from $4.5 trillion to $4.7 trillion in 2017 and an increase to the nominal growth rate of household production by 0.03 percentage point per year from 1946 to 2020.

Section 5: Data Prices, Real GDP, and Real Household Production

Measuring Data Prices over Time

The primary source used to measure tax form prices is a producer price index (PPI) published by the U.S. Bureau of Labor Statistics (BLS). The series PCU541211541211 tracks all output by offices of certified public accountants from 2003 onwards.[24] This paper uses a similar discontinued series before 2003, NDU5412115412118, as an extrapolator back to 1995. Like other service price indexes, BLS calculates its PPI for certified public accountants by asking survey respondents to estimate the revenue they might receive after performing specific services (Swick et al. 2006). Interested readers can get more information on these price indexes from a paper presented to the Voorburg Group (Borde and Garneau 2001). Before 1995, a weighted average of labor costs and equipment costs are used as proxies. Labor costs are extrapolated using the same QCEW wages and non-wage benefits described earlier. Equipment costs are extrapolated from BEA's published price indexes for equipment relevant to accounting (NIPA table 5.5.4, lines 4, 5, 8, and 9). This paper benchmarks that weighted average to the BLS PPI to construct a consistent price series from 1929 to 2020. This consistent time series shows an average labor productivity growth rate of 1.2 percentage points per year between 1929 and 2021. This productivity growth is almost identical to the 1.2 percentage points per year productivity growth rate estimated from the IRS' collections and costs (Jensen and Lagakos 2019). BLS does not publish a PPI for

24. BLS also publishes a specialized price index for tax preparation and planning (PCU 541211541215). It might seem that this index is more focused on the data studied in this paper and therefore should be used. In fact, that index excludes the initial record-keeping time which accounts for 80 percent of total tax form time (Little 1988).

either collection costs or expected bad debt. However, the basic production process for tax forms and individual credit reports is similar. Hence, this paper assumes that the two data types have similar costs.

This paper adjusts individual credit report quality for the shorter lifespans associated with FCRA. To remind readers, annual depreciation was 3.7 percent per year before 1970. If future revenues are discounted at a 7 percent real rate, then complete deletion of data in 7 years[25] is equivalent to newly created data quality being cut in half [1-(1-0.037)^7]*[(1-0.07)^7]. The impact on data quality is harder to measure when data users can infer deleted data from other sources, when some negative credit data are allowed to last longer than 7 years, or when individuals change their behavior in response to expected deletions. But data quality almost certainly fell noticeably due to the FCRA's 7-year rule.

This paper also adjusts individual credit report quality for recent restrictions on employers checking credit. In theory, one might treat this reduction in services as a reduction in productivity for the data usage sector. Instead, this paper models the reduction in services as a decrease in the data quality and therefore a reduction in the productivity of the data creation sector. Based on the papers reviewed in section 3, this paper estimates that a complete ban on credit checks for employees would reduce the services associated with credit data by 45 percent. The previous section used Friedberg et al. 2021 to estimate that the actual laws passed only restricted credit checks for 23 percent of potential employees. These recent restrictions on employer usage may be slightly offset by a recent increase in landlord usage of individual credit reports (Farivar 2021), but information on the exact impact of the increasing usage by landlords was not located. This paper therefore treats the passage of those restrictions on employer credit checks as a 12 percent [1/(1-0.45*0.23)-1] increase in the quality-adjusted price of data but does not adjust for the growth of landlord credit checks.

Revision to Overall GDP Prices and Household Production Prices

Capitalizing data impacts overall GDP prices through three separate channels. Most importantly, average price growth changes when a new output category with its own price index is added to the national accounts. In addition, BEA currently uses input prices to impute government output prices. As a result of this imputation, measured government output prices change when financial data services are included as an input to government services. Finally, measured PCE prices change when the consumer purchases related to financial data are reclassified from consumer services to the consumer durable of financial data if those two consumer products differ.

25. Chapter 7 bankruptcies are kept on record for ten years. In addition, the written text of the FCRA exempts large loans and high paying jobs. In practice, credit bureaus responded to the planned passage of FCRA by adopting "voluntary" guidelines that did not exempt large loans or high paying jobs (Senate Hearing 1969).

