The Spiderweb of Partnership Tax Structures

EMILY BLACK, Stanford University*
RYAN HESS, Oklahoma State University*
REBECCA LESTER, Stanford University †
JACOB GOLDIN, University of Chicago
DANIEL E. HO, Stanford University
MANSHEEJ PAUL, Stanford University
ANNETTE PORTZ, Internal Revenue Service

February 24, 2023

Using an interdisciplinary approach combining expertise in computer science and business taxation, we examine the complex tax structure of pass-through businesses, focusing on partnership entities. Partnerships control more than $30 trillion in assets, vastly outnumber U.S. public firms, and are responsible for three times the amount of U.S. tax non-compliance as corporations. However, the prior literature provides extremely little evidence explaining the pervasive use of such entities and which specific characteristics enable the lightly taxed nature of partnership business income. Using administrative U.S. tax data, we first create graphical organizational structures by tracing income through millions of partnership entities. We show that 85 percent of partnership groups are simple structures composed of one single partnership owned directly by individual taxpayers. In contrast, the most complex structures resemble "webs," characterized by multiple tiers of ownership and clusters of overlapping partners. Second, we harness the power of machine learning models to study partnership non-compliance. We demonstrate that network characteristics are informative for predicting non-compliance. Finally, we show that more flexible machine learning models—random forest classifiers—significantly outperform simpler statistical methods. These models reveal important nonlinear relationships between firm characteristics and noncompliance, sharply contrasting with linear models predominantly used in the corporate space. Thus, beyond adding to the nascent literature explaining the prevalent use of partnerships, we provide new insights about how to correctly specify models that explain the under-reporting of U.S. business tax.

Keywords: Private firms, partnerships, tax noncompliance, machine learning
JEL Codes: H24, H25, D85

We thank John Guyton, Thomas Hertz, Ron Hodge, Larry May, Keisha Miller, Eric Tressler, Alex Turk, and John Worsley from the IRS for providing us with the relevant data, sharing their expertise, and supporting our research. We also appreciate comments from Erin Towery, Steve Utke, and participants at the IRS/TPC Joint Research Conference on Tax Administration. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors and do not necessarily reflect the views or the official positions of the U.S Department of the treasury or the Internal Revenue Service.

*Both authors contributed equally to this work.
†Supervisor and corresponding author; rlester@stanford.edu. Mailing address: 655 Knight Way, Stanford, CA 94305.

1 Introduction

Partnerships are an increasingly important component of the U.S. economy, rising from 28% of U.S. business entities in 2003 to 40% by 2020; see Figure 1 (IRS 2021). A number of factors have contributed to this rapid rise, including favorable tax treatment and the flexibility that partnerships offer for owners in allocating business profits, expenses, and various tax attributes. At the same time, audit rates for partnerships have plummeted. These facts, in conjunction with the disproportionate use of partnerships by the highest income taxpayers, have resulted in sustained concern about the role of partnerships in facilitating tax avoidance and evasion (Versprille 2020, Iacurci 2021, Burns 2022). However, despite the importance of partnerships and the policy interest in understanding their role in tax planning, some of even the most basic descriptive facts about partnerships remain largely unknown. In this paper, we use confidential, anonymized IRS administrative records to i) provide descriptive evidence about the prevalence and complexity of U.S. partnership organizations, and ii) study how the unique organizational flexibility afforded by partnerships facilitates tax non-compliance.

Prior work focuses primarily on corporate tax avoidance, with the literature including hundreds of studies on this topic (reviewed in Hanlon et al. 2010, Wilde et al. 2018). In contrast, academic working studying partnerships comprises only a small fraction of the business tax literature, even though pass-through business noncompliance is estimated to be three times as large as corporate noncompliance.[1] One explanation for the lack of evidence on partnerships relates to the lack of data: most partnerships are private and thus not subject to the same disclosure requirements as publicly-traded corporate entities. Furthermore, the majority do not voluntarily produce financial statements that permit measurement of firms' tax choices (Allee et al. 2009; Lisowsky et al. 2020). A

[1] Pass-through business income, which captures non-compliance by both S corporations and partnerships, is estimated to be $110 billion for the tax years 2011-2013; see Figure 1 of Publication 1415 (Rev. 9-2019). "Federal Tax Compliance Research: Tax Gap Estimates for Tax Years 2011-2013.".

second explanation for the lack of evidence relates to complexity; partnership structures can include multiple entities across many different tiers and types of owners. Understanding how partnership businesses are organized and the role of tax planning in these businesses requires a complete picture of the chain of entities and owners.

In this paper, we use an interdisciplinary approach that combines computer science expertise with business tax knowledge to study the organizational structures of U.S. partnerships. To identify which business entities are connected to a partnership, either directly or indirectly, we graphically map the network of ownership relations between business entities and their individual owners, tracing income through millions of partnership entities and multiple tiers of ownership. This graphical analysis produces a visualization of each partnership organization, which we define as a business group consisting of at least one partnership and all connected owners. The analysis yields what we believe to be the first comprehensive dataset of Partnership-Related Businesses (PRBs) in the United States.

From the data, we report several novel descriptive facts about PRBs and their organizational structures. We find that the vast majority of partnerships – approximately 85% – are directly owned by individuals; see Figure 2. We coin these organizations "simple" partnerships. Among this group of simple partnerships, most (72% of all partnerships) are owned by exactly two individuals. The average simple partnership reports approximately $356,000 of sales and $27,000 of ordinary income.

The remaining 15% of partnerships employ a wide range of structures, such as the ones depicted in Figures 3 through 6. The most complicated structures resemble "spiderwebs" with groups of related entities and clusters of overlapping partners. These structures are strikingly different from corporate structures, which are often tiered and characterized by several layers of entities underneath one parent entity. We refer to this group of partnerships as "complex" PRBs. The average complex partnership reports approximately $1.5 million of sales and $86,000 of ordinary income.

We observe different patterns of reported income among complex organizations as compared to the simple partnerships. For example, while operating income predictably increases with number of partners in a simple partnership, we observe a less monotonic relation between operating income and partner count among complex PRBs. These patterns are consistent with the use of such complex structures for shifting profits and losses among group members to minimize overall tax obligations.

Having provided new descriptive insights about the use of partnerships in contemporary U.S. business structures, we next study whether partnership organizational characteristics can help predict partnership tax avoidance. We posit that these complex organizational structures are a distinct feature of partnerships and a central mechanism contributing to the lightly taxed nature of these firms. For example, partnership complexity can cause difficulties in "determining the relationships and allocations of income and losses" reported to partners (GAO 2014). One recent paper examines the use of pass-through entities within a subset of public C corporations, finding that the use of such entities is indeed associated with greater levels of corporate tax avoidance (Agarwal et al. 2021). However, it is unclear whether these findings extend to the broader population of partnerships in our data, which are primarily private and thus responsive to differing reporting and tax incentives. To the extent that partnership structures are driven by non-tax considerations, such as legal protection or owners' non-tax preferences, we would find little relation between the level of complexity and measures of non-compliance.

We distill characteristics of each PRB structure into specific measures ("features"), such as the number of partners, the tiers of ownership, and the length of ownership chains. Drawing on outcomes from the near-universe of IRS partnership audits between 2013 and 2015, we show that inclusion of these features into machine learning algorithms can help predict which audits are likely to detect non-compliance. Specifically, we observe two key results. First, for random forest machine learning models trained to predict non-compliance, incorporating PRB network characteristics increases model accuracy by approximately 2 percentage points. This is a substantial gain, given the high resource demands of complex audits and the potential revenue at stake. Second, there are substantial performance gains for predicting non-compliance when using flexible non-linear models (random forest), as compared to simpler linear models (OLS or logistic regression), regardless of whether network characteristics are incorporated.

Our work contributes to the literature in several ways. Despite the widespread use of partnerships and popularity among businesses and high net worth individuals, the literature includes relatively little information about these companies and the role of tax in these firms' decisions. Cooper et al. 2016 examine partnerships and S corporations and document that income earned through the partnership sector is relatively low-taxed, with estimated effective tax rates of 15.9 percent. Their work and other papers provide some evidence that the lightly taxed nature of partnership income is attributed to tax rate differences between the corporate and non-corporate sectors, higher levels of preferentially taxed investment income reported within partnership structures, and ownership by tax-exempt and foreign individuals (Smith et al. 2021, Kopczuk et al. 2020, Cooper et al. 2016). However, beyond this work, there is extremely little evidence. Thus, the first contribution of our paper is providing new descriptive evidence about the amount, type, and use of partnership structures in the United States. Our graphical depictions of partnership organizations provide visual evidence of complex structures and the nature of entity structures used by these private companies.

Second, we construct measures of partnership complexity from the network structures and demonstrate that these novel measures help explain the relatively low levels of taxation on income earned through this sector. Thus, we further contrast partnership businesses from the hundreds of studies on corporate avoidance and non-compliance, showing the distinct importance of flow-through characteristics. In doing so, we also extend work on the corporate side that studies the use of flow through businesses in a corporate structure (Agarwal et al. 2021).

Third, by using machine learning systems in this context, we shed light on the nature of the relationships between partnership characteristics and tax non-compliance. Specifically, we document important non-linearities in the partnership space. The fact that random forest models substantially outperform linear models calls into question the use of linear OLS analyses as a tool for understanding non-compliance for these businesses. The presence of such a nonlinear relation raises questions about limitations of OLS as an analytic tool beyond the partnership space.

Finally, the paper offers policy relevant evidence. The complexity of partnership structures makes them an apt vehicle for business tax planning. Thus, addressing non-compliance within this sector is critical for addressing questions of tax under-reporting; indeed, the Biden administration intends that the $80 billion recently promised to the IRS to focus on "enforcement activity aimed at high-wealth taxpayers, large corporations, and partnerships" (CBO 2021). We provide evidence useful for enhancing tax policies and IRS tax administration capabilities for these businesses.

2 Data

2.1 Sample Construction

We first identify all partnerships in the IRS data for 2013 through 2015 (10.9 million entities). We compare this number to the total number of partnerships reported in the IRS 2013-2015 Data Book and confirm that our initial dataset contains the near universe of firms. We obtain data about each partnership, including income and expense items reported on Form 1065, U.S. Return of Partnership Income. Also using Schedule K-1s from the partnership return, we obtain information on each partner's ownership percentage, capital account balance, and distributive portion of each partnership line item. Finally, we obtain information about whether each partner is an individual, C corporation, S corporation, or trust.

We impose two sample restrictions necessary for creating partnership organizational structures. First, we retain only those partnerships that report any activity during the tax year. That is, we drop any partnerships for which there are no income or loss amounts reported on the partnership tax return, as we require non-zero income to estimate ownership percentages. We also drop any partnership for which the ownership percentages do not sum to 100%. This latter step ensures that we account for all owners and all income. While necessary, it substantially reduces the sample size, primarily due to an extremely large network structure (greater than 2 million nodes) that we cannot otherwise unpack without de-anonymized data. The resulting sample includes over 7.0 million partnership firm-years. Table 1a provides the sample selection.

2.2 Creation of Organizational Structures

We next create visualizations of partnership organizations using graphical imaging software (Python NetworkX), which permits analysis of the structure and activity of complex networks. We start by graphing one node for each partnership entity. We then graph the owners of each partnership as additional nodes and connect the nodes with lines or "edges." If the owner nodes are tax-paying entities (i.e., individuals or corporations), we consider this the boundary of the partnership and do not graph any further lines of ownership. However, if the owner nodes are flow-through entities, such as S corporations or additional partnerships, we also graph nodes for those entities. We then look through the entities to further graph those entities' ownership. We iteratively progress through the ownership chain until we arrive at the final taxpaying entity for each partnership. This process produces network graphs for each partnership organization that depicts overlapping and common ownership when present.

We classify a partnership organization as either "simple" or "complex" based on both the number of partnerships and the type of partners included in each distinct business group. We define simple partnership organizations as a single partnership wholly owned by individual taxpayers. Figure 2 provides examples of three simple partnership structures. Orange triangles represent the partnership entities; blue circles indicate the individual owners, where each circle is sized proportionate to the amount of partnership income reported to that partner. The arrow indicates ownership and points to the partner directly owning each entity. Panel A presents a partnership group with two owners; Panels B and C present groups with four and eight owners, respectively. Table 1a shows that we classify approximately 6.0 million, or 85 percent, of firm-year observations as simple partnerships.

Complex partnerships include all other partnership organizations. For example, we categorize any business group with multiple partnership entities as complex partnerships, even if all partners are individuals. Figure 3a provides an example of this type of complex organization. We also categorize business groups with one partnership entity but a mix of partner types as a complex partnership. Figure 3b provides an example of this group, with the green node denoting partners that are a C corporation, S corporation, or trust. The most complex organizational groups include both multiple partnerships and multiple partner types. Approximately 1.0 million partnerships in our data are included in over 600,000 complex partnership organizations.

