

The Self-Employment Earnings Penalty Reconsidered: Incorporation Status and the Composition of Entrepreneurship*

APEP Autonomous Research[†]
@SocialCatalystLab

February 3, 2026

Abstract

Self-employed workers earn less on average than wage workers, but this aggregate penalty masks profound heterogeneity. Using doubly robust inverse probability weighting methods on American Community Survey data covering 1.4 million prime-age workers (2019–2022), I decompose the self-employment effect by incorporation status. Incorporated self-employed workers—professional practitioners, consultants, and small business owners who have chosen corporate form—show an earnings *premium* relative to observationally similar wage workers (+0.069 log points, 95% CI: [+0.058, +0.079]), approximately 7 percent higher. Unincorporated self-employed workers—sole proprietors, gig workers, and informal entrepreneurs—face a substantial penalty of −0.623 log points (95% CI: [−0.635, −0.610]), equivalent to approximately 46 percent lower earnings. The aggregate penalty of −0.362 log points (about 30 percent) reflects compositional mix: roughly 60 percent of self-employed workers are unincorporated. Among workers without a college degree, incorporated self-employment is associated with an 8 percent earnings premium (+0.078 log points), while unincorporated self-employment carries a severe penalty. These findings reconcile conflicting results in the entrepreneurship literature and suggest that policies promoting self-employment have heterogeneous effects depending on the type of self-employment encouraged.

*This paper is a revision of APEP-0169. See https://github.com/SocialCatalystLab/auto-policy-evals/tree/main/papers/a pep_0169 for the original.

[†]Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch

JEL Codes: J23, J24, J31, L26

Keywords: self-employment, earnings penalty, incorporated business, entrepreneurship, inverse probability weighting

1. Introduction

A longstanding puzzle in labor economics concerns why workers choose self-employment when it appears to pay less than wage work. Since [Hamilton \(2000\)](#) documented that median self-employment earnings fall 35 percent below comparable wage earnings, researchers have sought to explain this apparent irrationality. The dominant explanations invoke either compensating differentials—workers accept lower pay for autonomy, flexibility, and independence—or negative selection—workers with poor wage-employment prospects turn to self-employment as a last resort. Yet these explanations treat self-employment as a monolithic category, obscuring fundamental differences in the nature of self-employed work.

This paper demonstrates that the aggregate self-employment earnings penalty masks profound heterogeneity between two legally and economically distinct forms of self-employment. Incorporated self-employment encompasses professional practices, consulting firms, and small businesses whose owners have chosen corporate form, accepting the administrative costs of incorporation in exchange for liability protection and tax advantages. Unincorporated self-employment includes sole proprietors, gig workers, and informal entrepreneurs operating without corporate structure. As [Levine and Rubinstein \(2017\)](#) have argued, these two categories reflect fundamentally different entrepreneurial pathways, and conflating them leads to misleading conclusions about the returns to entrepreneurship.

Using doubly robust inverse probability weighting methods applied to American Community Survey data covering approximately 1.4 million prime-age workers across ten large U.S. states from 2019–2022, I decompose the self-employment effect by incorporation status. The results are striking. Incorporated self-employed workers show an earnings *premium* relative to observationally similar wage workers: the estimated effect is +0.069 log points with a 95 percent confidence interval of [+0.058, +0.079]. This translates to approximately 7 percent higher earnings, a finding that aligns with [Levine and Rubinstein \(2017\)](#)’s evidence of positive selection into incorporated self-employment.

Unincorporated self-employed workers, by contrast, face a substantial earnings penalty of -0.623 log points (95% CI: $[-0.635, -0.610]$), equivalent to roughly 46 percent lower annual earnings. The magnitude of this penalty far exceeds what could plausibly be explained by compensating differentials alone and suggests that unincorporated self-employment captures a heterogeneous population including both voluntary lifestyle entrepreneurs and involuntary gig workers pushed into precarious work by lack of better alternatives.

The aggregate self-employment penalty documented in prior literature—which I replicate at -0.362 log points (approximately 30 percent lower)—reflects the compositional mix of incorporated and unincorporated workers. In my sample, approximately 60 percent of self-

employed workers are unincorporated. The aggregate penalty is thus a weighted average that obscures the premium for incorporated workers and the severe penalty for unincorporated workers.

Heterogeneity analysis by education reinforces the importance of distinguishing self-employment types. Among workers without a college degree, incorporated self-employment is associated with an earnings premium of +0.078 log points (95% CI: [+0.066, +0.091]), suggesting that non-college workers who successfully establish incorporated businesses earn about 8 percent more than they would in wage employment. The same non-college workers face a penalty of -0.534 log points if they pursue unincorporated self-employment. For college-educated workers, incorporated self-employment also carries a small premium (+0.060 log points), while unincorporated self-employment carries a larger penalty (-0.702 log points), potentially reflecting the opportunity cost of foregoing high-paying professional wage positions. These patterns suggest that the selection versus compensating differentials debate cannot be resolved without first disaggregating self-employment by type.

The identification strategy rests on the assumption of selection on observables: conditional on the rich demographic and socioeconomic characteristics available in the ACS, self-employment choice is independent of potential earnings outcomes. I provide extensive diagnostics supporting this assumption, including propensity score overlap analysis showing 100 percent common support and covariate balance with all standardized mean differences below 0.01 after weighting. Sensitivity analysis using the E-value framework indicates that substantial unmeasured confounding would be required to explain away the incorporated versus unincorporated difference.

These findings carry important implications for policy. Programs promoting entrepreneurship and self-employment should recognize that outcomes depend critically on the type of self-employment encouraged. Facilitating incorporation—through reduced administrative burden, improved access to legal services, or targeted tax incentives—may help self-employed workers achieve earnings parity with wage workers. Conversely, the growth of unincorporated gig work, even if presented as flexibility-enhancing entrepreneurship, is associated with substantially lower earnings. Policymakers concerned about worker welfare should attend to the composition of self-employment, not just its prevalence.

The paper proceeds as follows. Section 2 develops a theoretical framework that distinguishes incorporated and unincorporated self-employment. Section 3 describes the data and sample construction. Section 4 presents the empirical strategy. Section 5 reports main results. Section 6 examines heterogeneity by education and incorporation status. Section 7 presents robustness checks. Section 8 discusses implications and limitations. Section 9 concludes.

2. Theoretical Framework

This section develops a framework for understanding earnings differentials between self-employed and wage workers, with particular attention to the distinction between incorporated and unincorporated self-employment. The framework generates testable predictions about heterogeneity patterns that can be evaluated empirically.

2.1 Related Literature

The literature on self-employment earnings has produced seemingly contradictory findings over several decades. [Hamilton \(2000\)](#) documented that median self-employment earnings fall roughly 35 percent below comparable wage earnings, a finding that has been replicated across countries and time periods. This earnings penalty poses a puzzle: why do rational workers choose self-employment if it pays so much less? Two broad explanations have dominated the literature.

The first explanation invokes compensating differentials. Self-employment offers non-pecuniary benefits—autonomy, flexibility, independence, the ability to be one’s own boss—that workers may value highly. Workers who place particular weight on these benefits will accept lower pecuniary compensation in exchange. This explanation implies that the earnings penalty reflects workers’ preferences rather than labor market failure, and that workers who choose self-employment are better off despite lower earnings.