Figure 7: Revision to GDP Prices from Capitalizing Tax Forms

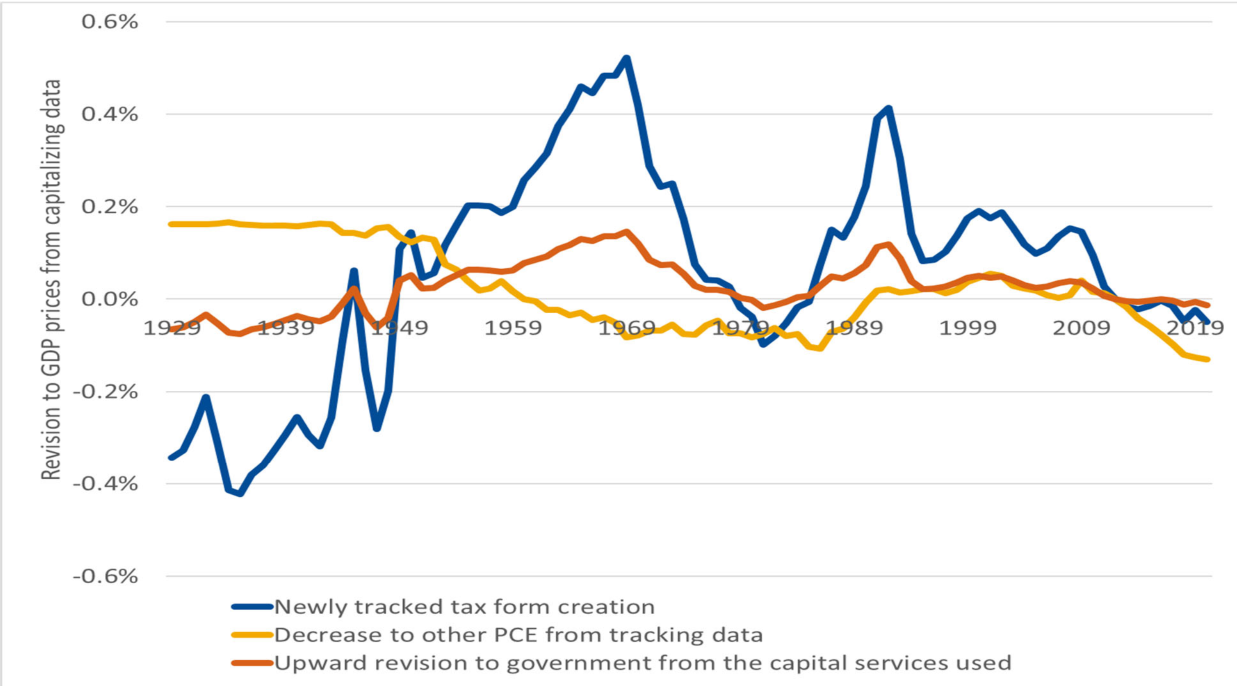


Figure 8: Revision to GDP Prices from Capitalizing Individual Credit Reports

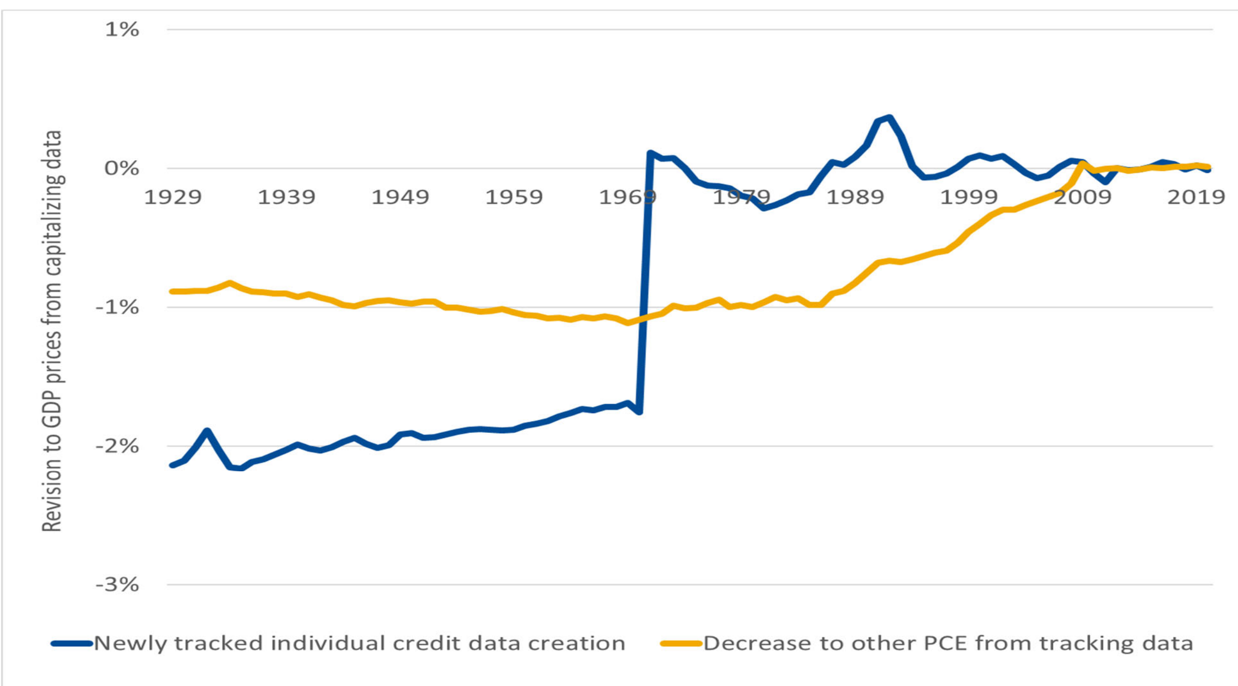


Figure 8 shows that overall GDP prices experienced a large jump in 1971. This jump is entirely due to this paper's assumption that the FCRA raised quality adjusted prices of individual credit reports. This price jump may be important for understanding some of the changes that occurred during the 1970s. Tracking data also impacts household production prices. The relative impact is slightly different because household production prices have risen more rapidly than overall GDP prices (Bridgman et al. 2022). However, the overall trends for household creation of data are not that different from the trends for business creation of data shown in figures 7 and 8. In order to save space, this paper does not show graphs for household production prices.

To be clear, the price jump associated with FCRA may have benefits that offset the higher prices. Individuals value their privacy (Acquisti et al. 2016) and might have enjoyed significant nonmonetary benefits from the deletion of old data. In addition, individual credit reports could have negative externalities. For example, data sharing can reduce competition in the financial sector and therefore raise average prices (He et al. 2022). These policy issues are beyond the scope of GDP and therefore do not directly impact the published national accounts.

Revision to real GDP and real Household Production

Figure 9: Revision to GDP Quantities from Including Tax Forms

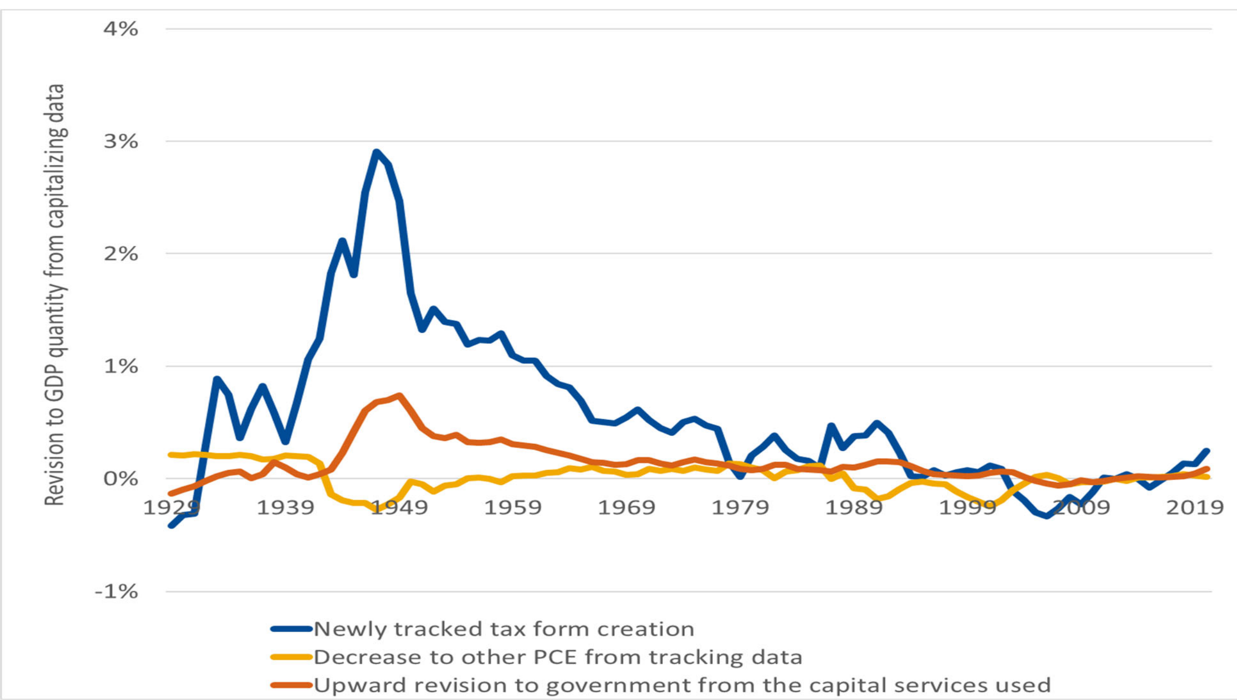
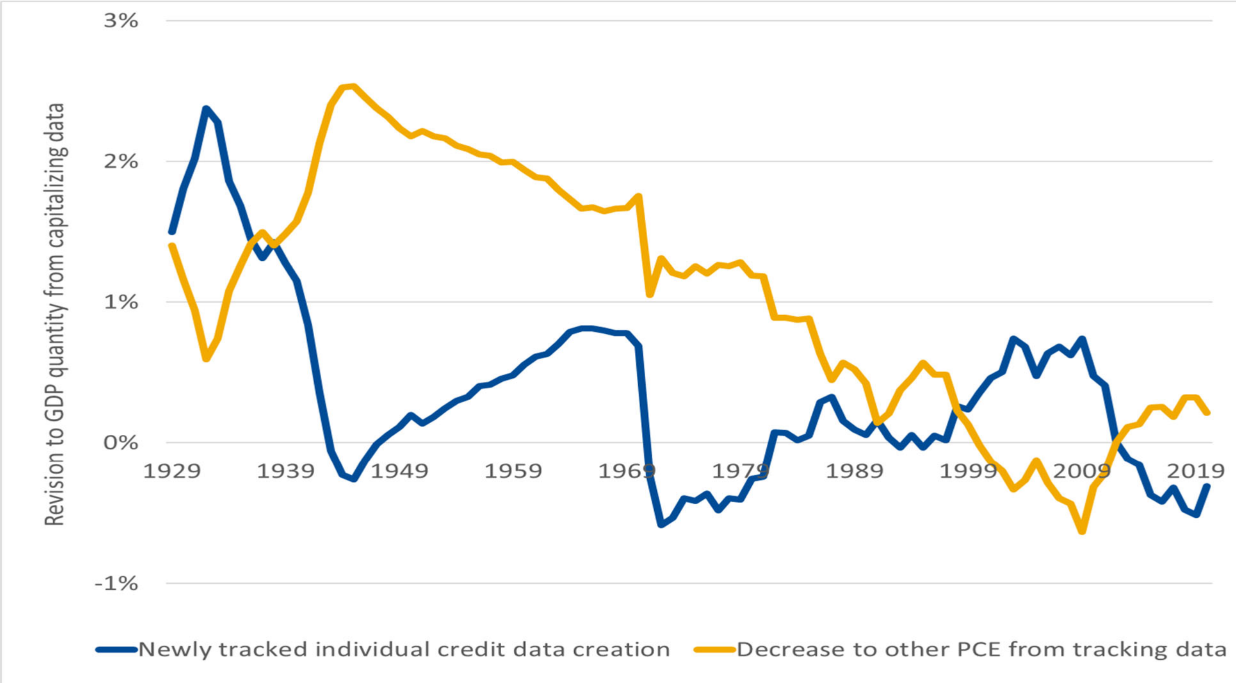


