Demand for Privacy from the U.S. Government: Evidence from the Census Income Question*

Zoë Cullen
Harvard Business School

Tom Nicholas
Harvard Business School

January 22, 2026

Abstract

We provide evidence on the demand for privacy at scale, studying millions of wage earners asked to report income in the contested 1940 U.S. Census, which first included an income question. Around 5% of active wage-earners—1.2 million individuals—refuse to report, with substantial variation across counties. We assess the drivers of withholding by exploiting quasi-exogenous variation in enumerator characteristics and prior exposure to unanticipated publication of government tax list data. Withholding patterns reflect fears of identifiable disclosure. Income differentiated individuals withhold when inequality is high, when the data collector's wealth is distant from theirs, and when tax-list data has been made public in the past. We find evidence of pressure to comply among vulnerable residents but little evidence of intrinsic privacy motives. Our results suggest that privacy-sensitive nonresponse can distort inequality measures and highlight the importance of understanding behavioral responses to disclosure risks in modern data policy.

JEL Classification: D83, D84, D91, C93, J16, J31, M12.
Keywords: inequality, income, privacy, government, U.S. Census.

*Cullen: zcullen@hbs.edu, Rock Center 210, Boston, MA 02163, United States. Nicholas: tnicholas@hbs.edu, Rock Center 321, Boston, MA 02163, United States. We are extremely grateful to Dan Marcin for providing access to his newspaper tax list data (ICPSR 36640), to Price Fishback for data on New Deal expenditures and to Tim Guinnane, Jesse Shapiro and Noam Yuchtman for detailed comments and suggestions. The Division of Research and Faculty Development at Harvard Business School provided financial support.

1. INTRODUCTION

The U.S. Census Bureau, and the data it collects, have been the subject of political debate throughout U.S. history (Bouk, 2022). While Census data serve as a critical public good, concerns over surveillance and confidentiality can deter disclosure of sensitive information. In 2018, the U.S. Census announced it would use differential privacy, forgoing statistical precision in published statistics in exchange for greater identity protection. Despite the active policy debate, we lack evidence on the drivers of the demand for privacy at scale.

We study the individual demand for privacy from the U.S. Census at scale. We exploit the 1940 Census, which for the first time asked respondents to report annual wage income. Contemporary debates were extensive. Proponents emphasized the value of income data for understanding Depression-era labor markets, while opponents argued that such questions constituted unjustified intrusions into private affairs (United States Senate, 1940). To assuage concerns, the Census isolated the question about exact earnings amounts at the very end of the survey and allowed respondents to record their income privately for the enumerator to mail the sealed form to the Census Bureau. This unusual survey design creates a rare opportunity to observe data withholding among individuals already identified as wage earners earlier in the survey when asked about wage-earning status, hours, and occupation.

We measure non-disclosure as a zero or missing response to the earnings question, and we validate this interpretation using the original 1940 manuscript schedules, which record when income was entered as zero, left blank or provided on the sealed form. These distinctions are preserved in the 1940 complete-count Census data we use (Ruggles et al., 2022).

We characterize patterns of withholding and the motivations behind resistance to government data collection. We test three leading mechanisms emphasized in privacy theory—social concerns arising from public visibility, intrinsic demands for privacy, and government misuse (Acquisti et al., 2016). Each mechanism predicts a distinct pattern of withholding behavior. Public visibility concerns predict earners who are highly differentiated from others withhold as social consequences are largest for them.[1] By contrast, intrinsic privacy motives predict withholding among parties who defend privacy as a normative right: in our context, the Republican party. Concerns about government misuse of the data predict vulnerable or state-persecuted groups withhold; for example, when the government is motivated to seize the assets of a group or penalize targeted populations.[2] Our data allow us to distinguish

[1]The personal stakes of public observation of own earnings rise with differentiation, or income inequality—a fact documented in several settings (Luttmer, 2005; Perez-Truglia, 2020; Cullen and Perez-Truglia, 2022). Costs to revealing own income rise with perceived inequality owing to factors including self-image, social-image, resentment and competition (Luttmer, 2005; Perez-Truglia, 2020; Cullen and Perez-Truglia, 2022).

[2]Calo (2011) distinguishes between subjective privacy harm arising from perceived observation and objective harm arising from informational misuse.

these mechanisms by observing which individuals withhold, and how withholding responds to perceptions of income differentiation, data leaks and risks of persecution.

We find strong support in the data for concern over public visibility. Among the most active labor market group of wage earners aged 25 to 65, 5.1% withhold their earnings data (1.2 million individuals). Non-disclosure is especially pronounced among the predicted top and bottom deciles of the earnings distribution, where it is roughly twice as high as in the middle of the distribution. This pattern is stark among the ultra-wealthy. Among 104 such individuals identified in U.S. Treasury records, we find those with wage income withhold their earnings from the U.S. Census at four times the rate of similarly aged workers.[3]

We begin by showing that county-level inequality strongly predicts income withholding exploiting variation in the 90/10 inequality ratio across counties within the same state and within the same occupation. Non-disclosure rates rise monotonically across deciles of the 90/10 income ratio with the top and bottom decile responding most strongly as the income ratio grows. The top decile experienced approximately 38% higher non-disclosure rates than the bottom decile. As a placebo, we show that local inequality does not predict nonresponse to the Census education question, which is far less privacy-sensitive than income. Furthermore, the linear relationship between inequality and non-disclosure strengthens as the peer group narrows within occupation. This strong relationship suggests withholding incentives increase as ones data become differentiable from ones peers, consistent with theories of social comparison (Luttmer, 2005; Perez-Truglia, 2020; Cullen and Perez-Truglia, 2022).

In contrast, we find little evidence that intrinsic privacy motives drive withholding. The Republican vote share, associated with privacy-oriented opposition to Census expansion, is only weakly correlated with non-disclosure. The gradient in non-disclosure across deciles of the Republican vote share is economically small, statistically indistinguishable from zero, and far flatter than the inequality gradient. Counties in the top decile of the Republican vote share do not exhibit higher withholding rates than those in the bottom decile. The magnitudes are also small when we use local religiosity and government spending as proxies for intrinsic preferences related to trust, community norms, or attitudes toward redistribution.

To further test how withholding responds to perceived earnings differentiation, we use the quasi-random assignment of Census enumerators to measure how non-disclosure responds to distance in housing wealth between the subject and data collector. The enumerator represented the first person with potential to access the data, and as such, a relevant audience. Although income withholding responds to socially salient features of the enumerator-respondent interaction, these interactions are the mechanism through which the federal government col-

[3]This occurs despite the fact that enumerators are top-coding incomes at $5,000 for approximately the top 1% of earners.

lected data, making social exposure an integral component of privacy from the state.

We observe a sample of 1940 Census enumerators using the occupation string reported in the Census data and use their home and rental values to estimate the level of capitalized and reported housing wealth inequality between them and each of their subjects. Enumerators typically enumerated in areas nearby their own residence, so enumerators and respondents were likely to have visibility into each others' housing wealth. A 10% increase in the housing wealth gap between the subject and the enumerator, based on reported housing values in the Census, is associated with a 3.6% increase in the non-disclosure rate in our specification with demographic controls. We cannot reject symmetric effects when the wealth gap favors the enumerator or subject, highlighting that differentiation itself, rather than relative advantage or disadvantage, drives withholding. We find no corresponding housing-wealth effect when we estimate a placebo treatment by randomly assigning another enumerator of the same gender residing in another enumeration district in a different state.

To explore the extent to which data withholding responds to potential data publicity, we use variation stemming from exposure to the unanticipated release of federal income tax returns under the short-lived 1924 Revenue Act. At the time, the Internal Revenue Service collected income data from around 7% of the population, and newspapers in some cities, but not others, published the names, addresses, and tax payments of high earners. Because publication varied across locations and only individuals above certain ages in 1940 would have been old enough to have seen the lists, exposure differs sharply by place and cohort. We measure this exposure using archival records on which newspapers printed the tax returns (Marcin, 2014), combined with historical circulation data in Gentzkow et al. (2014), and test whether exposure to this earnings data leak predicts non-disclosure in the 1940 Census.

We find that newspaper circulation in counties where individual tax returns had been published is associated with increased non-disclosure among age groups directly exposed to the lists, especially among the college-educated, consistent with their greater newspaper readership. Younger cohorts for whom the episode would have been less visible display negligible differences across exposed and unexposed counties. Our findings corroborate evidence across settings that behavior responds strongly to past economic events, with effects that can persist for decades (Malmendier and Wachter, 2024). In our case, past data leaks predict future data withholding from the government.

Finally, we find little support for fears of government misuse as a motive for withholding. We distinguish privacy concerns from confidentiality assurances by defining privacy as discomfort with state observation rather than fear of public disclosure (Ruggles and Magnuson, 2023). We show individuals who faced the highest risk of state persecution and asset confiscation were *less* likely to withhold their income using two within-group comparisons. Following

Arellano-Bover (2022), we construct predicted internment risk at World War II camps for Japanese residents. For Mexican residents, identified using the Mexican designation in the 1930 Census and traced to 1940, we measure perceived enforcement risk using the per capita prevalence of local law-enforcement authorities in light of the large-scale deportation and repatriation campaigns targeting Mexican communities during the Great Depression (Lee et al., 2022). In both cases, withholding declines as the risk of persecution rises, a pattern consistent with compliance pressures outweighing fears of data misuse.

In sum, we find that concerns over public observation and personal differentiation are the dominant drivers of withholding, with little evidence supporting partisan or intrinsic privacy motives. We find patterns counter to what a government-misuse motive would imply for vulnerable groups. Although average withholding is modest, the resulting distortions are sizable because withholding is concentrated among individuals at the extremes of the income distribution. We use differences in non-disclosure across exposed and unexposed counties from our tax list analysis to construct counterfactual income distributions following DiNardo et al. (1996), showing that withholding understates measured inequality. Our findings highlight the importance of incorporating behavioral responses to social exposure into the interpretation of survey data and the design of modern data protection policies.

Our paper proceeds as follows. We next describe related literature and situate our contribution. Section 2 provides institutional context around the 1940 Census expansion, the assignment of enumerators, and the release of individual income tax records. Section 3 describes the details of our data construction and sources. Section 4.1 reports our evidence on the relationship between non-disclosure and inequality. Section 4.2 examines inequality between the subject and enumerator for a sharp test of how inequality impacts non-disclosure. Sections 4.3 through 4.8 examine how exposure to the publicized tax lists shaped perceptions of disclosure risk and distorted reported income distributions. Section 5 analyzes withholding among groups facing heightened government enforcement risk. Finally, we conclude.

1.1 *Contribution to literature*

Our work relates to several strands of the interdisciplinary literature on privacy and survey data quality. Prior research has documented key distortions in government-collected data. Serrato and Wingender (2016) and Chi (2022) show that Census population data diverge from true counts at the local level as people migrate and age, affecting funding allocations and firm-level investment decisions. A complementary literature finds that nonresponse in the Current Population Survey is concentrated in the tails of the income distribution (Bollinger et al., 2019), and Meyer and Sullivan (2023) demonstrates that tail mismeasurement can distort inequality trends. We show that this U-shaped pattern of withholding appears in historical

Census data, and we identify the mechanisms that generate this behavioral response.

Our paper contributes to the economics of privacy demands, which emphasizes that privacy concerns depend on the social context and perceived audience of disclosure (Nissenbaum, 2010; Acquisti et al., 2016; Tucker, 2023). Goldfarb and Tucker (2012) show that even in anonymous surveys some respondents refuse to reveal income. Experimental evidence shows privacy concerns vary with perceived relative standing and weaken when traits are average or socially favorable (Huberman et al., 2005), while small informational or incentive changes can override privacy preferences (Athey et al., 2017; Adjerid et al., 2013).

Incentives may also distort reporting behavior. Budd and Guinnane (1991) document that some individuals in the 1911 Irish Census exaggerated their ages to qualify for old-age pensions. Farrell (2012) argues that transparency alone is insufficient for privacy protection because identical disclosures produce heterogeneous responses across contexts. More generally, when pecuniary incentives to misreport for material gain are strong, verification and enforcement substantially limit strategic misreporting (Alpysbayeva et al., 2024).

We contribute to this literature by measuring data withholding in a context where non-disclosure carried formal legal penalties (including fines and imprisonment), but where enforcement against respondents was rare, and where concealment primarily affected social exposure and data quality rather than immediate material payoffs. Related work looks at employees' willingness to share information about their salaries with their co-workers in return for rewards, finding the majority would pay to conceal their information but some would pay to share it. Those who choose to conceal it are, on average, those who perceive themselves as relatively high earners (Cullen and Perez-Truglia, 2023).

Our paper also relates to the interdisciplinary literature on interactions between the surveyor and the respondent. Prior work shows that people are more likely to act on information delivered by a surveyor with shared traits (Durantini et al., 2006; Dolan et al., 2012). We examine respondents' willingness to share truthful information as a function of enumerator characteristics, including relative wealth. Bouk (2022) suggests that the low share of Black enumerators may have contributed to the undercount of Black Americans noting that this channel is hard to test: the under-counting could be "bias of the enumerator or act of resistance." Our setting allows us to measure acts of resistance in isolation.

We also speak to the large literature on the causes of distrust in the U.S. government. Citizens tend to distrust government the more it regulates (Aghion et al., 2010), and the late nineteenth and early twentieth centuries saw the rise of the regulatory state (Glaeser and Shleifer, 2003). Distrust, driven by perceptions of corruption, political bias, or fears of surveillance, reduces willingness to comply with laws or provide information (Levi and Stoker, 2000). In our context, trust in government was arguably at an all-time high, particularly in areas benefiting from New Deal spending (Caprettini and Voth, 2022). This high-trust environment likely contributed to the high response rates to the 1940 income question, making the withholding we observe especially informative about underlying privacy motives.

Lastly, we contribute to research on differential privacy in official statistics (Ruggles, 2024). Bowen (2024) emphasizes that the implications of changing how statistical agencies balance privacy protection and data accuracy remain poorly understood. Abowd and Schmutte (2019) propose using social welfare theory to balance these competing objectives for Census data. Our findings show that the perceived personal stakes of sharing information about income can be substantial, and that individuals respond primarily to concerns about social exposure in local environments. These distortions occur before privacy protections apply, through for example interactions with enumerators, underscoring that post-collection noise injection cannot address privacy concerns that drive non-disclosure.

2. BACKGROUND AND INSTITUTIONAL SETTING

In this section we outline central aspects of the historical background, focusing on the new income question in the 1940 census, the publication of lists of top earners by newspapers during the 1920s, and privacy demands and the fear of leaks of Census data.

2.1 *The Income Question*

Understanding how the 1940 Census introduced income reporting is important because the design created observable opportunities for withholding. The 16th Decennial Census was taken in the shadow of the Great Depression, and policymakers viewed income data as essential for understanding labor markets, housing conditions, and internal migration. On April 1, 1940, more than 120,000 enumerators collected information on over 131 million residents across 143,000 enumeration districts. While data protection is central to Census data collection efforts, the practice of in-person enumeration and the novel inclusion of wage-related questions made income reporting a particularly sensitive component of the survey.

The 1940 Census marked a turning point in the federal government's efforts to collect earnings information.[4] Enumerators were instructed to interview the most authoritative household member to obtain accurate personal information, and just under half were women, often housewives, salespeople, or clerks recruited temporarily for the Census (Bouk, 2022).

[4] Consumer expenditure surveys had collected some income information but not at scale. Population surveys mostly gathered data on property or capital assets (Dray et al., 2022). The Census of Agriculture had asked farmers about income from agricultural products as early as 1840, and the 1915 Iowa State Census collected occupational wages (Goldfield, 1958; Goldin and Katz, 2000), but urban coverage was sparse. International precedents were also limited: Britain asked whether individuals were wage earners but not their earnings, whereas the 1930 Swedish Census and the 1931 Canadian Census collected wage data.

As an indication of the significance of what became known as "the income question" enumerators asked it of all individuals rather than drawing a subsample, even though the Bureau had begun experimenting with sampling in 1940. Wage and salary income was also being reported to the Social Security Administration by this time (Goldfield, 1958), and according to Igo (2018), few Americans expressed privacy concerns, with many enthusiastic about the adoption of Social Security numbers. This combination of universal income collection, evolving administrative uses of wage data, and high public trust meant that enumerators played a central role in the disclosure process. As local, visible, and socially salient figures, they shaped the context in which respondents chose whether to reveal their earnings, which we exploit in our enumerator-respondent wealth gap test (see Section 4.2).

Questions 32 and 33 on the census form asked each respondent for their annual wage or salary and whether they had received $50 or more in non-wage income during the prior calendar year. Annual earnings were top-coded at $5,000 (roughly the top 1% of the wage distribution), although enumerators sometimes entered exact amounts above this level. The income questions were intentionally placed near the end of the interview to encourage response (Goldfield, 1958). Enumerators were also instructed to reassure respondents about confidentiality. The Census Bureau's *Instructions to Enumerators* booklet emphasized courtesy and persuasion when individuals hesitated, and directed enumerators to explain that all information was strictly confidential, available only to sworn employees, and used solely for statistical purposes. The manual further reminded enumerators that they were entitled to a truthful answer and underscored their own legal obligation to secrecy.

According to the Census Act of 1929, refusal to answer a Census question was a misdemeanor punishable by a fine of up to $100, imprisonment for up to 60 days, or both. In practice, however, the Bureau did not enforce these penalties. As the Director testified in 1940, "we do not use that feature of our law because the people of the United States have had confidence in the Bureau of the Census for a great many years, and they cooperate with us in these reports" (United States Senate, 1940). By contrast, the Bureau did impose penalties on its own staff for breaches of confidentiality. Failure to protect the privacy of Census returns could be charged as a felony, carrying up to two years' imprisonment, a fine of up to $1,000, or both. Together, these features imply that income withholding in 1940 reflects voluntary privacy behavior rather than fear of legal penalties.

Safeguarding personal data was widely discussed in the press and in government in the lead-up to the 1940 Census (see Appendix A1). In March 1939, *The New York Herald* announced that 'Uncle Sam is Getting Much More Inquisitive' as the new questions were tested in trial counties. Much of the press attention, however, emerged in early 1940 amid Congressional debate sparked by Republican Senator Charles Tobey's resolution to delete the income question. In February, *The Chicago Tribune* reported that 'Census Snooping Stirs Senate Storm' and in March *The Christian Science Monitor* covered a Republican-controlled New Jersey Assembly resolution opposing 'unnecessary snooping.' That same month, *The New York Times* covered an exchange in which President Roosevelt dismissed Tobey's objections as an 'obviously political move' to derail income collection. Roosevelt himself reported income in the $5,000+ category, though he did not specify whether he received non-wage income under Question 33 despite being a stock market investor.

In response to this debate, the Census Bureau modified its collection procedure for the income question. Individuals who declined to disclose their earnings verbally could write the amount on a confidential P-16 form, seal it in an envelope, and return it to the enumerator for mailing to the Census Bureau in Washington, D.C. This procedure allows us to interpret blank or missing income entries as instances of non-disclosure. Goldfield (1958) notes that only about 200,000 of the 15 million confidential forms printed were ultimately used, and that subject to the relatively rudimentary statistical analysis available at the time, the resulting income data were regarded as "reasonably accurate" but "somewhat underreported."

2.2 *Publication of Top Incomes*

Unwanted publicity of federal tax returns under the 1924 Revenue Act foreshadowed the controversy surrounding the 1940 Census income question. As extensively documented by Marcin (2014), newspapers across the country published long lists of taxpayers and their tax payments, turning what had been a private activity into public information (see Appendix A2). The lists published in 1924 reported incomes earned in 1923, and those published in 1925 reported incomes earned in 1924. This episode provides a rare historical precedent for large-scale disclosure of earnings, and a natural source of variation in exposure to public income revelation.

In October 1924, the *Boston Post* noted the large impact these disclosures had on public perceptions of income. Because incomes could be inferred from taxes paid, the lists revealed, for example that Jack Dempsey then world heavyweight boxing champion earned more than J.P. Morgan; that steel magnate Charles M. Schwab earned less than expected; and that actress Gloria Swanson received $120,000 a year (around $2 million today). The *Post* wrote "there was no greater surprise in the whole list to Boston people than the income reported by Judge Louis B. Brandeis of the United States Supreme Court", who had warned about newspapers and privacy concerns a few decades earlier in his highly influential 1890 *Harvard Law Review* article, "The Right to Privacy" with Samuel Warren. "No one looked upon Justice Brandeis as a rich man" the *Post* observed, "but he must be, since the tax of $9,508.22 shows he must have an income of around $55,000 a year" (almost $1 million today).

In September 1925, *The New York Times* ran multiple lists of 'Downtown Manhattan's Contributions to New York's Big Share of Federal Tax' reporting that Edward S. Harkness (a philanthropist whose wealth had derived from his father's investment in Standard Oil) paid $1,351,708 and that prominent industrialists such as J.D. Rockefeller Jr. ($6,277,669), Henry Ford ($2,608,808), Andrew Mellon ($1,882,600), and J.P. Morgan ($574,379) were among the nation's top taxpayers. Anna Thompson Dodge, widow of Horace E. Dodge, also appeared with a payment of $993,028 (Marcin, 2014).[5]

Publication of these lists often depended on quasi-random administrative choices. Local tax collection offices did not uniformly interpret the 1924 law. While some prepared lists for newspapers, others withheld them, and still others questioned the legality of publication. We exploit this geographic variation in publication to examine how exposure to public income disclosure affected non-disclosure in the 1940 Census (see Section 4.3).

2.3 *Expressed Fear of Census Data Misuse*

Public concern about the misuse of Census information has a long history. The Reapportionment Act of 1929 explicitly prohibited the publication of any Census tabulation that could identify an individual or establishment, reflecting the need for public trust in data confidentiality. More recently, the Census Bureau's adoption of differential privacy responds to the risk that reconstruction algorithms can indirectly re-identify individuals from published statistics (Dinur and Nissim, 2003).