Table 1b shows the count of the organizations by "Simple" and "Complex." The vast majority of simple partnership observations in our sample (4.3 million firm-years) have two individual partners, representing 72 percent of simple partnership observations. The frequency of simple partnerships generally declines as the number of partners increases, with approximately 600 firm-years reporting over 50 partners. We observe similar patterns among the complex partnership organizations: most complex organizations include partnerships with only two owners, and the count of partnerships again declines with the number of partners. Table 1c provides further details about complex partnerships. Over half of complex partnership observations contain only one partnership entity ((n=381,000), meaning that these business groups are categorized as complex due to having a mix of partner types (e.g., individuals, corporations, etc.). The remaining complex partnerships include multiple entities and a mix of partner types.

3 Descriptive Characteristics of Partnerships

3.1 *Industry Affiliation*

This section presents new descriptive evidence about the types of businesses using partnership structures and the amount of income reported by these firms. We start by presenting information about the industry affiliation of firms included in our sample in Table 2a. We present sample counts by two-digit NAICS codes self-reported on the partnership tax return. Columns (1) and (2) show that almost half the sample (47.7 percent) is in the real estate industry. We do not observe similar concentrations in any other industries, with the next largest counts in Professional Trade (7.8 percent), Retail Trade (5.9 percent), Agriculture/Forestry/Fishing (5.7 percent), and Construction (5.0 percent).

We observe this large real estate presence for both simple and complex partnerships in Columns (3)-(4) and Columns (5)-(6), respectively, with a slightly higher proportion among complex partnerships (52.6 percent) as compared to simple partnerships (46.8 percent). Complex partnerships also have a higher proportion of firms in the finance and insurance industry as compared to simple partnerships (9.3 percent versus 3.9 percent). This difference likely relates, in part, to the use of complex partnerships by the private equity industry.

3.2 *Type of Income Reported*

Partnerships allow for different allocations of income depending on the type and nature of income earned. This flexibility in allocating income may be particularly helpful for real estate and investment partnerships, which generate certain types of "passive" income subject to special tax rules. To shed light on the types and amounts of income reported in partnerships, we next present average levels of income by size of the organization (based on the number of partners) in Table 2b. Columns (1)-(3) present information for simple partnerships; Columns (4)-(6) provide information for complex organizations.

We first report Total Income in Columns (1) and (4) for simple and complex partnerships, respectively. Total Income is equal to Ordinary Business Income (Schedule K, Line 1) plus Investment income (Schedule K, Lines 5-9).[2] Simple partnerships with two individual partners report average income of $62,000. Income generally increases with the number of partners; among the biggest simple partnerships (with 30-50 partners), average income is $1.5-$17.9 million. Complex organizations in Column (4) have higher levels of average total income; for example, organizations with two partners report average income of $114,000. As with simple partnerships, income is increasing in the number of partners, although the largest partnerships appear to report *lower* levels of income when comparing to Column (1) ($2.8 million versus $17.9 million).

Columns (2) reports Operating Income for simple partnerships. We again observe amounts increasing with the number of partners, and the levels appear larger than Investment Income in Column (3). Investment Income in Column (3) varies less and generally remains in the range of $30,000-$40,000, except in the largest partnerships.

Columns (5) and (6) repeat the analysis for complex partnerships. While we observe similar patterns for Operating Income in Column (5), we note two differences as compared to simple partnerships. First, levels of Investment and Operating Income appear more similar for these organizations. Furthermore, we observe less variation in the levels of Investment Income across partnership groups,

[2]Ordinary Business Income is equal to a partnership's Revenues, less Deductions, all of which are considered to be derived in the course of an operating business. The other line items include Rental Real Estate income (Line 2), Other gross rental income (Line 3), Interest Income (Line 5), Dividends and equivalents (Line 6), Royalties (Line 7), Net short-term capital gain (Line 8), and Net long-term capital gain (Line 9).

with amounts not exceeding $150,000 even among the largest organizations. One explanation relates to flexibility afforded by partnerships: complex structures facilitate special income allocations that minimize income and/or optimize tax losses across related parties. We test this in Section 4.

3.3 *Financial and Network Characteristics*

Table 2c provides additional descriptive statistics using financial measures and network characteristics distilled from the organizational structures described in Section 2. Columns (1) and (2) present information for the full sample; Columns (3)-(4) and (5)-(6) present information for subsamples of simple and complex partnerships, respectively.

Partnerships report $522,000 of average gross receipts (*SALES*). Average compensation paid to employees and owners includes $65,000 of *SALARY AND WAGE* and $12,500 of *GUARANTEED PAYMENTS*, respectively.[3] *FOREIGN TAX* is equal to the amount of foreign tax paid by the partnership and reflects the multinational presence of a company; the amount is low ($193) because most partnerships report no value for this line item. Sample firms have external debt financing based on mean *INTEREST EXP* of $5,500, and they also own fixed assets based on average *DEPRECIATION* of $10,600. *ORDINARY INCOME* earned from operations is $36,000, whereas average income attributable to *RENTAL* activities is $13,500. Other investment income earned from *DIVIDENDS*, long-term capital gains (*LTCG*), and gains from the sale of assets used in business (*SEC 1231*) equal $1,400, $4,700, and $27,200, respectively.

The next items relate to partnership organizational structures. *IN DEG* measures the number of Schedule K-1s a partnership is issued, meaning that the partnership owns another partnership. The average *IN DEG* of 0.05 means that most partnerships do not own another partnership, reflecting that most partnerships having a simple structure. *OUT DEG* measures the number of Schedule K-1s

[3] Guaranteed payments are compensation amounts paid to partners, whereas salaries and wages include compensation paid to non-owner employees. The guaranteed payment is in addition to the proportionate share of income or loss that owners earn each year by virtue of their partnership stake.

issued to partners; the value of 2.72 means the average partnership has 2.72 direct owners. We count the type of entities within the organizational structure, including the number of *TAXABLE PARTNERS* (C corporations and individuals), *PARTNERSHIPS*, *S CORPS*, *TRUSTS*; average values are 3.02, 1.42, 0.08, and 0.09, respectively. Finally, we report the average degree of separation (*DoS*), which is defined as the maximum path between a partnership entity to a tax-paying entity within an organization. That is, DoS is a measure of the length of the ownership chain for a partnership. The value of 1.16 means that the average distance between a partnership and its ultimate owner is 1.16 links (i.e., it is directly owned, again reflecting that most partnerships are simple).

A comparison of the values across simple and complex organizations reveals that complex partnerships report higher mean and variance across each of the financial measures. For example, complex organizations report four times the amount of *SALES*. Consistent with the use of complex structures for investment purposes, these organizations report six times the dollar value of *LTCG* and ten times the amount of *DIVIDENDS*. By construction, *IN DEG*, *S CORPS*, and *TRUSTS* are equal to zero for simple structures, and *OUT DEG* equals *TAXABLE PARTNERS*. In contrast, complex structures have 3.6 direct partners but many more indirect partners based on the average value of 5.6 for *TAXABLE PARTNERS*. The average *DoS* is 2.1 links.

We provide additional graphical evidence about partnership structures in Figure 4, which depicts three complex organizations that each contain four partnerships. Panel A shows an organization with two DoS; Panel B (C) presents organizations with three (four) DoS. For example, in Panel A, each of the four partnerships is owned by a combination of individuals and another partnership in the group. Thus, the longest path to the ultimate owners for each of these partnerships is two, given that income is reported first to another partnership and then to the individual partner.

Figure 5 provides additional figures that permit a comparison of structures based on the number of partnerships and the DoS within these entities. The organizations in Panel A and B each contain

9 partnerships. Despite having the same number of partnerships, the structures have very different DoS: the structure in Panel A has two DoS, meaning that all nine partnerships are ultimately owned by the two blue nodes, whereas the structure in Panel B has five DoS. The structure in Panel B includes twelve individual owners (blue nodes) and one entity owner (green node). Panels C and D (E and F) depict additional structures that contain 14 partnerships (19 partnerships) and show increasingly complex web-like organizations characterized by a large number of partnership entities and greater DoS between the entities and the taxpayer. Figure 6 shows two of the most complex structures that include 20 partnership entities, with two (eight) DoS in Panel A (B).

A comparison of the simple structures in Figure 2 to the most complex structures in Figure 6 demonstrates vastly different use of the partnership entity form. Why these more complex structures exist and the economic factors motivating such structures are large and open questions in the literature. Section 4 uses characteristics of these structures to study partnership non-compliance.

3.4 *Size Distribution*

Table 2d presents additional information about the relative size and activity of partnerships. We use the total amount of *SALES* as a proxy for size and report the count of observations based on differing bins of receipts. Column (1) shows that over 4.1 million partnerships (60 percent) report no gross receipts, likely due to instead generating income required to be separately reported on Schedule K. Even among those firms reporting positive *SALES*, almost all (1.2 million observations) report amounts less than $100,000. Few observations (n=7,800) report sales over $50.0 million. Similar patterns hold for simple and complex partnerships, although a higher proportion of complex partnerships report $0 of gross receipts (70.6 percent in Column (6) as compared to 56.5 percent in Column 4). We observe relatively higher proportions of complex partnerships in the higher income bins, although the counts are still low.

Table 2e reports a similar summary based on Schedule K income (excluding Ordinary Income from Line 1 of Schedule K). Across the full sample, as well as the subsample of simple and complex partnerships, we observe that approximately three-quarters of the partnerships report either $0 or less than $100,000 of income on Schedule K.

In summary, this section provides new information about the amount, type, and complexity of U.S. partnerships based on descriptive statistics and visualizations of partnership network structures. While most partnerships are simple in nature and report relatively small amounts of income, we find that, even among simple partnerships, some organizations have a large number of partners and amount of reported operating income. The figures depicting complex partnerships suggest some extremely complicated ownership chains, possibly contributing to the differing amounts and types of income reported by those organizations.

4 Organizational Complexity and Partnership Tax Avoidance

4.1 *Research Design*

This section provides empirical evidence to answer three questions about partnerships. First, to what extent are firm characteristics identified in the prior literature – but for corporate entities – predictive for partnership tax avoidance? Second, are network characteristics about partnership structures informative in improving model accuracy and predicting tax non-compliance? Third, what tools are most useful in answering these questions: traditional linear models such as OLS, or machine learning approaches which have the capacity to map nonlinear relationships between input and output?

To address these research questions, we focus on the subsample of audited partnerships. As a first attempt to answer our first two research questions, and to connect our work with the prior literature in the corporate literature, we by implementing OLS models with and without network characteristics. First we estimate the following OLS model:

$$\Pr(Assessment)_{it} = \beta_1 IncDed_{it} + \beta_2 Audit_{it} + \beta_3 Network_{it} + \epsilon_{it}, \tag{1}$$

where $\Pr(Assessment)_{it}$ is an indicator equal to one if firm i in tax year t was assessed an amount of tax due, conditional upon audit. Amounts are obtained from the IRS audit outcomes database. To address the first research question, we first include $IncDed_{it}$, which are measures ("features") for the tax return income and deduction line items described in Section 3. We also include $Audit_{it}$, which are measures related to the IRS tax audit conducted. Specifically, we include the rank of the IRS agent conducting the audit (*AGENT RANK*) to control for experience of the agent. The term ϵ_{it} is an additive error term. We compare both the coefficient estimates and the performance of this model to prior work to assess the predictive ability of this model for partnership entities.

To address the second research question related to the informativeness of network characteristics, we augment our model to include measures that capture the partnership network structures ($Network_{it}$). These include *IN DEG*, *OUT DEG*, and *DoS*. We also include counts of the types of owners: *TAXABLE PARTNERS* (for individuals and C corporations), *PARTNERSHIPS*, *S CORPS*, and *TRUSTS*. However, we find that OLS models are insufficiently flexible tools to provide a reliable understanding of the relationships between firm characteristics or network characteristics and non-compliance conditional on audit.

Upon finding this result, we turn using to machine learning methodologies as another tool to answer the first two research questions. We again create models with and without network features to predict non-compliance. As machine learning models are built and used differently from OLS models traditionally used in the tax literature—namely, ML models are built to predict outcomes on *unseen data* rather than estimating relationships within known data samples—we train two types of models, one of which serves as a bridge between traditional accounting and machine learning approaches. Specifically, we train *logistic regression* models, which are linear statistical models built and trained in the machine learning paradigm, but with similar descriptive power to OLS models, and *random forest models*, which are highly flexible machine learning models capable of mapping nonlinear relationships between input and output. Since logistic regression models are also based on a linear function, they can be seen as the machine learning counterpart to OLS. Thus, by comparing the performance of these models, we can shed light on whether linear modeling approaches such as OLS or logistic regression are sufficiently powerful to understand the relationships between partnership characteristics and noncompliance, or if other more flexible models are necessary—which we find is the case.