Figure 10: Revision to GDP Quantities from Capitalizing Individual Credit Reports



Figures 9 and 10 show that including financial data changes real GDP growth noticeably. Including tax forms raises growth before World War II, lowers growth immediately after World War II, and then has little impact on measured growth after 1970. In contrast, including individual credit reports has little impact on growth before 1970 or after 1971—but lowers growth in 1971 by 1.6 percentage points.

The real GDP results shown in figures 9 and 10 do not account for inequality. One recent paper argued that deleting negative credit information could be a good way to reduce inequality (Jansen et al. 2022) and another recent paper argued that banning employers from checking credit may help financially distressed individuals find stable employment (Friedberg et al. 2021). In addition, poor credit is often correlated with low income (Beer et al. 2018) and chronic diseases (Dean et al. 2019). So, the FCRA may have benefited financially distressed individuals, individuals with low income, and individuals with chronic diseases even as it decreased overall GDP. The inequality associated with credit scores may be particularly pernicious if naïve consumers are lured by banks into overborrowing (Agarwal et al. 2022) rather than voluntarily choosing to overborrow. BLS publishes information on the distribution of consumption based on the consumer expenditure survey data (Garner et al. 2022) and BEA publishes information on the distribution of income based on the Current Population Survey (Fixler et al. 2020). Future research might use microdata from the Federal Reserve Bank of New York's consumer credit panel or other sources to examine how tracking financial data would impact measured inequality.

Figure 11: Revision to Household Production Quantities from Including Tax Forms

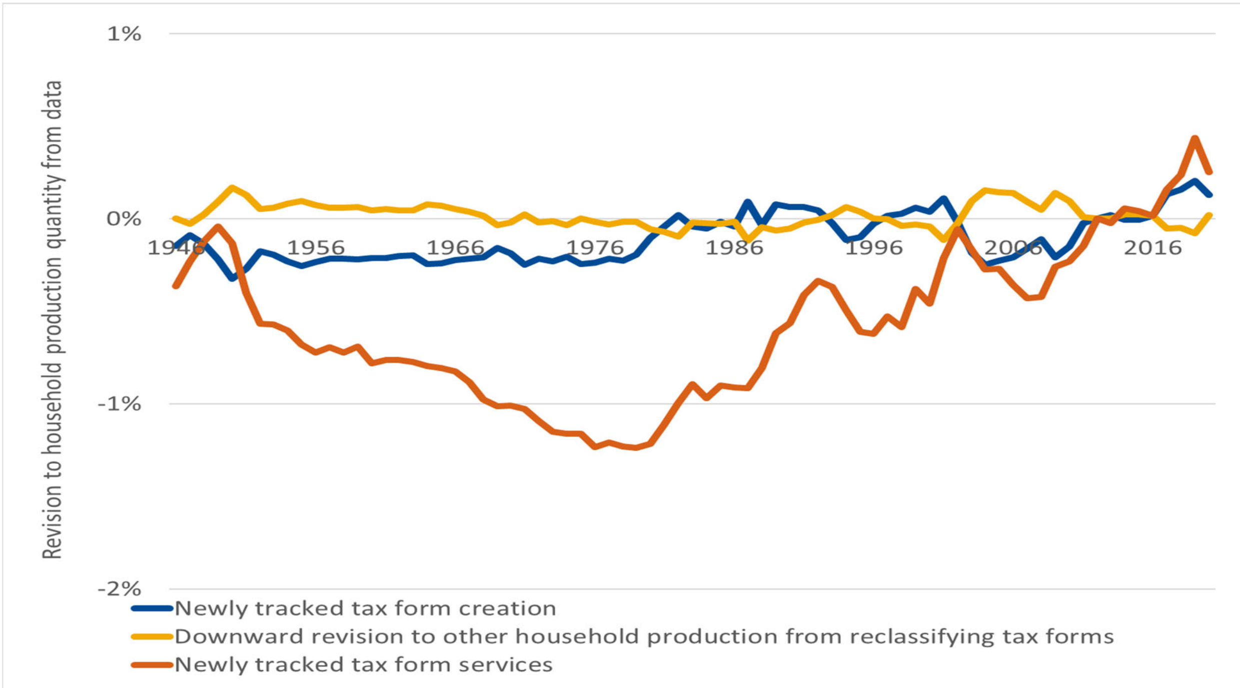
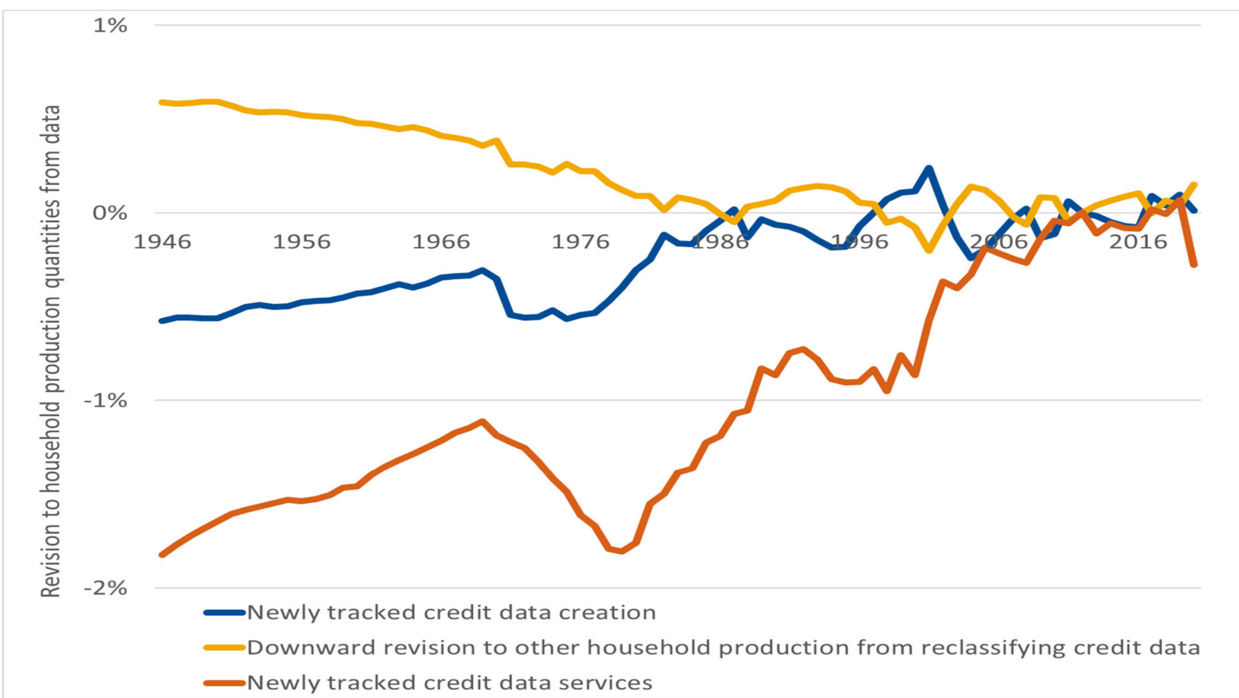


Figure 12: Revision to Household Production Quantities from Including Individual Credit Reports



Figures 11 and 12 show that including financial data increases real household production growth after 1975. By themselves, the revision to real growth shown in those figures is not enough to change the qualitative trends in household production reported in BEA's satellite account (Bridgman et al. 2022). However, household data creation may be much broader than just the two types of financial data studied in this paper. If other types of long-lived data have also experienced the same relative increase, then the overall revision associated with including household data may be enough to change post-1975 qualitative trends. This is a topic for future research.