Despite these safeguards, there are notable historical episodes in which Census information was used by government authorities. During World War II, the Census Bureau provided both aggregated and individual-level data to the U.S. military, contributing to the identification and incarceration of Japanese and Japanese-Americans under Executive Order 9066 (Seltzer and Anderson, 2001). Mexican residents also faced elevated perceived enforcement and deportation risk following the large-scale repatriation campaigns of the 1930s, which targeted families of Mexican descent across the United States (Balderrama and Rodríguez, 2006). Interned families of Japanese decent often faced forced sales of property at discounted prices and significant economic losses.[6] This historical backdrop motivates our empirical tests of perceived government threat, in which we examine withholding among Japanese Americans and Mexican residents who faced elevated risks of state action in 1940.

Debate about invasions of privacy and the handling of government data occurred promi-

[5] The Dodge brothers' firm later became a division of the Chrysler Corporation after their deaths in 1920.

[6] Although these disclosures were not known to respondents in 1940 (Pearl Harbor occurred after enumeration was complete) contemporaneous geopolitical tensions created an atmosphere in which Japanese immigrants were already subject to surveillance and raids by federal agents (Roxworthy, 2008). For labor market impacts of internment, see Chin (2005); Saavedra (2021); Arellano-Bover (2022).

nently after World War II and to the present. The Second War Powers Act of 1942 temporarily nullified earlier confidentiality protections, underscoring the tension between statistical agencies and the uses to which their data may be put. Later, in 2004, it was revealed that the Census Bureau had legally shared information on Arab-American populations at the ZIP-code level with the Department of Homeland Security, prompting renewed commitment to confidentiality (El-Badry and Swanson, 2007). In 1950, Senator Joseph McCarthy (Republican, representing Wisconsin) produced his list of alleged communists employed by the U.S. government creating impetus for clearer definitions of privacy boundaries. During the McCarthy era the "Lavender Scare" resulted in the use of sexual questioning to purge "lavender lads" from careers in the State Department under the guise of a link between homosexuality and communism (Johnson, 2023). McCarthy himself appears in the 1940 Census. Despite working 52 weeks of the year in 1939 as a Circuit Judge in Wisconsin, his income is recorded as 0, with a 'yes' response to the question of whether he received non-wage income. His household's income information was provided by his sister-in-law, which we can identify because our data indicate who responded to the enumerator.

3. DATA CONSTRUCTION

Our main data source is the 1940 complete-count Census data which provides information on responses for over 131 million individuals (Ruggles et al., 2022). We incorporate various complementary datasets at the county-level including Republican vote shares as a proxy for intrinsic privacy preferences. We use data from Marcin (2014) to identify the locations where newspapers did and did not publish lists of top tax payers during the 1920s, and the reason why. And we use newspaper circulation data across locations from Gentzkow et al. (2014) to measure the potential reach of these tax lists across counties.

3.1 *Individual-level Data from the 1940 Census*

We use data from the complete-count 1940 Census, restricting our analysis to individuals who were in the labor force, who self-reported being at work, and who received wages or a salary, including those employed by government. By construction, this restriction captures predominantly non-farm workers, since most of the farm population was self-employed. Our main dataset contains 24.9 million individuals aged 25-65.

We also identify full-time workers using questions asked early in the enumeration schedule. Respondents were asked the number of weeks worked in 1939 and the usual hours worked per week. Restricting to those who reported working 52 weeks and at least 40 hours per week yields a secondary sample of full-time wage earners so we can examine withholding among individuals with stable labor force attachment and comparable exposure to income reporting norms. At the end of the schedule, two questions captured a person's income (12 months ending December 31, 1939) from wages or salary and whether the person received $50 or more in income from sources other than wages or salary. As Bouk (2022) notes, the second item offered respondents a way to "hide" by signaling that they had additional income without revealing the amount, since no further details on non-wage income were collected.

As noted in the case of Joseph McCarthy above, a key advantage of the Census manuscripts is that we observe which member of the household spoke to the enumerator, allowing us to restrict attention to income reports provided by that specific individual. We also observe the estimated value of the respondent's home (or monthly rent, which we capitalize), enabling us to place individuals in the distribution of housing wealth. Importantly, we observe the same information for enumerators who self-identify as such in the occupation string in the Census data, allowing us to construct respondent-enumerator wealth gaps used in our analysis.

Finally, the Census provides a rich set of individual and household characteristics which we use as controls, including age, sex, race, marital status, educational attainment, occupation, hours and weeks worked as well as homeownership status. These variables allow us to flexibly control for demographic and labor-market heterogeneity.

3.2 *Validating the Measurement of Income Non-Disclosure*

Because our analysis relies on interpreting zero or missing values for the income question in the Census as intentional withholding, we validate this measure directly against the original schedules. Although missing income values do not appear in the original 5% samples of the 1940 data, they are recorded in the 100% complete-count data which is derived from transcriptions of the original manuscript schedules that IPUMS subsequently cleans, standardizes, and harmonizes. We therefore conduct an audit comparing IPUMS-coded values with the original manuscript schedules. We draw a random sample of 100 individuals aged 16-80 whose income is recorded as missing in the IPUMS 100% file and inspect their corresponding entries on the schedules. In the vast majority of cases (95.0% for ages 16-80 and 96.2% for ages 25-65) the manuscript also shows no income amount, confirming that missing values in IPUMS almost always reflect genuine non-disclosure rather than transcription error. Only a small number of cases appear to involve misclassification.

A small fraction of schedules contain a 'C' in the income box, indicating that the respondent chose to submit earnings using the confidential P-16 form rather than disclose them to the enumerator (see Appendix A3). These cases (three in our sample), are consistent with the Census Bureau's estimate that around 2% of respondents used sealed envelopes. We consider these cases to represent intentional non-disclosure and are coded as such. Taken together, our audit demonstrates that our non-disclosure measure accurately reflects respondent choices not to reveal their income rather than measurement or processing error.

3.3 *County-level Data*

We assemble several county-level measures, including inequality, government spending, political affiliation, religiosity, and education. Income inequality in 1940 remained pronounced, with the top decile accounting for over 45% of national income (Piketty and Saez, 2003; Kopczuk et al., 2010), but it varied considerably across counties. We measure inequality using multiple complementary approaches. First, we compute the county-level 90/10 income ratio based on wage and salary income reported in the Census. Top coding affected only about the top 1% of earners, so this measure captures meaningful differences between richer and poorer residents. Second, we estimate *predicted* income for each individual using location, occupation, housing value, age, gender, and years of schooling, and construct a parallel county-level 90/10 ratio from this series (see Appendix Figure A4). Finally, following Cowell and Flachaire (2023), we compute the mean logarithmic deviation (MLD), which has desirable statistical properties and responds monotonically to increases in top incomes. To bypass income censoring entirely, we also calculate the MLD using housing values..

To measure exposure to federal redistribution, we incorporate county-level New Deal spending from Fishback et al. (2003).[7] Caprettini and Voth (2022) show that the intensity of these programs shaped patriotic sentiment during World War II, suggesting that federally funded local investment may influence trust in, and willingness to cooperate with, the federal government. At the same time, because the federal government justified income collection partly on its relevance for allocating welfare spending, high-income respondents in high-spending counties may have had incentives to withhold their incomes.

Because opposition to the 1940 income question ran along party lines, we measure partisan orientation using county-level Republican vote share in the 1938 Congressional elections (Clubb et al., 2006). Political attitudes also vary strongly with geography since rural and urban voters tend to differ systematically in moral values and views of government (Enke, 2020). We incorporate data from Haines (2010) on the urban share of the population and on educational attainment (the share of adults aged 25 and older completing high school). Religiosity, which empirical work links to trust and social cohesion (McCleary and Barro, 2006), is measured using the 1936 Census of Religious Bodies. Although some denominations were underreported, Stark (1992) documents that the Census Bureau implemented procedures to improve coverage and that the resulting figures align well with independent enumerations.

[7] We use the Fishback-Kantor-Wallis aggregation of New Deal spending categories. Counties combined in the original data for New York, Missouri, and Virginia are separated using 1930 population shares.

Finally, we control for county size and economic structure using population counts and value added in manufacturing from Haines (2010).

3.4 *Inequality Between Respondents and Enumerators*

At the individual-level, we also have a measure of the salience of local inequality. We know both who the respondent to the enumerator was in the household and we can identify the sample of enumerators who would have visited the house, consisting of neighboring individuals who reported they were a 'census taker' or a 'census enumerator' in the Census occupation string. This allows us to investigate responses to the income questions as a function of socioeconomic gaps between respondents and enumerators. Appendix A5 provides an example of a population schedule showing the information available in the IPUMS data.

4. RESULTS: EVIDENCE OF WITHHOLDING

4.1 *Inequality Predicts Withholding*

Among individuals 16-80 in age, 6.19% of individuals who reported earning a salary or wage to the Census enumerator did not disclose their actual income, equivalent to about 2 million individuals. This drops to 5.1% for 25-65 year olds.[8] County-level non-disclosure, shown in Figure 1 and based on 1940 spatial boundaries, varies substantially from 5.8% at the 25th percentile to 9.6% at the 75th percentile.[9] We present further descriptive statistics in Appendix Table A1 of individuals reporting zero, missing and positive incomes.

We find that non-disclosure is highest among those that we predict would have incomes in the top and bottom income deciles. In Figure 2, we illustrate the relationship between predicted income rank based on occupation, housing value, age, location, gender, and years of schooling and non-disclosure. The stark U-shaped pattern reveals that non-disclosure rates among those in the top and bottom decile are nearly twice as high as among the rest of the population, a pattern consistent with modern evidence that earnings nonresponse is heavily concentrated in the tails of the distribution (Bollinger et al., 2019). We observe the same U-shaped pattern when we use capitalized house values (Panel b), actual house values (Panel c), or rental values (Panel d). Moreover, the degree of inequality in one's environment

[8] Among individuals who worked 52 weeks of the year and 40 hours—the most active labor market group—3.7% did not disclose their incomes. For a 5% sample of the population of wage earners the Census Bureau at the time estimated the non-disclosure rate was 4.9%, (5.4% among men and 3.5% among women) (Bureau of the Census, 1943).

[9] The top five states by mean non-disclosure are South Dakota (11.8%), Oklahoma (11.0%), Mississippi (10.9%), Tennessee (10.3%), and Missouri (9.8%). The bottom five states, where individuals appear more willing to disclose, are Virginia (5.7%), California (5.6%), New Hampshire (5.5%), Maine (5.2%), and Massachusetts (5.2%).

significantly magnifies the rates of non-disclosure. This suggests that the U-shape is not an artifact of survey design, but rather reflects behavioral motives tied to income position.

As suggested by Figure 2, the super-rich exhibited a much higher tendency toward non-disclosure than wage earners in general. For 104 individuals on the list of the largest taxpayers compiled by Brandes (1983), 88 of whom we can link to the 1940 Census, 57.5% reported zero wage earnings or left the response blank when queried by the enumerator (see Appendix Table A2). While some members of this group lived primarily as rentiers, non-disclosure remained high even among those who reported working in 1939 with 26.5% failing to disclose their earnings, or 20.5% among those aged 25-65. These rates contrast sharply with the 5.1% non-disclosure rate observed for the broader population of wage earners aged 25-65.

In Figure 3, Panel (a), we display binned scatter plots showing the sharp increase in non-disclosure rates as county-level inequality rises. The relationship is robust across multiple measures of inequality. Panel (a) shows a positive association between non-disclosure and the 90/10 reported income ratio: as the ratio rises from 5 to 6, average non-disclosure increases from around 4% to 5%, and rises to about 6% as the ratio reaches 10. Panel (b) documents a similar gradient using the predicted 90/10 ratio. Panel (c) employs the mean log deviation (MLD) of self-reported income, and Panel (d) displays the MLD of housing wealth as an alternative measure of inequality. Across all four measures, non-disclosure increases monotonically with inequality in the respondent's county.

We further examine the relationship between non-disclosure and inequality within 228 narrowly defined occupations in the 1940 Census occupational classification system. Panel (e) shows that non-disclosure rates rise even more steeply as inequality increases within these peer groups. When the occupational 90/10 ratio is around 3, non-disclosure is approximately 2.5%, but it exceeds 10% when the ratio reaches 10.

By contrast, other characteristics of the environment such as political stance, religiosity, and local redistribution, have small and statistically insignificant associations with non-disclosure. Panels (f) and (g) show that neither county-level New Deal spending nor Republican vote share meaningfully co-vary with non-disclosure rates, despite contemporary political debates emphasizing intrinsic privacy values. Similarly, Panel (h) shows little relationship between non-disclosure and the county-level share that is religious, even though religion has been linked in other contexts to stronger norms of honesty (McCleary and Barro, 2006).

In Table I, we show that the correlation between county-level inequality and withholding is robust to an extensive set of controls at the individual, county, and occupation levels, summarized in Table A1. Using millions of observations on wage earners, we estimate the following linear probability model at the individual level i, where $Privacy_i$ is an indicator equal to one if the enumerator recorded zero or missing income for that individual:

$$Privacy_i = \alpha\, Inequality_c + \beta\, \mathbf{X}_c + \delta\, \mathbf{Z}_i + \kappa_{state} + \phi_{occupation} + \epsilon_i, \quad (1)$$

Our key parameter of interest is α, the coefficient on county-level inequality. We measure inequality using several metrics: the 90/10 ratio of reported incomes (Panel A), the 90/10 ratio of predicted incomes (Panel B), the mean logarithmic deviation (MLD) of income (Panel C), the MLD of housing wealth (Appendix Table A4), and the within-occupation 90/10 income ratio (Appendix Table A5). If the value of privacy increases in more unequal environments, we would expect $\alpha > 0$. The vector $\mathbf{X}_c$ includes standardized county-level covariates for New Deal spending per capita, the Republican vote share, religiosity, educational attainment, the urban share, county population, and manufacturing value added. The vector $\mathbf{Z}_i$ contains individual characteristics from the Census. We also include state fixed effects and occupation fixed effects at three levels of aggregation (11 major categories, 22 subcategories, and 228 granular occupations). Together, these specifications control flexibly for local area and individual characteristics to mitigate concerns about omitted variable bias.

We find a clear and robust relationship between inequality and non-disclosure. In Table I, a one standard deviation increase in the county-level 90/10 income ratio is associated with a 21% increase in non-disclosure relative to the mean in our baseline specification (column 1). The effect declines to 10.5% among full-time workers (column 7) and to 5.8% when we restrict the sample to self-respondents who personally provided their income information to the enumerator (column 8). Restricting the sample to self-respondents removes any confounds associated with household reporting dynamics, if individuals conceal incomes from other members of the household and not just the government. In terms of magnitudes the 90/10 coefficient stands out compared to New Deal spending, the share religious, or the Republican vote share as a measure of intrinsic privacy preferences. While Table I consolidates the coefficients on individual controls, the full set of estimates is in Appendix Table A3.[10]

Figure 4 Panel (a) further illustrates results using decile indicators for the 90/10 ratio and for Republican vote share. Privacy preferences rise linearly across deciles of the 90/10

[10]These results show women withhold more often than men, though the pattern reverses for full-time workers (column 7). Following Moehling (2001), relative demand for privacy may have been greater among women whose partners were unemployed, as this could have afforded more control over household finances. Household heads (mostly men) are more likely to disclose, whereas the divorced and separated are less likely to do so relative to married respondents. Singles generally disclose at higher rates, though less so among full-time workers (column 7). Whites (about 90% of the population in 1940) exhibit greater demand for privacy than minority respondents, while immigrants are more likely to disclose. College-educated individuals display a strong preference for non-disclosure, possibly reflecting concern over relative income revelation or the idea of privacy as a luxury good. We also find a positive relationship between housing wealth and non-disclosure, though the magnitudes are modest. In column 3, a one standard deviation increase in capitalized housing value is associated with a 1.6% increase in non-disclosure.

county-level income ratio, with the top decile experiencing 37.7% higher non-disclosure rates (relative to the mean) than the bottom decile. By contrast, coefficients on the Republican vote share deciles hover around zero, so the magnitude of the inequality gradient far exceeds any effect associated with intrinsic privacy. These patterns are consistent with the personal stakes of revealing income being higher in more unequal environments.

Panel (b) presents a placebo exercise using nonresponse to the education question (highest grade of school completed), which is asked earlier in the Census, is much less privacy-sensitive, and is therefore more likely to reflect recall or reporting error (Goldin, 1998). The nonresponse rate to this question is 1.6% among 25-65 year old wage earners. If local inequality operates through privacy concerns, it should predict income withholding but not education nonresponse. Unlike income non-disclosure, education nonresponse exhibits no monotonic gradient across income deciles and no effect at the top of the distribution. Although the income decile indicators are jointly significant, the estimated coefficients are small and non-monotonic, and a direct comparison rejects equality of the income and education gradients. These results support the interpretation that the strong inequality decile gradient in Panel (a) reflects income-specific privacy concerns rather than general nonresponse behavior.

Two additional sets of regressions in Table I show the robustness of the baseline relationship between inequality and privacy from Table I. Panel B re-estimates the models using predicted income to construct the 90/10 ratio, thereby avoiding any mechanical bias from censoring in reported incomes. The effect sizes are slightly smaller, consistent with classical measurement error attenuating coefficients. In Panel C, using the MLD of income inequality, the effects grow larger. For full-time workers (column 7), a one standard deviation increase in income inequality is associated with a 13.4% increase in non-disclosure, compared to a 10.5% increase in Panel A and an 8.5% increase in Panel B.

Appendix Table A4 addresses the concern that income-based inequality measures could themselves be distorted by non-disclosure, particularly in the tails of the distribution. Using a measure independent of self-reported incomes, namely the MLD of housing values, we find economically consistent effects. For full-time workers (column 7), a one standard deviation increase in housing-wealth inequality is associated with a 7.1% increase in non-disclosure.

Our strongest results arise when we measure inequality within narrowly defined occupations. Using within-occupation reported incomes to compute the 90/10 ratio, we find much larger responses. In Appendix Table A5, column 7, a one standard deviation increase in within-occupation inequality is associated with a 32.3% increase in non-disclosure among full-time workers and a 33.4% increase for the subsample of individuals who directly responded to the enumerator (column 8). Our estimates remain stable across alternative control sets, including when we additionally control for the occupation's median wage as a proxy for employment turnover (Appendix Table A6).[11] Overall, these findings highlight the central role of relative income position in shaping disclosure behavior.

4.2 *Withholding Rises When Enumerators Differ Economically from Respondents*

As an individual-level test of how privacy demands respond to the salience of local inequality, we exploit the quasi-random assignment of Census enumerators to households and the resulting variation in the wealth gap between respondents and enumerators. Enumerators were typically drawn from nearby neighborhoods and conducted interviews inside respondents' homes, making differences in housing wealth and social standing readily apparent during the interaction. We use data on the housing wealth and socioeconomic characteristics of both respondents and enumerators to test whether being visited by a Census worker with differential wealth to the respondent increases the probability of non-disclosure.

Enumerators were required to be U.S. citizens with at least a high school education, possess legible handwriting, and pass a formal aptitude test designed to mimic completion of the population schedule (Thomson, 1940). Although the names of enumerators appear at the top of manuscript schedules, no dataset identifying them exists. We therefore construct a sample of enumerators by searching the occupation string in the IPUMS data for individuals whose reported occupation includes both 'census' and 'enumerator' or 'census' and 'taker.' This procedure yields 1,023 enumerators from 48 states and the District of Columbia. Enumeration personnel included door-to-door canvassers as well as area managers and district supervisors who coordinated activities, trained enumerators, and consolidated returns. We cannot distinguish these roles in the data, but our keyword-based approach most likely captures door-to-door enumerators as the individuals who interacted directly with respondents. Descriptive statistics are show in Appendix Table A7.

Since a wealthy individual and poor individual are both statistically more likely to have a larger inequality gap with the 'typical' enumerator than a middle income individual, we take an additional step to isolate variation stemming from the particular enumerator assignment. For comparison, we draw a placebo enumerator designated as a random enumerator of the same gender from an enumeration district in a different state. We compare the effects of wealth distance between the subject and realized enumerator on non-disclosure, with the effect of differences between the subject and placebo enumerator, yielding our preferred statistical test of how face-to-face inequality with an enumerator impacts non-disclosure.

Specifically, we estimate the following linear probability specification at the individual level i where *Privacy* is a binary variable coded 1 for non-disclosure, as before, and the

[11] If turnover is higher in lower-wage occupations, individuals may be less able to recall their earnings, increasing the likelihood of reporting zero or missing income.

superscripts S, E and E^* refer to subject, enumerator and placebo enumerator respectively:

$$Privacy_i = \underbrace{\gamma_1\, log\left(\frac{House^S}{House^E}\right)_i}_{\text{Actual Enumerator}} + \underbrace{\gamma_2\, log\left(\frac{House^S}{House^{E^*}}\right)_i}_{\text{Placebo Enumerator}} + \omega\,\mathbf{I}_i + \pi\,\mathbf{G}_i + \nu\,\mathbf{D}_d + \zeta_{county} + \epsilon_i. \quad (2)$$

Our main coefficient of interest is γ_1 capturing the wealth gap between subject and enumerator while controlling for the effect of the placebo treatment through γ_2. Our causal test of the impact of face-to-face inequality on non-disclosure is $\gamma_1 - \gamma_2 = 0$ corresponding to the null hypothesis that the difference between these coefficients is zero.