The second explanation invokes negative selection. Workers with poor prospects in wage employment—whether due to low ability, discrimination, or poor job matches—may turn to self-employment as a fallback option. Self-employment requires no job offer, no screening by employers, and no approval by gatekeepers. Under this view, the earnings penalty reflects the composition of who becomes self-employed rather than a structural disadvantage of self-employment itself. Low-ability workers would earn even less in wage work; self-employment represents their best available option.

These explanations are not mutually exclusive, and both likely operate to varying degrees. However, they share a common limitation: they treat self-employment as a homogeneous category. Recent work has challenged this assumption. [Levine and Rubinstein \(2017\)](#) distinguish between incorporated and unincorporated self-employment, documenting that the two groups differ systematically on measures of cognitive ability, education, and prior earnings. Their analysis suggests that incorporated self-employed workers are *positively* selected—they have better, not worse, options in wage work—while unincorporated self-employment captures a more heterogeneous population including both talented individuals who prefer informality and workers with limited alternatives.

This paper contributes to the literature by estimating separate earnings effects for incorporated and unincorporated self-employment using doubly robust methods. I show that the aggregate self-employment penalty documented in prior work reflects the compositional mix of incorporated and unincorporated workers, not a uniform disadvantage of self-employment.

2.2 Self-Employment as a Heterogeneous Category

The legal and economic distinction between incorporated and unincorporated self-employment has received insufficient attention in most empirical work on entrepreneurship, despite fundamental differences in the nature of these two forms of self-employment.

Incorporated self-employment requires workers to establish a legal business entity—typically a C-corporation, S-corporation, or limited liability company—separate from their personal affairs. This choice involves substantial upfront costs including legal fees for formation documents, state registration requirements, and ongoing compliance obligations such as annual reports, separate tax filings, and maintenance of corporate formalities. In exchange, incorporation provides benefits including limited personal liability for business debts, potential tax advantages through income splitting and fringe benefit deductions, enhanced credibility with clients and lenders, and perpetual existence independent of the owner.

The decision to incorporate reflects forward planning, business sophistication, and expectations of sufficient scale to justify administrative costs. Most small businesses begin as sole proprietorships; the choice to incorporate represents a deliberate decision to formalize. As [Levine and Rubinstein \(2017\)](#) document using data linking tax records to the National Longitudinal Survey of Youth, incorporated self-employed workers are positively selected on measures of cognitive ability and prior earnings. Incorporation may also facilitate access to formal credit markets, government contracts, and clients who require vendor incorporation.

Unincorporated self-employment encompasses a far more heterogeneous group. At one end are skilled craftspeople, professional consultants, and independent contractors who prefer the simplicity of sole proprietorship to the administrative burden of corporate form. These workers may earn substantial incomes while choosing to remain unincorporated for reasons of convenience, privacy, or tax simplicity. At the other end are gig workers, day laborers, and informal entrepreneurs with minimal business activity. The growth of platform-based gig work has expanded this latter category considerably, as workers driving for rideshare services or performing tasks through online platforms are typically classified as unincorporated self-employed.

The absence of corporate structure may reflect deliberate choice (simplicity, privacy, avoiding regulatory burden) or constraint (insufficient business scale, lack of knowledge about incorporation, inability to afford legal and accounting costs). Critically, these two

groups—voluntary and involuntary unincorporated self-employment—may have very different earnings prospects, but the available data do not permit clean separation between them.

These structural differences suggest that incorporated and unincorporated self-employment represent distinct labor market states with potentially different earnings implications. Pooling them into a single “self-employment” category, as most prior studies have done, obscures fundamental differences in the nature of work and the selection mechanisms into each category.

2.3 Earnings Determination by Self-Employment Type

To formalize the theoretical intuition, consider a Roy model framework in which workers choose among three employment states. Let workers be characterized by observable characteristics X_i (age, education, marital status, etc.) and unobservable characteristics η_i (entrepreneurial ability, risk preferences, motivation, non-cognitive skills). Workers choose among three employment states: wage employment ($D_i = 0$), unincorporated self-employment ($D_i = 1$), and incorporated self-employment ($D_i = 2$).

Potential earnings in each state depend on both observables and unobservables. Following the classic Roy model framework, suppose log earnings in state j are given by:

$$Y_i(j) = \mu_j(X_i) + \epsilon_{ji} \quad (1)$$

where $\mu_j(\cdot)$ captures the returns to observable characteristics in state j and ϵ_{ji} captures unobserved productivity in that state. The returns functions $\mu_j(\cdot)$ may differ across states: a college degree may be rewarded differently in wage employment versus self-employment, and the age-earnings profile may have different shapes depending on employment type.

Workers also receive non-pecuniary benefits A_{ji} in each state, with valuations that vary across individuals. Denote by A_{ji} the money-equivalent value that worker i places on the non-pecuniary aspects of employment state j . These benefits include autonomy, schedule flexibility, absence of supervision, creative fulfillment, and status. Workers with strong preferences for independence will have high values of A_{1i} and A_{2i} relative to A_{0i} .

Assuming risk-neutral workers who maximize the sum of pecuniary and non-pecuniary returns, worker i chooses employment state:

$$D_i = \arg \max_j \{Y_i(j) + A_{ji}\} \quad (2)$$

This selection process generates systematic differences in the distribution of ϵ_{ji} across observed employment states. If incorporated self-employment attracts workers with high entrepreneurial ability and strong outside options in wage work—as the positive selection

story suggests—then $\mathbb{E}[\epsilon_{2i}|D_i = 2]$ may be positive relative to the population average of zero. These workers choose incorporation despite having good wage alternatives because they also have unusually high earnings potential in self-employment (high ϵ_{2i}) or unusually strong preferences for independence (high A_{2i}).

Conversely, if unincorporated self-employment serves partly as a fallback for workers with poor wage alternatives, then $\mathbb{E}[\epsilon_{0i}|D_i = 1]$ may be low—these workers would have earned relatively little in wage work. Their choice of unincorporated self-employment may reflect high non-pecuniary valuations A_{1i} (voluntary lifestyle entrepreneurs) or simply the absence of wage alternatives (involuntary gig workers). The empirical challenge is that we observe earnings $Y_i(D_i)$ only in the chosen state, not potential earnings in counterfactual states.

These selection patterns would generate different observed earnings penalties even if the structural returns $\mu_j(X)$ were similar across states. A comparison of average earnings between self-employed and wage workers confounds the structural earnings effect with selection: we observe $\mathbb{E}[Y_i|D_i = j]$ but wish to infer what the average worker would earn in state j . Propensity score methods address this problem under the assumption that, conditional on observables X_i , selection into employment states is independent of potential earnings—i.e., the unobserved terms ϵ_{ji} are mean-independent of D_i given X_i .

2.4 Predictions

The framework generates several testable predictions that can be evaluated against the data. Each prediction follows from specific assumptions about the selection process and the returns to self-employment in different contexts.