Finally, to provide more reliable evidence for whether network characteristics are predictive of noncompliance, we compare the performance of machine learning models with and without network characteristics.

4.2 *Sample of Audited Partnerships*

We estimate Eq. (1) on the sample of partnerships that have been selected for audit. Table 1, Panel A shows that this sample includes approximately 16,000 audited partnerships, of which 83.0 percent (or 13,200) are simple partnership organizations.

Table 3a presents information similar to Table 1, Panel A and confirms that the audited subsample exhibits the same pattern as the overall sample. That is, similarly large proportions of simple and complex organizations have only two partners (85 percent and 47 percent, respectievly); furthermore, the number of organizations declines as the number of partners per organization increases.[4] Table 3b shows that half of complex organizations (n=1,415) include two partnerships.

[4]Throughout the tables, we suppress some data to conform with IRS disclosure requirements. We indicate these items with asterisks.

Table 4 presents similar summary statistics for the audited subsample as presented in Table 2 for the full population. We present statistics for the full sample, as well as subsamples based on whether an amount was assessed or not. Panel A presents industry statistics. As in Table 2, we observe that the largest concentration of audited firms occurs in the real estate industry. However, the proportion is lower (19 percent) as compared to the full population (47 percent). We again observe that Professional Trade, Retail Trade, Agriculture/Forestry/Fishing, and Construction are again represented as some of the larger industries. The Accommodation/Food Services industry appears as a new industry represented.

Table 4b presents summary statistics about the type of income reported in the simple and complex audited partnerships. Given the low number of entities, some amounts are suppressed to avoid IRS disclosure issues. Two findings stand out. First, we observe that several of these amounts are losses, suggesting audits around the amount, type, or allocation of losses within a partnership group. The trends holds for both simple and complex partnership organizations. Second, within this sample, the amount of income can be sizeable, increasing beyond \$100 million for the largest partnership.

Table 4, Panels C, D, and E provide additional descriptive evidence about audited firms. We observe a much larger amount of *SALES* as compared to the population (over \$3.8 million), with a large standard deviation (\$85.0 million). Other financial characteristics are similarly larger than the population, with the exception of *ORDINARY INCOME* and *RENTAL*, both of which are negative numbers indicating losses. The IRS is likely more interested in auditing loss partnerships than C corporations because partnership can directly offset taxable income at the individual level.

4.3 Estimating Likelihood of Audit Adjustment with OLS

We start by estimating Eq. (1) using OLS in Table 5. We observe that three financial characteristics exhibit statistically significant relations with the probability of assessment: *SALES* is negatively related to assessments, whereas *ORDINARY INCOME* and *DEPRECIATION* is positively associated. Investment-related items, such as *LTCG*, *SEC1231*, and *GUARANTEED PAYMENTS* are negatively associated with the likelihood of assessment, whereas *AGENT RANK* is positively associated.

These results suggest that including characteristics previously identified in the corporate literature can help explain tax non-compliance among partnerships. However, we note that the overall explanatory power of this model is low. The R-squared is 0.004, suggesting that the use of these characteristics can explain only a tiny portion of the assessment decision.

We next include network characteristics to understand, when using a similar OLS model, if these characteristics help predict the likelihood of assessments. We indeed find that inclusion of these characteristics provides incremental explanatory power. The R-squared statistic triples, and four characteristics exhibit a significant relation with the likelihood of assessment: *DoS*, *IN DEG*, *OUT DEG*, and *TRUSTS*. Interestingly, *DoS*, *OUT DEG*, and *TRUSTS* are *negatively* associated with the likelihood of audit, suggesting either that complex organizations are not necessarily less compliant, or that the complexity of organizations is not fully considered during the audit process.

While the OLS model gives some indication that network features may be helpful in understanding the mechanism of partnership noncompliance, overall, the low R-squared values suggest that OLS may not be a sufficiently powerful tool to reliably predict noncompliance in the partnership space. To create a more powerful tool to predict noncompliance—and to get a more reliable answer to our research questions—we turn to machine learning techniques.

4.4 Predicting Likelihood of Audit Adjustments

Having established a baseline comparable to the prior literature, we next turn to using other estimation approaches, namely ML logistic regression models and Random Forest models.

We build each class of algorithm to estimate whether or not there is an adjustment. More precisely, the model is trained to predict a binary output $\hat{y} \in \{0, 1\}$, where 1 indicates a positive adjustment and 0 indicates no adjustment or negative adjustment. This is type of prediction problem is referred to as a binary classification problem. Importantly, as machine learning models are built, or trained, to predict an outcome on unseen data, rather than to describe relationships between inputs and outputs within a known sample, we allow the machine learning models to learn patterns from one random subset of the audited partnerships, called the training set, and we test their predictive performance on a disjoint subset of the data, called the test set. We report all ML model performances on the test set.

4.5 Model Types

As mentioned previously, we train two different types of models with the machine learning paradigm: logistic regression and random forest models. As logistic regression models are linear, and thus similar to OLS models, they serve as a way to provide an like for like comparison between more traditional (linear) approaches of understanding tax data and complex machine learning methods such as random forests.

Logistic regression models predict whether or not there is an audit adjustment by attempting to separate partnerships with and without an audit adjustment with a linear function— in this sense, it is similar to an OLS model. However, the loss function that the model optimizes to find this linear function is different—while OLS minimizes mean squared error to fit its line, logistic regression uses maximum likelihood estimation. Further, in order for the model to generalize well to unseen data, an *regularization penalty* (ℓ_2) is added to prevent the model's coefficients from *over-fitting*, or over-specializing to the data which the model is trained on. Finally, in order to optimize the model's performance on a the *classification task* (selecting a binary output, 0 or 1) as opposed to estimating a continuous value, the linear function's output is mapped to a sigmoid function to concentrate scores on values closer to zero or one. The final output is then thresholded to give a binary value. Thus, this is similar to an OLS model in that it is linear, but differs in its optimization method and final output.

Random forest models are ensembles of decision trees whose predictions are aggregated to determine the overall model's prediction. Decision trees are models that iteratively separate data based on thresholded feature values in order of highest information gain, relative to the prediction task. Random forest models thus have great flexibility to pick up nonlinear and local patterns in the data.

4.5.1 Model Training

To ensure model performance is not dependent upon the specific subset of data that was randomly selected over which to calculate performance, we calculate all model performance over 10 trials, each with randomly sampled train, validation, and test sets. In each trial, we use the validation set to pick the best-performing model of each type across a selection of models trained across a randomized grid of hyperparameters. For each trial, we separate the sample into 11,901 observations for training and validation (75%) and 3,968 test (25%) observations. Within this split, we performed three-fold cross validation to select the best hyperparameters, with a training set of 7,934 (66%) and a validation set of 3,967 (33%). We evaluate model performance on the test set, i.e., data that was completely unseen to the model during training and validation. For each of the models described below, we report the average and standard deviation of test set accuracy for each model over the ten trials.

4.5.2 Linear verses Random Forest Models

We find that more flexible random forest models are better able to predict probability of adjustment than linear models, especially in areas of the distribution where the models are *confident* in their predictions. In Columns (1) and (3) of Table 6, we report mean accuracies of baseline models over the top 10% of most confident predictions, over ten trials with random train-test splits. Machine learning classification models are able to report a raw score indicative of confidence in its decision in addition to a zero or one output of whether there is noncompliance or not. In logistic regression, this is simply the unthresholded output of the model, and in random forest models, the confidence is the percentage of random trees that give the positive outcome. A raw output of 0.5 connotes no confidence, and outputs close to zero or one indicate high confidence in a negative or positive outcome, respectively.

We see that, when calculating accuracy over the top 10% of confident predictions for each model, random forest models leads to a 27% increase in accuracy. This is derived from comparing the Test Accuracy of 79.7 percent reported at the bottom of the table in Column (3) to the Test Accuracy of 53.1 percent in Column (1).

We graph the accuracy of linear and random forest classification baseline models over different percentages of the distribution where the partnerships are ranked in order of the model's confidence in prediction in Figure 7. The blue line plots the accuracy of the logistic regression, and the orange line is the random forest classification. Over the entire distribution, we can see the accuracy increase is approximately 10% based on the average wedge between the two lines.

Notably, find that the performance increase for using random forest models over traditional linear models is greater when looking at the subset of predictions where each model has the most confidence. We can see this result in Table 6 and graphed in Figure 8: while accuracy over the entire dataset is relatively low for both models—52-55% and 62-64% for linear and random forest models, respectively–in the top 5% of predictions, the random forest model can perform with 87% accuracy. This contrasts with the logistic regression model, which remains around 55%. This is a remarkable difference in performance. The differences in accuracy possibly reflect that, while many of the partnerships with and without adjustments are indistinguishable from each other based on the audited data, there exists a subset of partnerships with adjustments that reliably stands out from the rest—but *only* detected in a nonlinear model.

The fact that accuracy is so high at the top of the confidence distribution is promising for the use of partnership noncompliance prediction models in practice, as limitations of IRS budgets prevent investigation of all partnerships marked as possibly noncompliant by the model. Instead, the most confident recommendations from such a model could inform audit selection, representing a sample with a higher guarantee of assessment.

These findings suggest the relationships between tax planning and observable partnership features are complex, necessitating use of models that can detect and predict non-linear patterns to identify tax non-compliant partnerships.

4.5.3 Inclusion of Network Features

Next, we turn to the impact of including characteristics of the partnership network, or network features, as inputs to the machine learning models. The results from the OLS model in Table 5 suggest that the inclusion of these characteristics can improve model performance. Consistent with this, we observe that including network features in the high-performing random forest model leads to performance increases for both linear and random forest models in Table 6. However, impacts on the logistic regression model appear to be mildly negative, but too noisy to draw a clear impact.

We report the results for the Random Forest model in 6, Columns (3) and (4). Including network features in Random Forest Models increases accuracy on the top 10% of predictions from 79.7% to 81.9%, or about 2.2%. Figure 7 plots the feature importance that corresponds with Column (4). Note that Random Forest models do *not* create coefficients that are comparable to linear models such as logistic regression and OLS. Instead, feature importance captures the average accuracy decrease observed from randomly perturbing a feature value over the dataset; here, this feature importance is derived from perturbations over five trials for each model. Thus, these numbers represent a measure each feature's contribution to the overall accuracy of the model. While the network characteristics contribute to the model performance, we note that the importance of these features is lower than several other characteristics of the firm or the audit, such as *ORDINARY INCOME* and *AGENT RANK*. However, given that deriving feature importances in random forests are largely heuristic and susceptible to errors due to colinearities, and only displaces the each feature's importance over the entire model rather than in areas of the distribution (i.e. regions of high confidence) where the model performs exceptionally well, we place more import on the performance improvements of the model when adding network features rather than feature importance metrics.

While the feature importance suggests some slight improvement across the entire sample, we find that including network features has a sizeable improvement in model performance at the very top of the confidence distribution. That is, when focusing on the top 1%-5% most confident predictions, including network features increases model performance between 5-10%. This is a notable improvement in performance and demonstrates the importance of including such network characteristics when predicting non-compliance.

Figure 8, Panel C depicts the random forest model performance across the distribution of firms. The graph shows that the accuracy is highest in areas of high confidence, regardless of whether the model includes or excludes network features. We do observe the incremental improvement even across the entire population when including full features based on the position of the blue line relative to orange.

Notably, among the firms at the very top of the confidence distribution, we observe even greater improvement from inclusion of network characteristics. Figure 8, Panel D provides a zoomed in graph of the model performance for the top 10% of the population. This graph provides a clearer visualization of the improved performance at the different budget levels, showing that the performance is 7.0 percentage points different at the very top of the distribution (97.25% with features versus 9.25% without).

The findings in this section thus suggest that the characteristics of complex structures are helpful for explaining non-compliance, particularly in non-linear models that are capable of detecting non-linearities in the data.

5 Conclusion

Our work provides some of the first evidence about tax planning in partnership entities. To do so, we obtain access to the population of U.S. partnerships for three tax years and graph their network structures. We then use machine learning algorithms to understand non-compliance among partnerships, incorporating network features into our tests.

Our work has substantial implications for research on business tax avoidance. The fact that random forest models outperform linear techniques at predicting noncompliance is of dual importance for both tax administration and academic research. First, it offers promise for tax administration and enforcement models. Second, it demonstrates the limitations of relying on tools such as OLS to understand the relationship between partnerships, their network structures, and tax planning aims. However, while machine learning algorithms are an incredible predictive tool, they can be difficult to interpret[Carvalho et al. 2019; Burkart et al. 2021]. Thus, we acknowledge that, while machine learning may be helpful for assisting in noncompliance prediction, they may be challenging for understanding the underlying mechanisms driving tax planning.