Figures 9 to 12 do not find much impact on either real GDP growth or real household production growth in recent years. This null result is a stark contrast to other papers that suggest capitalizing data increases GDP growth (Calderon and Rassier 2022) (Mitchell et al. 2022). The difference between this paper and those papers can be easily explained by the type of data that is studied. Other researchers have studied data that is lightly regulated, requires very little labor to create, and that has benefited enormously from computerization. In contrast, the two types of financial data studied in this paper are heavily regulated, very labor intensive to create, and have benefited only as much as the overall economy from computerization.

Section 6: Impact of Financial Data on Business Productivity

The productivity calculations presented in this paper are based on existing industry-level production accounts that track labor, capital services, and intermediate inputs for 61 separate private business sector industries (Garner et al. 2020). These production accounts calculate productivity for each industry separately and then aggregate across the entire economy with Domar weights. This paper calculates revisions to productivity for each industry separately. Each industry has its own outputs and inputs, and so the relative impact of tracking financial data can be very different across industries. For example, both cellphone providers and hospitals implicitly lend money to customers by giving them services upfront and billing them afterwards. If customers do not pay, both industries report the unpaid bill as a negative item on the customers' credit report. Accordingly, this paper assumes that both of these industries produce individual credit data as a secondary product. The price of cellphone hardware and cellphone services have both been falling steadily (Aizcorbe et al. 2019) even as the price of individual credit reports has been stable. Hence, real value-added growth in the telecommunications industry falls when a portion of its output is deflated with the price of data creation rather than the price of cellphone services. In contrast, the price of healthcare services has been rising even as the price of individual credit reports has been stable. Accordingly, real value-added growth in the hospital industry rises when a portion of its output is deflated with the price of data creation rather than the price of healthcare services. This paper then calculates how those revisions change total factor productivity (TFP) in each industry and then aggregates each industry's TFP revision using revised Domar weights to calculate a total TFP revision for the overall private business sector.

This paper makes five revisions to each of the existing industry-level production accounts. First, data that are given without compensation to an owner, worker, or government are considered part of value-added and therefore increase measured business output. For example, the value of a W-2 form that is given to an employee who then uses it to apply for a personal loan is now tracked as output of the business sector. Second, data that are given to a customer as part of a service bundle are shifted from that industry's primary output to the newly recognized secondary output of data creation services. For example, the value of the credit data which are reported by automobile loan companies are shifted from the bank's primary output of financial services to the newly recognized secondary output of credit report creation services. Third, newly included financial data services received from customers, business owners, and employees increase the measured intermediate inputs of data users. For example, the value of the data service associated with a W-2 form that is submitted to an automobile loan company by a loan applicant is now tracked as an intermediate input of the automobile loan company. Fourth, measured output for data users increases by discounts on the list price that they use to implicitly pay customers for access to customer financial data. For example, the value of the financial services provided by an automobile loan company is increased by the implicit payment for the data services associated with an applicant's W-2 form. Fifth, capital services for data users decreases by the portion of capital compensation that is classified as payment for financial data of business owners and labor services for data users decreases by the portion of labor compensation that is classified as payment for financial data of workers. For example, the proprietors' income from a small business is split between general proprietors' income and services associated with the small business owner's individual credit report. Businesses use both data created by businesses and data created by households, and so both types of data services are included identically in productivity account revisions described above.

Figure 13: Impact on Business Productivity from Including Tax Forms

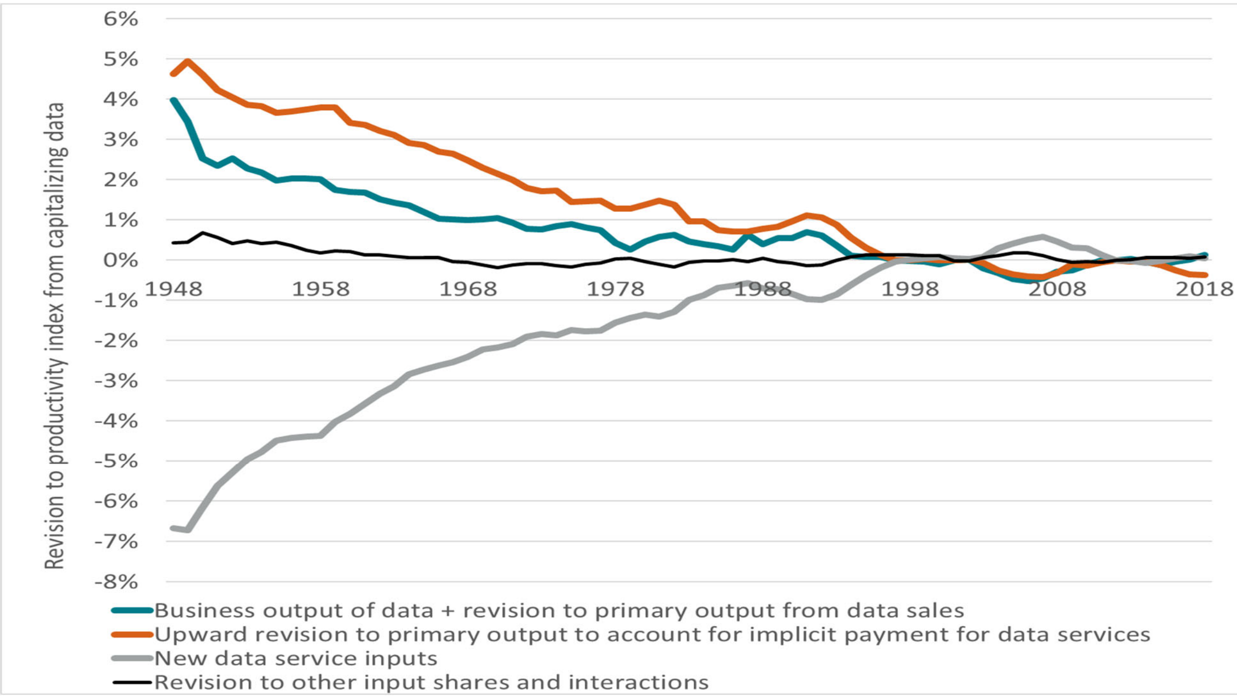


Figure 13 shows that including tax forms had three offsetting impacts on growth before 1980. On the one hand, real output of tax forms grew slower than overall output before 1980. This slower growth reduces overall output growth and therefore reduces measured TFP growth between 1948 and 1980 by 0.11 percentage point per year. On the other hand, real usage of tax form services grew slower than overall GDP. This slower growth has two separate impacts. On the intermediate input side, the slow growth of tax form services reduces overall intermediate input growth and therefore increases measured TFP growth between 1948 and 1980 by 0.17 percentage point. On the output side, the slow growth of tax form services reduces the growth of output that is implicitly bartered for tax form services, overall output growth, and measured TFP growth between 1948 and 1980 by 0.10 percentage point per year. Revisions to the other input shares have a minor additional effect. The net impact of all four revisions is a slight decrease to measured growth before 1980.