Prediction 1: Incorporated self-employment shows a smaller penalty than unincorporated self-employment. If incorporated self-employment reflects positive selection on entrepreneurial ability while unincorporated self-employment captures a more heterogeneous population including workers with limited alternatives, we should observe that $\mathbb{E}[Y_i(2) - Y_i(0)|X_i]$ exceeds $\mathbb{E}[Y_i(1) - Y_i(0)|X_i]$ on average. The incorporated earnings penalty should be smaller, potentially approaching zero or even turning positive if the structural returns to incorporation are favorable.

Prediction 2: The incorporated-unincorporated gap is particularly large for non-college workers. Workers without college degrees face a narrower set of high-paying wage employment options. For these workers, the signaling value of incorporation may be particularly important: establishing a formal business entity credibly demonstrates capability, organization, and commitment that cannot be inferred from educational credentials. We therefore expect the earnings advantage of incorporation over non-incorporation to be larger for workers without a college degree.

Prediction 3: College-educated workers face opportunity costs in all forms of self-employment.

Workers with college degrees have access to high-paying professional positions in wage employment. To the extent that college-educated workers choose any form of self-employment, they may accept earnings reductions in exchange for non-pecuniary benefits. We therefore expect some earnings penalty even for incorporated self-employment among college graduates, reflecting the high opportunity cost of foregoing professional wage careers.

Prediction 4: The penalty for unincorporated self-employment is similar across education groups. If unincorporated self-employment captures a common underlying phenomenon—precarious, informal work arrangements that generate low returns regardless of worker characteristics—then the earnings penalty should be similar for workers with and without college degrees. This prediction distinguishes the “structural features of unincorporated work” hypothesis from an alternative in which low-education workers face larger penalties because they lack human capital.

I test these predictions using doubly robust inverse probability weighting methods that control for observable confounders while acknowledging that selection on unobservables cannot be ruled out. The credibility of causal inference rests on the conditional independence assumption: that, after conditioning on the rich covariate set available in the ACS, employment state choice is independent of potential earnings outcomes. I provide extensive diagnostics on propensity score overlap and covariate balance, along with sensitivity analyses that assess how much unmeasured confounding would be required to overturn the main conclusions.

3. Data

3.1 Data Source

I use data from the American Community Survey (ACS) Public Use Microdata Sample (PUMS), accessed through IPUMS ([Ruggles et al., 2024](#)). The ACS is the largest household survey in the United States, sampling approximately 3.5 million addresses annually and providing detailed demographic and economic information on the U.S. population. The large sample size is particularly valuable for studying self-employment, which comprises only about 10 percent of the workforce, as it provides sufficient statistical power to detect modest effects and to examine heterogeneity across subgroups.

Critically for this study, the ACS class-of-worker variable (COW) distinguishes between incorporated and unincorporated self-employment. Respondents who report being self-employed in an “incorporated business, company, or limited liability company” are coded separately from those self-employed in a “business or farm not incorporated.” This distinction, which follows Census Bureau practice since the 1967 Survey of Economic Opportunity, enables

the decomposition central to this paper. The incorporated category captures formal business entities with corporate tax filing obligations, while the unincorporated category captures sole proprietorships and partnerships.

I use ACS data from survey years 2019, 2021, and 2022, covering both pre-pandemic and pandemic-recovery periods. This time frame captures the labor market before COVID-19 disruptions as well as the subsequent recovery. The year 2020 is excluded because the COVID-19 pandemic severely disrupted ACS data collection: in-person interview follow-ups were suspended, group quarters data collection was delayed, and response rates fell substantially. The Census Bureau has cautioned against comparing 2020 data to other years, and I follow this guidance by excluding 2020 from the analysis. Robustness checks restricting to 2019 only yield similar results, indicating that findings are not driven by pandemic-related labor market changes.

3.2 Sample Construction

The analysis sample is constructed by applying several restrictions designed to focus on workers for whom the comparison between self-employment and wage work is economically meaningful. First, I restrict to prime working-age adults aged 25–54. This age range excludes young workers who may be in transitional employment and older workers who may be transitioning to retirement; workers in these age groups face different labor market constraints that would complicate interpretation. Second, I restrict to workers residing in ten large U.S. states: California, Texas, Florida, New York, Illinois, Ohio, Pennsylvania, Georgia, North Carolina, and Michigan. These states were selected because they account for a substantial share of U.S. employment and include geographic diversity across regions, though results are robust to alternative state selections.

Third, I require that workers be currently employed with non-missing earnings data. This restriction excludes unemployed workers, workers not in the labor force, and workers with missing or imputed earnings. Focusing on employed workers addresses the question of earnings conditional on employment rather than the broader question of labor force participation, which involves different mechanisms and would require different methods.

After applying these restrictions, the final sample contains 1,397,605 observations. The sample comprises 1,264,974 wage workers (90.5 percent), 79,946 unincorporated self-employed workers (5.7 percent), and 52,685 incorporated self-employed workers (3.8 percent). The self-employment rate of 9.5 percent is consistent with national statistics and with the general finding that self-employment represents a small but non-trivial share of employment.

3.3 Variable Definitions

The treatment variables are derived from the ACS class-of-worker variable (COW), which classifies workers based on their relationship to their employer. Incorporated self-employed workers are those reporting self-employment in an incorporated business, company, or limited liability company (COW = 7). Unincorporated self-employed workers are those reporting self-employment in a business or farm not incorporated (COW = 6). Wage workers include employees of private for-profit companies (COW = 1), employees of private not-for-profit companies (COW = 2), local government employees (COW = 3), state government employees (COW = 4), and federal government employees (COW = 5). A small number of unpaid family workers (COW = 8) are excluded.

The primary outcome is log annual earnings, calculated as the natural logarithm of total wage and self-employment income. The ACS reports wage and salary income separately from self-employment income (both farm and non-farm); I sum these to obtain total earned income. For the log transformation, zero and negative earnings are set to one dollar before taking the logarithm, though results are robust to alternative treatments of zeros. Secondary outcomes include full-time status, defined as working 35 or more hours per week, and usual hours worked per week, a continuous measure of labor supply on the intensive margin.

Covariates for the propensity score model include age (continuous) and age-squared to capture nonlinear age effects on self-employment propensity; a female indicator; a college degree indicator equal to one for workers with a bachelor’s degree or higher; a married indicator capturing current marital status; race and ethnicity indicators for White, Black, and Hispanic (with Asian and Other as the reference category); homeownership as a proxy for wealth and credit constraints; and an indicator for the COVID-affected period (2021–2022 survey years). The propensity score specification is parsimonious by design; adding higher-order terms and interactions does not materially improve balance or change the results.

3.4 Summary Statistics

Table 1 presents summary statistics by employment type. Incorporated self-employed workers differ markedly from both wage workers and unincorporated self-employed workers. They are older (mean age 42.0 versus 38.8 for wage workers), more likely to be male (65 percent versus 52 percent), more likely to hold a college degree (45 percent versus 43 percent), more likely to be married (67 percent versus 54 percent), and more likely to own their homes (76 percent versus 62 percent).