This raises may questions for future research. For example, are there other methods of understanding nonlinear relationships between partnership characteristics and tax planning? Would these methods be appropriate for other business entities, such as S corporations? Can efforts in the world of machine learning explainability be harnessed to use machine learning systems both as a method for predicting outcomes such as noncompliance, but understanding its mechanism as well? We look forward to future research to address these important and unanswered questions.

The evidence from this study will contribute to the literature by providing information about a large sector of U.S. firms that are relatively understudied in the academic literature. Furthermore, the evidence is useful to tax authorities in advancing their capabilities for administering and enforcing U.S. tax law.

References

Agarwal, Ashish, Shannon Chen, and Lillian F Mills (2021). “Entity Structure and Taxes: An Analysis of Embedded Pass-Through Entities”. In: *The Accounting Review* 96.6, pp. 1–27.

Allee, Kristian D and Teri Lombardi Yohn (2009). “The demand for financial statements in an unregulated environment: An examination of the production and use of financial statements by privately held small businesses”. In: *The Accounting Review* 84.1, pp. 1–25.

Burkart, Nadia and Marco F Huber (2021). “A survey on the explainability of supervised machine learning”. In: *Journal of Artificial Intelligence Research* 70, pp. 245–317.

Burns, Tobias (2022). “Business incomes, partnerships rife with unpaid taxes, experts say”. In: *The Hill* November 16.

Carvalho, Diogo V, Eduardo M Pereira, and Jaime S Cardoso (2019). “Machine learning interpretability: A survey on methods and metrics”. In: *Electronics* 8.8, p. 832.

CBO (2021). “The Effects of Increased Funding for the IRS”. In: *Publication 57444*.

Cooper, Michael, John McClelland, James Pearce, Richard Prisinzano, Joseph Sullivan, Danny Yagan, Owen Zidar, and Eric Zwick (2016). “Business in the United States: Who Owns It, and How Much Tax Do They Pay?” In: *Tax Policy and the Economy* 30.1, pp. 91–128.

GAO (2014). “Large partnerships: with growing number of partnerships, IRS needs to improve audit efficiency.” In: *GAO Publication No. 14-732*.

Hanlon, Michelle and Shane Heitzman (2010). “A review of tax research”. In: *Journal of Accounting and Economics* 50.2-3, pp. 127–178.

Iacurci, Greg (2021). “IRS crackdown: Biden administration says it can raise $700 billion by targeting tax cheats”. In: *CNBC* May 20.

IRS (2021). “Internal Revenue Service Data Book”. In: *IRS Publication 55–B*.

Kopczuk, Wojciech and Eric Zwick (2020). “Business incomes at the top”. In: *Journal of Economic Perspectives* 34.4, pp. 27–51.

Lisowsky, Petro and Michael Minnis (2020). “The silent majority: Private US firms and financial reporting choices”. In: *Journal of Accounting Research* 58.3, pp. 547–588.

Smith, Matthew, Danny Yagan, Owen Zidar, and Eric Zwick (2021). *The Rise of Pass-Throughs and the Decline of the Labor Share*. Tech. rep. National Bureau of Economic Research.

Verspille, Allyson (2020). “New Teams, Hiring Part of IRS Plan to Crack Down on Partnerships”. In: *Bloomberg Tax* October 28.

Wilde, Jaron H and Ryan J Wilson (2018). “Perspectives on corporate tax planning: Observations from the past decade”. In: *The Journal of the American Taxation Association* 40.2, pp. 63–81.

Appendix A

Variable Definitions

Dependent Variable	
ADJUST [0/1]	Indicator equal to one if an IRS audit resulted in an adjustment.
ADJUST	Amount of IRS audit adjustment.
Partnership features	
DEPRECIATION	Depreciation expense reported on Form 1065, line 16c.
DIVIDENDS	Ordinary dividends reported on Schedule K, line 6a.
FOREIGN TAX	Total foreign taxes paid reported on Form 1065, schedule K, line 21.
GUARANTEED PAYMENTS	Guaranteed payments to partners reported on Form 1065, line 10.
INTEREST EXP	Interest expense reported on Form 1065, line 15.
LTCG	Net long-term capital gains reported on Schedule K, line 9a.
ORDINARY INCOME	Ordinary business income reported on Form 1065, line 22 is negative.
RENTAL	Net rental real estate income reported on Schedule K, line 2.
SALARY AND WAGE	Salaries and wages reported on Form 1065, line 9.
SALES	Gross receipts reported on Form 1065, line 1a.
SEC 1231	Net Section 1231 gains reported on Schedule K, line 10.
Audit features	
AGENT RANK	The promotion level associated with the examining officer conducting the audit.
Organization features	
DoS	The maximum number of entities through which income flows before it is reported to a taxable partner.
IN DEG	Number of K-1s received by a partnership entity.
OUT DEG	Number of K-1s issued by a partnership entity.
PARTNERSHIPS	Number of partnership entities within an organization.
S CORPS	Number of S corporations within an organization.
TAXABLE PARTNERS	Number of individuals and C corporations within an organization.
TRUSTS	Number of trusts within an organization.

Appendix B

Data Construction

Network Graph

- Network graphs are constructed for each tax year.
- We exclude a K-1 form if either the payer or the payee has TIN flagged as invalid. Invalid TINs can occur for a variety of reasons, such as the SSN/EIN is not found in corresponding databases or the name control could not be matched.
  - Other TINs identified as invalid are also removed (e.g., 0, 123456789, and 999999999).
- We construct a NetworkX MultiDiGraph from the IRS 1065 K-1 table.
- We add the payer and the payee as nodes on the graph for each queried form. Each node is identified by a (TIN, TIN_TYPE) tuple, which is used as the hash for the node.
- Each edge (K-1) is recorded as amended if it is either an amended or a corrected original form. Otherwise it is recorded as original.

Collapse (Payer, Payee)s with two edges into one edge

1. Identical Edges
   Two edges are identical if they have the same payer, payee, return type, cycle and entry dates, and amounts.
2. Prioritize Amended Edges
   We keep the amended edge for if there are two identical edges and one is amended and one is original.
3. Filing Dates for Ordered Edges We keep the later entry date and cycle date if there are two identical edges.
4. Identical Edges with Different Amounts
   We keep the edge with the smaller (larger) reported income/gain (loss) because tax non-compliance strategies generally require taxpayers to minimize (maximize) their income/gains (losses).

Collapse (Payer, Payee)s with more than two edges into one edge

1. Identical Edges
   Edges are identical if they have the same payer, payee, return type, cycle and entry dates, and amounts.
2. Prioritize Amended Edges
   We keep the amended edge for if there are multiple identical edges and one is amended and the remaining are original.
3. Filing Dates for Ordered Edges
   We keep the later entry date and cycle date if there are multiple identical edges.

4. Identical Edges with Different Amounts
   We keep the edge with the smaller (larger) reported income/gain (loss) because tax non-compliance strategies generally require taxpayers to minimize (maximize) their income/gains (losses).

Directed Graph of Partnerships

- After these data cleaning steps, the MultiDiGraph can be reduced to a DiGraph, and any pair of nodes has at most one edge.

Ownership

1. We remove partnerships reported as being 100 percent owned by a single individual based on capital ownership.
2. We calculate profit and loss ownership using the income and loss items reported on schedule K-1s
   (a) For example, in a two partner partnership, if one partner receives $100 in income and the other receives $200 in income, we compute their profit percentages as 33% and 67% respectively.
   (b) We compute several measures of profit and loss based off combinations of Schedule K-1 line items to account for special allocations.

Inactive Partnerships

- Due to the ownership calculation requirements outlined above, we do not include partnerships with no income/loss activity in our sample. In future analysis, we will examine what portion of these partnership carry significant assets.

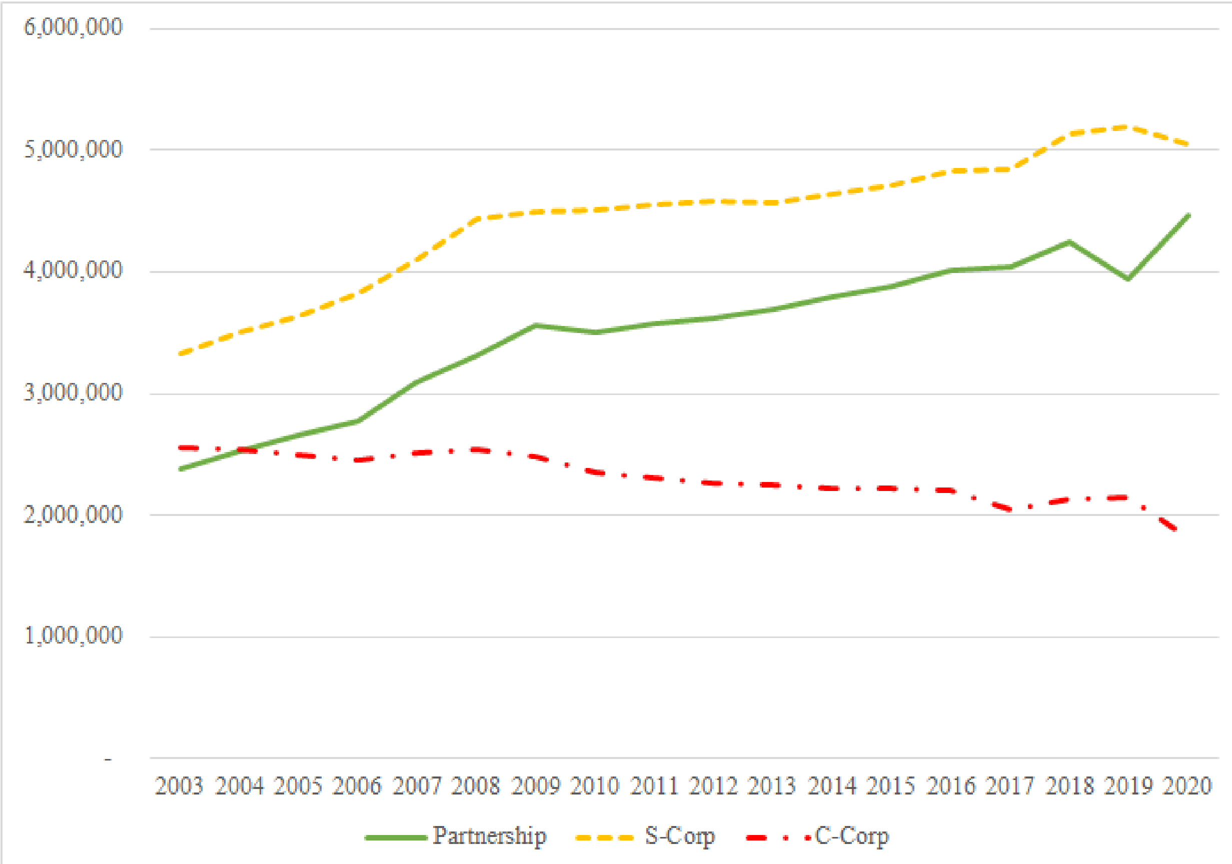
Publicly Traded Partnerships

- We do not currently analyze PTPs separately. We plan to identify and separately analyze/drop these partnerships in the future.