However, the small net revision shown in figure 13 is very sensitive to changes in the assumptions made. One major issue is that calculated financial data services depend enormously on the assumed share of tax forms owned by governments, the depreciation rate on privately owned tax forms, and the rate of return on privately owned tax forms. In addition, the result that revisions to other input shares have a minor effect on measured TFP is sensitive to the treatment of required tax forms. Because the productivity revisions shown in figure 13 are so uncertain, this paper does not discuss them further.

Figure 14: Impact on Business Productivity from Including Individual Credit Reports

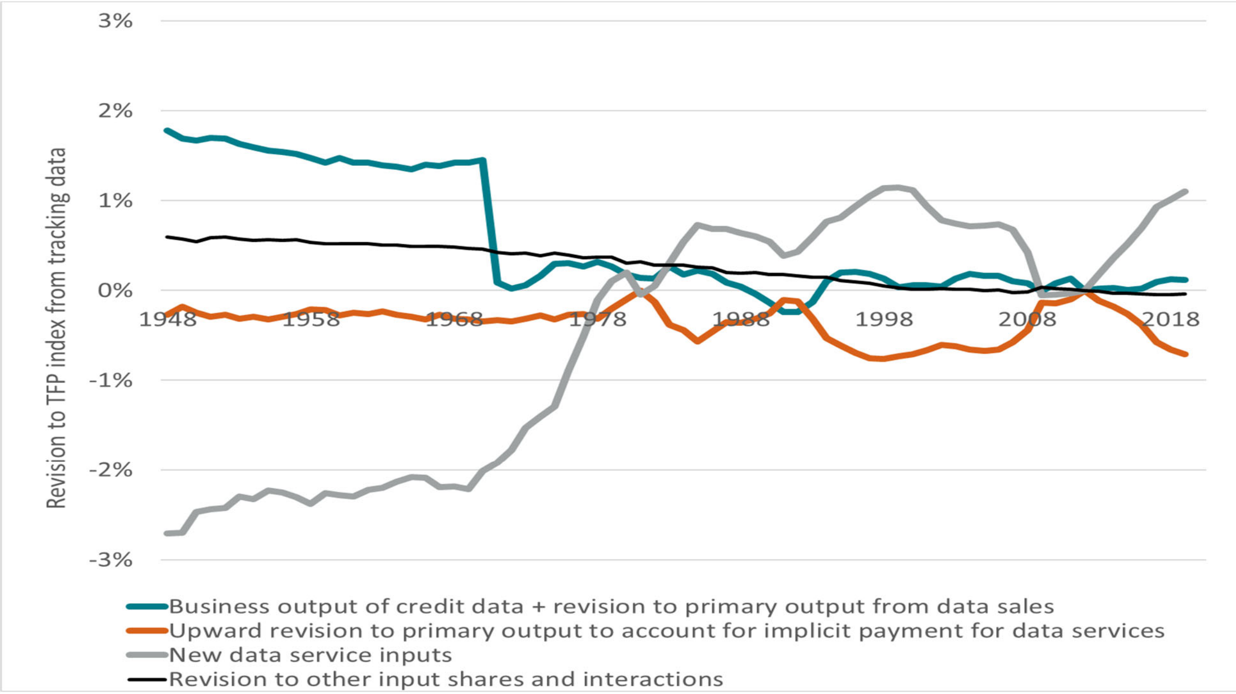


Figure 14 shows that including individual credit reports has two offsetting impacts on productivity in the 1970s. On the one hand, real output of credit data fell suddenly in 1971 due to the passage of the FCRA. This sudden reduction in output is treated as a halving in the productivity of the individual credit data creation sector and a 1.4 percentage point reduction in overall productivity. On the other hand, real usage of credit data fell gradually during the 1970s due to implementation of the FCRA. On the intermediate input side, the gradual fall reduced overall intermediate input growth and increased measured TFP growth between 1970 and 1980 by 0.22 percentage point per year. Unlike figure 13, these two revisions are much less sensitive to changes in the productivity formulas. The price of credit data usage rose when real usage fell and so there is little change in the real growth of output which is bartered for credit data services and little change to measured TFP from that channel. The net impact of these four revisions is a 0.10 percentage point per year increase to productivity growth in the 1970s.

Figure 14 shows that the FCRA may have been important enough to explain macroeconomic trends. It is true that the 1970s revisions associated with individual credit reports shown in figure 14 explain only 8 percent of the total 1970s slowdown (0.10 percent point per year relative to a total slowdown of 1.21 percentage point per year). However, many other types of data were impacted by the FCRA's requirement to delete negative information after 7 years. Employers share workers' criminal histories via background check firms like Retail Credit Co. (Brenton 1964) and security firms like Pinkerton (Friedman 1907). Insurers share health data via the Medical Information Bureau (Greenberg 2022) and claims data via the Comprehensive Loss Underwriting Exchange (Ashton 2022). In addition, other provisions in the FCRA restricted credit bureaus from collecting or sharing unsubstantiated gossip (Fink 1972) and other laws like the 1977 Fair Debt Collection Practices Act added to the further restrictions on data sharing (Griffith 1999). If each of those other four data types (criminal, health, insurance, and gossip) explain the same proportion of the total slowdown as individual credit reports, then 40 percent (8 percent for one data type * 5 data types) of the 1970s slowdown might be explained by the FCRA.

It is possible that online data can partially substitute for the credit bureau data that was regulated by the FCRA. Employers in many states are able to view criminal records directly (Finlay 2009) without going through background check firms or security firms. Search engines like Google allow insurers to find any data on health or previous claims that has been publicly reported (Huddleston 2021). Social media sites like Facebook even allow businesses to access and collect unsubstantiated gossip about workers and customers without involving a credit bureau (Maurer 2018). Data exhaust may also substitute for the tax forms and individual credit bureau data studied in this paper. Data exhaust is naturally generated during ordinary business operations, and therefore often has a minimal creation cost (O'Leary and Storey 2020). The minimal creation cost of data exhaust is currently balanced out by inconsistent formats and numerous errors that require extensive checking (Krishnan 2022). The current digital profiles that data brokers create from online data exhaust are very low quality (Neumann et al. 2019). In contrast, all tax forms use a standard format that is carefully checked by the Federal tax agency (IRS 2023a) and all individual credit reports use a standard format that it is carefully checked by the credit bureau (Datalinx 2022). This standardization and checking currently make tax forms and individual credit reports much more expensive to create than data exhaust. Data scientists have also started using complex online footprints to augment individual credit reports (Berg et al. 2019). If new software and other technologies are able to use cheap data exhaust in place of expensive standardized data, then productivity for data users would increase noticeably. These substitutes are a topic for future research.

Conclusion

This paper presented case studies of two types of financial data: tax forms and individual credit reports. These two types of data are simple enough that they existed long before computers (Brenton 1964) and can be managed by workers with only a high school degree (Bureau of Labor Statistics 2022a) (Weedmark 2021), but are so valuable that they had a creation cost of $1 trillion in 2017. Including financial data increases real GDP growth between 1929 and 1950 by 0.12 percentage point per year and decreases real GDP growth in 1970 by 1.6 percent point.

Bibliography

Acquisti, A., C. Taylor, and L. Wagman, (2016) "The Economics of Privacy," *Journal of Economic Literature* 54(2), pages 442–92.

Agarwal, S., A. Presbitero, A. Silva, and C. Wix (2022) "Who Pays for Your Rewards? Redistribution in the Credit Card Market," *SSRN Working Paper* 4412641.

Aizcorbe, A., D. Byrne, and D. Sichel (2019) "Getting Smart About Phones: New Price Indexes and the Allocation of Spending Between Devices and Service Plans in Personal Consumption Expenditures," *NBER Working Paper* 25645.

Ashton, J. (2022) "Small Steps to Major Change," posted sometime in 2022 on stories.relx.com and accessed in October 2022.

Attiat, F. and D. Ott, (1969) "Private Costs of Filing Returns," chapter in *Studies in Substantive Tax Reform*, edited by Willis, A., American Bar Foundation.