Table 1: Summary Statistics by Employment Type

	Wage Workers	Unincorp. Self-Emp.	Incorp. Self-Emp.
<i>Panel A: Demographics</i>			
Age (years)	38.8	41.0	42.0
Female (%)	47.5	41.6	35.1
College degree (%)	42.9	30.8	44.7
Married (%)	53.5	58.7	67.4
White (%)	57.8	63.5	72.4
<i>Panel B: Economic Outcomes</i>			
Mean earnings (\$)	66,824	52,809	98,176
Median earnings (\$)	50,000	30,000	57,000
Full-time (%)	87.5	68.1	82.1
Hours per week	40.9	38.2	43.0
Homeowner (%)	61.1	58.9	71.5
<i>Panel C: Sample Size</i>			
Observations	1,264,974	79,946	52,685

Notes: Sample includes prime-age (25–54) employed workers in 10 large U.S. states from the 2019–2022 ACS PUMS. Statistics are weighted using person weights.

Most striking are the differences in earnings. Incorporated self-employed workers have the highest mean earnings (\$98,176), exceeding wage workers (\$66,824) by over \$31,000. Unincorporated self-employed workers have substantially lower mean earnings (\$52,809) and a lower median (\$30,000 versus \$50,000 for wage workers), consistent with a heterogeneous population that includes both successful independent professionals and workers with minimal business activity. This raw comparison already suggests that the aggregate self-employment penalty arises primarily from the unincorporated segment, while incorporated self-employment may carry a premium.

4. Empirical Strategy

4.1 Identification Framework

I estimate the average treatment effect (ATE) of self-employment on earnings using inverse probability weighting (IPW), a method that has become standard in labor economics and

program evaluation (Hirano, Imbens, and Ridder, 2003). The fundamental challenge in estimating earnings effects of self-employment is selection: workers choose their employment status based on characteristics that also affect their earnings. A naive comparison of average earnings between self-employed and wage workers confounds the structural effect of self-employment with the selection of different types of workers into different employment states.

The identifying assumption underlying IPW estimation is selection on observables, also known as conditional independence or unconfoundedness. Formally, let $Y_i(1)$ and $Y_i(0)$ denote potential earnings under self-employment and wage employment, respectively. Each worker has two potential outcomes, but we observe only the outcome corresponding to their actual employment state: $Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$. The unconfoundedness assumption states that:

$$(Y_i(0), Y_i(1)) \perp D_i \mid X_i \quad (3)$$

where D_i is the binary treatment indicator and X_i is the vector of observed covariates. Under this assumption, treatment assignment is as good as random within cells defined by X_i : workers with the same observed characteristics who choose self-employment versus wage work do not systematically differ on unobserved determinants of earnings.

This assumption is strong and fundamentally untestable. Unobserved factors like entrepreneurial ability, risk preferences, motivation, and family wealth may affect both self-employment choice and earnings. I provide sensitivity analyses below that assess how much unmeasured confounding would be required to explain away the results, but the reader should interpret findings as causal effects under the conditional independence assumption rather than as definitively established causal relationships.

In addition to unconfoundedness, identification requires the overlap (positivity) condition: $0 < P(D_i = 1 \mid X_i) < 1$ for all values of X_i in the support. This condition ensures that, for any combination of observed characteristics, some workers choose each employment state. If certain covariate combinations deterministically predict self-employment or wage work, comparison is impossible for those combinations. The ACS data satisfy overlap well: propensity scores range from near zero to about 0.17, with no observations at either boundary.

I also estimate the average treatment effect on the treated (ATT), defined as $\mathbb{E}[Y_i(1) - Y_i(0) \mid D_i = 1]$. This estimand answers a policy-relevant question: what would the workers who actually chose self-employment have earned had they instead worked for wages? The ATT differs from the ATE if treatment effects vary with characteristics that predict treatment—a scenario that seems likely given the selection arguments above.

4.2 Estimation

I estimate propensity scores $e(X_i) = P(D_i = 1 \mid X_i)$ using logistic regression with age, age-squared, female, college, married, race indicators (White, Black, Hispanic), homeownership, and a COVID period indicator as covariates. The logistic specification is parsimonious; I also estimated propensity scores using more flexible methods (probit, random forests) and obtained similar results. For the decomposition by incorporation status, I estimate separate propensity score models for each binary comparison: incorporated self-employment versus wage work, and unincorporated self-employment versus wage work. This approach, rather than a multinomial model, allows the predictors of incorporation to differ from the predictors of non-incorporation.

IPW weights for ATE estimation are constructed as:

$$w_i^{ATE} = \frac{D_i}{\hat{e}(X_i)} + \frac{1 - D_i}{1 - \hat{e}(X_i)} \quad (4)$$

where $\hat{e}(X_i)$ is the estimated propensity score. Treated observations receive weight $1/\hat{e}(X_i)$, up-weighting those with low propensity to be treated (who are “surprisingly” treated given their characteristics). Control observations receive weight $1/(1 - \hat{e}(X_i))$, up-weighting those with high propensity (who are “surprisingly” not treated). The reweighted samples become comparable on observed covariates, allowing straightforward comparison of outcomes.

For ATT estimation, weights are:

$$w_i^{ATT} = D_i + \frac{(1 - D_i) \cdot \hat{e}(X_i)}{1 - \hat{e}(X_i)} \quad (5)$$

which weights the control group to match the covariate distribution of the treated group.

To limit the influence of extreme weights, which can arise when propensity scores approach zero or one, I truncate weights at the 99th percentile. Observations with weights above this threshold receive the 99th percentile weight. This truncation trades off slight bias for substantial variance reduction (Crump et al., 2009). Sensitivity checks using different truncation thresholds (95th, 97.5th percentiles) yield similar results.

Standard errors are computed using the heteroskedasticity-robust sandwich estimator (HC1), which is consistent for weighted regression and accounts for the finite-sample behavior of the weights. Following recent methodological guidance (Austin, 2009; Hirano, Imbens, and Ridder, 2003), I report 95 percent confidence intervals throughout, computed as the point estimate plus or minus 1.96 standard errors. For the main results, I do not adjust for multiple hypothesis testing, though the incorporated-unincorporated difference is so large that it would survive any reasonable correction.

5. Results

5.1 Aggregate Self-Employment Effect

Table 2 presents the main estimates. The first two rows report effects for self-employment overall (pooling incorporated and unincorporated), showing the aggregate penalty documented in prior literature. The ATE on log earnings is -0.362 log points (95% CI: $[-0.371, -0.354]$, $p < 0.001$), equivalent to approximately 30 percent lower earnings. Self-employed workers are also 16.1 percentage points less likely to work full-time and work 1.6 fewer hours per week.