We measure the subject-enumerator wealth gap as the log ratio of the respondent's house value to that of the enumerator, and construct the same measure for the placebo enumerator, $\left(\frac{House^S}{House^{E^*}}\right)i$. We estimate equation 2 using both capitalized and reported housing values. The vector $\mathbf{I}_i$ includes the log of the respondent's own house value (ensuring that the wealth-gap coefficient does not simply pick up the effect of own wealth on withholding) along with demographic controls for race, marital status, immigration status, education, age, and age squared. To account for homophily that may influence disclosure, $\mathbf{G}_i$ includes differences in age and years of education between the respondent and the enumerator, and between the respondent and the placebo enumerator. We also report specifications with and without restricting the sample to respondent-enumerator pairs of matching gender, since gender differences may affect reporting behavior. Finally, $\mathbf{D}_d$ includes enumeration-district means of log housing wealth, educational attainment, and age to capture unobserved local characteristics relevant for privacy decisions. All regressions include county fixed effects, and standard errors are clustered at the household level.

In column 1 of Table II we find that a 10% increase in the capitalized housing wealth gap between the subject and the enumerator is associated with an increase of $\gamma_1 \times \ln(1.10)$=0.00022 in the probability of non-disclosure or about 0.53% relative to the mean, or 0.56% in column 2 when we control for the placebo enumerator—the effect we would expect to see by chance. Under our test of the importance of face-to-face inequality, $\gamma_1 - \gamma_2 > 0$ (p-value=0.0005). With classical measurement error in the capitalized housing wealth series we would expect these coefficients to reflect lower bound estimates. Indeed, in columns 3 and 4 we find much larger effects when we use reported house values. In column 4 a 10% increase in the housing wealth gap is associated with a 2.5% increase in non-disclosure relative to the mean non-disclosure rate of 4.3%. The difference between actual and placebo coefficients again exhibits statistical significance ($\gamma_1 - \gamma_2 > 0$, p-value=0.0266).

In columns 5 to 8 we replicate the results in columns 1 to 4 using only subject-enumerator pairs of the same gender. We find consistent results and effect sizes that are slightly larger. In column 8, for example, a 10% increase in the housing wealth gap is associated with a 3.6% increase in non-disclosure relative to the mean, compared to the corresponding effect of 2.5% in column 4 when we do not gender match on subject-enumerator pairs. One explanation for the increased tendency to withhold information when we condition on gender-matching could be attributed to the importance of the reference group (Cullen and Perez-Truglia, 2022). As individuals become more closely aligned with their peers, social concerns around earnings information become more pronounced. This would be consistent with the larger magnitudes that we find in our occupation-level results, as discussed in Section 4.1 above.

In Appendix Table A8 we replicate columns 1, 3, 5 and 7 of Table II, splitting our sample by whether the subject housing wealth is greater than or less than the enumerator's housing wealth. We cannot reject that the sensitivity of non-disclosure to the housing wealth gap is symmetric regardless of who is wealthier, subject or enumerator. This suggests respondents react to the magnitude of the wealth gap rather than the direction of the difference.

4.3 *Exposure to Historical Data Release and Privacy Behavior*

We next examine whether changes in perceived risk of public release of personal income data affected disclosure behavior. Our empirical setting exploits the brief window surrounding the 1924 Revenue Act, during which newspapers in some counties published names and tax payments of individual federal taxpayers. We assess whether individuals with greater likelihood of having observed this unexpected release of private income information displayed systematically different propensities to report their earnings in 1940.

We implement two complementary research designs. First, we estimate a triple-differences specification comparing cohorts that differed in age at the time of publication across counties that did and did not publish the tax lists. Older individuals were more likely to have been in the labor market and consuming newspapers when the lists appeared, generating differential exposure by cohort. Second, we measure the intensity of exposure using newspaper circulation per capita in list-publishing counties. If public disclosure altered beliefs about the likelihood of income becoming publicly accessible, non-disclosure should rise with both cohort exposure and circulation intensity.

4.4 *Identifying Counties with Published Individual Tax Returns*

We identify counties where the lists were published or not published among states that include a top 50 city by population size due to our data source on newspaper circulation (see Figure A6). Hence, we compare non-disclosure rates in counties like Hartford County, Connecticut where both the *Hartford Courant* and the *Hartford Times* published lists with control counties like Fairfield County, Connecticut where the lists were not published. According to Marcin, a key determinant in whether a county widely published the list of tax returns is whether or not the local tax registrar could implement the request in a timely way (while the 1924 Revenue Act stood). Appendix Tables A9 and A10 provide summary statistics and a balance table showing very few economically meaningful correlations between county characteristics and tax list circulation. New Deal spending per capita is higher in counties where the lists were published.[12] These counties also had a higher rate of home ownership, but no significant differences can be observed between these counties and those without published lists in both the level of wage and non-wage income reported in the 1940 Census. As a check against differential reporting quality, Appendix Figure A7 shows that age-heaping is low and statistically indistinguishable between newspaper-list and non-list counties.

4.5 *Estimating Tax List Exposure Using Age at Time of Publication*

To proxy individual exposure to the publication of income tax returns, we classify individuals who were age 40 or older at the time of the 1940 Census as having had greater exposure to the 1924-1925 newspaper lists than those who were younger. Individuals aged 40 in 1940 would have been 24 in 1924, just entering the labor market, eligible to vote, and plausibly reading newspapers regularly if they were sufficiently educated to do so. Because age misreporting and other factors make a precise cutoff undesirable, we treat the 40+ threshold as a coarse proxy rather than a sharp research design choice. Furthermore, we will show that our results are robust when using five-year age bins from ages 25 to 65.

In our regression framework, we code an indicator equal to one if an individual was age 40 or older in 1940 and zero otherwise. Table III presents linear probability estimates of privacy demands. Column 1 shows that, on average, non-disclosure rates do not differ significantly between counties that did and did not publish the tax lists. However, the patterns change markedly when we consider older cohorts. Column 2 shows that individuals aged 40-65 have significantly higher non-disclosure rates in counties that published the lists. In column 3, the interaction between age and news-list counties is imprecisely estimated, but becomes statistically significant in column 4 when we allow covariates to vary flexibly by age, following the recommendations in Feigenberg et al. (2023). The interaction coefficient in column 4 implies that non-disclosure among 40-65 year olds in news-list counties is approximately 6.4% higher relative to the mean ($0.00276 \div 0.043$).

Weighting by 1920 county population strengthens these results. The interaction term

[12] This is largely driven by counties in New York state receiving substantial federal relief and recovery aid during the Great Depression.

becomes larger and more precisely estimated (column 5), and the specification allowing age-specific covariates (column 6) implies a 13.5% increase in non-disclosure among 40-65 year olds in news-list counties (0.00579 ÷ 0.043). Figure 5 further shows that the effects concentrate sharply among individuals 40 and older, the group most likely to have remembered the original tax-return publications. When we estimate the specifications in columns 4 and 6 using five-year cohort dummies, we observe a distinct increase in non-disclosure beginning around age 40. Weighted and unweighted estimates track closely for individuals under 40 but diverge thereafter, consistent with stronger treatment effects in more populated counties where circulation and spillovers associated with the lists would have been more extensive.

4.6 *Estimating Tax List Exposure Using Newspaper Circulation*

To measure exposure to published tax lists, we follow Gentzkow et al. (2014) in treating news markets as operating at the county level. Newspaper circulation in the 1920s was highly localized even for national outlets and newspapers typically printed lists of local taxpayers.[13] County-level circulation therefore provides a meaningful proxy for the extent to which the release of individual tax returns became publicly visible in a given area, allowing us to capture geographic variation in perceived disclosure risk.

We estimate own-county and cross-county exposure to newspaper circulation of tax lists. We proceed in two steps. First, let N^T represent the set of all newspapers publishing lists in counties and $Circulation(N_n^T, C_c^T)$ the circulation of the n-th newspaper in list county C_c^T, then the total circulation of newspapers in a single list county is $\sum_{n \in N^T} Circulation(N_n^T, C_c^T)$. We construct an exposure measure by scaling this total circulation by the county population in 1920. For ease of interpretation, counties with zero circulation are assigned an exposure value of zero, and positive exposure values are grouped into low, medium, and high terciles. We also estimate specifications weighted by 1920 county population. Appendix Figure A8 plots the distribution of scaled circulation and the tercile cutoffs.

Appendix Table A11 reports summary statistics across circulation terciles and shows little systematic sorting across counties. Of the 31 county and individual-level variables we examine, only two county characteristics and five individual characteristics differ monotonically across the low, medium, and high circulation terciles.[14] In all regressions, we control for this full set of observables, ensuring that estimated effects of circulation intensity are not driven by underlying differences in county or individual attributes.

[13] Around 70% of the circulation of *The New York Times*, for example, occurred in the states of New York and New Jersey.

[14] The county-level variables are New Deal spending per capita and the share religious; the individual-level variables are indicators for being single, being an immigrant, being a homeowner, and measures of house value and capitalized house value.

We estimate the following linear probability models:

$$Privacy_i = \sum_{j=1}^{3} \lambda_j Circulation_c + \lambda_4 Age_i^{40-65} + \sum_{j=5}^{7} \lambda_j (Circulation_c \times Age_i^{40-65}) \\ + \beta\, \mathbf{X}_c + \delta\, \mathbf{Z}_i + \xi\, \mathbf{W}_i + \kappa_{state} + \phi_{occupation} + \epsilon_i, \tag{3}$$

where the circulation indicators denote whether county c falls into the low, medium, or high tercile of circulation among counties that published the lists. The omitted category consists of counties with no tax-list publication and individuals aged 25-39 in those counties. Our key coefficients of interest are λ_5, λ_6, and λ_7, which capture how the relationship between non-disclosure and circulation intensity differs for older cohorts. As in equation 1, we progressively add individual controls $\mathbf{Z}_i$, county-level characteristics $\mathbf{X}_c$, and a complete set of two-way interactions $\mathbf{W}_i$ along with state and occupation fixed effects.

As a second test, we relax the assumption that exposure operates strictly through local news markets and incorporate multi-county newspaper circulation patterns using data from Gentzkow et al. (2014). Newspapers based in list counties often circulated across multiple counties, including non-list counties. For example, *The Hartford Courant* circulated across all eight Connecticut counties, while *The Hartford Times* circulated across six; *The New York Times* circulated in at least 245 counties nationwide and was a major publisher of the lists. Accounting for these cross-county flows allows us to test whether exposure to tax-list publication outside one's own county influenced subsequent income-disclosure behavior.

Formally, let C denote counties in states with a top 50 city and N^T be the set of all newspapers publishing the lists. Let $\widehat{Circulation}(N_n^T, C_c)$ denote the circulation of the n-th newspaper in the c-th county, such that the total circulation of all newspapers in each county is given by $\sum_{n \in N^T} \widehat{Circulation}(N_n^T, C_c)$. We scale this by each county's 1920 population, and group counties into low, medium, and high circulation terciles (Appendix Figure A8).[15] Although exposure to taxpayer lists originating elsewhere may be less personally relevant than local publication, it may still heighten perceptions of the likelihood that income data could become publicly accessible. We therefore interact these multi-way circulation terciles with the indicator for list-publishing counties ($Newslist$) and with the age 40-65 indicator capturing individuals of working age at the time of publication.

We estimate the following linear probability models:

[15] Our initial exposure measure distinguished non-circulation counties from circulation counties and then grouped the latter into terciles (four categories in total). The multi-way measure uses terciles only (three categories).

$$
\begin{aligned}
Privacy_i = &\sum_{k=1}^{2} \psi_k \widehat{Circulation}_{c,k} + \psi_3 Age_i^{40-65} + \psi_4 Newslist_c \\
&+ \psi_5 Age_i^{40-65} \times Newslist_c \\
&+ \sum_{k=6}^{7} \psi_k (\widehat{Circulation}_{c,k-5} \times Age_i^{40-65}) \\
&+ \sum_{k=8}^{9} \psi_k (\widehat{Circulation}_{c,k-7} \times Newslist_c) \\
&+ \sum_{k=10}^{11} \psi_k (\widehat{Circulation}_{c,k-9} \times Age_i^{40-65} \times Newslist_c) \\
&+ \beta \mathbf{X}_c + \delta \mathbf{Z}_i + \xi \mathbf{W}_i + \kappa_{state} + \phi_{occupation} + \epsilon_i
\end{aligned} \tag{4}
$$

where $\widehat{Circulation}_{c,1}$ and $\widehat{Circulation}_{c,2}$ denote medium and high tercile circulation counties, respectively, with low-circulation counties serving as the omitted reference category. Our main coefficients of interest are ψ_{10} and ψ_{11}, the triple interaction terms. These coefficients capture how non-disclosure differs for individuals who were of working age at the time of publication (ages 40-65 in 1940) and who lived in counties with medium or high circulation exposure, relative to younger cohorts in low exposure counties.

4.7 *Magnitude of Exposure Effects*

We report the interaction terms λ_5, λ_6, and λ_7 from equation 3 graphically in Figure 6. Following the structure of our baseline results in Table I, Panel (a) begins with a specification including age controls and state fixed effects, and Panels (b)-(f) sequentially add county-level controls, demographic controls, occupation fixed effects, and sample restrictions. Across all specifications, we consistently find higher non-disclosure rates among individuals most exposed to the circulation of the tax lists relative to those with little or no exposure. We observe no discernible effects in low and medium circulation counties, suggesting that a sufficiently high level of newspaper circulation was necessary to shift privacy behavior.

In Panel (a), non-disclosure is (0.001111 $\div$ 0.043)=25.8% higher among 40-65 year olds in the highest exposure counties using basic controls, with similar magnitudes in Panel (b) (26.0%), Panel (c) (27.0%), and Panel (d) (24.6%) after incorporating additional county characteristics, demographic controls, and occupation fixed effects. Among 40-65 year olds in the full-time worker sample (Panel (e)), non-disclosure is (0.00657 $\div$ 0.027)=24.3% higher in high-circulation counties; for 40-65 year olds who directly interacted with the enumerator (Panel (f)), the effect is (0.01138 $\div$ 0.043)=26.5%. If the publication of the tax lists altered beliefs about the likelihood of income disclosure, we would anticipate stronger effects among individuals more inclined to consume newspapers and better able to process the implications of publication. Consistent with this, restricting the sample to college-educated individuals in Panel (g) yields a ($0.00560 \div 0.054$)=28.0% increase in non-disclosure among 40-65 year olds in list counties. Allowing covariates to vary flexibly by age (Panel (h)) reduces the estimate to 16.9%. Appendix Figure A9 shows robustness to restricting to individuals who remained in newspaper-list counties between 1920 and 1940, as only 3-5% of 40-65 year olds moved states in the five years prior to the 1940 Census (Rosenbloom and Sundstrom, 2003).

Finally, Figure A10 presents estimates of equation 4, which incorporates multi-county newspaper circulation patterns. Our focus is on the triple interaction terms ψ_{10} and ψ_{11} corresponding to $\widehat{Circulation} \times Newslist \times Age : 40 - 65$. Panel (a) reports baseline differences across medium and high circulation counties, showing no effect of circulation intensity on non-disclosure when interactions are excluded. Panel (b) introduces an interaction with the 40-65 age group and likewise shows no detectable effect. In Panels (c) and (d), however, once we account for whether the county itself published the lists, we find statistically significant effects. Non-disclosure is 12.6% higher among 40-65 year olds in high-exposure counties (Panel (c)), rising to 14.5% after including a full set of two-way interactions in Panel (d).

Taken together, these results provide suggestive evidence that demand for privacy in the 1940 Census responded to prior exposure to the public release of federal income information, particularly in locations with high intensity coverage. Exposure may have increased perceptions of the risk that personal data could become publicly accessible and illustrated the potential implications to individuals of such disclosure.

4.8 *Counterfactual Distributions and Survey Data Distortion*

Our results show that exposure to the historical publication of income tax lists increased withholding in the 1940 Census. We now examine how this withholding affects inference about the *reported* income distribution by constructing counterfactuals. We estimate what the distribution of reported incomes in newspaper list counties would have looked like had individuals reported their earnings like observationally similar individuals in non-list counties. Comparing the actual and counterfactual distributions allows us to quantify how withholding distorts the observable wage distribution under more muted privacy concerns.

Among approaches for constructing counterfactual distributions, we employ the kernel reweighting method of DiNardo et al. (1996) (DFL), which is tractable given the large number of fixed effects in our setting. We first estimate a propensity score indicating whether an individual resides in a non-list (coded 1) or list (coded 0) county, using age, state and full occupation fixed effects, and all individual and county level covariates used throughout the analysis. The propensity score is then used to define the region of common support and to reweight the kernel density functions. Following DFL, we reweight individuals in non-list counties so that their covariate distribution matches that of individuals in list counties. Our interest is then in the gap between the actual and counterfactual income distributions.

Figure 7 displays the actual and counterfactual income distributions for individuals aged 25-65 in list counties. The counterfactual density places more mass in the lower ranges, suggesting meaningful distortions in the composition of reported incomes. Because the 1940 Census top-coded wage income at $5,000 and DFL relies on observed values, differences in the upper tail are less visually pronounced, even though withholding among high earners also compresses the right tail, as we show in Section 2.2 for the ultra wealthy. The 90/10 ratio in list counties is 7.52, compared with 8.00 in the counterfactual distribution, implying that privacy leads to understated income inequality in the reported data. These patterns are consistent with the population correlations between predicted income and non-disclosure in Figure 2 and with modern survey evidence showing that nonresponse disproportionately suppresses both low earners and very high earners (Bollinger et al., 2019).

5. PRIVACY BEHAVIOR AND EXPOSURE TO STATE ENFORCEMENT RISK

Finally, beyond concerns about public disclosure, individuals may also withhold information if they fear that the government collecting the data could later use it for punitive purposes. In this case, withholding would be more common among groups facing a higher perceived risk of state enforcement actions, such as asset seizure or detention. As noted in Section 2.3 the period surrounding the 1940 Census provides two key contexts in which such fears may have been particularly relevant. Japanese residents faced growing scrutiny and, shortly thereafter, mass internment and confiscation of property during World War II. Although these events postdate the Census, contemporary accounts indicate widespread anxiety and anticipation. Mexican residents, likewise, lived under the shadow of the large-scale deportation and repatriation campaigns of the 1930s, which involved detention and economic losses.

We therefore examine withholding among Japanese and Mexican wage earners in the 1940 Census. Japanese residents are identified directly in 1940, while Mexican residents are traced from their recorded origin in the 1930 Census to the 1940 Census using the links provided by Abramitzky et al. (2020). To proxy perceived enforcement risk, we use data from Arellano-Bover (2022) to construct a predicted probability of internment for Japanese individuals based on demographic characteristics. We proxy enforcement risk for Mexicans using the local intensity of law enforcement, measured as the per-capita prevalence of policemen, detectives, marshals, constables, sheriffs, and bailiffs identified from occupation codes in the Census. If individuals feared that reported income could facilitate asset seizure or other punitive actions, withholding should rise with perceived enforcement risk. Alternatively, those at greatest risk of state scrutiny may disclose more, not less, due to compliance pressures.

In both cases, greater exposure to state enforcement risk is associated with *lower* rates of non-disclosure. For Japanese residents a one standard deviation increase in predicted internment risk is associated with a 1.0 to 1.5 percentage point reduction in non-disclosure (Table IV). For Mexican residents, a one standard deviation increase in local law-enforcement presence predicts a decline in withholding of around 0.7 to 1.1 percentage points across specifications (Table V). These patterns also hold when we restrict the sample to individuals living in the western United States, the region most directly affected by mass internment and repatriation efforts. Our findings suggest that heightened vulnerability to state action is associated with compliance rather than concealment, indicating that concerns about government misuse of Census data were outweighed by other incentives to respond, rather than serving as a primary driver of withholding behavior.

6. CONCLUSION

Debate over the protection of privacy is fundamental to society in an age where personal data has value to governments, firms and researchers but individuals have concerns about intrusions or the potential for misuse. We have examined privacy concerns in Census data using the unusual circumstances around the 1940 Census expansion, a period marked by particularly intense public and political debate over the collection of income information. Millions of U.S. residents were asked to disclose their incomes at the very end of the survey, after having revealed whether they were wage earners earlier on. We can therefore observe patterns of withholding at national scale. Our analysis has explored social concerns over data publicity, intrinsic or partisan motives for privacy, and concern over government misuse.

We find meaningful resistance through non-disclosure where local inequality was pronounced and in places where personal data collected by the federal government had been released publicly in the past. Our preferred interpretation of these results is that individuals form expectations about the private stakes of revealing their personal information. Private stakes are higher when the likelihood of a leak is higher, and when the respondent is more highly differentiated from their reference group (e.g., on either extreme end of a dispersed income distribution). We find little evidence that privacy concerns are related to politics, or redistribution policies and that vulnerable groups tend to comply more. We also find that resistance leads government statistics to underestimate the true extent of inequality.

We highlight three implications for contemporary data policy. First, distortions to the distribution of reported incomes arise before any statistical disclosure limitation is applied. Modern differential privacy frameworks, which introduce noise at the publication stage, cannot correct for selective non-response that has already reshaped the underlying microdata due to fears of social exposure, enumerator visibility, or the threat of data leaks. Second, the structure of withholding we observe with missing low and high earners closely resembles patterns documented in modern surveys (Bollinger et al., 2019; Goldfarb and Tucker, 2012). This suggests that the mechanisms we identify are not confined to a unique historical episode but reflect a more general behavioral response to perceived disclosure risk. Third, the relatively modest level of withholding in 1940 might reflect the unusually high trust in government at the time, implying that our estimates could be lower-bound effects. In settings with less trust today, similar disclosure concerns may generate substantially larger distortions.

Overall, our results highlight that privacy preferences do affect the quality of administrative and statistical data and that past breaches or public releases of identifiable information can have long-lived consequences for public cooperation with data collection efforts.