Disregarded Entities

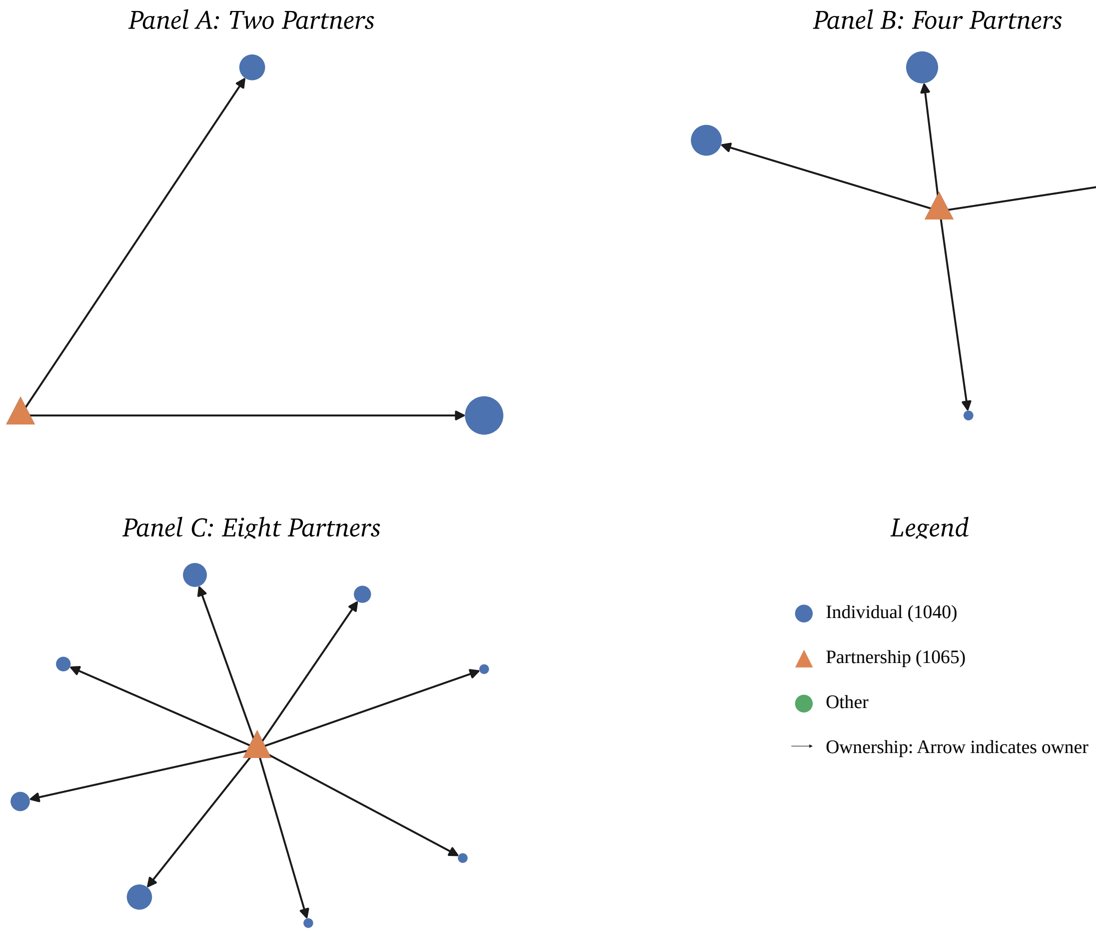
- There are partnerships that issue K-1s to disregarded entities. These partnerships represent an important population in identifying tax non-compliance. However, it is currently not possible to determine what portion of these K-1s are ultimately reported on a taxable entity's return.
- In future analysis, we plan to pull the data from individual Schedule E forms to estimate the portion of unreported disregarded entities.

Figure 1: Growth of Business Structures



Notes: This figure illustrates the number of business returns filed between 2003 and 2020. Partnerships do not include disregarded entities that are not required to file a form 1065 with the IRS. Over this time period, partnership filings have increased the most, both in terms of the number of returns and the rate of increase: partnership (S-corp) returns filed increased by 87.7% (51.5%) at an average annual rate of 3.9% (2.5%). C-corp returns filed decreased by 28.9% at an average annual rate of -1.9%.

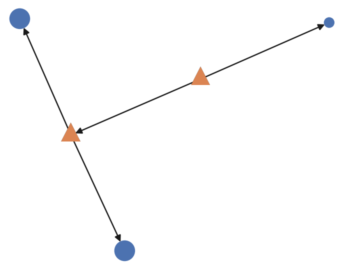
Figure 2: Examples of Simple Partnership Structures



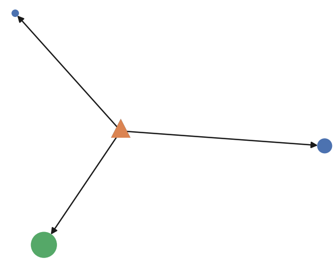
Notes: This figure illustrates partnership organization structures and depicts the relationship between the partnership and its partners. All panels show a simple partnership organization, which is a group composed of one partnership entity owned directly by individuals. Panel A presents a simple partnership organization with two owners; Panel B (C) present an organization with four (eight) partners. Blue circles denote individual partners; orange triangles denote the partnership entity; the arrow indicates ownership and points to the direct partner owner. The blue nodes are sized based on the proportion of the organization's total income reported to each partner.

Figure 3: Examples of Complex Partnership Structures

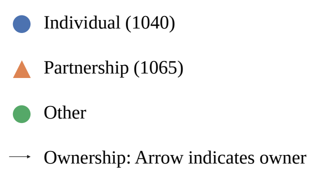
Panel A: Multiple Partnerships



Panel B: Multiple Entity Types

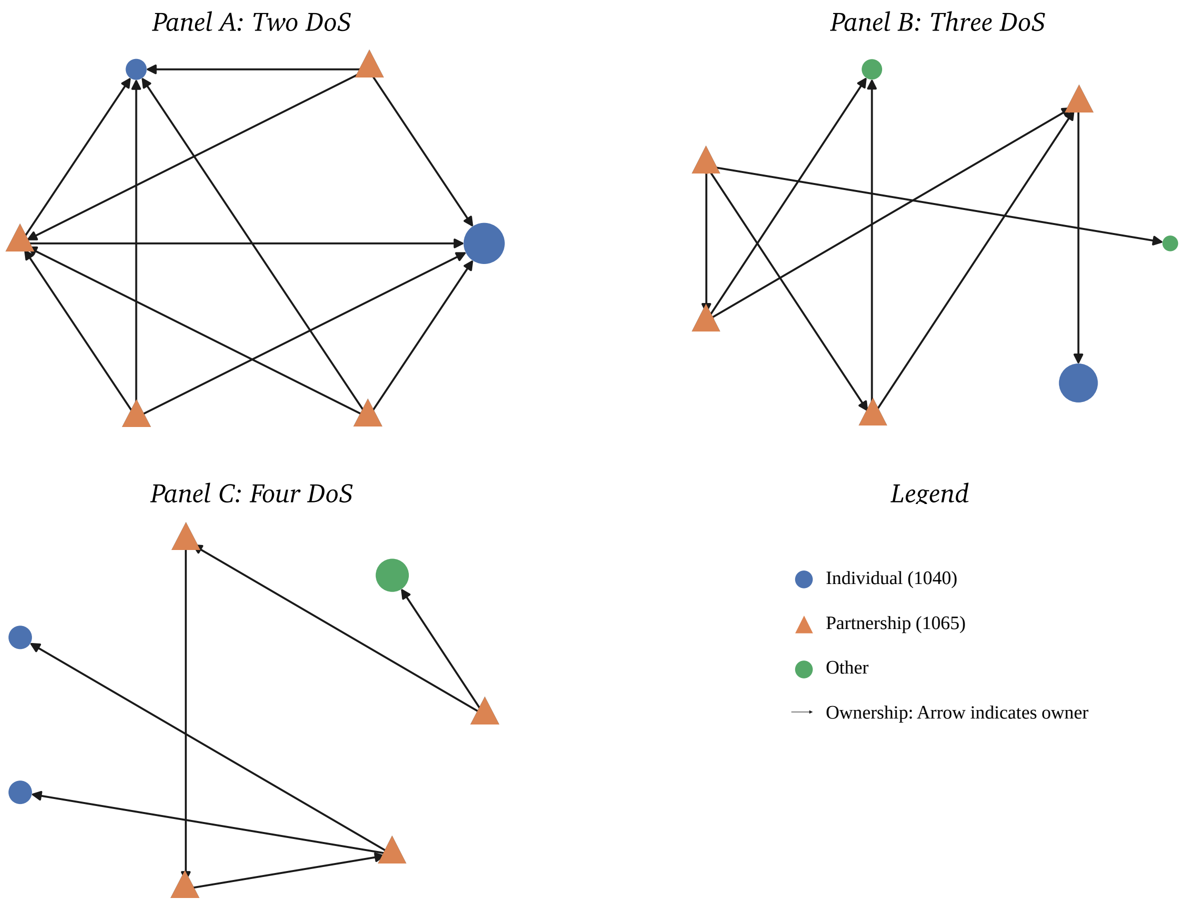


Legend



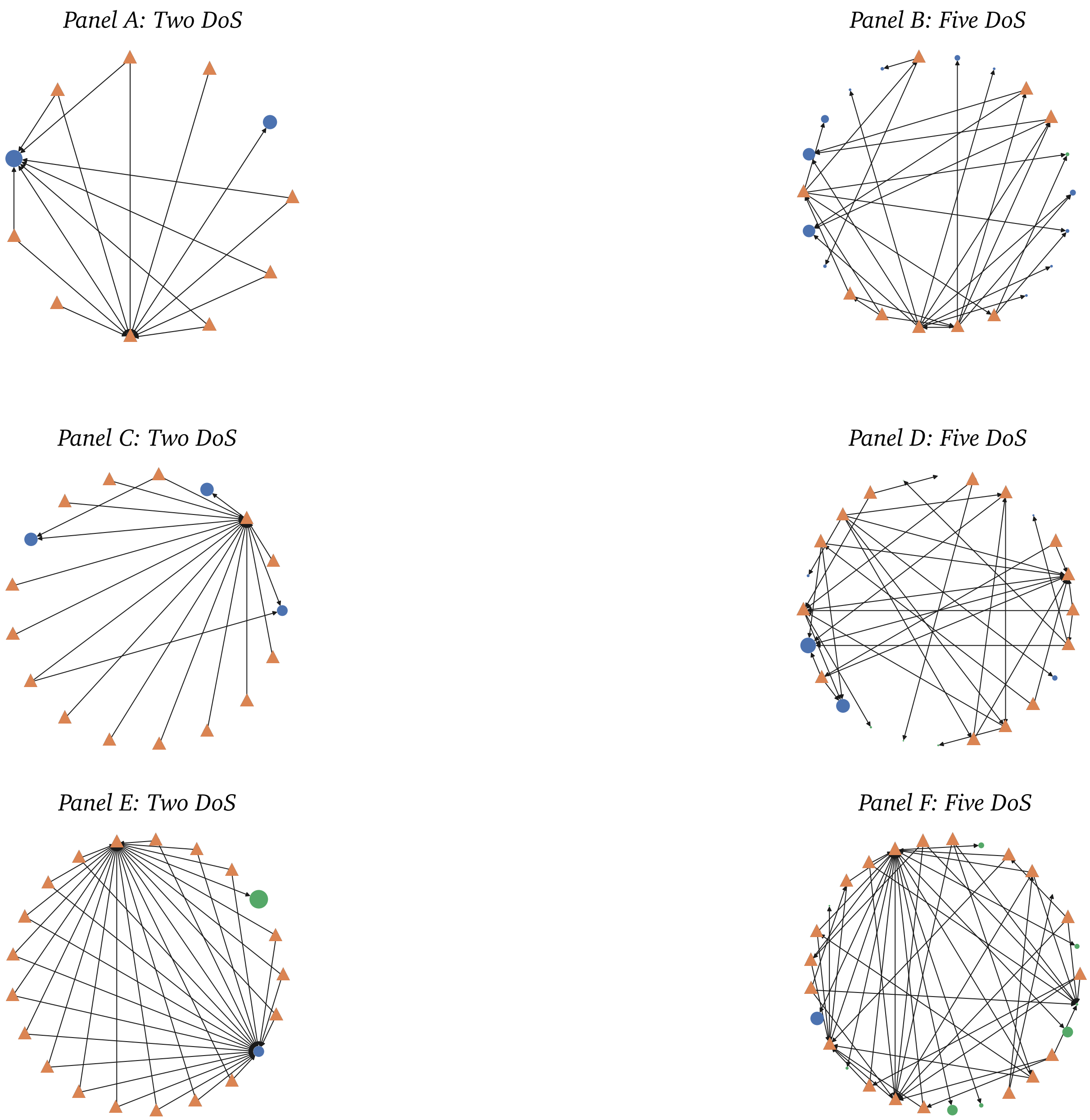
Notes: This figure illustrates partnership organization structures and depicts the relationship between the partnership and its partners. Both panels show a complex partnership organization, which is an organization (i) composed of multiple partnerships and/or (ii) directly owned by partners who are not individuals. Panel A presents a complex partnership organization composed of two partnerships with three distinct individual owners; Panel B presents an organization with one partnership organization directly owned by two individual owners and one entity. Blue circles denote individual partners; green circles indicate entity owners (C corporation, S corporation, and trusts); orange triangles denote the partnership entity. The arrow indicates ownership and points to the direct partner owner. The blue and green nodes are sized based on the proportion of the organization's total income reported to each partner.

Figure 4: Complex Partnership Structures with Contrasting Degrees of Separation



Notes: This figure illustrates partnership organization structures and depicts the relationship between the partnership and its partners, comparing organizations with differing Degrees of Separation. The panels show complex partnership organizations, which are groups (i) composed of multiple partnerships and/or (ii) directly owned by partners who are not individuals. Panel A presents a complex partnership organization with two degrees of separation; Panel B (C) denotes three (four) degrees of separation. Blue circles denote individual partners; green circles indicate entity owners (C corporation, S corporation, and trusts); orange triangles denote the partnership entity. The arrow indicates ownership and points to the direct partner owner. The blue and green nodes are sized based on the proportion of the organization's total income reported to each partner.