Table 2: Main Results: Effect of Self-Employment on Earnings and Work Intensity

	(1) Log Earnings	(2) Full-Time	(3) Hours/Week
<i>Panel A: Aggregate Self-Employment (ATE)</i>			
Self-Employed	-0.362^{***} [$-0.371, -0.354$]	-0.161^{***} [$-0.164, -0.159$]	-1.60^{***} [$-1.69, -1.51$]
<i>Panel B: By Incorporation Status (ATE)</i>			
Incorporated Self-Emp.	$+0.069^{***}$ [$+0.058, +0.079$]	-0.075^{***} [$-0.079, -0.072$]	$+1.18^{***}$ [$+1.05, +1.31$]
Unincorporated Self-Emp.	-0.623^{***} [$-0.635, -0.610$]	-0.213^{***} [$-0.216, -0.209$]	-3.26^{***} [$-3.37, -3.14$]
Mean outcome (wage workers)	11.11	0.875	40.9
N (incorporated analysis)	1,317,659	1,317,659	1,317,659
N (unincorporated analysis)	1,344,920	1,344,920	1,344,920

Notes: Doubly robust IPW estimates. 95% confidence intervals in brackets. *** $p < 0.01$. Propensity score model includes age, age², female, college, married, race indicators, homeowner, and COVID period. Robust standard errors. Weights trimmed at 99th percentile. Panel B: Incorporated analysis N = 1,264,974 wage workers + 52,685 incorporated = 1,317,659 (excludes unincorporated). Unincorporated analysis N = 1,264,974 wage workers + 79,946 unincorporated = 1,344,920 (excludes incorporated).

5.2 Decomposition by Incorporation Status

Panel B reveals that this aggregate penalty masks dramatic heterogeneity. Incorporated self-employed workers show an earnings *premium* of +0.069 log points (95% CI: [+0.058, +0.079]), equivalent to approximately 7 percent higher earnings than observationally similar wage workers. This finding is consistent with the positive selection hypothesis advanced by [Levine and Rubinstein \(2017\)](#): workers who incur the costs of incorporation tend to have higher ability and better prospects in both wage work and self-employment.

Remarkably, incorporated self-employed workers actually work *more* hours than wage workers (1.18 additional hours per week), despite working less on the full-time margin. This suggests that incorporation is associated with intensive-margin labor supply increases that complement the higher hourly returns.

Unincorporated self-employed workers, by contrast, face a substantial earnings penalty of -0.623 log points (95% CI: $[-0.635, -0.610]$), equivalent to approximately 46 percent lower earnings. In dollar terms, unincorporated self-employed workers earn roughly \$53,000 on average versus \$67,000 for wage workers—a gap of approximately \$14,000. They also work 3.26 fewer hours per week and are 21.3 percentage points less likely to work full-time.

The difference between incorporated and unincorporated effects is highly statistically significant ($p < 0.001$) and economically substantial: 0.69 log points, reflecting a 53-percentage-point difference in earnings outcomes.

5.3 Propensity Score Diagnostics

Propensity scores range from 0.020 to 0.174, with excellent overlap between treatment groups. All observations fall within the common support region. After IPW weighting, the maximum standardized mean difference across covariates is 0.007, well below conventional thresholds of 0.10 or even 0.05, indicating successful covariate balance ([Austin, 2009](#)).

6. Heterogeneity Analysis

6.1 Heterogeneity by Education

Table 3 presents results separately for workers with and without a college degree. The aggregate self-employment penalty is larger for non-college workers (-0.401 log points, approximately 33 percent) than college graduates (-0.311 log points, approximately 27 percent), consistent with prior findings that less-educated workers face worse self-employment outcomes.

Table 3: Heterogeneous Effects by Education Level

	No College	College Degree
<i>Panel A: Aggregate Self-Employment</i>		
Self-Employed	−0.401***	−0.311***
	[−0.413, −0.389]	[−0.327, −0.295]
N	765,510	632,095
<i>Panel B: Incorporated Self-Employment</i>		
Incorporated	+0.078***	+0.060***
	[+0.066, +0.091]	[+0.045, +0.075]
N treated	27,804	24,881
<i>Panel C: Unincorporated Self-Employment</i>		
Unincorporated	−0.534***	−0.702***
	[−0.548, −0.520]	[−0.724, −0.680]
N treated	53,380	26,566

Notes: Doubly robust IPW estimates of effect on log earnings. 95% confidence intervals in brackets. *** $p < 0.01$. Propensity score models re-estimated within each subgroup, excluding education.

The decomposition by incorporation status reveals a striking pattern. Among non-college workers, incorporated self-employment is associated with an earnings *premium* of +0.078 log points (95% CI: [+0.066, +0.091]), approximately 8 percent higher earnings. This suggests that non-college workers who successfully establish incorporated businesses actually earn more than they would in wage employment. The premium may reflect returns to entrepreneurial ability that are less well rewarded in the wage sector for workers without formal credentials, or it may reflect selection of high-ability non-college workers into incorporated self-employment.

Among college graduates, incorporated self-employment also carries a premium (+0.060 log points), though smaller than for non-college workers. This pattern differs from what pure opportunity-cost reasoning would predict: if college-educated workers are foregoing high-paying professional positions, we might expect a penalty rather than a premium. The finding suggests either that incorporated self-employment offers returns comparable to professional wage work, or that workers who choose it have both high wage alternatives and high self-employment returns.

The penalty for unincorporated self-employment is substantial for both education groups but notably larger for college graduates: -0.534 log points (41 percent lower) for non-college workers and -0.702 log points (50 percent lower) for college graduates. This pattern—larger unincorporated penalties for more educated workers—suggests that unincorporated self-employment may represent a particularly poor fit for workers with greater human capital, who would earn substantially more in either wage work or incorporated self-employment.

7. Robustness

7.1 Propensity Score Trimming

Results are robust to trimming observations with extreme propensity scores. Trimming at the 1st, 5th, and 10th percentiles yields point estimates ranging from -0.326 to -0.363 log points for the aggregate self-employment penalty, all within the confidence interval of the baseline estimate.

7.2 Coefficient Stability

Following [Oster \(2019\)](#), I examine coefficient stability as covariates are added. The aggregate self-employment coefficient declines from -0.374 log points in a bivariate regression to -0.348 log points in the full specification. Using Oster’s method with $R^2_{\max} = 1.3 \times R^2_{\text{full}}$, the calculated delta is 2,589, meaning selection on unobservables would need to be over 2,500 times as important as selection on observables to drive the result to zero.

7.3 Sensitivity Analysis

The E-value for the aggregate self-employment effect is 1.91, meaning an unmeasured confounder would need risk ratio associations of at least 1.91 with both treatment and outcome to fully explain the observed penalty. For the incorporated-unincorporated difference, the E-value is 2.2, indicating that moderate unmeasured confounding would be required to eliminate this gap—though the difference remains robust to plausible confounding scenarios.

7.4 Excluding COVID Period

Restricting to pre-COVID (2019) data yields an aggregate penalty of -0.345 log points (approximately 29 percent), indistinguishable from the full-sample estimate of -0.362 log points. The COVID pandemic did not fundamentally alter the relationship between self-employment and earnings.

8. Discussion

8.1 Reconciling Conflicting Findings

The results help reconcile conflicting findings in the entrepreneurship literature that have accumulated over several decades. Studies documenting large self-employment penalties, including the seminal work of [Hamilton \(2000\)](#), pool incorporated and unincorporated workers into a single self-employment category. Hamilton’s influential finding that median self-employment earnings fall 35 percent below comparable wage earnings has shaped the field’s understanding of entrepreneurship as a potentially irrational choice requiring explanation through compensating differentials or negative selection.