References

Abowd, J. M. and I. M. Schmutte (2019). An economic analysis of privacy protection and statistical accuracy as social choices. *American Economic Review 109*(1), 171–202.

Abramitzky, R., L. Boustan, and M. Rashid (2020). Census linking project: Version 1.0 [dataset]. Data retrieved from, https://censuslinkingproject.org.

Acquisti, A., C. Taylor, and L. Wagman (2016). The economics of privacy. *Journal of Economic Literature 54*(2), 442–92.

Adjerid, I., A. Acquisti, L. Brandimarte, and G. Loewenstein (2013). Sleights of Privacy: Framing, Disclosures, and the Limits of Transparency. *Proceedings of the Ninth Symposium on Usable Privacy and Security*, 1–9.

Aghion, P., Y. Algan, P. Cahuc, and A. Shleifer (2010). Regulation and Distrust. *The Quarterly Journal of Economics 125*(3), 1015–1049.

Alpysbayeva, D., A. Alstadsæter, W. Kopczuk, S. Markussen, and O. Raaum (2024). Misreporting in the norwegian business cash support scheme. *International Tax and Public Finance 31*, 1463–1493.

Arellano-Bover, J. (2022). Displacement, diversity, and mobility: Career impacts of japanese american internment. *The Journal of Economic History 82*(1), 126–174.

Athey, S., C. Catalini, and C. Tucker (2017). The Digital Privacy Paradox: Small Money, Small Costs, Small Talk. *NBER Working Paper No. 23488*.

Balderrama, F. E. and R. Rodríguez (2006). *Decade of Betrayal: Mexican Repatriation in the 1930s.* University of New Mexico Press.

Bollinger, C. R., B. T. Hirsch, C. M. Hokayem, and J. P. Ziliak (2019). Trouble in the tails? what we know about earnings nonresponse 30 years after lillard, smith, and welch. *Journal of Political Economy 127*(5), 2143–2185.

Bouk, D. (2022). *Democracy's Data: The Hidden Stories in the U.S. Census and How to Read Them.* MCD.

Bowen, C. M. (2024). Government data of the people, by the people, for the people: Navigating citizen privacy concerns. *Journal of Economic Perspectives 38*(2), 181–200.

Brandes, S. D. (1983). America's super rich, 1941. *The Historian 45*(3), 307–323.

Budd, J. W. and T. Guinnane (1991). Intentional age-misreporting, age-heaping, and the 1908 old age pensions act in ireland. *Population Studies 45*(3), 497–518.

Bureau of the Census (1943). Bureau of the Census: Sixteenth Census of the United States, 1940.

Calo, M. R. (2011). The boundaries of privacy harm. *Indiana Law Journal 86*(3), 1131–1161.

Caprettini, B. and H.-J. Voth (2022). New Deal, New Patriots: How 1930s Government Spending Boosted Patriotism During World War II. *The Quarterly Journal of Economics 138*(1), 465–513.

Chi, F. (2022). Information waves and firm investment. *Working Paper*.

Chin, A. (2005). Long-Run Labor Market Effects of Japanese American Internment during World War II on Working-Age Male Internees. *Journal of Labor Economics 23*(3), 491–526.

Clubb, J. M., W. H. Flanigan, and N. H. Zingale (2006). Electoral Data for Counties in the United States: Presidential and Congressional Races, 1840-1972 [ICPSR 8611,dataset]. Technical report, ICPSR.

Cowell, F. A. and E. Flachaire (2023). Inequality measurement and the rich: Why inequality increased more than we thought. *Review of Income and Wealth 0*(0), 1–24.

Cullen, Z. and R. Perez-Truglia (2022). How Much Does Your Boss Make? The Effects of Salary Comparisons. *Journal of Political Economy 130*(3), 766–822.

Cullen, Z. and R. Perez-Truglia (2023). The Salary Taboo: Privacy Norms and the Diffusion of Information. *Journal of Public Economics 222*, 104890.

DiNardo, J., N. M. Fortin, and T. Lemieux (1996). Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica 64*(5), 1001–1044.

Dinur, I. and K. Nissim (2003). Revealing information while preserving privacy. New York, NY, USA. Association for Computing Machinery.

Dolan, P., M. Hallsworth, D. Halpern, D. King, R. Metcalfe, and I. Vlaev (2012). Influencing behaviour: The mindspace way. *Journal of economic psychology 33*(1), 264–277.

Dray, S., C. Landais, and S. Stantcheva (2022). Wealth and property taxation in the united states. Working paper, Harvard University.

Durantini, M. R., D. Albarracin, A. L. Mitchell, A. N. Earl, and J. C. Gillette (2006). Conceptualizing the influence of social agents of behavior change: A meta-analysis of the effectiveness of hiv-prevention interventionists for different groups. *Psychological bulletin 132*(2), 212.

El-Badry, S. and D. A. Swanson (2007). Providing census tabulations to government security agencies in the united states: The case of arab americans. *Government Information Quarterly 24*(2), 470–487.

Enke, B. (2020). Moral values and voting. *Journal of Political Economy 128*(10), 3679–3729.

Farrell, J. (2012). Can privacy be just another good? *Journal on Telecommunications and High Technology Law 10*(2), 251–264.

Feigenberg, B., B. Ost, and J. A. Qureshi (2023). Omitted Variable Bias in Interacted Models: A Cautionary Tale. *The Review of Economics and Statistics*, 1–47.

Fishback, P. V., S. Kantor, and J. J. Wallis (2003). Can the New Deal's three Rs be rehabilitated? A program-by-program, county-by-county analysis. *Explorations in Economic History 40*(3), 278–307.

Gentzkow, M., J. M. Shapiro, and M. Sinkinson (2014). Competition and ideological diversity: Historical evidence from us newspapers. *American Economic Review 104*(10), 3073–3114.

Glaeser, E. L. and A. Shleifer (2003). The rise of the regulatory state. *Journal of Economic Literature 41*(2), 401–425.

Goldfarb, A. and C. Tucker (2012). Shifts in privacy concerns. *American Economic Review 102*(3), 349–53.

Goldfield, E. D. (1958). *Decennial Census and Current Population Survey Data on Income*, pp. 37–62. Princeton University Press.

Goldin, C. (1998). America's graduation from high school: The evolution and spread of secondary schooling in the twentieth century. *Journal of Economic History 58*(2), 345–374.

Goldin, C. and L. F. Katz (2000). Education and income in the early twentieth century: Evidence from the prairies. *Journal of Economic History 60*(3), 782–818.

Haines, M. R. (2010). Historical, Demographic, Economic, and Social Data: The United States, 1790-2002 [ICPSR 2896,dataset]. Technical report, ICPSR.

Huberman, B., E. Adar, and L. Fine (2005). Valuating privacy. *IEEE Security Privacy 3*(5), 22–25.

Igo, S. (2018). *The Known Citizen: A History of Privacy in Modern America*. Harvard University Press.

Johnson, D. K. (2023). *The Lavender Scare: The Cold War Persecution of Gays and Lesbians in the Federal Government*. University of Chicago Press.

Kopczuk, W., E. Saez, and J. Song (2010). Earnings inequality and mobility in the united states: Evidence from social security data since 1937. *The Quarterly Journal of Economics 125*(1), 91–128.

Lee, J., G. Peri, and V. Yasenov (2022). The labor market effects of Mexican repatriations: Longitudinal evidence from the 1930s. *Journal of Public Economics 205*, 1–20.

Levi, M. and L. Stoker (2000). Political trust and trustworthiness. *Annual Review of Political Science 3*(1), 475–507.

Luttmer, E. F. P. (2005). Neighbors as Negatives: Relative Earnings and Well-Being. *The Quarterly Journal of Economics 120*(3), 963–1002.

Malmendier, U. and J. A. Wachter (2024). Memory of past experiences and economic decisions. In M. Kahana and A. Wagner (Eds.), *Handbook of Human Memory*. Oxford University Press.

Marcin, D. (2014). Essays on the revenue act of 1924. Phd thesis, University of Michigan.

McCleary, R. M. and R. J. Barro (2006). Religion and economy. *Journal of Economic Perspectives 20*(2).

Meyer, B. D. and J. X. Sullivan (2023). Consumption and income inequality in the united states since the 1960s. *Journal of Political Economy 131*(2), 247–284.

Moehling, C. M. (2001). Women's work and men's unemployment. *The Journal of Economic History 61*(4), 926–949.

Nissenbaum, H. (2010). *Privacy in Context: Technology, Policy, and the Integrity of Social Life.* Stanford, CA: Stanford University Press.

Perez-Truglia, R. (2020). The effects of income transparency on well-being: Evidence from a natural experiment. *American Economic Review 110*(4), 1019–54.

Piketty, T. and E. Saez (2003). Income Inequality in the United States, 1913-1998. *The Quarterly Journal of Economics 118*(1), 1–41.

Rosenbloom, J. L. and W. A. Sundstrom (2003). The decline and rise of interstate migration in the united states: Evidence from the ipums, 1850-1990. Working Paper 9857, National Bureau of Economic Research.

Roxworthy, E. (2008). *The Spectacle of Japanese American Trauma: Racial Performativity and World War II.* University of Hawaii Press.

Ruggles, S. (2024). When privacy protection goes wrong: How and why the 2020 census confidentiality program failed. *Journal of Economic Perspectives 38*(2), 201–26.

Ruggles, S., C. A. Fitch, R. Goeken, J. D. Hacker, M. A. Nelson, E. Roberts, M. Schouweiler, and M. Sobek (2022). IPUMS Ancestry Full Count Data: Version 3.3 [dataset]. Technical report, IPUMS.

Ruggles, S. and D. L. Magnuson (2023). It's none of their damn business: Privacy and disclosure control in the u.s. census, 1790-2020. *Population and Development Review 49*(3), 651–679.

Saavedra, M. (2021). Kenji or kenneth? pearl harbor and japanese-american assimilation. *Journal of Economic Behavior & Organization 185*, 602–624.

Seltzer, W. and M. Anderson (2001). The dark side of numbers: The role of population data systems in human rights abuses. *Social Research 68*(2).

Serrato, J. C. S. and P. Wingender (2016). Estimating local fiscal multipliers. *NBER Working Paper No. 22425*.

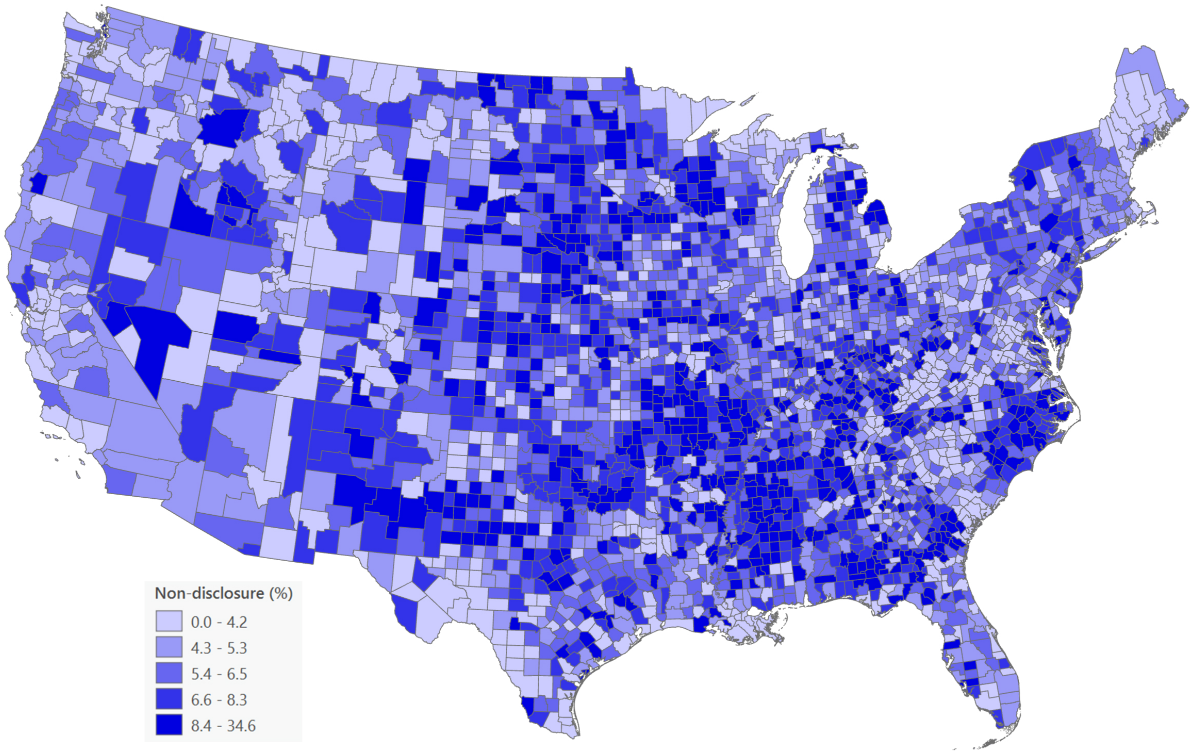
Stark, R. (1992). The reliability of historical united states census data on religion. *Sociological Analysis 53*(1), 91–95.

Thomson, C. A. H. (1940). Public relations of the 1940 census. *The Public Opinion Quarterly 4*(2), 311–318.

Tucker, C. (2023). The economics of privacy: An agenda. Working paper, MIT.

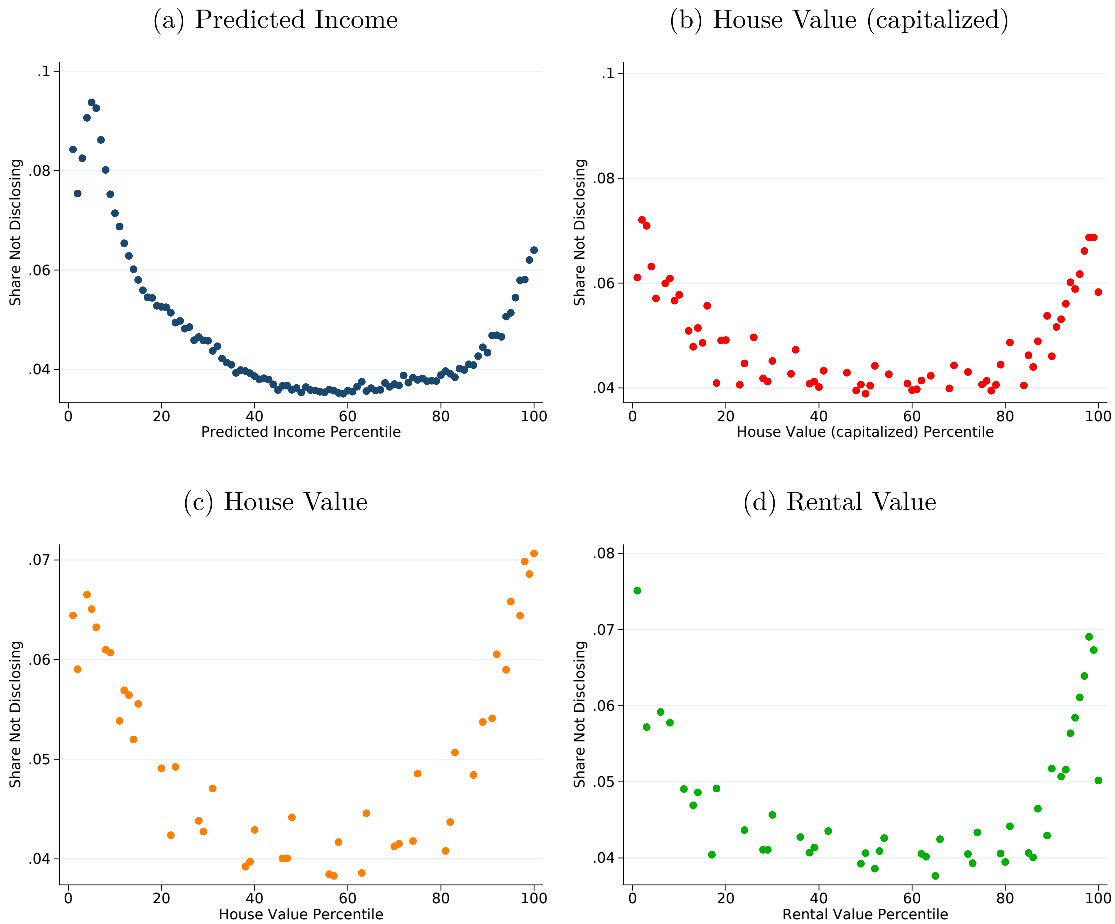
United States Senate (1940, February-March). Hearings Before the Subcommittee of a Committee on Commerce, United States 76th Congress: A Resolution Favoring the Deletion from the 16th Census Population Schedule of Inquiries Numbered 32 and 33, Relating to Compensation Received.

Figure 1: Income Privacy Across U.S. Counties



Notes: County-level means are from individual responses to the 1940 Census income question for wage earners aged 25-65. Non-disclosure is defined as failure to report income, given as zero or missing. Counties are shown using 1940 boundaries. Shading reflects county-level quintiles of the share of individuals withholding income.

Figure 2: Own Rank and Income Privacy Demands

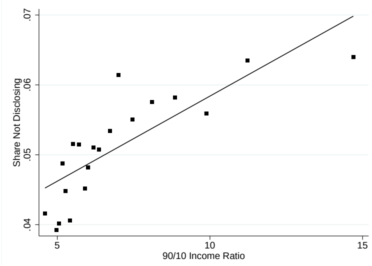


Notes: These figures show the mean income non-disclosure rate by income and wealth percentiles. The predicted income series used in Panel (a) is described in the notes to Appendix Figure A4. In Panel (b) we use capitalized house values using the actual house values and capitalized rental values whereas Panels (c) and (d) use actual house and rental values reported.

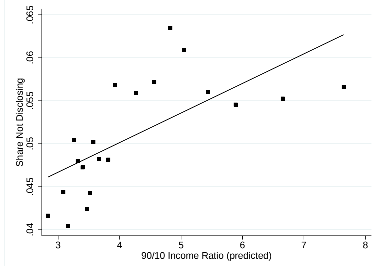
Figure 3: Population Predictors of Income Privacy Demands

Local Inequality vs. Income Privacy

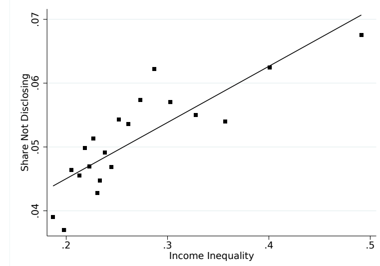
(a) 90/10 Income Ratio



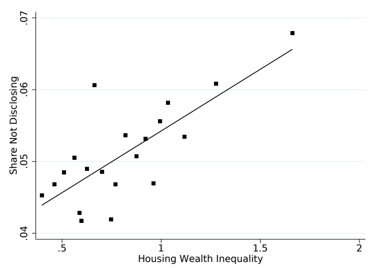
(b) Predicted 90/10



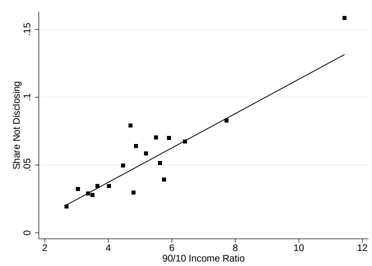
(c) Income Inequality (MLD)



(d) Housing Inequality (MLD)

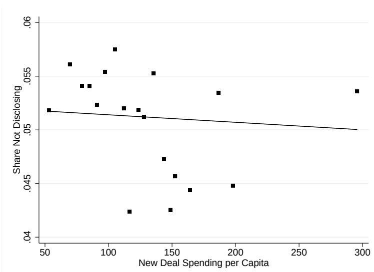


(e) Occupation 90/10

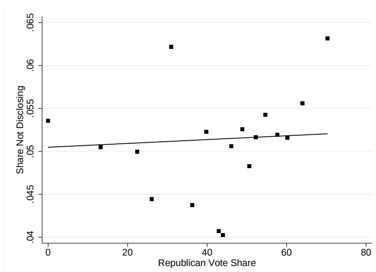


Political and Religious Views vs. Income Privacy

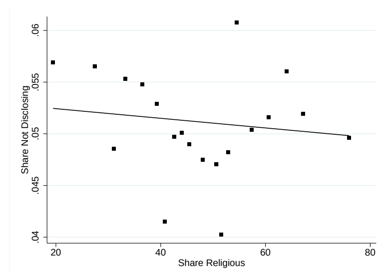
(f) New Deal per Capita



(g) Share Republican



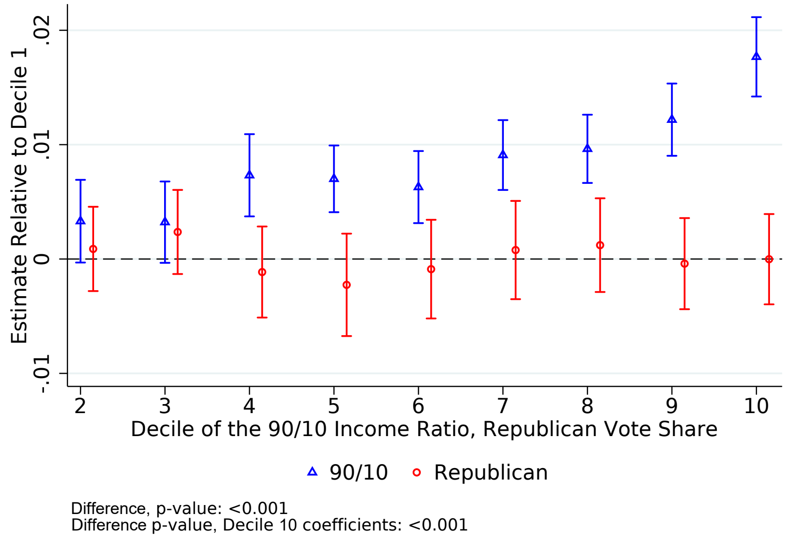
(h) Share Religious



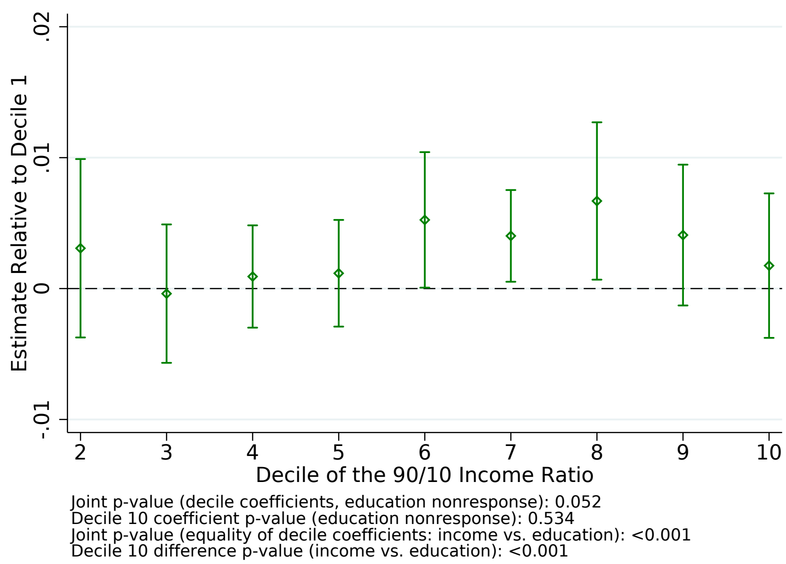
Notes: These figures show binned scatter plots. The predicted 90/10 income ratio is estimated using a predicted income series described in the notes to Appendix Figure A4. Housing and income inequality are the mean-log deviation of house values and incomes at the county-level respectively whereas the occupation 90/10 income ratio is the within occupation ratio using 228 US census occupation categories.