Baker, G. (2022) "10 Proof of Income Documents Landlords Use to Verify Income," posted at *Fitsmallbusiness* on August 18, 2022 and access January 17, 2023.

Ballance, J., R. Clifford, and D. Shoag (2020) "No More Credit Score": Employer Credit Check Bans and Signal Substitution," *Journal of Labor Economics* 63.

Bartik, A. and S. Nelson (2019) "Deleting a Signal: Evidence from Pre-Employment Credit Checks," *Becker Friedman Institution Working Paper* 2019-137.

Beer, R., F. Ionescu, and G. Li (2018) "Are Income and Credit Scores Highly Correlated?" posted at *FEDS Notes* on August 13, 2018, and accessed in October 2022.

Benzarti, Y. (2020) "Estimating the Costs of Filing Tax Returns and the Potential Savings from Policies Aimed at Reducing These Costs," chapter in *Tax Policy and the Economy*, University of Chicago Press.

Berg, T., V. Burg, A. Gombovic, and M. Puri (2019) "On the Rise of FinTechs: Credit Scoring Using Digital Footprints," *The Review of Financial Studies* 33(7), pages 2845–2897.

Borde, F. and G. Garneau (2001) "A Producer Price Index for Accounting Services," draft available at https://www.voorburggroup.org/Documents/2001%20orebro/papers/2001-036.pdf.

Bos, M., E. Breza, and A. Liberman, (2018) "The Labor Market Effects of Credit Market Information," *The Review of Financial Studies* 31(6), pages 2005–2037.

Brenton, M. (1964) *The Privacy Invaders*, Fawcett World Library, New York City, New York.

Bridgman, B., A. Craig, and D. Kanal, (2022) "Accounting for Household Production in the National Accounts: An Update 1965–2020," *Survey of Current Business* 2022 (2).

Buchak, G. (2019) "Financing the Gig Economy," *SSRN Working Paper* 3766338.

Bureau of Economic Analysis (2022) "Concepts and Methods of the U.S. National Income and Production Accounts," https://www.bea.gov/resources/methodologies/nipa-handbook/pdf/all-chapters.pdf, accessed January 17, 2023.

Bureau of Labor Statistics (2022a) *Occupational Outlook Handbook*, Bill and Account Collectors, at https://bls.gov/ooh/office-and-administrative-support/bill-and-account-collectors.htm, accessed in October 2022.

Bureau of Labor Statistics (2022b) *May 2021 National Occupational Employment and Wage Statistics*, at https://www.bls.gov/oes/current/oes_nat.htm, accessed in October 2022.

Bureau of Labor Statistics (2023) *Time Spent in Detailed Primary Activities and Percent of the Civilian Population Engaging in Each Activity, Averages Per Day Be Sex, Annual Averages 2003–2021* at https://www.bls.gov/tus/tables.htm, accessed in January 2023.

Burke, J., J. Jamison, D. Karlan, D. Mihaly, and J. Zinman (2019) "Credit Building or Credit Crumbling? A Credit Builder Loan's Effects on Consumer Behavior, Credit Scores And Their Predictive Power," *NBER Working Paper* 26110.

Calderon, J.B., and D. Rassier (2022) "Valuing Stocks and Flows of Data Assets for the U.S. Business Sector," Presentation to the BEA Advisory Committee Meeting on May 13.

Carpenter, C. (1910) "Banks are Almost Human," *The Houghton Line* 6(1).

Chan, T., N. Hamdi, X. Hui, and Z. Jiang (2021) "The Value of Verified Employment Data for Consumer Lending: Evidence from Equifax," *SSRN Working Paper* 3556554.

Chatterjee, S., D. Corbae, K. Dempsey, and J. Rios-Rull (2020) "A Quantitative Theory of the Credit Score," *NBER Working Paper* 27671.

Cole, T. (2009) "Congress Backs Delay of Digital TV Switch to June," *National Public Radio* article, posted February 4 2009 and accessed February 2023.

Cortes, K., A. Glover, and M. Tasci (2022) "The Unintended Consequences of Employer Credit Check Bans for Labor Markets," *The Review of Economics and Statistics* 104(5), pages 997–1009.

Coyle, D. (2022) "Socializing Data," *Daedalus* 151 (2), pages 348–359.

Coyle, D., and S. Diepeveen (2021) "Creating and Governing Social Value from Data," presented at the IARIW-ESCoE Conference in November 2021 and accessed in October 2022.

Dackehag, M., L. Ellegard, U. Gerdtham, and T. Nilsson (2019) "Debt and Mental Health: New Insights About the Relationship and the Importance of the Measure of Mental Health," *European Journal of Public Health* 29(3), pages 488–493.

Daly, L. (2021) "5 Costly Consequences of Having No Credit History," posted at *The Ascent, A Motley Fool Service* on July 26, 2021, and accessed in October 2022.

Datalinx (2022) "What is Metro 2 Format for Credit Reporting?" posted at *Datalinx.com* and accessed in October 2022.

Dean, L., E. Knapp, S. Snguon, Y. Ransome, D. Qato, and K. Visvanathan (2019) "Consumer Credit, Chronic Disease and Risk Behaviors," *Journal of Epidemiology & Community Health* 73(1).

DeNicola, L. (2020) "Can Businesses Run My Credit Without My Permission," posted at *Experian.com* on January 16, 2020, and accessed in October 2022.

Department of Commerce (1975) *Historical Statistics of the United States: Colonial Times to 1970*, Washington, D.C.: Department of Commerce, September 1975, accessed October 2022.

DeWitt, L. (2010) "The Development of Social Security in America," *Social Security Bulletin* 70(3).

Di Maggio, M., A. Ma, , and E. Williams, (2020) "In the Red: Overdrafts, Payday Lending and the Underbanked," *NBER Working Paper* 282420.

Di Maggio, M., D. Ratnadiwakara, and D. Carmichael (2022) "Invisible Primes: Fintech Lending with Alternative Data," *NBER Working Paper* 29840.

Diewert, W.E. (2004) "Issues in the Measurement of Capital Services, Depreciation, Asset Price Changes, and Interest Rates," Discussion Paper 04-11 http://papers.economics.ubc.ca/legacypapers/dp0411.pdf.

Dobbie, W., P. Goldsmith-Pinkham, N. Mahoney, and J. Song (2020) "Bad Credit, No Problem? Credit and Labor Market Consequences of Bad Credit Reports," *Journal of Finance* 75(5), pages 2377–2419.

Dunlap, J. (1951) "Inspection of Employer's Quarterly Federal Tax Returns, Form 941, by State Unemployment Compensation Commissions." *Supplementing Mimeograph* 6310 (C.B 1948-2, 149).

Dynarski, S., J. Scott-Clayton, and M. Wiederspan (2013) "Simplifying Tax Incentives and Aid for College: Progress and Prospects," *Tax Policy and the Economy* (27).

Eschner, K. (2017) "How America's First Adding Machine is Connected to 'Naked Lunch,' " posted on *Smithsonianmag.com* on August 21, 2017, accessed January 17, 2023.

Eurostat (2020) "Recording and Valuation of Data in National Accounts," Working paper 3.4 https://unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.20/2020/mtg1/3.4_Recording_of_Data_in_NA_Eurostat_June_2020_after_SG_comments.pdf.

Executive Office of the President of the United States (1981–2020) "*Information Collection Budget of the United States,*" posted at https://www.whitehouse.gov/omb/information-regulatory-affairs/reports/#ICB, accessed January 17, 2023.

Farboodi, M. and L. Veldkamp (2022) "A Model of the Data Economy," *NBER Working Paper* 28427.

Farboodi, M., D. Singal, L. Veldkamp, and V. Venkateswar (2022) "Valuing Financial Data," working paper posted at https://sites.google.com/site/maryamfarboodi/research?authuser=0&pli=1.