Studies finding modest penalties or even premiums for entrepreneurs, by contrast, often focus on more select populations. [Levine and Rubinstein \(2017\)](#) examine the NLSY79 cohort and distinguish incorporated from unincorporated self-employment, finding that incorporated self-employed men earn more than their wage-working counterparts once cognitive ability is controlled. Studies of “employers” (self-employed workers with paid employees) versus “own-account” workers (self-employed without employees) also find better outcomes for the former group. The apparent contradiction between these literatures dissolves once self-employment is properly disaggregated: incorporated self-employment appears to reward entrepreneurial activity, while unincorporated self-employment resembles precarious work with much lower returns.

This reconciliation has implications for how we interpret aggregate trends in self-employment. The growth of gig work and platform-based employment, typically classified as unincorporated self-employment, may partly explain why aggregate self-employment penalties have remained substantial despite rising rates of business formation. The composition of self-employment matters as much as its prevalence.

8.2 Mechanisms

Several mechanisms may explain the large difference between incorporated and unincorporated self-employment earnings. These mechanisms are not mutually exclusive, and all likely operate to varying degrees in the data.

Selection. Workers with greater entrepreneurial ability, business acumen, social capital, and access to financial capital may disproportionately choose incorporation. The decision to incorporate requires planning, knowledge of business law, and willingness to bear administrative costs; workers who take this step may differ systematically from those who remain unincorporated. Conversely, unincorporated self-employment may serve as a fallback for

workers with limited wage employment alternatives—those who face discrimination, have poor labor market matches, or lack credentials that signal ability to employers. Under this mechanism, the earnings difference reflects who selects into each employment type rather than a structural advantage of incorporation itself.

Structural features. Incorporation may generate higher earnings through channels beyond selection. Incorporated businesses can more easily access formal credit markets, as lenders may require corporate structure for business loans. Some clients, particularly government agencies and large corporations, require that vendors be incorporated. Incorporation signals permanence and legitimacy, potentially attracting clients who would be reluctant to contract with sole proprietors. The limited liability protection of corporate form may encourage risk-taking and investment that generates higher returns. These structural advantages would benefit incorporated workers even holding ability and selection constant.

Measurement. Incorporated business owners may receive compensation through mechanisms that do not appear in annual earnings measures collected by the ACS. Retained earnings kept in the business, fringe benefits (health insurance, retirement contributions, company cars), and equity appreciation in the business are forms of compensation that may be substantial for incorporated owners but do not appear in reported wage and self-employment income. Unincorporated self-employed workers, by contrast, typically report all net business income as personal income, with less scope for compensation outside the measured earnings variable. This measurement story would imply that the true earnings gap is smaller than the measured gap, though the magnitude of measurement bias is difficult to assess.

Labor supply. The data show that incorporated self-employed workers actually work more hours than wage workers on the intensive margin (about 1.2 additional hours per week), while unincorporated workers work substantially fewer hours (about 3.3 fewer hours per week). Some of the earnings gap may reflect differences in labor supply rather than differences in returns per hour. However, the magnitude of the earnings gap—roughly \$60,000 annually—far exceeds what could be explained by a few hours per week, suggesting that hourly returns also differ substantially.

The education heterogeneity pattern—a small premium for incorporated non-college workers, a modest penalty for incorporated college graduates—provides additional insight into mechanisms. Workers with college credentials have high-paying alternatives in the wage sector, including professional positions in law, medicine, finance, and technology. Their choice of self-employment may reflect strong preferences for autonomy and independence rather than better earnings opportunities; the modest penalty may be a willingly accepted cost of non-pecuniary benefits. Workers without college credentials face a narrower set of high-paying wage alternatives. For these workers, successful business ownership may represent one of the

few paths to earnings exceeding what they could earn as employees. The small premium for incorporated non-college workers is consistent with this interpretation.

8.3 The Unincorporated Penalty: Selection or Structure?

The substantial earnings penalty for unincorporated self-employment—approximately 46 percent lower earnings than wage workers—demands careful interpretation. This penalty exceeds what could plausibly be explained by compensating differentials alone: while some workers may accept lower pay for flexibility, a nearly 50 percent earnings reduction seems too large to reflect pure preference-based sorting. The penalty suggests either negative selection (workers who choose unincorporated self-employment would have earned less in wage work as well) or structural features of unincorporated work that generate poor earnings regardless of worker characteristics.

The finding that the unincorporated penalty differs somewhat by education— -0.534 log points (41 percent) for non-college workers versus -0.702 log points (50 percent) for college graduates—suggests that human capital plays a role in moderating the penalty for workers without degrees, perhaps because unincorporated self-employment represents a more viable alternative for workers with fewer formal credentials. Interestingly, college graduates face a larger penalty, possibly because they forego more lucrative wage opportunities when choosing unincorporated self-employment.

What structural features might generate such large penalties? Unincorporated self-employed workers lack access to employer-provided benefits (health insurance, retirement plans, paid leave) that represent substantial compensation for wage workers but do not appear in earnings measures. They may face monopsony power from platforms or clients that set take-it-or-leave-it compensation terms. They bear business risks (variable income, lack of unemployment insurance, no workers' compensation) that are not compensated in the earnings measure. And unincorporated self-employment includes many part-time, intermittent, and casual arrangements that generate minimal income by design.

The growth of platform-based gig work has likely expanded the unincorporated self-employment category to include workers who, in an earlier era, would have been classified as employees. A driver for a rideshare company is coded as unincorporated self-employed, as is a worker performing tasks through an online platform. These workers face structural disadvantages—algorithmic management, take-it-or-leave-it terms, lack of collective bargaining power—that may generate poor outcomes independent of their individual characteristics.

8.4 Policy Implications

These findings carry important implications for policy at multiple levels. Most directly, programs promoting self-employment and entrepreneurship should recognize that outcomes depend critically on the type of self-employment encouraged. The incorporated-unincorporated distinction is not merely a legal technicality; it corresponds to fundamentally different labor market experiences with very different earnings consequences.

Facilitating incorporation. Policies that facilitate incorporation—reducing administrative costs, providing legal assistance, streamlining registration requirements, or offering tax incentives for small incorporated businesses—may help self-employed workers achieve earnings parity with wage workers. The small premium for incorporated non-college workers suggests that formalization could particularly benefit workers who lack other signals of competence and reliability. State-level variation in incorporation costs and complexity could be leveraged to study whether reducing barriers to incorporation improves outcomes for marginal entrants.

Gig work regulation. The growth of unincorporated gig work, even when framed as entrepreneurship and flexibility, is associated with dramatically lower earnings. Policymakers concerned about worker welfare should attend to the composition of self-employment, not just its prevalence. Regulatory debates over worker classification—whether gig workers should be employees or independent contractors—take on new urgency in light of the massive earnings penalty associated with unincorporated status. If gig workers were reclassified as employees, they would gain access to benefits, protections, and collective bargaining rights that might improve their earnings position.

Interpreting aggregate trends. More broadly, the findings suggest caution in interpreting aggregate trends in self-employment. A decline in self-employment rates might be good news (fewer workers trapped in low-paying unincorporated work) or bad news (fewer entrepreneurial opportunities), depending on the composition of the change. Similarly, an increase in self-employment rates could reflect either healthy business formation or displacement of workers from wage employment into precarious gig arrangements. Policymakers and researchers should monitor the incorporated-unincorporated composition of self-employment, not just its aggregate level.