Figure 4: Privacy Demands by Income Decile

(a) Inequality and Republican Vote Share

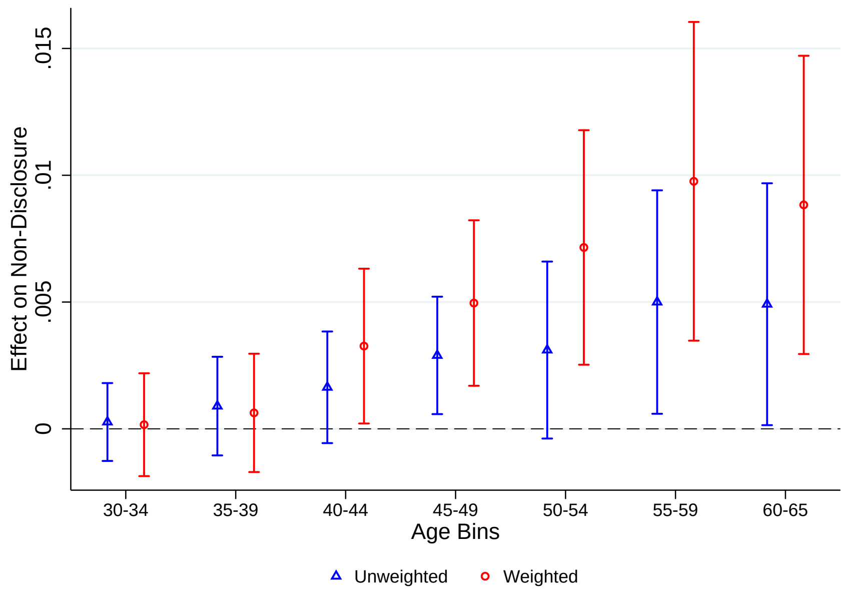


(b) Placebo: Education Nonresponse



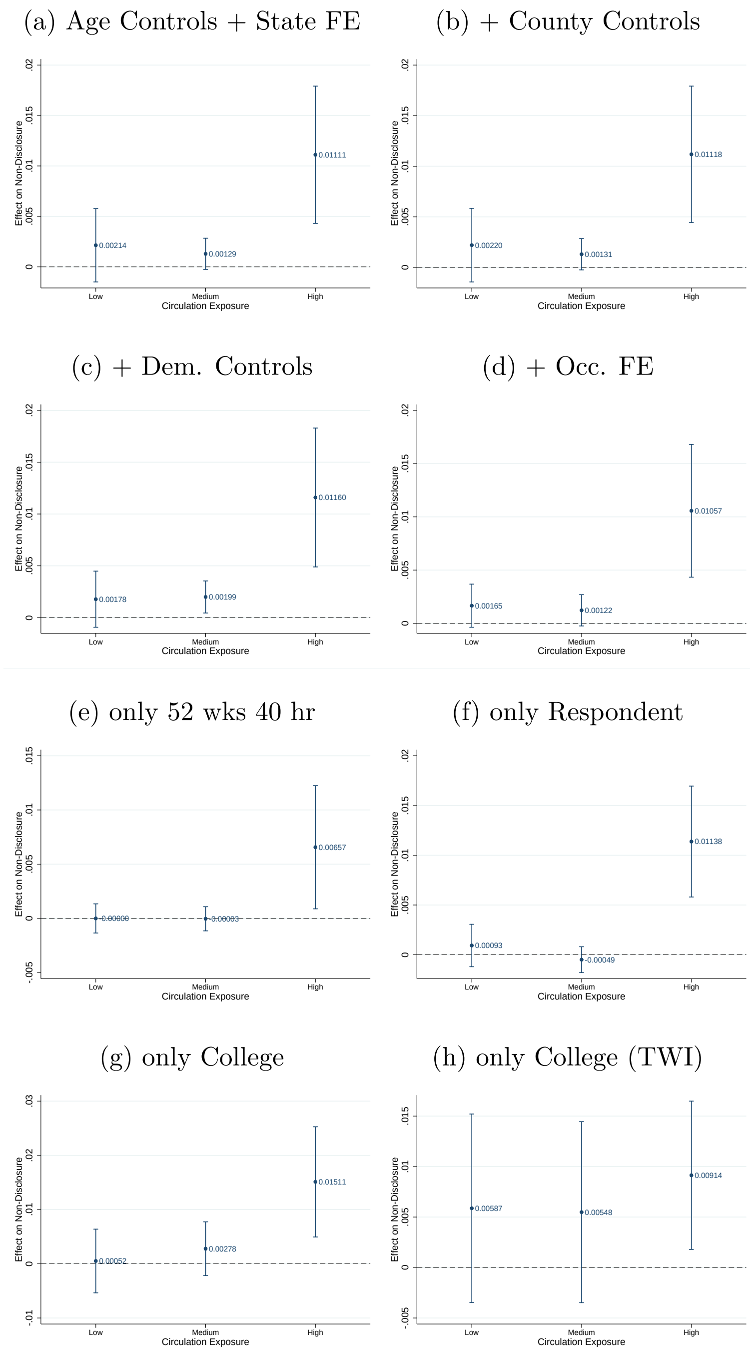
Notes: Panel (a) reports coefficients and 95% confidence intervals using indicators by decile of the 90/10 income ratio and the Republican vote share. Specification includes age, county and demographic controls, state fixed effects and full occupation fixed effects. The *p*-values are from tests of the null that the coefficients on the indicators by decile of the 90/10 income ratio are equal to the coefficients on the indicators for decile of the Republican vote share. Panel (b) reports analogous estimates using education nonresponse as the outcome variable. The *p*-values are from tests of whether the income decile coefficients are jointly zero and whether the top decile coefficient equals zero. Cross outcome tests examine whether the income decile coefficients, jointly and for the top decile, are equal between income and education nonresponse.

Figure 5: Privacy by Age Exposure in News List v. Non-News List Counties



Notes: This figures reports coefficients and 95% confidence intervals of privacy demands for individuals according to their age cohorts and exposure to newspaper lists. The coefficients are interactions between a newspaper list indicator and age cohorts following the specification in column 4 (unweighted) and column 6 (weighted) of Table III. The omitted category is 25-29 year-olds.

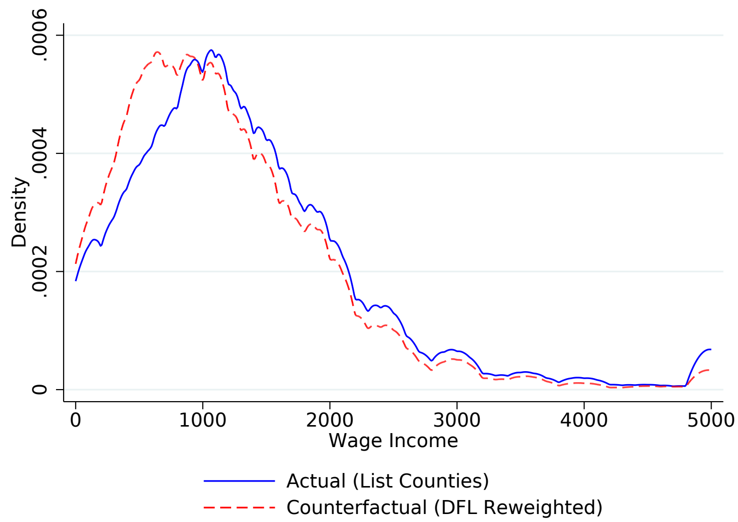
Figure 6: Privacy: Newspaper List Counties and Circulation Exposure



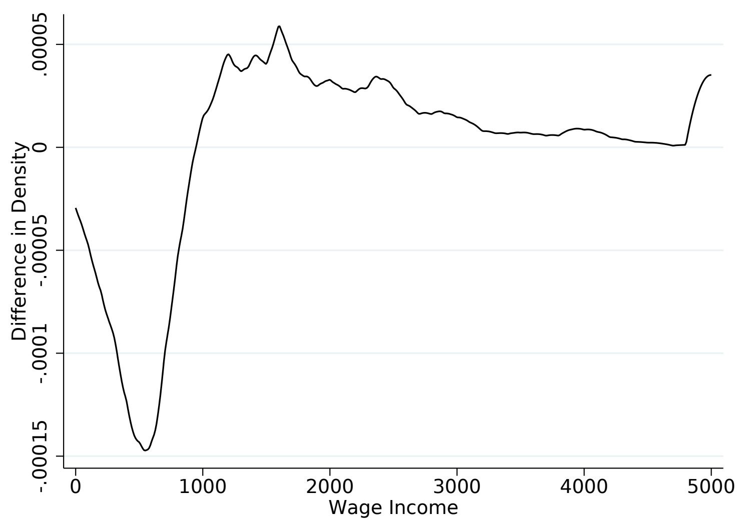
Notes: These figures reports coefficients and 95% confidence intervals of non-disclosure for individuals according to their newspaper list county status and age cohorts. These are plots of λ_5 λ_6 and λ_7 from equation 3. Panels (a)-(d) incorporate different controls and fixed effects sequentially so Panel (d) includes age controls, state fixed effects, county controls, demographic controls and occupation fixed effects. Panels (e)-(g) include these as well and Panel (h) adds two-way interactions (TWI). Regressions are weighted by each county's population in 1920. Standard errors clustered by county.

Figure 7: Actual and Counterfactual Income Distributions in Newspaper List Counties

(a) Actual and Counterfactual Densities



(b) Density Differences (Actual-Counterfactual)



Notes: These figures show the actual and counterfactual income distributions for newspaper-list counties. The counterfactual is constructed by applying DFL reweighting to respondents in non-list counties so that their observable characteristics match those of list-county individuals. The bottom panel plots the difference between the actual and counterfactual densities.

Table I: Privacy Demands and Inequality: 90/10 Income Ratio

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: 90/10 Income Ratio								
New Deal Spending per Capita (std)	-0.00011	0.00005	0.00011	0.00003	0.00002	0.00000	-0.00067*	0.00033
	(0.00038)	(0.00038)	(0.00037)	(0.00038)	(0.00037)	(0.00037)	(0.00036)	(0.00028)
Republican Vote Share (std)	0.00151**	0.00069	0.00059	0.00102	0.00092	0.00094	0.00126**	0.00108*
	(0.00077)	(0.00072)	(0.00070)	(0.00066)	(0.00065)	(0.00064)	(0.00058)	(0.00065)
Share Religious (std)	-0.00251***	-0.00055	-0.00070	-0.00096*	-0.00100**	-0.00100**	-0.00097**	-0.00064
	(0.00055)	(0.00054)	(0.00052)	(0.00052)	(0.00050)	(0.00050)	(0.00048)	(0.00048)
90/10 Income Ratio (std)	0.00989***	0.00708***	0.00653***	0.00432***	0.00401***	0.00367***	0.00366***	0.00271***
	(0.00063)	(0.00055)	(0.00052)	(0.00047)	(0.00047)	(0.00046)	(0.00050)	(0.00046)
Panel B: Predicted 90/10 Income Ratio								
New Deal Spending per Capita (std)	-0.00027	-0.00008	-0.00000	-0.00002	-0.00001	-0.00002	-0.00076**	0.00031
	(0.00041)	(0.00040)	(0.00039)	(0.00039)	(0.00039)	(0.00038)	(0.00039)	(0.00029)
Republican Vote Share (std)	0.00194**	0.00130*	0.00110	0.00123*	0.00109	0.00107	0.00150**	0.00113*
	(0.00081)	(0.00073)	(0.00071)	(0.00068)	(0.00067)	(0.00066)	(0.00060)	(0.00067)
Share Religious (std)	-0.00307***	-0.00089*	-0.00097*	-0.00101*	-0.00101**	-0.00098*	-0.00110**	-0.00056
	(0.00056)	(0.00054)	(0.00052)	(0.00053)	(0.00051)	(0.00051)	(0.00048)	(0.00049)
Predicted 90/10 Income Ratio (std)	0.00813***	0.00655***	0.00569***	0.00276***	0.00234***	0.00190***	0.00296***	0.00104**
	(0.00078)	(0.00064)	(0.00061)	(0.00057)	(0.00056)	(0.00054)	(0.00053)	(0.00053)
Panel C: Income Inequality (MLD)								
New Deal Spending per Capita (std)	-0.00001	0.00011	0.00016	0.00006	0.00005	0.00003	-0.00064*	0.00034
	(0.00036)	(0.00036)	(0.00036)	(0.00038)	(0.00037)	(0.00037)	(0.00036)	(0.00028)
Republican Vote Share (std)	0.00160**	0.00079	0.00066	0.00105	0.00094	0.00096	0.00128**	0.00109*
	(0.00070)	(0.00068)	(0.00067)	(0.00065)	(0.00064)	(0.00063)	(0.00056)	(0.00065)
Share Religious (std)	-0.00235***	-0.00066	-0.00080*	-0.00099*	-0.00102**	-0.00101**	-0.00100**	-0.00063
	(0.00051)	(0.00050)	(0.00048)	(0.00051)	(0.00049)	(0.00049)	(0.00046)	(0.00048)
Income Inequality (std)	0.01165***	0.00884***	0.00817***	0.00490***	0.00446***	0.00401***	0.00470***	0.00282***
	(0.00064)	(0.00059)	(0.00056)	(0.00054)	(0.00053)	(0.00052)	(0.00052)	(0.00051)
Observations	22281334	22281334	22281334	22281334	22281334	22281334	11599405	5383435
Clusters	3075	3075	3075	3075	3075	3075	3075	3075
Mean Dep Var.	0.047	0.047	0.047	0.047	0.047	0.047	0.035	0.047
Age Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographic Controls	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Main Occ. FE	No	No	No	Yes	No	No	No	No
Sub Occ. FE	No	No	No	No	No	No	No	No
Full Occ. FE	No	No	No	No	Yes	Yes	Yes	Yes
Sample	Full	Full	Full	Full	Full	Full	52 wks 40 hrs	Resp.

Notes: This table reports linear probability regression coefficients where the dependent variable is 1,0 for privacy (non-disclosure). Age controls are a linear and quadratic term in age. County controls are the 1940 population, manufacturing value-added, the share urban and the share of the population 25+ completing high school. Demographic controls are indicators for gender, household head, marriage status (married, divorced/separated single), race, immigrant and college attendance and a continuous variable for capitalized house values. Occupation fixed effects at the main (11), sub (22) and full (228) levels. Resp. (column 8) refers to the respondent only subsample of individuals who personally reported their income to the enumerator. Standard errors clustered by county in parentheses. *p<0.1, **p<0.05, ***p<0.01.

Table II: Privacy and the Subject-Enumerator Housing Wealth Gap

	Capitalized		Reported		Capitalized		Reported	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
House Gap Ratio (log)	0.00232***	0.00246***	0.01148***	0.01143***	0.00265***	0.00292***	0.01278**	0.01630***
	(0.00064)	(0.00064)	(0.00403)	(0.00443)	(0.00081)	(0.00081)	(0.00548)	(0.00590)
Placebo House Gap Ratio (log)		-0.00041		0.00162		-0.00001		0.00214
		(0.00055)		(0.00229)		(0.00080)		(0.00304)
Difference, p-value		0.0007		0.0293		0.0111		0.0181
Mean Dep Var.	0.042	0.042	0.043	0.043	0.042	0.042	0.043	0.043
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gender Matching	No	No	No	No	Yes	Yes	Yes	Yes
Clusters	181252	181252	18290	18290	124971	124971	13251	13251
Observations	276712	276712	27328	27328	163887	163887	17320	17320

Notes: This table reports linear probability regression coefficients where the dependent variable is 1,0 for privacy (non-disclosure). The right-hand side variables specify the log of the housing wealth gap between the subject and the enumerator or the subject and a randomly selected enumerator of the same gender drawn from an enumeration district in a different state. Controls include the log of the subjects own level of housing wealth, their age and years of education, gaps in age and education between subject and (random) enumerator and mean log housing wealth, age and educational attainment in an enumeration district. Standard errors clustered by household in parentheses. *p<0.1, **p<0.05, ***p<0.01.

Table III: Privacy and Exposure to Newspaper Tax Lists

	(1)	(2)	(3)	(4)	(5)	(6)
Newslist	0.00252	0.00252	0.00163	0.00124	0.00190	0.00082
	(0.00362)	(0.00362)	(0.00369)	(0.00363)	(0.00399)	(0.00389)
Age: 40-65		0.00264***	0.00114	-0.01312***	-0.00066	-0.01823***
		(0.00041)	(0.00115)	(0.00277)	(0.00138)	(0.00402)
Newslist × Age: 40-65			0.00188	0.00275**	0.00349*	0.00579***
			(0.00147)	(0.00122)	(0.00205)	(0.00163)
Observations	7772606	7772606	7772606	7772606	7772606	7772606
Clusters	49	49	49	49	49	49
Mean Dep Var.	0.043	0.043	0.043	0.043	0.043	0.043
Age Controls	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
County Controls	Yes	Yes	Yes	Yes	Yes	Yes
Demographic Controls	Yes	Yes	Yes	Yes	Yes	Yes
Full Occ. FE	Yes	Yes	Yes	Yes	Yes	Yes
Two-Way Interactions			No	Yes	No	Yes
Weighted			No	No	Yes	Yes

Notes: This table reports linear probability regression coefficients where the dependent variable is 1,0 for privacy (non-disclosure). Newslist is an indicator for counties where tax lists were published. Age controls are a linear and quadratic term in age. County controls are the 1940 population, manufacturing value-added, the share urban and the share of the population 25+ completing high school. Demographic controls are indicators for gender, household head, marriage status (married, divorced/separated single), race, immigrant and college attendance and a continuous variable for capitalized house values. Occupation fixed effects at the full (228) level. In columns 5 and 6 regressions are weighted by each county's population in 1920. Standard errors clustered by county in parentheses. *p<0.1, **p<0.05, ***p<0.01.

Table IV: Privacy by Probability of Internment (Japanese)

	(1)	(2)	(3)	(4)	(5)	(6)
Prob. Internment (std)	-0.0010	-0.0010	-0.0102*	-0.0112*	-0.0148*	-0.0147*
	(0.0043)	(0.0043)	(0.0060)	(0.0061)	(0.0079)	(0.0080)
Observations	8327	8327	8327	8327	8327	7742
Mean Dep Var.	0.063	0.063	0.063	0.063	0.063	0.064
Age Controls	No	Yes	Yes	Yes	Yes	Yes
County Controls	No	No	Yes	Yes	Yes	Yes
Individual Controls	No	No	No	Yes	Yes	Yes
Region FE	No	No	No	No	Yes	West Only

Notes: This table reports linear probability regression coefficients where the sample is individuals of Japanese origin in the 1940 census. The dependent variable is 1,0 for privacy (non-disclosure) run on the standardized probability of internment for each individual from Arellano-Bover (2022) based on their sex, birth year, and state. Individuals in all regions are included in columns 1 to 5 and individuals in the West of the U.S. only in column 6. Standard errors are clustered by county. *p<0.1, **p<0.05, ***p<0.01.

Table V: Privacy and Threat Risk (Mexican)

	(1)	(2)	(3)	(4)	(5)	(6)
Threat (std)	-0.0070***	-0.0069***	-0.0074***	-0.0076***	-0.0073**	-0.0111**
	(0.0015)	(0.0015)	(0.0028)	(0.0028)	(0.0029)	(0.0043)
Observations	22353	22353	22353	22353	22353	10952
Mean Dep Var.	0.027	0.027	0.027	0.027	0.027	0.025
Age Controls	No	Yes	Yes	Yes	Yes	Yes
County Controls	No	No	Yes	Yes	Yes	Yes
Individual Controls	No	No	No	Yes	Yes	Yes
Region FE	No	No	No	No	Yes	West Only

Notes: This table reports linear probability regression coefficients where the sample is individuals of Mexican origin in the 1930 census traced to the 1940 census using the links from Abramitzky et al. (2020). The dependent variable is 1,0 for privacy (non-disclosure) run on the standardized 'threat' exposure defined as the number of policemen, detectives, marshals and constables, sheriffs and bailiffs in a county scaled by the 1940 county population. Individuals in all regions are included in columns 1 to 5 and individuals in the West of the U.S. only in column 6. Standard errors are clustered by county. *p<0.1, **p<0.05, ***p<0.01.