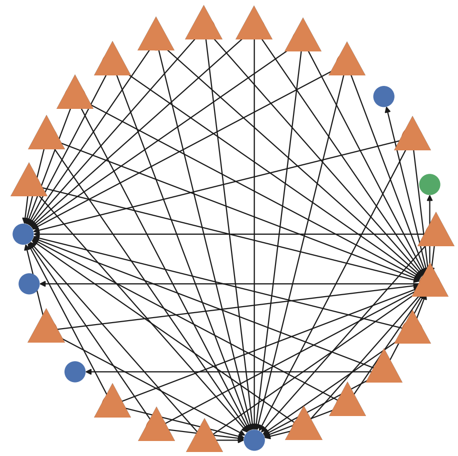
Figure 5: Complex Partnership Structures Comparing Degrees of Separation



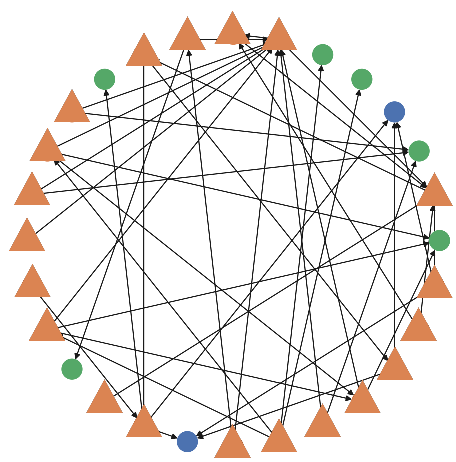
Notes: This figure illustrates partnership organization structures, comparing organizations with differing numbers of partnerships and degrees of separation. The panels show complex partnership organizations, which are groups (i) composed of multiple partnerships and/or (ii) directly owned by partners who are not individuals. Panels A and B present organizations with 9 partnerships; Panels C and D (E and F) present organizations with 14 (19) partnerships. Panels A, C, and E (B, D, and F) show organizations with two (five) degrees of separation. Blue circles denote individual partners; green circles indicate entity owners (C corporation, S corporation, and trusts); orange triangles denote the partnership entity. The arrow indicates ownership and points to the direct partner owner. The blue and green nodes are sized based on the proportion of the organization's total income reported to each partner.

Figure 6: Examples of Complex Partnership Structures with 20 Partnerships

Panel A: Two DoS



Panel B: 8 DoS



Notes: This figure illustrates partnership organization structures and depicts the relationship between the partnership and its partners. Both panels show a complex partnership organization, which is an organization (i) composed of multiple partnerships and/or (ii) directly owned by partners who are not individuals. Panel A presents a complex partnership organization composed of 20 partnerships with two Degrees of Separation ("DoS"); Panel B presents an organization composed of 20 partnerships with eight Degrees of Separation ("DoS"). Orange nodes denote the partnership entity; blue nodes denote individual partners; green nodes denote other entity types (C corporation, S corporation, or trust).

Figure 7: Random Forest Feature Importance

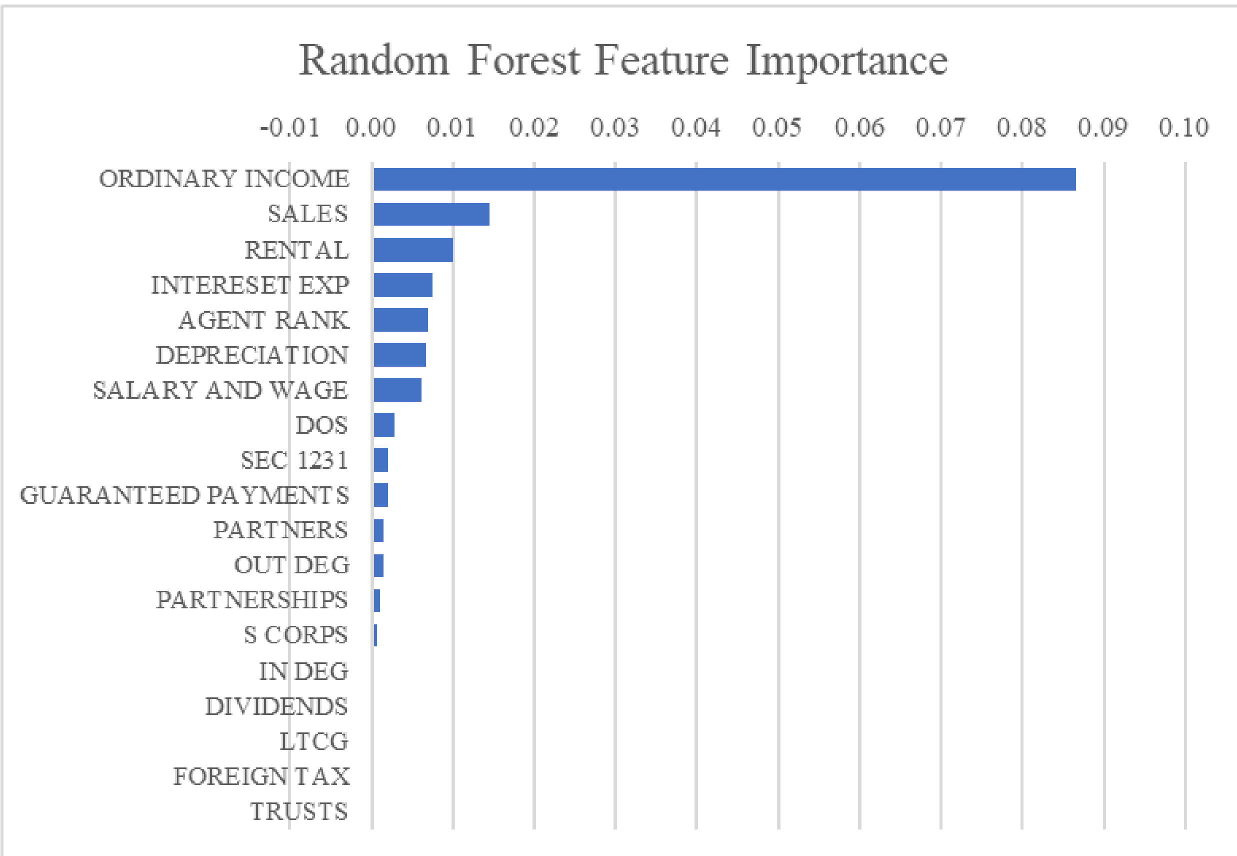
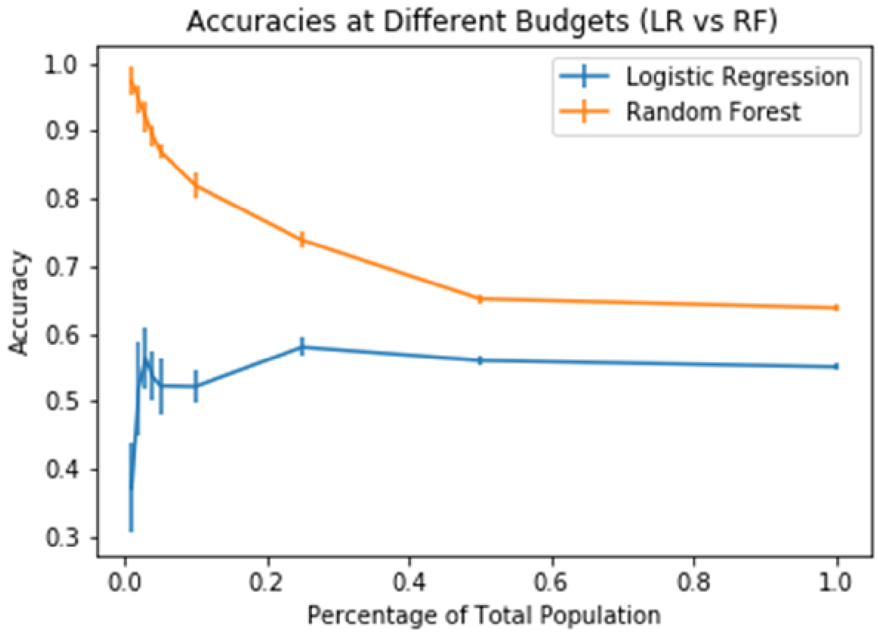


Figure 8: Machine Learning Graphs

Panel A



Panel B

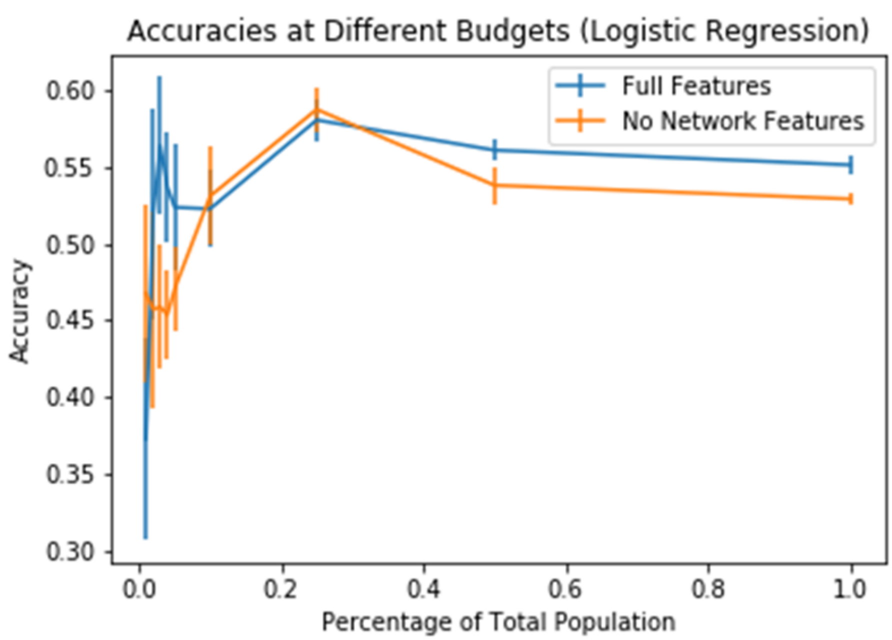
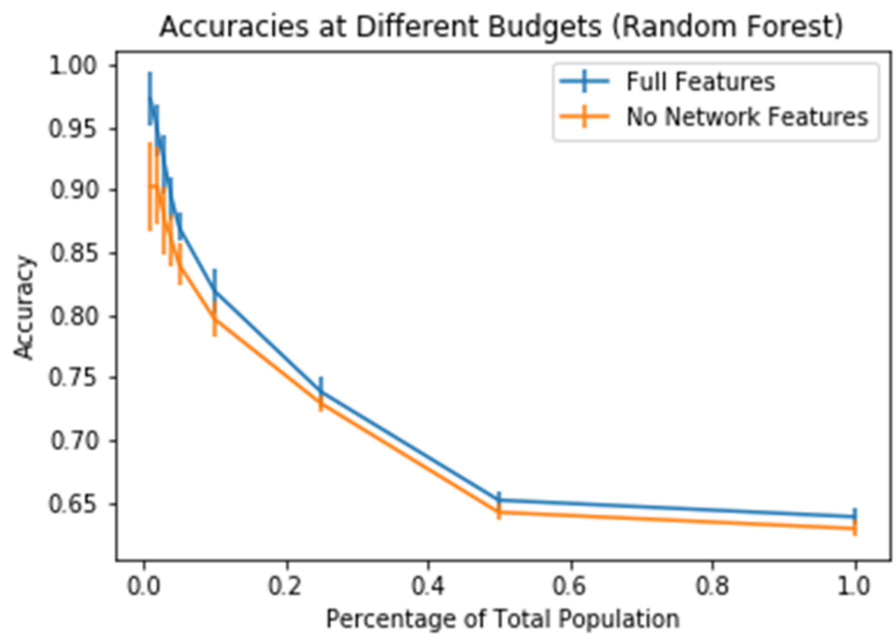


Figure 8: continued

Panel C



Panel D

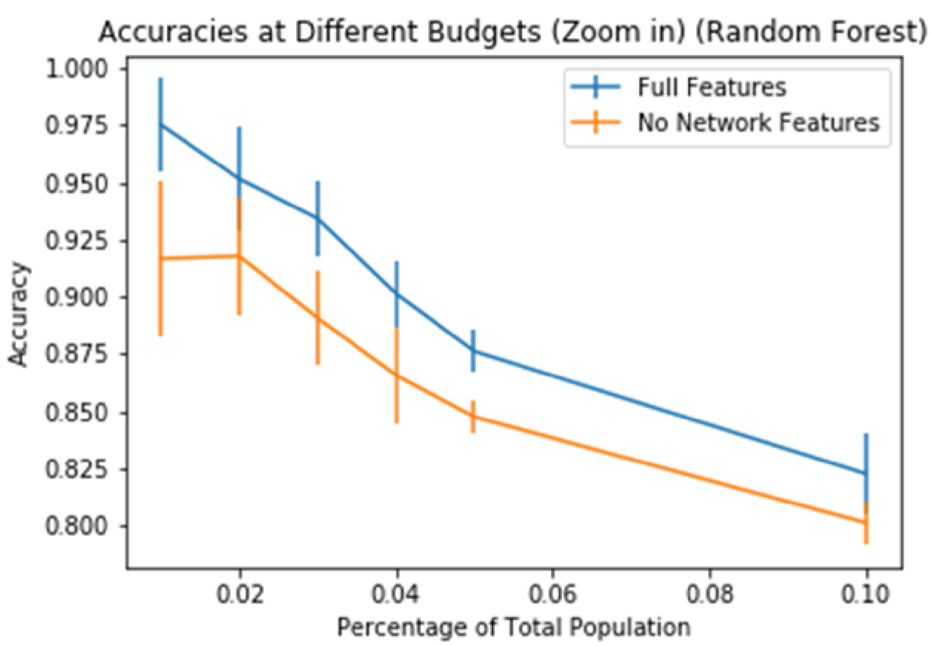


Table 1: Sample Composition

(a) *2013-2015 Sample Construction*

Population from IRS partnership database	10,915,532	
Drop: Partnerships with insufficient ownership data	3,875,554	
Including partnerships with no income activity		
All partnerships	7,039,978	
	Full Sample	Audited Sample
Partnerships in simple organizations	5,995,314	13,179
Partnerships in complex organizations	1,044,664	2,690
All partnerships	7,039,978	15,869

Sample consists of 7,039,978 partnership organization observations from 2013-2015 with sufficient data from the IRS for estimation. We separately identify our sample by simple and complex partnership organizations. Simple partnership organizations are single partnerships wholly owned by individuals; complex partnership organizations include all other structures.