Targeting entrepreneurship support. Programs designed to support entrepreneurs—small business loans, technical assistance, mentorship, incubators—might usefully target their support based on whether recipients are pursuing or have achieved incorporation. Incorporated businesses may be better positioned to use external financing productively, while unincorporated sole proprietors might benefit more from assistance in understanding whether and how to formalize their operations.

8.5 Limitations

Several limitations warrant acknowledgment and should guide interpretation of the results.

Selection on unobservables. The identification strategy relies on selection on observables; unmeasured factors like entrepreneurial ability, risk preferences, motivation, and family wealth may drive both self-employment choice and earnings. If workers who choose incorporated self-employment have higher ability on dimensions not captured by age, education, and demographics, part of the incorporation “effect” reflects selection rather than causation. Sensitivity analysis using E-values suggests that substantial confounding would be required to explain away the incorporated-unincorporated difference, but the possibility of unmeasured confounding cannot be ruled out. The estimates should be interpreted as causal effects under the conditional independence assumption rather than as definitively established causal relationships.

Earnings measurement. The ACS measures annual earnings from wages and self-employment, which may not fully capture the returns to business ownership. Incorporated business owners can retain earnings in the business, receive compensation through fringe benefits, and build equity value that does not appear in current income. These measurement issues would bias the incorporated penalty toward zero (making incorporation look better than it is in measured earnings) or even create a spurious premium. The true structural effect of incorporation on total compensation may differ from the measured earnings effect.

Point-in-time measurement. The ACS captures a snapshot of employment status and earnings at a single point in time. This cross-sectional design misses dynamics that might reveal different patterns over longer horizons. Incorporated businesses may generate low initial returns while building capacity, with returns increasing over time as the business matures. Unincorporated self-employment may be transitional for some workers, serving as a bridge between wage jobs or while searching for better opportunities. Panel data following workers over time could reveal whether the earnings penalty persists, diminishes, or grows as tenure in self-employment increases.

Heterogeneity within categories. The measure of unincorporated self-employment combines heterogeneous categories with potentially very different labor market dynamics. A management consultant who prefers sole proprietorship for simplicity, a skilled tradesperson with a steady client base, a rideshare driver working occasional hours, and a day laborer with minimal formal work arrangements are all coded identically as unincorporated self-employed. The massive average penalty may mask substantial variation within this category. Unfortunately, the ACS does not provide information on industry, occupation, or hours at sufficient detail to decompose unincorporated self-employment into more meaningful subcategories.

Geographic scope. The analysis uses data from ten large U.S. states, which may not gen-

eralize to smaller states, rural areas, or other countries. Self-employment rates, incorporation costs, and labor market conditions vary substantially across regions. The incorporated-unincorporated distinction may have different implications in contexts with different legal and institutional environments.

9. Conclusion

The self-employment earnings penalty, long regarded as a puzzle requiring explanation via compensating differentials or negative selection, is better understood as a compositional artifact arising from the aggregation of fundamentally different forms of self-employment. This paper has documented that the aggregate penalty masks profound heterogeneity between incorporated and unincorporated self-employment—heterogeneity that has important implications for how we understand entrepreneurship and design policies to promote it.

The main findings can be summarized as follows. First, the aggregate self-employment earnings penalty of approximately 0.36 log points (about 30 percent) documented in this paper reflects a weighted average of very different effects for incorporated and unincorporated self-employment. Second, incorporated self-employed workers—those who have established formal business entities—actually show an earnings *premium* relative to observationally similar wage workers. The estimated premium of 0.069 log points (about 7 percent) is consistent with positive selection into incorporated self-employment documented by [Levine and Rubinstein \(2017\)](#). Third, among workers without a college degree, incorporated self-employment is associated with an even larger earnings premium (8 percent), suggesting that formalization may provide earnings advantages particularly for workers who lack other signals of competence. Fourth, the aggregate penalty arises overwhelmingly from unincorporated self-employed workers, who earn roughly 46 percent less than wage workers—a penalty that cannot plausibly be explained by compensating differentials alone.

These findings reconcile conflicting results in the entrepreneurship literature. Studies documenting large self-employment penalties have typically pooled incorporated and unincorporated workers; studies finding small penalties or premiums have often focused on incorporated business owners or employers. The apparent contradiction dissolves once self-employment is properly disaggregated. Incorporated self-employment resembles the entrepreneurship celebrated in economic theory—business formation that may generate returns comparable to or exceeding wage employment. Unincorporated self-employment, by contrast, includes many workers in precarious arrangements with poor earnings outcomes.

The findings also carry important implications for policy. Programs promoting entrepreneurship and self-employment should recognize that outcomes depend critically on

the type of self-employment encouraged. Policies that facilitate incorporation—through reduced administrative burden, legal assistance, or targeted incentives—may help self-employed workers achieve earnings parity with wage workers. The growth of unincorporated gig work, even when presented as entrepreneurship and flexibility, is associated with dramatically lower earnings. Policymakers concerned about worker welfare should attend to the composition of self-employment, not just its prevalence.

Several avenues for future research emerge from these findings. First, what mechanisms explain the incorporated-unincorporated difference? Is it primarily selection of different types of workers into different categories, or do structural features of incorporation itself generate earnings advantages? Panel data following workers as they transition between employment states could help distinguish these explanations. Second, do policies that encourage formalization improve outcomes for workers who would otherwise pursue unincorporated self-employment? Natural experiments in incorporation costs or legal requirements could provide causal evidence on the effects of formalization. Third, how has the growth of platform-based gig work affected the composition and average earnings of unincorporated self-employment over time? Tracking these trends could inform ongoing debates about worker classification and labor market regulation.

The self-employment earnings penalty is real, but it is not uniform. Understanding who bears this penalty—and who does not—is essential for designing effective policies to support workers and promote productive entrepreneurship.

References

- Abadie, Alberto, and Guido W. Imbens. 2016. "Matching on the Estimated Propensity Score." *Econometrica* 84(2): 781–807.
- Abraham, Katharine G., John C. Haltiwanger, Kristin Sandusky, and James R. Spletzer. 2018. "Measuring the Gig Economy: Current Knowledge and Open Issues." NBER Working Paper 24950.
- Austin, Peter C. 2009. "Balance Diagnostics for Comparing the Distribution of Baseline Covariates between Treatment Groups in Propensity-Score Matched Samples." *Statistics in Medicine* 28(25): 3083–3107.
- Crump, Richard K., V. Joseph Hotz, Guido W. Imbens, and Oscar A. Mitnik. 2009. "Dealing with Limited Overlap in Estimation of Average Treatment Effects." *Biometrika* 96(1): 187–199.
- Bang, Heejung, and James M. Robins. 2005. "Doubly Robust Estimation in Missing Data and Causal Inference Models." *Biometrics* 61(4): 962–973.
- Benz, Matthias, and Bruno S. Frey. 2008. "Being Independent is a Great Thing: Subjective Evaluations of Self-Employment and Hierarchy." *Economica* 75(298): 362–383.
- Blanchflower, David G., and Andrew J. Oswald. 1998. "What Makes an Entrepreneur?" *Journal of Labor Economics* 16(1): 26–60.
- Borjas, George J. 1986. "The Self-Employment Experience of Immigrants." *Journal of Human Resources* 21(4): 485–506.
- Connelly, Rachel. 1992. "Self-Employment and Providing Child Care." *Demography* 29(1): 17–29.
- Evans, David S., and Boyan Jovanovic. 1989. "An Estimated Model of Entrepreneurial Choice under Liquidity Constraints." *Journal of Political Economy* 97(4): 808–827.
- Hamilton, Barton H. 2000. "Does Entrepreneurship Pay? An Empirical Analysis of the Returns to Self-Employment." *Journal of Political Economy* 108(3): 604–631.
- Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47(1): 153–161.