Online Appendix

Demand for Privacy from the U.S. Census

Zoë Cullen
Harvard Business School

Tom Nicholas
Harvard Business School

Figure A1: Newspaper Stories and Senate Hearing

An Official View of Proposed Questions for 1940 Census

Director Explains Bureau's Position on Controversial Inquiries, Taking Exception to The Times Articles Which Are Defended by the Writer

Census Snooping Stirs Senate Storm

INCOME QUERIES UNLAWFUL, SAYS FOE OF SCHEME

Senate Group Requests Ban On Census Income Questions

By a Staff Correspondent of The Christian Science Monitor

WASHINGTON, March 12—The Senate Commerce Committee today recommended that the Senate ask the Census Bureau to eliminate from the 1940 census two much-disputed questions on personal incomes.

tor Francis T. Maloney (D) of Connecticut, which would have expressed a "desire" of the Senate that the Director of the Census reconsider the importance of the personal income questions and give "thoughtful consideration to the current allegation that such

PRESIDENT CHARGES TOBEY ASKS PUBLIC TO BREAK THE LAW

First American Senator to Do So, Says Roosevelt Defending Census Query on Income

Hopkins Revises Census Querying To Meet Protests on Income Data

Compromise Order Permits an Objector to Fill in Blank, Unsigned, and Seal It in Franked Envelope for Mailing

Special to THE NEW YORK TIMES.

Census: Are Questions on Income Legal?

By a Staff Correspondent of The Christian Science Monitor

WASHINGTON, Feb. 28—Has the Census Bureau exceeded the authority granted it by Congress in including two questions on personal incomes in its 1940 schedule?

"These things present a great social and political problem, and the facts about them ought to be brought out," he told the Committee.

"It is our business," replied Senator White, "but has Congress made it our business?"

Senator Tobey, who marshalled

becoming "Paul Prys and Sally Snoops."

"It isn't necessary in a Republic such as ours, that every citizen should live in a gold fish bowl," he insisted. "There are just some things that are none of the public's business. If the Government keeps on encroaching on the

CENSUS AIDES BACK INCOME QUESTIONS AS WOMEN PROTEST

Director Reminds Senators No One Has Ever Been Jailed Yet for Failing to Answer

REVOLT PICTURED BY FOES

Prisons Will Overflow to Halt the 'Bureaucratic Snooping,' One Woman Asserts

1940 Census Will Go Into Economics

Employment Will Be Studied, Along With Income, Birth Rate

PREDICT ERASURE OF INCOME QUERY IN CENSUS PRYING

Senators Aroused Over Display of Burocracy.

BY CHESLY MANLY.
[Chicago Tribune Press Service.]

Uncle Sam Is Getting Much More Inquisitive

1940 Census Will Propound Some New Questions

WASHINGTON, March 3 (AP).—Are you working, and how much do you make?

Do you own your home, and how much is it worth?

Where were you and what were you doing ten years ago?

These are the new questions, it was learned today, that the census man will ask you next year. He wants to know a lot more than your age and birthplace.

A tentative draft of the 1940 questionnaire, prepared by the Census Bureau's chief statistician, includes these new questions, and further conferences may add a few more. About 100 leaders in business, labor and education have been asked to offer suggestions.

1940 CENSUS

HEARINGS

BEFORE A

SUBCOMMITTEE OF THE COMMITTEE ON COMMERCE UNITED STATES SENATE

SEVENTY-SIXTH CONGRESS

THIRD SESSION

ON

S. Res. 231

A RESOLUTION FAVORING THE DELETION FROM THE SIXTEENTH CENSUS POPULATION SCHEDULE OF INQUIRIES NUMBERED 32 AND 33, RELATING TO COMPENSATION RECEIVED

FEBRUARY 28, 29 AND MARCH 1, 1940

Notes: Newspaper headlines surrounding the 1940 income questions in the census, and the front cover of the documentary evidence summarizing the 1940 Senate Hearing on the proposed deletion of these questions.

Figure A2: Publication of Tax Lists


THE NEW YORK TIMES, WEDNESDAY, SEPTEMBER 2, 1925. 9

Income Tax Payments in 1925 by Individuals and Corporations on 1924 Income

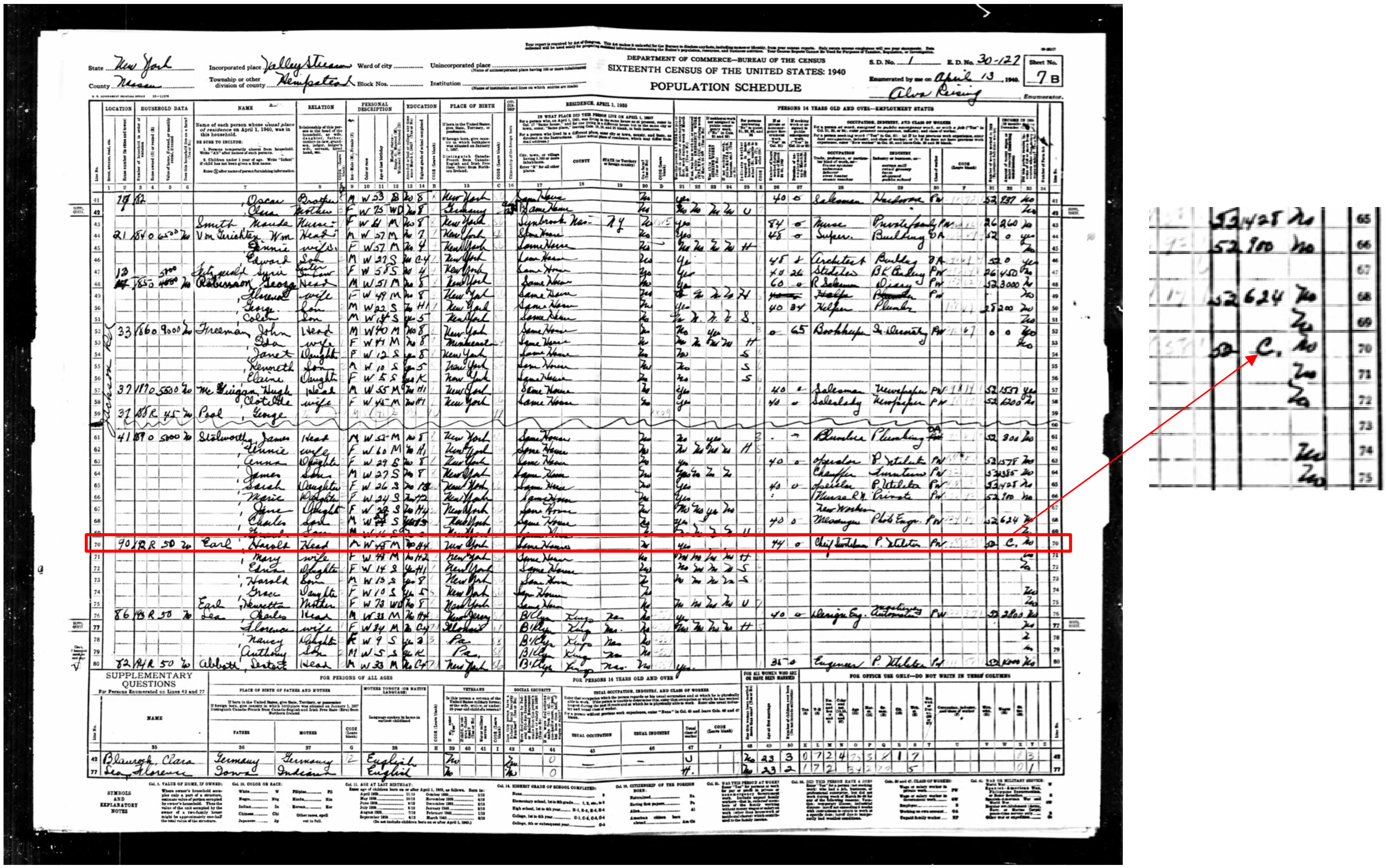
NEW YORK CITY

Second District

INDIVIDUALS

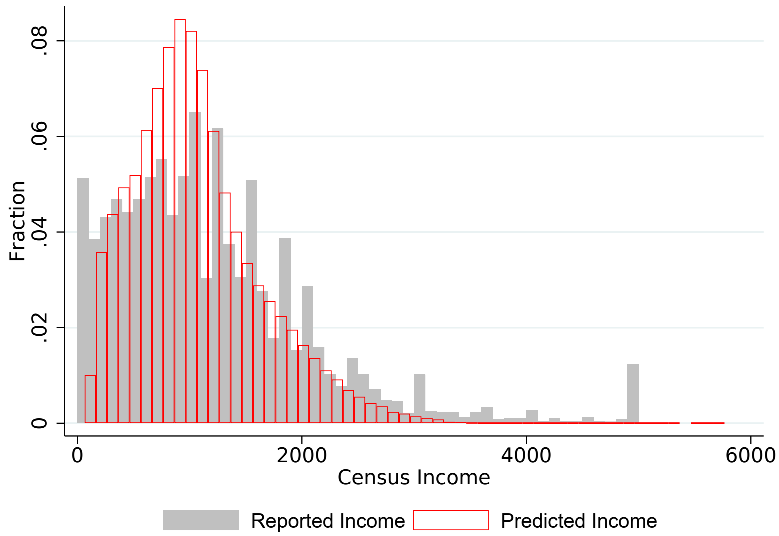
Notes: This shows an example of the tax lists published by *The New York Times*.

Figure A3: 1940 Census Population Schedule Where Individual had Used Confidential Form P-16



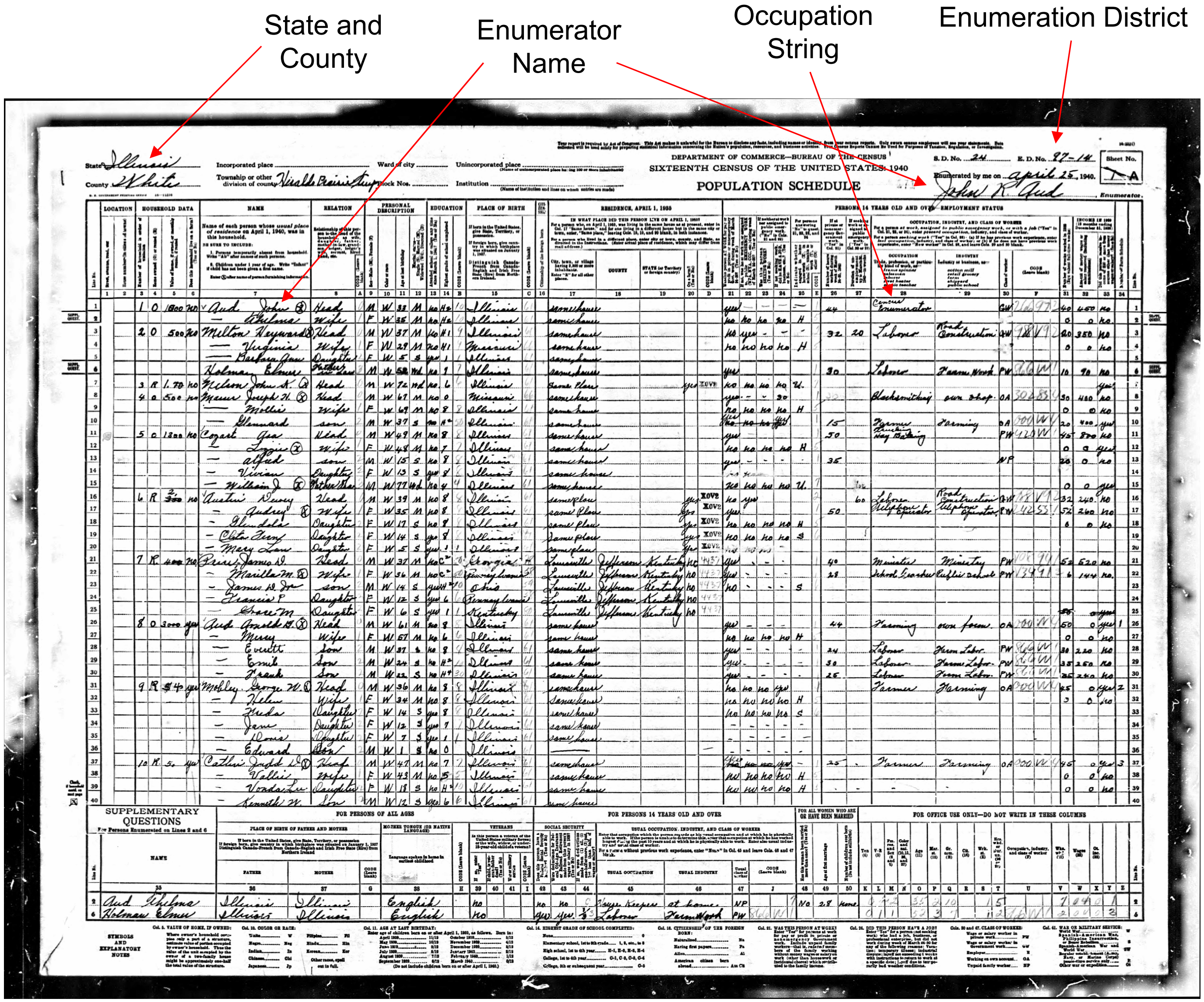
Notes: This shows an example of a Census population schedule illustrating that an individual had elected to use form P-16 to submit their income in a sealed envelope rather than stating it to the enumerator.

Figure A4: The Distribution of Reported and Predicted Income



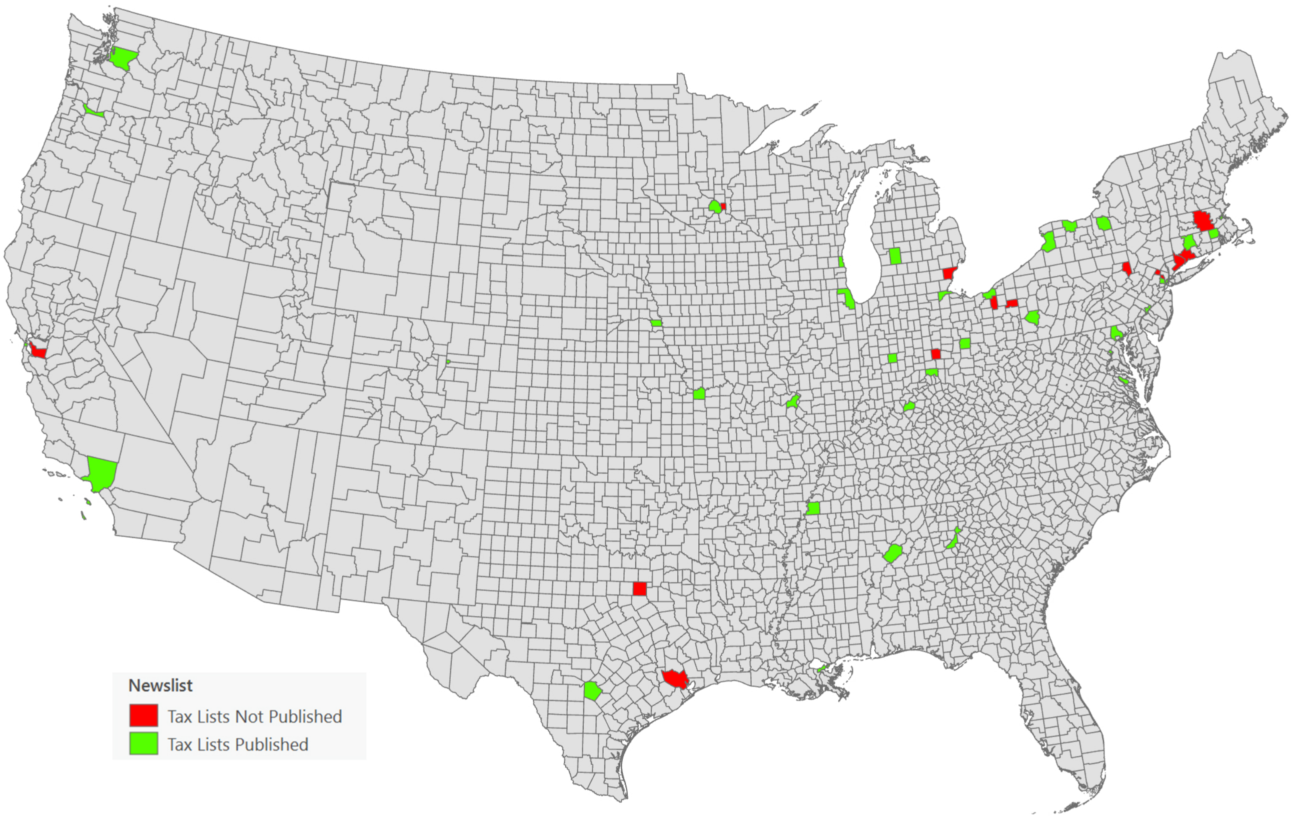
Notes: This figure plots kernel densities of reported income from the 1940 census and predicted income based on demographic characteristics. We first regress log income for individuals with positive reported incomes on their age, a quadratic in age, their capitalized house value, years of education, indicators for gender, race, and fixed effects for state and occupation. We then use the model to predict income for all individuals in the dataset including those for whom reported income is missing or reported as zero.

Figure A5: 1940 Census Population Schedule



Notes: This shows an example of a Census population schedule illustrating the information we use to identify enumerators.

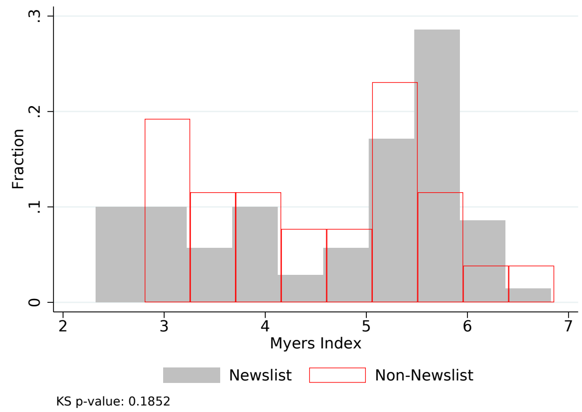
Figure A6: Counties where Tax Lists were Published



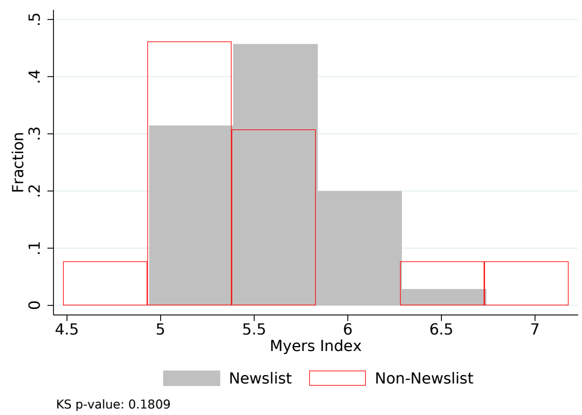
Notes: This map shows counties in which tax lists of prominent earners were publicly released or not, focusing on the 50 largest cities by population. Data are compiled from Marcin (2014).

Figure A7: Myers Index: Newspaper List versus Non-Newspaper List Counties

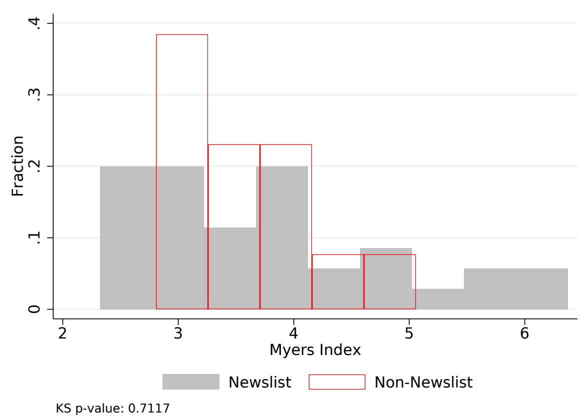
(a) Ages 25 to 65



(b) Ages 25 to 39



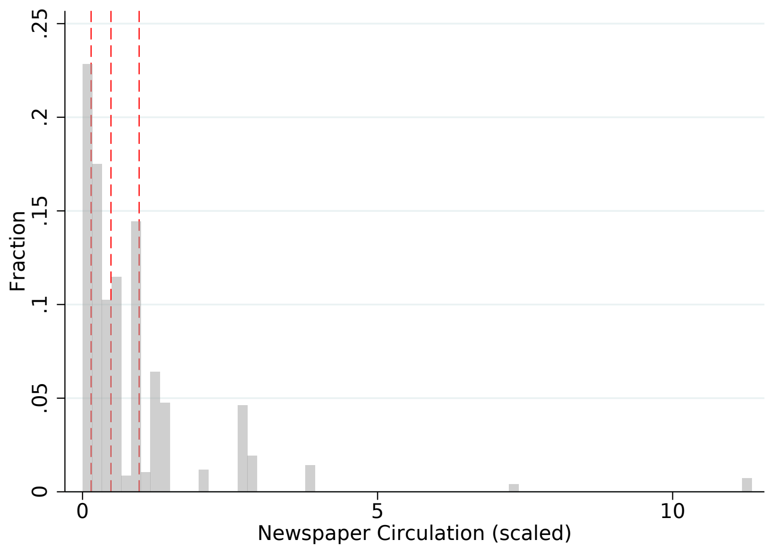
(c) Ages 40 to 65



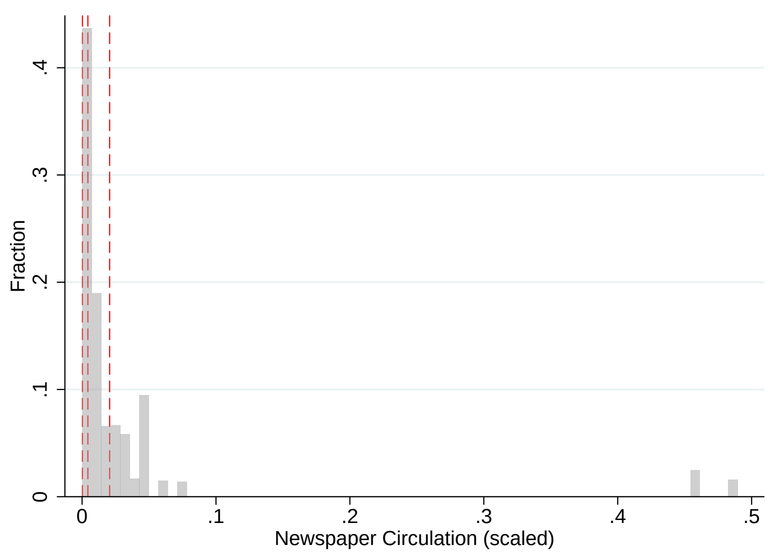
Notes: These figures show the distribution of Myers Index of age-heaping calculated for county-age cells in newspaper list and non-newspaper list counties. Age cells are 25 to 39 and 40 to 65 year olds. A Myers Index of 0 indicates no age-heaping (reported ages are evenly distributed across all final digits from 0 to 9) whereas a value of 90 indicates perfect heaping (every age is reported using only one final digit). Kolmogorov-Smirnov exact p-value reported under the null that the distributions are the same.