(b) *Partnership Organizations*

Simple Partnership Organizations		Complex Partnership Organizations	
Number of Partners	Number of Organizations	Number of Partners	Number of Organizations
2	4,328,169	2	277,132
3	856,051	3	104,186
4	436,712	4	72,534
5	159,494	5	45,630
6	80,292	6	28,891
7	39,763	7	18,843
8	26,596	8	13,121
9	16,475	9	9,958
10	12,274	10	7,465
11-20	33,569	11-20	26,683
21-30	4,099	21-30	5,385
31-40	820	31-40	1,764
41-50	384	41-50	700
51+	616	51+	1,293

Sample of simple partnership organizations consists of 5,995,314 partnerships from 2013-2015 with sufficient data from the IRS for estimation. Sample of complex partnership organizations consists of 1,044,664 partnerships in 613,585 partnership organizations from 2013-2015 with sufficient data from the IRS for estimation.

Table 1: continued

(c) *Complex Partnership Organizations by Number of Partnerships*

Number of Partnerships in Organization	Number of Organizations
1	381,876
2	156,557
3	38,283
4	15,108
5	7,619
6	4,375
7	2,658
8	1,691
9	1,276
10	876
11-20	2,652
21-30	389
31-40	111
41-50	48
51+	66

Sample of complex partnership organizations consists of 613,585 partnership organizations from 2013-2015 with sufficient data from the IRS for estimation.

Table 2: Descriptive Statistics: Full Population

(a) *Industry Composition*

NAICS code	Industry	All Partnerships: # Observations (1)	All Partnerships: Percent of total (2)	Simple Partnerships: # Observations (3)	Simple Partnerships: Percent of total (4)	Complex Partnerships: # Observations (5)	Complex Partnerships: Percent of total (6)
11	Agriculture and Forestry	396,700	5.69	346,405	5.84	50,295	4.84
21	Mining; Oil and Gas	62,445	0.90	44,178	0.74	18,267	1.76
22	Utilities	7,013	0.10	4,837	0.08	2,176	0.21
23	Construction	346,736	4.97	310,638	5.24	36,098	3.48
31-33	Manufacturing	139,469	2.00	121,713	2.05	17,756	1.71
42	Wholesale Trade	141,708	2.03	123,101	2.08	18,607	1.79
44-45	Retail Trade	413,918	5.94	387,248	6.53	26,670	2.57
48-49	Transportation and Warehousing	113,413	1.63	103,500	1.74	9,913	0.95
51	Information	94,845	1.36	81,869	1.38	12,976	1.25
52	Finance and Insurance	328,052	4.71	231,601	3.90	96,451	9.29
53	Real Estate and Rental	3,321,109	47.65	2,775,118	46.79	545,991	52.56
54	Professional Services	540,411	7.75	474,972	8.01	65,439	6.30
55	Management of Companies	43,559	0.62	20,796	0.35	22,763	2.19
56	Administrative; Waste Mgmt.	126,502	1.81	110,649	1.87	15,853	1.53
61	Educational Services	33,883	0.49	31,746	0.54	2,137	0.21
62	Health Care	155,971	2.24	126,904	2.14	29,067	2.80
71	Arts and Entertainment	143,274	2.06	129,707	2.19	13,567	1.31
72	Accommodation and Food	267,342	3.84	231,786	3.91	35,556	3.42
81	Other Services	279,439	4.01	263,375	4.44	16,064	1.55
92	Public Administration	47	0.00	37	0.00	10	0.00
	Other	84,142	0.21	75,134	0.19	9,008	0.30

Sample consists of 7,039,978 partnership organization observations from 2013-2015 with sufficient data from the IRS for estimation. Industry is defined by two-digit NAICS codes. The ’other’ category includes partnerships whose NAICS codes were listed as zero. We separately identify our sample by simple and complex partnership organizations. Simple partnership organizations are single partnerships wholly owned by individuals; complex partnership organizations include all other structures.

Table 2: continued

(b) Income Distribution by Number of Taxable Partners

	Simple Partnership Organizations			Complex Partnership Organizations		
Number of Partners	Total Income (1)	Operating Income (2)	Investment Income (3)	Total Income (4)	Operating Income (5)	Investment Income (6)
2	62,045	36,037	26,008	114,163	48,990	42,055
3	46,566	28,252	18,314	118,238	61,741	46,083
4	220,203	30,773	189,429	140,744	58,878	54,033
5	78,576	48,590	29,986	132,624	56,607	57,620
6	96,712	61,217	35,495	152,119	106,551	72,873
7	116,615	78,355	38,260	168,390	94,337	70,974
8	110,700	73,282	37,418	184,142	75,062	69,277
9	150,272	105,550	44,722	187,783	98,377	75,336
10	167,934	133,777	34,157	176,222	43,160	80,275
11-20	266,038	224,814	41,224	231,679	116,598	83,774
21-30	654,098	604,239	49,859	325,976	200,985	109,532
31-40	1,485,823	1,407,736	78,087	326,745	210,861	96,956
41-50	2,895,670	2,781,206	114,464	599,024	384,998	131,697
51+	17,867,571	17,421,852	445,719	2,779,192	2,526,432	141,949

This table presents descriptive statistics for the sample of 6.6 million partnership organizations from 2013-2015 with sufficient data. Total Income is the amount of income reported on Form 1065, U.S. Return of Partnership Income, Sch. K, Lines 1-3 and 5-11. Operating Income includes amounts reported on Line 1; Investment Income includes amounts reported on Lines 2-3 and 5-11. Profitable Partnerships are those reporting total positive income. We separately calculate the amount of profit based on Line 1 ("Operating Income") and based on Lines 3 through 11 ("Investment Income"), excluding guaranteed payments. Amounts shown are averages for all partnerships (top line) and based on the number of taxable partners in the organization.

Table 2: continued

(c) *Descriptive Statistics*

	All Partnerships (n=7,039,978)		Simple Partnerships (n=5,995,314)		Complex Partnerships (n=1,044,664)	
Variable	Mean (1)	SD (2)	Mean (3)	SD (4)	Mean (5)	SD (6)
SALES	521,528.15	12,410,794.12	356,517.18	5,526,852.46	1,468,524.06	29,353,616.34
SALARY AND WAGE	65,176.96	1,418,507.09	47,772.55	778,806.65	165,060.65	3,172,900.55
GUARANTEED PAYMENTS	12,485.39	209,715.57	11,487.54	171,514.69	18,212.05	357,100.68
FOREIGN TAX	193.42	47,654.78	59.32	16,032.33	963.02	117,593.68
INTEREST EXP	5,500.79	246,745.58	3,335.45	61,059.14	17,927.66	623,469.98
DEPRECIATION	10,627.03	804,916.49	6,691.82	148,746.29	33,211.13	2,058,773.90
ORDINARY INCOME	36,036.64	2,063,895.04	27,402.82	1,263,180.50	85,586.03	4,421,047.40
RENTAL	13,478.11	703,516.35	10,484.35	101,017.46	30,659.28	1,810,097.35
DIVIDENDS	1,372.01	144,050.46	553.75	38,506.62	6,067.96	362,356.42
LTCG	4,724.67	506,269.19	2,729.03	271,686.96	16,177.58	1,141,704.56
SEC 1231	27,264.47	46,811,453.00	26,997.89	50,720,176.28	28,794.39	1,863,159.45
IN DEG	0.05	0.40	-	-	0.33	0.99
OUT DEG	2.72	2.60	2.57	1.96	3.58	4.75
PARTNERS	3.02	4.49	2.57	1.96	5.59	10.29
PARTNERSHIPS	1.42	4.46	1.00	-	3.86	11.28
S CORPS	0.08	0.88	-	-	0.56	2.22
TRUSTS	0.09	0.78	-	-	0.63	1.94
DoS	1.16	0.48	1.00	-	2.10	0.70

This table presents descriptive statistics for the sample of partnerships from 2013-2015 with sufficient data. Variable definitions are included in Appendix A. Amounts shown are averages for all partnerships and based on the number of taxable partners in the organization.

Table 2: continued

(d) *Partnerships by Total Sales*

	All Partnerships		Simple Partnerships		Complex Partnerships	
	# Observations (1)	Percent of total (2)	# Observations (3)	Percent of total (4)	# Observations (5)	Percent of total (6)
$0	4,123,650	58.57%	3,385,720	56.47%	737,930	70.64%
$1-$99,999	1,244,242	17.67%	1,163,447	19.41%	80,795	7.73%
$100,000-$499,999	888,120	12.62%	816,380	13.62%	71,740	6.87%
$500,000-$999,999	300,794	4.27%	262,498	4.38%	38,296	3.67%
$1,000,000-$4,999,999	376,068	5.34%	299,492	5.00%	76,576	7.33%
$5,000,000-$49,999,999	99,299	1.41%	64,754	1.08%	34,545	3.31%
>=$50,000,000	7,805	0.11%	3,023	0.05%	4,782	0.46%

This table presents the number of partnerships in our sample by total sales. We separately show partnerships in simple and complex organizations.

(e) *Partnerships by Schedule K income (excluding ordinary income)*

	All Partnerships		Simple Partnerships		Complex Partnerships	
	# Observations (1)	Percent of total (2)	# Observations (3)	Percent of total (4)	# Observations (5)	Percent of total (6)
$0	4,751,180	67.49%	4,137,696	69.02%	613,484	58.73%
$1-$99,999	1,828,715	25.98%	1,547,604	25.81%	281,111	26.91%
$100,000-$499,999	383,010	5.44%	268,145	4.47%	114,865	11.00%
$500,000-$999,999	45,948	0.65%	26,059	0.43%	19,889	1.90%
$1,000,000-$4,999,999	27,325	0.39%	14,235	0.24%	13,090	1.25%
$5,000,000-$49,999,999	3,682	0.05%	1,550	0.03%	2,132	0.20%
>=$50,000,000	118	0.00%	25	0.00%	93	0.01%

This table presents the number of partnerships in our sample by total Schedule K items included in our analyses (RENTAL, DIVIDENDS, LTCG, and, SEC 1231). We separately show partnerships in simple and complex organizations.

Table 3: Partnership: Audited Partnerships

(a) *Partnership Organizations: Audited Sample*

Simple Partnership Organizations		Complex Partnership Organizations	
Number of Partners	Number of Organizations	Number of Partners	Number of Organizations
2	11,233	2	1,271
3	1,240	3	425
4	438	4	271
5	138	5	161
6	46	6	125
7	23	7	84
8	21	8	47
9	*	9	54
10	*	10	43
11-20	27	11-20	140
21-30	*	21-30	39
31-40	0	31-40	14
41-50	0	41-50	*
51+	*	51+	*

Sample of simple partnership organizations consists of 13,179 audited partnerships from 2013-2015 with sufficient data from the IRS for estimation. Sample of complex partnership organizations consists of 2,690 audited partnerships in partnership organizations from 2013-2015 with sufficient data from the IRS for estimation. An * indicates the information is suppressed to conform with IRS disclosure requirements.