Farivar, C. (2021) "Tenant Screening Software Faces National Reckoning," NBCnews.com, posted March 14 2021 and accessed February 4 2023.

Federal Accounting Standards Advisory Board (2022) "FASAB Handbook of Federal Accounting Standards and Other Pronouncements, as Amended" posted online at https://files.fasab.gov/pdffiles/2022_%20FASAB_%20Handbook.pdf, accessed January 17, 2023.

Federal Trade Commission (2007) "Credit-Based Insurance Scores: Impact on Consumers of Automobile Insurance," Report to Congress.

Federal Trade Commission (2012) *Credit Repair: How to Help Yourself*, posted online in November 2012 and accessed in October 2022.

Feldmann, J. (2017) "What Should Employers Be Aware of When Requesting W-2 Forms From Job Applicants?" posted on *Forbes.com* on June 6, 2017, and accessed January 17, 2023.

Fink, V. (1972) “Consumer Protection: Regulation and Liability of the Credit Reporting Industry,” *Notre Dame Law Review* 47(5), pages 1291–1304.

Finlay, K. (2009) “Effect of Employer Access to Criminal History Data on the Labor Market Outcomes of Ex-Offenders and Non-Offenders,” *Studies of Labor Market Intermediation*, University of Chicago Press.

Fishback, P. (2020) “Social Insurance and Public Assistance in the Twentieth-Century United States,” *Journal of Economic History* 80(2), pages 311–350.

Fixler, D., M. Gindelsky, and D. Johnson (2020) “Measuring Inequality in the National Accounts,” *BEA Working Paper* 2020-3.

Fox, M. (2021) “A Third of Americans Found Errors on Their Credit Reports. Here’s How to Fix Those Mistakes,” *CNBC.com*, posted June 11 2021 and accessed in February 2023.

Frankel, M. (2017) “The Average American Spends This Much Time on Their Tax Return. How Do You Compare?” posted at *The Motley Fool* on February 15, 2017, accessed January 18, 2023.

French, S. (2020) “The Best Ways for Immigrants to Build Credit,” posted at the *WireCutter of the New York Times* on April 8, 2020, and accessed in October 2022.

Friedberg, L., R. Hynes, and N. Pattison (2021) “Who Benefits from Bans on Employers’ Credit Checks?” *The Journal of Law and Economics* 64(4).

Friedman, M. (1907) *The Pinkerton Labor Spy*, Wilshire Book Co., New York City, New York.

Garner, C., J. Harper, M. Russell, and J. Samuels. (2020) “Integrated Industry-Level Production Account for the United States and the Sources of U.S. Economic Growth Between 1987 and 2018,” *Survey of Current Business* 100.

Garner, T., R. Martin, B., Matsumoto, and S. Curtin (2022) “Distribution of U.S. Personal Consumption Expenditures for 2019: A Prototype Based on Consumer Expenditure Survey Data,” *BLS Working Paper* 557.

Goodman, S., G. Li, A. Mezza, and L. Nathe (2021) “Developments in the Credit Score Distribution over 2020,” posted at *FEDS Notes* on April 30, 2021, and accessed in October 2022.

Gordon, E. (1949) “Income Tax Penalties,” *Tax Law Review* 5(November), pages 131-198.

Greenberg, B. (2022) “MIB Insurance Report: What You Need to Know,” posted at insurist.com on July 19th 2022 and accessed in October 2022.

Griffith, E. (1999) “The Fair Debt Collection Practices Act—Reconciling the Interests of Consumers and Debt Collectors,” *Hofstra Law Review* 28(1), pages 1–66.

Cole, R. (2014) “Credit Scores and Credit Market Outcomes: Evidence from the Survey of Small Business Finances and the Kauffman Firm Survey,” *Krahenbuhl Global Consulting* report SBAHQ-12-M-0168.

Graham, J., Kim, H., Li, S., and J. Qiu (2022) “Employee Costs of Corporate Bankruptcy,” *Federal Reserve Bank of Chicago Working Paper* 2022-09.

Gupta, S. and L. Mills (2003) “How Do Differences in State Corporate Income Tax Systems Affect Compliance Cost Burdens?” *National Tax Journal* 56(2), pages 355–371.

Gusner, P. (2021) “How Much Do Car Insurance Rates Go Up Because of Poor Credit?” posted at *Forbes Advisor* on December 6, 2021, and accessed in October 2022.

Hartley, P (2019) “What is Credentialing and How Does it Work?” posted at *Datalinx.com* on June 28, 2019, and accessed in October 2022.

Haselmann, R., C. Leuz, and S. Schreiber (2022) “Know Your Customer: Informed Trading by Banks,” *NBER Working Paper* 30521.

He, Z., J. Juang, and J. Zhou (2022) “Open Banking: Credit Market Competition When Borrowers Own the Data,” posted at https://conference.nber.org/conf_papers/f165322.pdf.

Herkenhoff, K., G. Phillips, and E. Cohen-Cole (2021) “The Impact of Consumer Credit Access on Self-Employment and Entrepreneurship,” *Journal of Financial Economics* 141(1), pages 345–371.

Highfill, T. and C. Surfield (2022) “New and Revised Statistics of the U.S. Digital Economy, 2005-2021,” BEA report, posted November 2022 and accessed February 2023.

Huddleston, C., (2021) “How Life Insurance Companies Get Intel On You,” posted at *Forbes Advisor* on March 26, 2021, and accessed in October 2022.

Huntley, C. (2019) "20+ Credit Repair Statistics & Facts for 2022 (Survey)" posted at *Finmasters.com* on October 11, 2019,[26] and accessed in October 2022.

Internal Revenue Service (2017) *11.3.34 Disclosure for Non-tax Criminal Violations*, posted at https://www.irs.gov/irm/part11/irm_11-003-034 on July 17, 2017, and accessed January 17, 2023.

Internal Revenue Service (2022a) "Common Tax Return Mistakes That Can Cost Taxpayers," posted at https://www.irs.gov/newsroom/common-tax-return-mistakes-that-can-cost-taxpayers and accessed January 17, 2023.

Internal Revenue Service (2022b) "Topic No. 305 Recordkeeping," posted at https://www.irs.gov/taxtopics/tc305 December 12, 2022, and accessed January 18, 2023.

Internal Revenue Service (2023a) "IRS Submission Processing Pipeline," accessed January 17, 2023, https://www.irsvideos.gov/Professional/IRSWorkProcesses/SubmissionProcessingPipeline.

Internal Revenue Service (2023b) "Your Rights as a Taxpayer," accessed March 21st 2023, https://www.irs.gov/pub/irs-pdf/p1.pdf.

Jansen, M., F. Nagel, C. Yannelis, and A. Zhang (2022) "Data and Welfare in Credit Markets," *NBER Working Paper* 30235.

Jensen, A. and D. Lagakos (2019) "Productivity in the U.S. Tax Collection, 1929 to 2017," unpublished manuscript.

Jones, A. and R. Kramer (2022) "The Relationship Between Financial Behaviours and Personality," https://bpb-eu-w2.wpmucdn.com/blogs.lincoln.ac.uk/dist/3/8345/files/2022/07/Academic-report.pdf

Jones, C., and C. Tonetti, (2020) "Nonrivalry and the Economics of Data," *American Economic Review* 110(9), pages 2819–58.

Jorda, O., K. Knoll, D. Kuvshinov, M. Schularick, and A. Taylor (2019) "The Rate of Return on Everything, 1870-2015," *The Quarterly Journal of Economics* 134(3), pages 1225-1298.

Jorgenson, D. (1963) "Capital Theory and Investment Behavior," *American Economic Review* 53(2), pages 247–259.