Figure A8: Histograms of Newspaper Circulation

(a) County Circulation

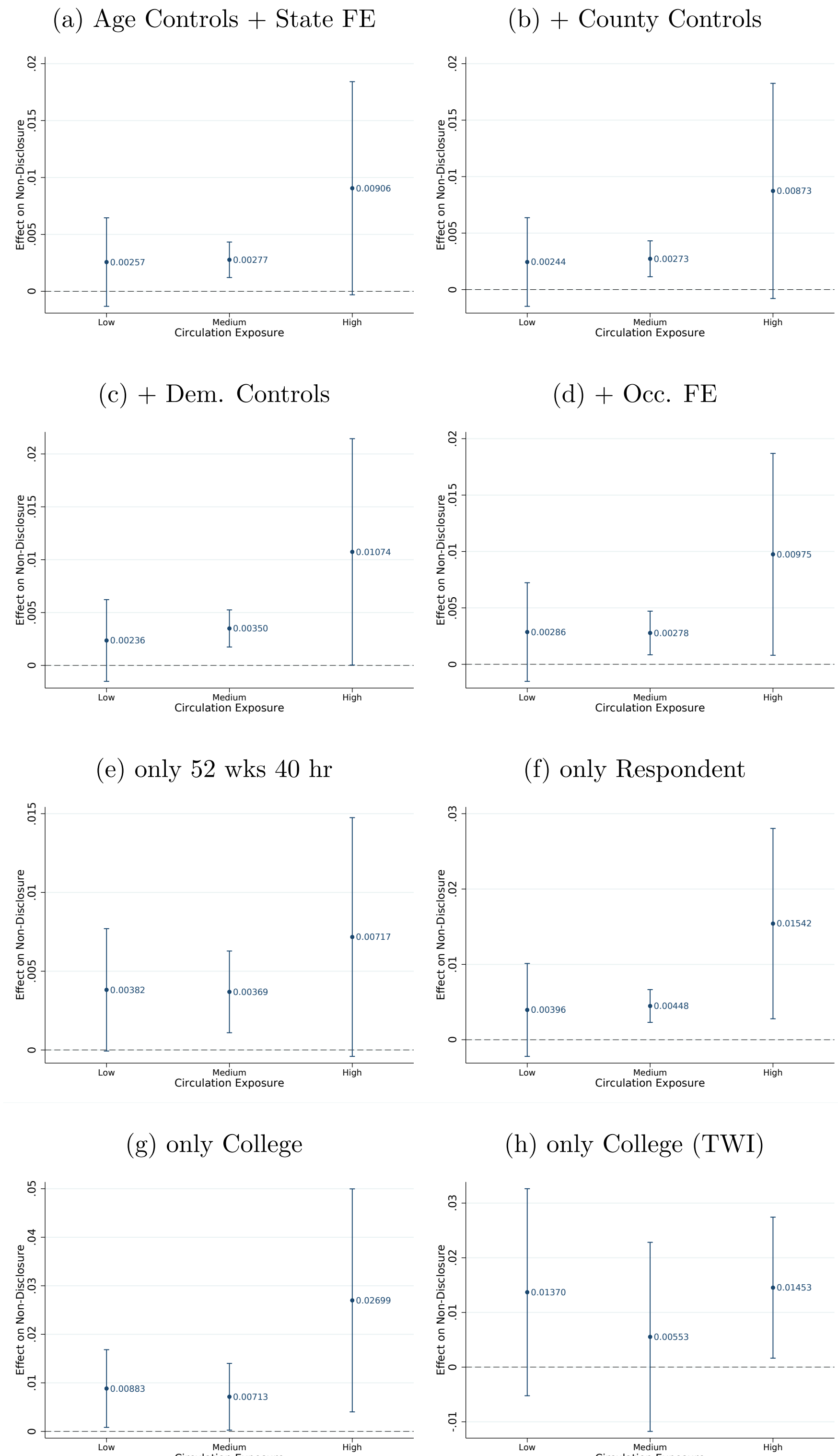


(b) County Circulation (Multi-Town)



Notes: This figure shows the distribution of newspaper circulation numbers scaled by 1920 county population. Panel (a) measures circulation in counties with newspapers that published the lists. Panel (b) allows for those newspapers to circulate in other counties as well. Vertical dashed lines represent splits by tercile. Panel (a) includes weekday and weekend circulation. Panel (b) uses weekday circulation based on the data from Gentzkow et al. (2014).

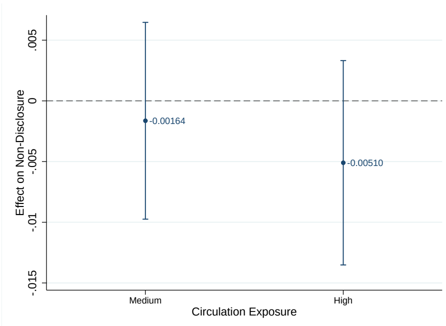
Figure A9: Replication of Figure 6 using only Stayers 1920-1940



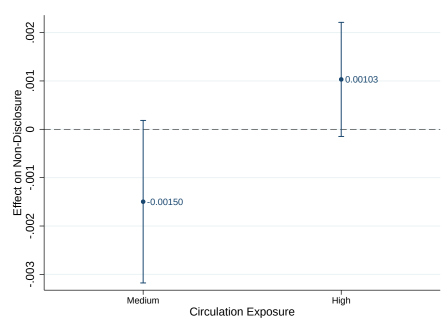
Notes: These figures replicate the results in Figure 6 using only individuals who remained in a newspaper list or non-newspaper list county across the 1920 to 1940 censuses. We use the male-only links provided by Abramitzky et al. (2020). A stayer is defined as an individual who remained in the same newspaper list county, moved to different newspaper list county in the same state, or a different newspaper list county in another state.

Figure A10: Privacy: Newspaper List Counties and Multi-Way Circulation Exposure

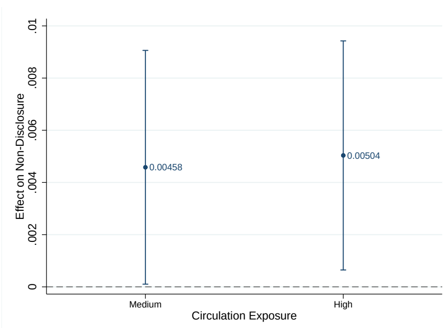
(a) No Interaction



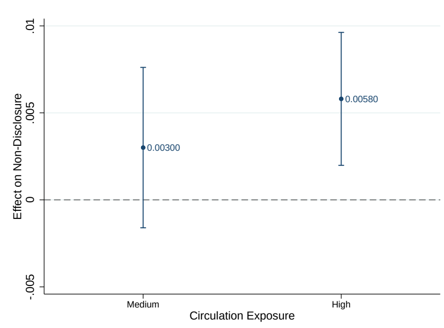
(b) $\widehat{Circ.} \times Age$



(c) $\widehat{Circ.} \times Age \times Newslist$



(d) $\widehat{Circ.} \times Age \times Newslist$ (TWI)



Notes: These figures reports coefficients and 95% confidence intervals of non-disclosure for individuals according to their newspaper list county status and age cohorts. Panel (a) includes the exposure indicator variables in the original, non-interacted form whereas Panel (b) interacts these variables with the indicator for age. Panels (c) and (d) are plots of the triple interactions, ψ_{10} and ψ_{11}, from equation 4 (i.e. interacting the exposure indicators with the indicator for age and the indicator for news list counties). All specifications include age controls, county controls, demographic controls, occupation and state fixed effects. TWI refers to two-way interactions. Regressions are weighted by each county's population in 1920. Standard errors clustered by county.

Table A1: Descriptives

	Missing Income		Zero Income		Income Reported	
	Mean	SD	Mean	SD	Mean	SD
Panel A: County-Level						
New Deal Spending per Capita	128.600	67.505	133.734	78.929	133.173	71.467
Republican Vote Share	40.887	21.716	35.747	22.411	37.556	21.551
Share Religious	47.972	14.169	45.997	14.354	47.047	14.031
Share Urban	61.969	31.015	61.742	32.546	66.184	30.161
Share Educated	24.173	7.304	24.469	7.916	25.077	7.558
County Population 1940 (Millions)	0.543	0.877	0.664	1.030	0.700	1.049
Manufacturing Value Added (Millions)	354.687	681.030	474.344	866.078	487.991	862.583
90/10 Income Ratio	7.088	2.532	7.453	2.855	6.968	2.531
Predicted 90/10 Income Ratio	4.274	1.251	4.427	1.332	4.239	1.281
Housing Wealth Inequality	0.865	0.334	0.828	0.324	0.805	0.302
Income Inequality	0.272	0.076	0.283	0.085	0.268	0.076
Panel B: Occupation-Level						
Occupation 90/10 Income Ratio	5.375	2.067	6.089	2.411	5.021	1.690
Panel C: Individual-Level						
Privacy	1.000	0.000	1.000	0.000	0.000	0.000
Age	40.709	10.781	41.280	11.331	39.599	10.546
Male = 1	0.751	0.432	0.650	0.477	0.755	0.430
Household Head = 1	0.579	0.494	0.508	0.500	0.649	0.477
Married = 1	0.724	0.447	0.741	0.438	0.774	0.418
Divorced/Separated = 1	0.023	0.151	0.032	0.176	0.023	0.149
Single = 1	0.253	0.435	0.227	0.419	0.203	0.402
White = 1	0.918	0.275	0.883	0.321	0.905	0.294
Immigrant = 1	0.106	0.308	0.146	0.353	0.139	0.346
Years of Education	9.591	3.713	8.853	3.814	9.093	3.602
College = 1	0.168	0.374	0.133	0.340	0.130	0.336
Home Owner = 1	0.530	0.499	0.582	0.493	0.578	0.494
House Value	4423.168	7543.008	3901.968	8834.087	3846.472	6875.483
Rental Value	89.912	478.343	58.375	347.402	62.106	361.948
Capitalized House Value	7795.931	42219.042	5708.250	32350.924	5931.850	33376.577
Weeks Worked	47.503	9.828	47.279	10.374	45.076	11.320
Hours Worked	43.956	11.785	44.480	15.206	43.095	11.536
Wage Income	.	.	0.000	0.000	1224.076	891.091
Non-Wage Income = 1	0.285	0.451	0.565	0.496	0.133	0.340

Notes: This table reports descriptive statistics for observations in the 1940 census for individuals aged 25 to 65 years who were in the labor force, who self-reported being at work, and who received wages or a salary, including those who worked in government. We define income non-disclosure as missing incomes and zero incomes. We estimate capitalized house values using the actual house values and capitalized rental values at 10%. The predicted 90/10 income ratio is estimated using a predicted income series described in the notes to Appendix Figure A4. The occupation 90/10 income ratio is the within occupation ratio using 228 US census occupation categories. The variables "Housing Wealth Inequality" and "Income Inequality" are the mean-log deviation of house values and incomes at the county-level respectively.

Table A2: The Super Rich and Non-Disclosure

Name	Net Income 1940	Weeks Worked 1939	Census Income	Census Non-Wage
John D. Rockefeller Jr	3,789,204	0	0	Yes
Clarence Dillon	129,019	52	0	Yes
Sid W. Richardson Jr	-264,498	52	5000	Yes
Reuben H. Fleet	291,013	52	5000	Yes
Richard K. Mellon	4,069,178	52	5000	Yes
Paul Mellon	5,074,832	52	0	Yes
Sarah M. Scaife	4,021,264	0	0	Yes
George L. Hartford	3,140,642	52	5000	Yes
Ailsa M. Bruce	2,074,634	0	0	Yes
Edsel B. Ford	3,483,889	52	0	No
Charles S. Chaplin	under 100,000	52	5000	Yes
Edgar Palmer	1,883,406	52	5000	Yes
Jeremiah Milbank Sr	211,628	52	5000	Yes
Katherine S. Milbank		0	0	Yes
Arthur V. Davis	2,054,765	0	0	Yes
John A. Hartford	2,819,498	52	5000	Yes
Minnie H. Reilly	3,029,144	Missing	Missing	Yes
Alfred P. Sloan Jr	2,169,154	45	5000	Yes
Irene J. Sloan		0	0	Yes
Lammot Du Pont	1,805,381	52	5000	Yes
Jessie B. Du Pont	1,785,279	Missing	Missing	Yes
Everette L. Degolyer	under 100,000	52	0	Yes
Nell V. Degolyer		0	0	No
Edward J. Noble	209,380	52	5000	Yes
William Du Pont Jr	1,458,160	52	5000	Yes
Alwin C. Ernst	1,303,815	52	5000	Yes
Charles S. Mott	1,623,670	Missing	Missing	Yes
Ethel M. Dorrance	2,152,426	Missing	Missing	Yes
James H. Cannon	339,754			
Mary S. Harkness	1,596,543	0	0	Yes
Henry Ford	2,933,531	52	0	Yes
Irenee Du Pont	1,702,128	Missing	Missing	Yes
Felix W. Zelcer	117,247			
Ignatius J. Miranda	122,757	52	5000	Yes
Alfred J. Miranda Jr	126,632	52	5000	Yes
Mary D. Biddle	1,310,094			
Gregory Ferend	under 100,000			
Robert S. Clark	1,101,090			
Allen G. Oliphant	under 100,000	52	5000	Yes
Anita M. Blaine	1,046,439			
Walter P. Murphy	950,436	35	5000	Yes
Mills Bennett	118,582	0	0	Yes
William R. Coe	1,244,800	0	0	Yes
Lammot D. Copeland	1,109,660	0	5000	Yes
Marie H. Robertson	1,357,449			
Robert R. M. Carpenter	1,125,524	52	5000	Yes
George H. Hartford II	1,432,434			
Evelyn Mendelssohn	1,138,971			
Garfield A. Wood	under 100,000	52	0	Yes
Helen H. Whitney	1,079,321	0	0	Yes
Josephine H. McIntosh	1,301,990	0	0	Yes
Marion D. Scott	1,025,286	0	0	Yes
Joan W. Payson	357,543	0	0	Yes
Alexis F. Du Pont	875,502	0	0	Yes

Name	Net Income 1940	Weeks Worked 1939	Census Income	Census Non-Wage
William T. Grant	1,246,739	52	0	Yes
Samuel H. Kress	1,657,698	52	5000	Yes
Cartter T. Lupton	645,054	52	5000	Yes
Ella B. Kearney	246,906	0	0	Yes
Henry B. Du Pont Jr	953,829	52	5000	Yes
Jessie W Donahue	1,260,734			
Rudolf J. Schaefer	672,878	52	5000	Yes
Lucia M. Schaefer		0	0	No
Frederick M. E. Schaefer	616,211	52	5000	Yes
Eugene Du Pont Jr	697,475	52	Missing	Yes
Harry P. Bingham	930,782	52	0	Yes
Alexander Smith	163,021			
Clifford Mooers	under 100,000			
Thomas M. O'Connor Jr	826,945	52	0	Yes
Marjorie P. Davies	851,741	Missing	Missing	Missing
Katherine D. Butterworth	584,471	0	0	Yes
Carl G. Swebilius	under 100,000	52	5000	Yes
Hulda Swebilius		Missing	Missing	No
Joseph E. Widener	1,475,478	0	0	Yes
Francis N. Bard	623,735	52	5000	Yes
Philip K. Wrigley	860,257	52	5000	Yes
Edith H. Harkness	763,455			
Donaldson Brown	918,183	52	5000	Yes
Barclay Douglas	126,001			
Josephine H. Douglas				
Edwin A. Link	236,207	52	5000	Yes
George A. Adam	397,370			
Josiah K. Lilly Sr	818,883	0	0	Yes
John D. Jackson	767,878	52	5000	Yes
Raymond Pitcairn	706,051	0	0	Yes
Edward S. Moore	371,272	52	0	Yes
Robert W. Woodruff	763,187	52	5000	Yes
Mahlon D. Thatcher Jr	269,247	50	5000	Yes
William T. Rawleigh	534,490	52	Missing	Yes
Hugh R. Sharp	1,007,876	52	5000	Yes
Sarah G. Kenan	361,062	0	0	Yes
Abby A. Rockefeller	616,440	0	0	Yes
Stanley R. McCormick	464,400	0	0	Yes
E. F. Stokes	381,877			
Doris D. Cromwell	690,665	0	0	Yes
Glenn L. Martin	533,852	52	5000	Yes
Edward H. Moore	under 100,000	Missing	Missing	Missing
Walker P. Inman	772,073	0	0	Yes
Alta R. Prentice	726,204	0	0	Yes
Cora T. Burnett	687,694	0	0	Yes
Sydney M. Shoenberg	773,189	52	5000	Yes
Eli Lilly	757,626	52	5000	Yes
Leonie B. Guggenheim	719,416	0	0	Yes
Miguel J. Ossorio	789,699	52	5000	Yes
William R. Kenan Jr	381,121	52	0	Yes

Notes: We matched the list of the super rich from Brandes (1983), as compiled by the U.S. Treasury, to the 1940 Census. Names in red text are unmatched. Some individuals might have been abroad at the time of enumeration. For example, Robert S. Clark, inheritor of the Singer Sewing Machine fortune, resided in Normandy, France for part of the year, following his marriage to a French actress in 1919. Net Income is from the list itself which also includes incomes for 1941. Sid W. Richardson, an oil industry magnate, is reported as having a negative net income in 1940 but a net income of $3,948,794 in 1941. The data on weeks worked, census income (top-coded at $5,000) and census non-wage income are from the 1940 Census. Highlighted cells show our definition of non-disclosure as zero or missing responses to the income question. The list often includes individuals and their spouses because certain states allowed income sharing within the household for federal tax purposes.

Table A3: Privacy Demands and Inequality: Demographic Controls

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
New Deal Spending per Capita (std)	-0.00011	0.00005	0.00011	0.00003	0.00002	0.00000	-0.00067*	0.00033
	(0.00038)	(0.00038)	(0.00037)	(0.00038)	(0.00037)	(0.00037)	(0.00036)	(0.00028)
Republican Vote Share (std)	0.00151**	0.00069	0.00059	0.00102	0.00092	0.00094	0.00126**	0.00108*
	(0.00077)	(0.00072)	(0.00070)	(0.00066)	(0.00065)	(0.00064)	(0.00058)	(0.00065)
Share Religious (std)	-0.00251***	-0.00055	-0.00070	-0.00096*	-0.00100**	-0.00100**	-0.00097**	-0.00064
	(0.00055)	(0.00054)	(0.00052)	(0.00052)	(0.00050)	(0.00050)	(0.00048)	(0.00048)
90/10 Income Ratio (std)	0.00989***	0.00708***	0.00653***	0.00432***	0.00401***	0.00367***	0.00366***	0.00271***
	(0.00063)	(0.00055)	(0.00052)	(0.00047)	(0.00047)	(0.00046)	(0.00050)	(0.00046)
Male=1			-0.00334***	0.00456***	-0.00176**	-0.00454***	0.00412***	-0.01618***
			(0.00087)	(0.00074)	(0.00074)	(0.00076)	(0.00049)	(0.00075)
Household Head=1			-0.02769***	-0.02641***	-0.02630***	-0.02531***	-0.01215***	-0.01917***
			(0.00049)	(0.00041)	(0.00039)	(0.00035)	(0.00040)	(0.00043)
Divorced/Separated=1			0.00556***	0.00356***	0.00378***	0.00212***	0.00035	-0.00311***
			(0.00055)	(0.00052)	(0.00050)	(0.00049)	(0.00040)	(0.00059)
Single=1			-0.00403***	-0.00666***	-0.00447***	-0.00335***	0.00195***	-0.00920***
			(0.00053)	(0.00049)	(0.00048)	(0.00045)	(0.00038)	(0.00057)
White=1			0.00087	0.01118***	0.01092***	0.00706***	0.00178***	0.00949***
			(0.00088)	(0.00157)	(0.00159)	(0.00145)	(0.00066)	(0.00244)
Immigrant=1			-0.00286***	-0.00182***	-0.00207***	-0.00210***	0.00016	-0.00061
			(0.00071)	(0.00067)	(0.00061)	(0.00053)	(0.00069)	(0.00043)
College=1			0.00685***	0.00247***	0.00146***	0.00238***	0.00290***	-0.00160***
			(0.00096)	(0.00055)	(0.00042)	(0.00033)	(0.00035)	(0.00044)
Capitalized House Value (std)			0.00075***	0.00035***	0.00029***	0.00024***	0.00043***	-0.00008
			(0.00011)	(0.00010)	(0.00010)	(0.00009)	(0.00008)	(0.00014)
Observations	22281334	22281334	22281334	22281334	22281334	22281334	11599405	5383435
Clusters	3075	3075	3075	3075	3075	3075	3075	3075
Mean Dep Var.	0.047	0.047	0.047	0.047	0.047	0.047	0.035	0.047
Age Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographic Controls	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Main Occ. FE	No	No	No	Yes	No	No	No	No
Sub Occ. FE	No	No	No	No	No	No	No	No
Full Occ. FE	No	No	No	No	Yes	Yes	Yes	Yes
Sample	Full	Full	Full	Full	Full	Full	52 wks 40 hrs	Resp.