Table 3: continued

(b) *Complex Partnership Organizations by Number of Partnerships: Audited Sample*

Complex Partnership Organizations	
Number of Partnerships in Organization	Number of Organizations
1	1,415
2	464
3	341
4	137
5	98
6	49
7	39
8	20
9	20
10	23
11-20	67
21-30	*
31-40	*
41-50	*
51+	*

Sample of complex partnership organizations consists of 2,690 partnership organizations from 2013-2015 with sufficient data from the IRS for estimation. An * indicates the information is suppressed to conform with IRS disclosure requirements.

Table 4: Descriptive Statistics: Audited Population

(a) *Industry Composition*

		All Partnerships		Positive Adjustment		No Positive Adjustment	
NAICS code	Industry	# Observations (1)	Percent of total (2)	# Observations (3)	Percent of total (4)	# Observations (5)	Percent of total (6)
11	Agriculture and Forestry	2,225	14.11	942.00	11.68	1,283.00	16.66
21	Mining; Oil and Gas	101	0.64	43.00	0.53	58.00	0.75
22	Utilities	42	0.27	7.00	0.09	35.00	0.45
23	Construction	1,118	7.09	553.00	6.86	565.00	7.34
31-33	Manufacturing	796	5.05	289.00	3.58	507.00	6.58
42	Wholesale Trade	699	4.43	305.00	3.78	394.00	5.12
44-45	Retail Trade	1,291	8.19	729.00	9.04	562.00	7.30
48-49	Transportation and Warehousing	598	3.79	345.00	4.28	253.00	3.29
51	Information	292	1.85	109.00	1.35	183.00	2.38
52	Finance and Insurance	641	4.07	309.00	3.83	332.00	4.31
53	Real Estate and Rental and Leasing	3,060	19.41	1,540.00	19.10	1,520.00	19.74
54	Professional Services	1,285	8.15	753.00	9.34	532.00	6.91
55	Management of Companies	278	1.76	174.00	2.16	104.00	1.35
56	Administrative; Waste Mgmt. 365	2.32	241.00	2.99	124.00	1.61	
61	Educational Services	69	0.44	32.00	0.40	37.00	0.48
62	Health Care and Social Assistance	487	3.09	284.00	3.52	203.00	2.64
71	Arts, Entertainment, and Recreation	392	2.49	200.00	2.48	192.00	2.49
72	Accommodation and Food Services	1,240	7.87	706.00	8.75	534.00	6.93
81	Other Services	751	4.76	482.00	5.98	269.00	3.49
92	Public Administration	-	-	-	-	-	-
	Other	139	0.22	88.00	0.26	51.00	0.18

Sample consists of 15,869 audited partnership observations from 2013-2015 with sufficient data from the IRS for estimation. Industry is defined by two-digit NAICS codes. The 'other' category includes partnerships whose NAICS codes were listed as zero. We separately identify our sample by simple (8,131 observations) and complex partnerships (7,738 observations). Simple partnership organizations are single partnerships wholly owned by individuals; complex partnership organizations include all other structures.

Table 4: continued

(b) Income Distribution by Number of Taxable Partners

	Simple Partnership Organizations			Complex Partnership Organizations		
Number of Partners	Total Income (1)	Operating Income (2)	Investment Income (3)	Total Income (4)	Operating Income (5)	Investment Income (6)
2	(180,220)	(181,176)	956	(1,211,583)	(984,157)	194,294
3	(331,755)	(306,838)	(24,917)	1,697,446	1,220,759	70,998
4	(98,182)	(77,733)	(20,450)	(566,116)	(359,296)	(139,628)
5	(3,455,782)	(309,617)	(3,146,165)	118,299	36,643	104,303
6	399,585	223,828	175,757	188,660	20,622	246,385
7	(527,451)	(604,630)	77,179	992,300	790,293	(78,403)
8	714,225	(273,931)	988,156	(51,282)	(125,947)	114,200
9	*	*	*	(1,668,611)	(1,640,903)	294,930
10	*	*	*	(639,488)	(651,238)	100,933
11-20	(390,755)	(301,994)	(88,761)	2,598,438	995,231	805,052
21-30	*	*	*	1,092,330	598,853	157,841
31-40	-	-	-	(3,628,899)	(402,256)	(2,422,132)
41-50	-	-	-	*	*	*
51+	*	*	*	*	*	*

Sample consists of 15,869 audited partnership observations from 2013-2015 with sufficient data from the IRS for estimation. Total Income is the amount of income reported on Form 1065, U.S. Return of Partnership Income, Sch. K, Lines 1-3 and 5-11. Operating Income includes amounts reported on Line 1; Investment Income includes amounts reported on Lines 2-3 and 5-11. Profitable Partnerships are those reporting total positive income. We separately calculate the amount of profit based on Line 1 ("Operating Income") and based on Lines 3 through 11 ("Investment Income"), excluding guaranteed payments. Amounts shown are averages for all partnerships (top line) and based on the number of taxable partners in the organization. An * indicates the information is suppressed to conform with IRS disclosure requirements.

Table 4: continued

(c) *Descriptive Statistics*

Variable	Audited Partnerships Mean (1)	Audited Partnerships SD (2)	Positive Adjustment Mean (3)	Positive Adjustment SD (4)	No Positive Adjustment Mean (5)	No Positive Adjustment SD (6)
SALES	3,885,171.73	85,008,350.83	2,830,639.07	35,257,130.03	4,993,262.32	116,237,955.63
SALARY AND WAGE	397,284.53	6,290,611.41	316,323.20	3,756,325.96	482,357.75	8,143,257.12
GUARANTEED PAYMENTS	34,407.65	862,621.87	19,249.15	261,380.62	50,336.02	1,205,711.64
FOREIGN TAX	943.92	77,909.68	309.11	16,097.59	1,610.96	110,340.18
INTEREST EXP	48,482.10	725,269.90	47,502.01	903,903.04	49,511.97	469,264.78
DEPRECIATION	233,930.46	8,614,348.84	273,556.65	11,454,646.32	192,291.73	3,782,416.48
ORDINARY INCOME	(202,814.42)	18,121,856.70	(233,900.28)	18,747,717.12	(170,149.76)	17,439,967.02
RENTAL	(5,915.61)	677,944.19	(998.16)	484,122.08	(11,082.81)	834,403.74
DIVIDENDS	3,861.70	201,954.65	4,103.59	268,566.20	3,607.52	88,609.34
LTCG	15,431.49	2,030,368.04	(15,540.55)	2,085,994.17	47,976.55	1,969,700.62
SEC 1231	110,117.79	5,340,627.64	35,234.23	872,983.72	188,804.56	7,594,755.87
AGENT RANK	12.12	0.98	12.17	0.92	12.08	1.03
IN DEG	0.11	0.61	0.10	0.56	0.12	0.67
OUT DEG	2.29	2.18	2.19	1.09	2.39	2.91
PARTNERS	2.64	3.71	2.48	3.09	2.81	4.26
PARTNERSHIPS	1.41	2.13	1.35	2.13	1.47	2.13
S CORPS	0.08	0.61	0.07	0.69	0.10	0.52
TRUSTS	0.05	0.54	0.03	0.33	0.07	0.70
DOS	1.19	0.53	1.15	0.46	1.24	0.59

This table presents descriptive statistics for the sample of audited partnerships from 2013-2015 with sufficient data. Variable definitions are included in Appendix A. Amounts shown are averages for all partnerships and based on the number of taxable partners in the organization.

Table 4: continued

(d) *Partnerships by Total Sales*

	All Partnerships		Positive Adjustment		No Positive Adjustment	
	# Observations (1)	Percent of total (2)	# Observations (3)	Percent of total (4)	# Observations (5)	Percent of total (6)
$0	6,170	38.88%	2,910	35.79%	3,260	42.13%
$1-$99,999	2,416	15.22%	1,429	17.57%	987	12.76%
$100,000-$499,999	2,654	16.72%	1,679	20.65%	975	12.60%
$500,000-$999,999	1,222	7.70%	668	8.22%	554	7.16%
$1,000,000-$4,999,999	2,162	13.62%	984	12.10%	1,178	15.22%
$5,000,000-$49,999,999	1,102	6.94%	394	4.85%	708	9.15%
>=$50,000,000	143	0.90%	67	0.82%	76	0.98%

This table presents the number of partnerships in our sample by total sales. We separately show partnerships in simple and complex organizations.

(e) *Partnerships by Schedule K income (excluding ordinary income)*

	All Partnerships		Positive Adjustment		No Positive Adjustment	
	# Observations (1)	Percent of total (2)	# Observations (3)	Percent of total (4)	# Observations (5)	Percent of total (6)
$0	13,410	84.50%	7,073	86.99%	6,337	81.89%
$1-$99,999	1,408	8.87%	617	7.59%	791	10.22%
$100,000-$499,999	648	4.08%	296	3.64%	352	4.55%
$500,000-$999,999	172	1.08%	67	0.82%	105	1.36%
$1,000,000-$4,999,999	163	1.03%	57	0.70%	106	1.37%
$5,000,000-$49,999,999	*	*	*	*	*	*
>=$50,000,000	*	*	*	*	*	*

This table presents the number of audited partnerships in our sample by total Schedule K items included in our analyses (RENTAL, DIVIDENDS, LTCG, and, SEC 1231). We separately show partnerships in simple and complex organizations. An * indicates the information is suppressed to conform with IRS disclosure requirements.

Table 5: Determinants of Audit Adjustment

	(1)		(2)	
	OLS		OLS	
	ADJUSTt		ADJUSTt	
	coefficient	standard error	coefficient	standard error
SALES	-0.0142	0.0070	-0.0158	0.0070
SALARY AND WAGE	-0.0050	0.0060	0.0001	0.0060
GUARANTEED PAYMENTS	-0.0083	0.0040	-0.0016	0.0040
FOREIGN TAX	-0.0044	0.0040	-0.0040	0.0040
INTEREST EXP	-0.0002	0.0040	0.0024	0.0040
DEPRECIATION	0.0112	0.0060	0.0118	0.0060
ORDINARY INCOME	0.0142	0.0070	0.0137	0.0070
RENTAL	0.0014	0.0040	0.0019	0.0040
DIVIDENDS	-0.0075	0.0060	-0.0038	0.0060
LTCG	-0.0139	0.0060	-0.0107	0.0060
SEC 1231	-0.0075	0.0040	-0.0060	0.0040
AGENT RANK	0.0237	0.0040	0.0193	0.0040
IN DEG			0.0119	0.0050
OUT DEG			-0.0230	0.0080
PARTNERS			0.0021	0.0060
PARTNERSHIPS			0.0058	0.0100
S CORPS			-0.0005	0.0050
TRUSTS			-0.0071	0.0050
DOS			-0.0457	0.0060
Observations	15,869		15,869	
Adjusted R2	0.003		0.011	

This table presents the prediction models for audit adjustment using an OLS model. Standard errors are reported in the column next to the coefficients. All variables are defined in Appendix A.

Table 6: Determinants of Audit Adjustment

	(1)		(2)	
	Logistic		Random Forest	
	ADJUST [0/1]t	ADJUST [0/1]t	ADJUST [0/1]t	ADJUST [0/1]t
	Coefficients		Feature Importance	
SALES	-0.1479	-0.1301	0.0140	0.0145
SALARY AND WAGE	0.0156	0.0365	0.0015	0.0061
GUARANTEED PAYMENTS	-0.1320	-0.0696	0.0038	0.0020
FOREIGN TAX	-0.3012	-0.2474	0.0132	0.0000
INTEREST EXP	0.0077	0.0177	0.0105	0.0076
DEPRECIATION	0.0935	0.0918	0.0057	0.0068
ORDINARY INCOME	0.0867	0.0764	0.0148	0.0865
RENTAL	0.0341	0.0370	0.0127	0.0100
DIVIDENDS	-0.0894	-0.0487	0.0164	0.0002
LTCG	-0.2116	-0.1629	0.0148	0.0001
SEC 1231	-0.1678	-0.1143	0.0114	0.0020
AGENT RANK	0.0980	0.0813	0.0144	0.0070
IN DEG		0.0435		0.0003
OUT DEG		-0.1594		0.0014
PARTNERS		0.0105		0.0015
PARTNERSHIPS		0.1005		0.0011
S CORPS		0.0170		0.0007
TRUSTS		-0.0205		0.0000
DOS		-0.2034		0.0027
Training set observations	11,901	11,901	11,901	11,901
Test set observations	3,968	3,968	3,968	3,968
K-fold cross validation training	7,934	7,934	7,934	7,934
K-fold cross validation test	3,967	3,967	3,967	3,967
Training Accuracy at 10% budget	0.546	0.535	0.988	0.997
Test Accuracy at 10% budget	0.531	0.523	0.797	0.819
Generalization Error	0.015	0.013	0.191	0.177
Model F1 score at 10% budget	0.693	0.686	0.887	0.900

This table presents the prediction models for audit adjustment. We present a logistic Ridge regression and a Random Forest model. We use a 75%/25% split for our training and test sets. Partnership features have been standardized. We report model accuracy for the training and test set.