26. The title includes "2022," but the study is definitely from 2019 because it was mentioned in news articles then.

Kiviat, B. (2019) "The Art of Deciding with Data: Evidence from How Employers Translate Credit Reports Into Hiring Decisions," *Socio-Economic Review* 17(2), pages 283–309.

Krishnan, S. (2022) "What is Data Cleaning? How Data Cleaning Can Be Done?," posted at *Geek Culture* on September 30, 2022, and accessed in October 2022.

Lee, D., and W. van der Klaauw (2010) "An Introduction to the FRBNY Consumer Credit Panel," *Federal Reserve Bank of New York Staff Reports* 479.

Levy, R., L. Cummings, J. Hogarth, K. Asrow, and T. Ladha (2016) "Secured Credit Cards: Innovating at the Intersection of Savings and Credit," *Center for Financial Services Innovation Report*.

Little, A. (1988) "Development of Methodology for Estimating Taxpayer Paperwork Burden: Final Report to Department of Treasury, Internal Revenue Service," *Coopers & Lybrand*.

Manzi, M. (2022) "How Much is Your Personal Information Worth on the Dark Web," posted at *Geekflare* on March 4, 2022, and accessed in October 2022.

Marquez, R. (2002) "Competition, Adverse Selection, and Information Dispersion in the Banking Sector," *The Review of Financial Studies* 15(3), pages 901–926.

Martin, A. (2022) "What Documentation is Required for Personal Loans," posted on Bankrate.com on October 10, 2022, accessed January 17, 2023.

Maurer, R. (2018) "Screening Candidates' Social Media May Lead to TMI, Discrimination Claims," posted at *shrm.org* on April 23, 2018, and accessed in October 2022.

Mitchell, J., M. Lesher, and M. Barberis, (2022) "Going Digital Toolkit Note: Measuring the Economic Value of Data," OECD Directorate for Science, Technology and Innovation.

Moon, J. (2020) "Credit Sesame Survey: The Cost of Poor Credit is More than Just Financial," posted at *Credit Sesame* on December 17, 2020, and access in October 2022.

Motola, C. (2020) "Why You Can't Get Equipment Financing with No Credit Check & What To Do Instead," posted at *MerchantMaverick* on December 16th and accessed on March 24th 2023.

National Center for O*Net Development (2022) *O*Net Online*, https://www.onetcenter.org/database.html accessed in October 2022.

National Society of Accountants (2020) “2020-2021 Income and Fees of Accountants and Tax Preparers in Public Practice,” posted in 2020 and accessed February 4, 2023.

Neumann, N., C. Tucker, and T. Whitfield (2019) Frontiers: How Effective is Third-Party Consumer Profiling? Evidence from Field Studies,” *Marketing Science* 38(6), pages 918–926.

Newman, S., and B. Frontino (2015) “Recent Developments in Data Furnisher Liability Under the Fair Credit Reporting Act,” *The Banking Law Journal* (January), pages 21–26

O‘Leary, D., and V. Storey (2020) “Discovering and Transforming Exhaust Data to Realize Managerial Value” *SSRN Working Paper* 3740610.

Office of the Chief Counsel for Advocacy (1979) “Government Paperwork and Small Business: Problems and Solutions,” *Small Business Administration*.

Orman, S. (2019) “Give Yourself Credit," podcast posted at *Suzeorman.com* on April 28, 2019, accessed in October 2022.

Pew Research Center (2013) “A Third of Americans Say They Like Doing Their Income Taxes,” reported posted on April 11, 2013, and accessed January 17, 2023.

Posner, R. (1968) “Natural Monopoly and Its Regulation,” *Stanford Law Review* (21), pages 548–637.

Rassier, D., Kornfeld, R., and Strassner, E. (2019) “Treatment of Data in National Accounts,” *Paper prepared for BEA Advisory Committee*, posted in May 2019 and accessed in February 2023.

Redford, E. (1941) “History of the National Retail Credit Association,” Honors Theses available at https://scholar.utc.edu/cgi/viewcontent.cgi?article=1294&context=honors-theses.

Samuelson, P. (1954) “The Pure Theory of Public Expenditure,” The Review of Economics and Statistics 36(4), pages 387–389.

Santucci, L. (2016) “The Secured Credit Card Market,” Payments Card Center Discussion Paper.

Senate Hearings (1969) “A Bill to Enable Consumers to Protect Themselves Against Arbitrary, Erroneous, and Malicious Information,” *Hearings Before the Subcommittee on Financial Institutions of the Committee on Banking and Currency*, testimony given on May 19 to May 23.

Sergeyev D., L. Chen, and Y. Gorodnichenko (2023) “The Economics of Financial Stress,” NBER Working Paper 31285.

Soldz, S. and G. Vaillant (1999) “The Big Five Personality Traits and the Life Course: A 45-Year Longitudinal Study,” *Journal of Research in Personality* 33, pages 208-232.

Soloveichik, R. (2023) “Private Funding of ‘Free’ Data: A Theoretical Framework,” unpublished manuscript available upon request.

Statistics Canada (2019) “The Value of Data in Canada: Experimental Estimates,” posted on July 10, 2019, at https://www150.statcan.gc.ca/n1/pub/13-605-x/2019001/article/00009-eng.htm and accessed in October 2022.

Swick, R., D. Bathgate, and M. Horrigan (2006) “Service Producer Price Indices: Past, Present, and Future,” FESAC paper https://www.bls.gov/advisory/fesacp1060906.pdf.

Sylla, R. (2002) *Ratings, Rating Agencies and the Global Financial System*, “A Historical Primer on the Business of Credit Rating,” Springer, Boston, Massachusetts.

Taxpayer Advocate Service (2008) “The Complexity of the Tax Code,” posted at https://www.irs.gov/pub/tas/08_tas_arc_msp_1.pdf, accessed on January 17, 2023.

Taxpayer Advocate Service (2022) “TAS Tax Tip: What If I Receive an IRS Noticed that Says Something is Wrong with My 2021 Tax Return?” posted online March 23, 2022, accessed January 17, 2023.

Thorndike, J. (2022) “Timelines in Tax History: From ‘Class Tax’ to ‘Mass Tax’ During World War II,” posted at *Taxnotes.com* on September 19, 2022, and accessed January 17, 2023.

Trainor, S. (2015) “The Long, Twisted History of Your Credit Score,” *Time*, July 22, 2015.

Trebbi, F. and M. Zhang (2022) “The Cost of Regulatory Compliance in the United States,” *NBER Working Paper Series* 30691.

Turner, M., R. Varghese, and P. Walker (2011) *U.S. Consumer Credit Reports: Measuring Accuracy and Dispute Impacts*, Policy & Economic Research Council Report.

Ulzheimer, J. (2022) “Does Disputing a Credit Report Hurt Your Credit Score,” posted at *Mintlife* on May 31, 2022, and accessed in October 2022.

United Nations Statistics Division. (2008). *Updated System of National Accounts 2008*. Accessed October 2022. http://unstats.un.org/unsd/nationalaccount/sna2008.asp.

Walczak, S., and G. Borkan (2016) “Personality Type Effects on Perceptions of Online Credit Card Payment Services,” *Journal of Theoretical and Applied Electronic Commerce Research* 11(1).

Weedmark, D. (2021) “How to Become an H&R Block Tax Preparer,” posted at *Chron.* on January 19th 2021 and accessed March 21st 2023.

Whiting, E. (2022) “Do Credit Repair Companies Work?” posted at *CreditRepair.com* on May 16, 2022, and accessed in October 2022.

Wu, T. (2021) “Can I Open a Business Bank Account with Just an EIN,” posted at *Nearside.com* on October 27, 2021, and accessed January 17, 2023.

Zuckerman, M., B. DePaulo, and R. Rosenthal (1981) Verbal and Nonverbal Communication of Deception. In L. Berkowitz (Ed.) *Advances in Experimental Social Psychology* Vol. 14, pages 1–59. Academic Press.