Notes: This table reports linear probability regression coefficients where the dependent variable is 1,0 for privacy (non-disclosure). It shows the coefficients on the demographic controls which are otherwise condensed for reporting purposes in Table I. Age controls are a linear and quadratic term in age. County controls are the 1940 population, manufacturing value-added, the share urban and the share of the population 25+ completing high school. Occupation fixed effects at the main (11) sub (22) and full (228) levels. Resp. (column 8) refers to the respondent only subsample of individuals who personally reported their income to the enumerator. Standard errors clustered by county in parentheses. *p<0.1, **p<0.05, ***p<0.01.

Table A4: Privacy Demands and Inequality: MLD Housing Inequality

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
New Deal Spending per Capita (std)	-0.00012	0.00011	0.00016	0.00006	0.00005	0.00003	-0.00066*	0.00033
	(0.00043)	(0.00040)	(0.00039)	(0.00039)	(0.00038)	(0.00038)	(0.00037)	(0.00029)
Republican Vote Share (std)	0.00239**	0.00043	0.00040	0.00090	0.00081	0.00084	0.00118*	0.00102
	(0.00101)	(0.00082)	(0.00079)	(0.00071)	(0.00068)	(0.00067)	(0.00063)	(0.00067)
Share Religious (std)	-0.00272***	-0.00012	-0.00032	-0.00072	-0.00078	-0.00081	-0.00082	-0.00049
	(0.00067)	(0.00064)	(0.00061)	(0.00058)	(0.00056)	(0.00055)	(0.00053)	(0.00050)
Housing Wealth Inequality (std)	0.00507***	0.00363***	0.00347***	0.00263***	0.00266***	0.00251***	0.00247***	0.00203***
	(0.00052)	(0.00047)	(0.00045)	(0.00042)	(0.00041)	(0.00040)	(0.00038)	(0.00039)
Observations	22281334	22281334	22281334	22281334	22281334	22281334	11599405	5383435
Clusters	3075	3075	3075	3075	3075	3075	3075	3075
Mean Dep Var.	0.047	0.047	0.047	0.047	0.047	0.047	0.035	0.047
Age Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographic Controls	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Main Occ. FE	No	No	No	Yes	No	No	No	No
Sub Occ. FE	No	No	No	No	No	No	No	No
Full Occ. FE	No	No	No	No	Yes	Yes	Yes	Yes
Sample	Full	Full	Full	Full	Full	Full	52 wks 40 hrs	Resp.

Notes: This table reports linear probability regression coefficients where the dependent variable is 1,0 for privacy (non-disclosure). Age controls are a linear and quadratic term in age. County controls are the 1940 population, manufacturing value-added, the share urban and the share of the population 25+ completing high school. Demographic controls are indicators for gender, household head, marriage status (married, divorced/separated single), race, immigrant and college attendance and a continuous variable for capitalized house values. Occupation fixed effects at the main (11), sub (22) and full (228) levels. Resp. (column 8) refers to the respondent only subsample of individuals who personally reported their income to the enumerator. Standard errors clustered by county in parentheses. *p<0.1, **p<0.05, ***p<0.01.

Table A5: Privacy Demands and Inequality: Occupation 90/10 Income Inequality

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
New Deal Spending per Capita (std)	-0.00002	0.00010	0.00012	0.00003	0.00001	-0.00068*	0.00029
	(0.00037)	(0.00038)	(0.00038)	(0.00039)	(0.00038)	(0.00037)	(0.00030)
Republican Vote Share (std)	0.00206***	0.00098	0.00029	0.00081	0.00083	0.00121*	0.00115*
	(0.00074)	(0.00069)	(0.00071)	(0.00069)	(0.00068)	(0.00064)	(0.00068)
Share Religious (std)	-0.00201***	-0.00032	-0.00036	-0.00065	-0.00067	-0.00077	-0.00034
	(0.00052)	(0.00050)	(0.00050)	(0.00052)	(0.00051)	(0.00050)	(0.00049)
Occupation 90/10 Income Ratio (std)	0.02197***	0.02151***	0.02122***	0.01794***	0.01300***	0.01131***	0.01571***
	(0.00049)	(0.00048)	(0.00054)	(0.00037)	(0.00037)	(0.00044)	(0.00060)
Observations	22281334	22281334	22281334	22281334	22281334	11599405	5383435
Clusters	3075	3075	3075	3075	3075	3075	3075
Mean Dep Var.	0.047	0.047	0.047	0.047	0.047	0.035	0.047
Age Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County Controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Demographic Controls	No	No	Yes	Yes	Yes	Yes	Yes
Main Occ. FE	No	No	No	Yes	No	No	No
Sub Occ. FE	No	No	No	No	Yes	Yes	Yes
Sample	Full	Full	Full	Full	Full	52 wks 40 hrs	Resp.

Notes: This table reports linear probability regression coefficients where the dependent variable is 1,0 for privacy (non-disclosure). The occupation 90/10 ratio is calculated using reported incomes in 228 occupational categories. Age controls are a linear and quadratic term in age. County controls are the 1940 population, manufacturing value-added, the share urban and the share of the population 25+ completing high school. Demographic controls are indicators for gender, household head, marriage status (married, divorced/separated single), race, immigrant and college attendance and a continuous variable for capitalized house values. Occupation fixed effects at the main (11) and sub (22) levels. Resp. (column 7) refers to the respondent only subsample of individuals who personally reported their income to the enumerator. Standard errors clustered by county in parentheses. *p<0.1, **p<0.05, ***p<0.01.

Table A6: Privacy Demands and Inequality: Occupation 90/10 Income Inequality and Churn

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
New Deal Spending per Capita (std)	-0.00010	0.00003	0.00007	0.00002	0.00001	-0.00068*	0.00029
	(0.00038)	(0.00038)	(0.00038)	(0.00039)	(0.00038)	(0.00037)	(0.00030)
Republican Vote Share (std)	0.00190**	0.00091	0.00038	0.00080	0.00083	0.00120*	0.00115*
	(0.00075)	(0.00070)	(0.00070)	(0.00069)	(0.00068)	(0.00064)	(0.00068)
Share Religious (std)	-0.00214***	-0.00036	-0.00046	-0.00068	-0.00068	-0.00078	-0.00034
	(0.00054)	(0.00052)	(0.00051)	(0.00052)	(0.00052)	(0.00050)	(0.00049)
Occupation 90/10 Income Ratio (std)	0.02616***	0.02581***	0.02570***	0.01955***	0.01406***	0.01228***	0.01606***
	(0.00062)	(0.00062)	(0.00066)	(0.00044)	(0.00050)	(0.00052)	(0.00075)
Observations	22281334	22281334	22281334	22281334	22281334	11599405	5383435
Clusters	3075	3075	3075	3075	3075	3075	3075
Mean Dep Var.	0.047	0.047	0.047	0.047	0.047	0.035	0.047
Age Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County Controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Demographic Controls	No	No	Yes	Yes	Yes	Yes	Yes
Main Occ. FE	No	No	No	Yes	No	No	No
Sub Occ. FE	No	No	No	No	Yes	Yes	Yes
Sample	Full	Full	Full	Full	Full	52 wks 40 hrs	Resp.
Churn Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports linear probability regression coefficients where the dependent variable is 1,0 for privacy (non-disclosure). The occupation 90/10 ratio is calculated using reported incomes in 228 occupational categories. The specifications control for employment churn, measured as the median reported income for each of the in 228 occupational categories. Age controls are a linear and quadratic term in age. County controls are the 1940 population, manufacturing value-added, the share urban and the share of the population 25+ completing high school. Demographic controls are indicators for gender, household head, marriage status (married, divorced/separated single), race, immigrant and college attendance and a continuous variable for capitalized house values. Occupation fixed effects at the main (11) and sub (22) levels. Resp. (column 7) refers to the respondent only subsample of individuals who personally reported their income to the enumerator. Standard errors clustered by county in parentheses. *p<0.1, **p<0.05, ***p<0.01.

Table A7: Subject, Enumerator Descriptives

	Subjects		Enumerators	
	Mean	SD	Mean	SD
Privacy	0.045	0.208	0.542	0.498
Age	39.840	10.583	39.442	15.919
Male = 1	0.740	0.439	0.490	0.500
Household Head = 1	0.626	0.484	0.371	0.483
Married = 1	0.751	0.432	0.706	0.456
Divorced/Separated = 1	0.030	0.171	0.019	0.136
Single = 1	0.219	0.413	0.275	0.447
White = 1	0.938	0.242	0.942	0.233
Immigrant = 1	0.154	0.361	0.144	0.351
Years of Education	9.508	3.520	9.215	3.458
College = 1	0.151	0.358	0.118	0.322
Home Owner	0.583	0.493	0.557	0.497
House Value	4049.479	6782.993	3745.144	6484.312
Rental Value	63.636	382.305	61.989	380.324
Capitalized House Value	6132.909	35283.148	5794.083	34315.518
Weeks Worked	45.421	11.166	42.601	13.824
Hours Worked	43.240	11.341	43.775	13.275
Wage Income	1260.028	892.968	524.976	804.069
Non-Wage Income = 1	0.152	0.359	0.228	0.419

Notes: This table reports descriptive statistics for a sample of individuals we identify as a ‘census taker’ or a ‘census enumerator’ in the Census occupation string in the IPUMS data.

Table A8: Privacy and the Subject-Enumerator Housing Wealth Gap: Robustness

	(1)	(2)	(3)	(4)
	Panel A: Main Estimates			
House Gap Ratio (log)	0.00232***	0.01148***	0.00265***	0.01278**
	(0.00064)	(0.00403)	(0.00081)	(0.00548)
Observations	276712	27328	163887	17320
Mean Dep Var.	0.042	0.043	0.042	0.043
Clusters	181252	18290	124971	13251
	Panel B: Subject Housing Wealth $>$ than Enumerator			
House Gap Ratio (log)	0.00367***	0.00219	0.00413**	-0.01300
	(0.00139)	(0.00914)	(0.00181)	(0.01280)
Observations	124992	11330	72846	6792
Mean Dep Var.	0.042	0.046	0.042	0.049
Clusters	79787	7445	54491	5191
	Panel C: Subject Housing Wealth $\leq$ than Enumerator			
House Gap Ratio (log)	0.00199**	0.00175	0.00268***	0.00395
	(0.00086)	(0.00593)	(0.00102)	(0.00751)
Observations	151716	15992	91030	10515
Mean Dep Var.	0.043	0.042	0.041	0.038
Clusters	101461	10839	70469	8047
County FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Gender Matching	No	No	Yes	Yes
House Values	Capitalized	Reported	Capitalized	Reported
p-value (H_0: Panel B=Panel C)	0.3043	0.9673	0.4839	0.2485

Notes: This table reports linear probability regression coefficients where the dependent variable is 1,0 for privacy (non-disclosure). The right-hand side variables specify the log of the housing wealth gap between the subject and the enumerator. Controls include the log of the subjects own level of housing wealth, their age and years of education, gaps in age and education between subject and enumerator and mean log housing wealth, age and educational attainment in an enumeration district. Panel A shows the main estimates from columns 1, 3, 5 and 7 of Table II. Panel B shows estimates where the subjects' housing wealth is greater than the enumerator's housing wealth. Panel C shows estimates where the subjects' housing wealth is less than or equal to the enumerator's housing wealth. Standard errors clustered by household in parentheses. *p<0.1, **p<0.05, ***p<0.01.

Table A9: Newspaper List Descriptives

	Non-Newslist		Newslist	
	Mean	SD	Mean	SD
Newspaper Circulation (scaled)	0.000	0.000	1.030	1.336
Panel A: County-Level				
New Deal Spending per Capita	112.252	27.509	146.591	30.094
Republican Vote Share	37.352	17.122	38.435	14.382
Share Religious	48.959	11.932	51.532	11.967
Share Urban	86.958	8.726	90.585	10.971
Share Educated	27.829	6.240	28.095	6.873
County Population 1940 (Millions)	0.929	0.752	1.626	1.318
Manufacturing Value Added (Millions)	1069.696	1138.629	1235.662	1117.527
90/10 Income Ratio	5.725	1.413	5.970	1.279
Predicted 90/10 Income Ratio	3.632	0.876	3.874	0.890
Housing Wealth Inequality	0.602	0.111	0.754	0.211
Income Inequality	0.228	0.046	0.242	0.041
Panel B: Occupation-Level				
Occupation 90/10 Income Ratio	4.766	1.575	4.849	1.683
Panel C: Individual-Level				
Privacy	0.039	0.195	0.043	0.204
Age	39.770	10.367	40.040	10.523
Male = 1	0.765	0.424	0.726	0.446
Household Head = 1	0.646	0.478	0.630	0.483
Married = 1	0.775	0.418	0.748	0.434
Divorced/Separated = 1	0.027	0.163	0.027	0.161
Single = 1	0.198	0.398	0.226	0.418
White = 1	0.930	0.256	0.913	0.281
Immigrant = 1	0.206	0.405	0.192	0.394
Years of Education	9.273	3.461	9.421	3.466
College = 1	0.124	0.329	0.137	0.344
Home Owner	0.559	0.496	0.644	0.479
House Value	4634.610	8619.573	4992.775	8459.222
Rental Value	58.625	325.969	74.111	394.560
Capitalized House Value	5977.258	29833.206	7503.927	38369.461
Weeks Worked	45.043	11.146	46.118	10.878
Hours Worked	41.604	10.486	42.682	10.455
Wage Income	1367.998	912.949	1343.914	948.200
Non-Wage Income = 1	0.134	0.341	0.135	0.342

Notes: This table shows descriptive statistics by newspaper list locations. Newspaper list counties are those with a city in the top 50 cities by population size in the US in 1920 where lists of top earners were published between 1924 and 1925. Non-list counties are those with a city in the top 50 cities by population size in the US in 1920 in which lists were not published. Newspaper circulation numbers are scaled by 1920 county population.

Table A10: Newspaper List Balance Table

Variable	(1) Non-Newslist	(2) Newslist	(3) Difference (inc. FE)
New Deal Spending per Capita	112.252	146.591	25.311***
	(27.509)	(30.094)	(7.545)
Republican Vote Share	37.352	38.435	-1.732
	(17.122)	(14.382)	(2.619)
Share Religious	48.959	51.532	0.820
	(11.932)	(11.967)	(2.309)
Share Urban	86.958	90.585	1.584
	(8.726)	(10.971)	(3.019)
Share Educated	27.829	28.095	1.346
	(6.240)	(6.873)	(1.072)
County Population 1940 (Millions)	0.929	1.626	0.490*
	(0.752)	(1.318)	(0.283)
Manufacturing Value Added (Millions)	1,069.696	1,235.662	137.469
	(1,138.629)	(1,117.527)	(225.457)
90/10 Income Ratio	5.725	5.970	0.078
	(1.413)	(1.279)	(0.175)
Age	39.770	40.040	0.046
	(10.367)	(10.523)	(0.146)
Male = 1	0.765	0.726	-0.009*
	(0.424)	(0.446)	(0.005)
Household Head = 1	0.646	0.630	0.001
	(0.478)	(0.483)	(0.007)
Married = 1	0.775	0.748	-0.013
	(0.418)	(0.434)	(0.008)
Divorced/Separated = 1	0.027	0.027	0.000
	(0.163)	(0.161)	(0.001)
Single = 1	0.198	0.226	0.012
	(0.398)	(0.418)	(0.009)
White = 1	0.930	0.913	-0.003
	(0.256)	(0.281)	(0.011)
Immigrant = 1	0.206	0.192	0.009
	(0.405)	(0.394)	(0.021)
Years of Education	9.273	9.421	0.026
	(3.461)	(3.466)	(0.098)
College = 1	0.124	0.137	0.001
	(0.329)	(0.344)	(0.003)
Home Owner	0.559	0.644	0.053***
	(0.496)	(0.479)	(0.017)
House Value	4,634.610	4,992.775	290.069
	(8,619.573)	(8,459.222)	(229.196)
Rental Value	58.625	74.111	11.225*
	(325.969)	(394.560)	(6.526)
Capitalized House Value	5,977.258	7,503.927	1,053.878*
	(29,833.207)	(38,369.461)	(525.319)
Wage Income	1,367.998	1,343.914	-19.661
	(912.948)	(948.200)	(35.064)
Non-Wage Income = 1	0.134	0.135	0.005
	(0.341)	(0.342)	(0.005)
Observations	1,552,005	6,220,601	7,772,606

Notes: This table shows a balance table of descriptive statistics by newspaper list locations. Newspaper list counties are those with a city in the top 50 cities by population size in the US in 1920 where lists of top earners were published between 1924 and 1925. Non-list counties are those with a city in the top 50 cities by population size in the US in 1920 in which lists were not published. The difference in means test includes fixed effects for state and occupation.

Table A11: Newspaper List Descriptives by Tercile of Circulation

	Non-Newslist		Low		Medium		High	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Newspaper Circulation (scaled)	0.000	0.000	0.311	0.097	0.768	0.191	2.253	1.926
Panel A: County-Level								
New Deal Spending per Capita	112.252	27.509	144.812	29.128	145.038	28.105	150.560	32.858
Republican Vote Share	37.352	17.122	40.596	14.507	42.145	11.618	31.710	14.538
Share Religious	48.959	11.932	48.731	11.922	52.718	6.437	53.991	15.263
Share Urban	86.958	8.726	87.599	8.414	94.015	7.015	90.949	15.359
Share Educated	27.829	6.240	28.455	8.452	26.115	4.116	29.686	6.333
County Population 1940 (Millions)	0.929	0.752	1.550	0.890	2.417	1.765	0.901	0.605
Manufacturing Value Added (Millions)	1069.696	1138.629	933.928	419.038	1911.393	1475.328	928.910	1003.559
90/10 Income Ratio	5.725	1.413	6.015	1.180	5.469	0.788	6.433	1.593
Predicted 90/10 Income Ratio	3.632	0.876	3.760	0.640	3.637	0.734	4.270	1.153
Housing Wealth Inequality	0.602	0.111	0.803	0.244	0.719	0.157	0.726	0.202
Income Inequality	0.228	0.046	0.239	0.032	0.230	0.030	0.257	0.054
Panel B: Occupation-Level								
Occupation 90/10 Income Ratio	4.766	1.575	4.833	1.669	4.686	1.569	5.041	1.794
Panel C: Individual-Level								
Privacy	0.039	0.195	0.044	0.205	0.038	0.191	0.048	0.214
Age	39.770	10.367	40.060	10.561	40.033	10.512	40.021	10.485
Male = 1	0.765	0.424	0.743	0.437	0.731	0.443	0.699	0.459
Household Head = 1	0.646	0.478	0.646	0.478	0.630	0.483	0.609	0.488
Married = 1	0.775	0.418	0.765	0.424	0.751	0.432	0.721	0.448
Divorced/Separated = 1	0.027	0.163	0.029	0.168	0.023	0.150	0.027	0.162
Single = 1	0.198	0.398	0.206	0.405	0.226	0.418	0.252	0.434
White = 1	0.930	0.256	0.916	0.278	0.942	0.234	0.881	0.324
Immigrant = 1	0.206	0.405	0.153	0.360	0.208	0.406	0.229	0.420
Years of Education	9.273	3.461	9.459	3.449	9.326	3.388	9.469	3.567
College = 1	0.124	0.329	0.137	0.344	0.129	0.335	0.146	0.353
Home Owner	0.559	0.496	0.584	0.493	0.647	0.478	0.720	0.449
House Value	4634.610	8619.573	4799.540	7462.205	4990.369	6757.870	5375.401	11687.538
Rental Value	58.625	325.969	75.796	409.484	72.974	392.867	73.373	379.576
Capitalized House Value	5977.258	29833.206	7307.050	37907.840	7426.552	38169.651	7844.664	39172.718
Weeks Worked	45.043	11.146	45.691	11.162	46.396	10.609	46.390	10.759
Hours Worked	41.604	10.486	42.305	10.214	42.287	9.987	43.611	11.181
Wage Income	1367.998	912.949	1324.279	921.806	1383.911	949.246	1328.002	979.737
Non-Wage Income = 1	0.134	0.341	0.138	0.345	0.138	0.344	0.129	0.335

Notes: This table shows descriptive statistics by newspaper list locations split by tercile of newspaper circulation intensity (Low, Medium or High). Newspaper list counties are those with a city in the top 50 cities by population size in the US in 1920 where lists of top earners were published between 1924 and 1925. Non-list counties are those with a city in the top 50 cities by population size in the US in 1920 in which lists were not published. Newspaper circulation numbers are scaled by 1920 county population.