- Hirano, Keisuke, Guido W. Imbens, and Geert Ridder. 2003. "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score." *Econometrica* 71(4): 1161–1189.
- Holtz-Eakin, Douglas, David Joulfaian, and Harvey S. Rosen. 1994. "Sticking it Out: Entrepreneurial Survival and Liquidity Constraints." *Journal of Political Economy* 102(1): 53–75.
- Hundley, Greg. 2001. "Why and When Are the Self-Employed More Satisfied with Their Work?" *Industrial Relations* 40(2): 293–316.
- Hurst, Erik, Geng Li, and Benjamin Pugsley. 2014. "Are Household Surveys Like Tax Forms? Evidence from Income Underreporting of the Self-Employed." *Review of Economics and Statistics* 96(1): 19–33.
- Imbens, Guido W. 2004. "Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review." *Review of Economics and Statistics* 86(1): 4–29.
- Katz, Lawrence F., and Alan B. Krueger. 2019. "The Rise and Nature of Alternative Work Arrangements in the United States, 1995–2015." *ILR Review* 72(2): 382–416.
- Levine, Ross, and Yona Rubinstein. 2017. "Smart and Illicit: Who Becomes an Entrepreneur and Do They Earn More?" *Quarterly Journal of Economics* 132(2): 963–1018.
- Moskowitz, Tobias J., and Annette Vissing-Jorgensen. 2002. "The Returns to Entrepreneurial Investment: A Private Equity Premium Puzzle?" *American Economic Review* 92(4): 745–778.
- Oster, Emily. 2019. "Unobservable Selection and Coefficient Stability: Theory and Evidence." *Journal of Business and Economic Statistics* 37(2): 187–204.
- Robins, James M., Andrea Rotnitzky, and Lue Ping Zhao. 1994. "Estimation of Regression Coefficients When Some Regressors Are Not Always Observed." *Journal of the American Statistical Association* 89(427): 846–866.
- Roy, Andrew D. 1951. "Some Thoughts on the Distribution of Earnings." *Oxford Economic Papers* 3(2): 135–146.
- Ruggles, Steven, Sarah Flood, Matthew Sobek, Daniel Backman, Annie Chen, Grace Cooper, Stephanie Richards, Renae Rogers, and Megan Schouweiler. 2024. IPUMS USA: Version 15.0 [dataset]. Minneapolis, MN: IPUMS.
- VanderWeele, Tyler J., and Peng Ding. 2017. "Sensitivity Analysis in Observational Research: Introducing the E-Value." *Annals of Internal Medicine* 167(4): 268–274.

A. Data Appendix

A.1 Variable Definitions

Variable	ACS Variable	Definition
Incorporated SE	COW	= 1 if Class of Worker is 7 (self-employed, incorporated business)
Unincorporated SE	COW	= 1 if Class of Worker is 6 (self-employed, not incorporated)
Wage worker	COW	= 1 if Class of Worker is 1–5 (private or government employee)
Earnings	WAGP	Wages, salary, and self-employment income in past 12 months
Log earnings	–	= $\log(\text{WAGP} + 1)$
Full-time	WKHP	= 1 if usual hours worked per week ≥ 35
Hours	WKHP	Usual hours worked per week
Age	AGEP	Age in years
Female	SEX	= 1 if sex = 2 (female)
College	SCHL	= 1 if educational attainment ≥ 21 (bachelor’s degree)
Married	MAR	= 1 if marital status = 1 (married, spouse present)
White	RAC1P, HISP	= 1 if White alone and not Hispanic
Black	RAC1P, HISP	= 1 if Black alone and not Hispanic
Hispanic	HISP	= 1 if Hispanic origin, any race
Asian	RAC1P, HISP	= 1 if Asian alone and not Hispanic
Homeowner	TEN	= 1 if tenure = 1 or 2 (owned with or without mortgage)
COVID period	YEAR	= 1 if survey year is 2021 or 2022
Weight	PWGTP	Person weight

A.2 Propensity Score Model

Table 5 reports coefficients from the logistic regression used to estimate propensity scores for the aggregate self-employment treatment.

Table 5: Propensity Score Model Estimates

Variable	Coefficient	SE
Intercept	−4.521	0.045
Age	0.098	0.002
Age ²	−0.0008	0.00002
Female	−0.412	0.008
College	0.142	0.009
Married	0.298	0.009
White	0.186	0.012
Black	−0.324	0.016
Hispanic	−0.089	0.014
Asian	−0.156	0.018
Homeowner	0.245	0.010
COVID period	0.032	0.009
N	1,397,605	
Pseudo R ²	0.042	

Notes: Logistic regression of aggregate self-employment indicator on covariates.

B. Robustness Appendix

B.1 Trimming Sensitivity

Table 6: Sensitivity to Propensity Score Trimming

Trim Threshold	Estimate	95% CI	N
Baseline (99th % weights)	−0.362	[−0.371, −0.354]	1,397,605
Trim at 1%	−0.363	[−0.372, −0.354]	1,397,605
Trim at 5%	−0.363	[−0.372, −0.354]	1,265,784
Trim at 10%	−0.326	[−0.337, −0.315]	591,353

Notes: Trim threshold indicates propensity scores outside [threshold, 1−threshold] are excluded. Baseline uses IPW with weights truncated at 99th percentile. Full sample N = 1,397,605 (all employed: 1,264,974 wage + 79,946 unincorp + 52,685 incorp) used for aggregate self-employment effect.

B.2 Coefficient Stability

Table 7: Coefficient Stability Analysis

Specification	Estimate	SE	R ²	ΔR^2
(1) Minimal (Age, Female)	−0.374	0.003	0.055	–
(2) + Demographics	−0.350	0.003	0.194	0.139
(3) + Full controls	−0.348	0.003	0.199	0.005

Notes: Sequential OLS regressions adding covariates. Oster (2019) $\delta = 2,589$, indicating selection on unobservables would need to be over 2,500 times as important as selection on observables to explain away the result.

Acknowledgements

This paper was autonomously generated as part of the Autonomous Policy Evaluation Project (APEP).

Contributors: @SocialCatalystLab

First Contributor: <https://github.com/SocialCatalystLab>

Project Repository: <https://github.com/SocialCatalystLab/auto-policy-evals>