

Tax Evasion and Money Laundering: Insights from Textual and Contextual Analysis

Bea Chiang

Department of Accounting & Information Systems,
School of Business, The College of New Jersey, USA

[Doi: 10.60154/jaepp.2026.v27n2p27](https://doi.org/10.60154/jaepp.2026.v27n2p27)

Submitted: 17 January 2026

Accepted: 26 July 2026

Published: 20 August 2026

Copyright 2026 Author(s)

Under [Creative Commons CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)

OPEN ACCESS

Cite As:

Chiang, B. (2026). *Tax Evasion and Money Laundering: Insights from Textual and Contextual Analysis*. Journal of Accounting, Ethics & Public Policy, JAEPP, 27 (2), 27.

<https://doi.org/10.60154/jaepp.2026.v27n2p27>

Abstract

This study examines whether tax evasion operates as a predicate offense for money laundering by integrating an economic model of concealment with an analysis of 4,534 IRS Criminal Investigation (IRS-CI) prosecutions from 2021 through mid-2025. It addresses the gap between the Financial Action Task Force (FATF) framework, which treats serious tax crimes as predicate offenses for money laundering, and US law, which often treats underreported taxes as noncriminal. Building on the Allingham-Sandmo (1972) model and recent work by Kemsley and Morgan (2022, 2024), the study introduces a convex concealment cost function and distinguishes standalone tax evasion from sequential evasion-and-laundering behaviors.

Based on an analysis of 4,535 IRS CI prosecution news releases, cases were classified into three categories: tax evasion charges, money laundering charges, and joint prosecutions involving both offenses. The results show that 2,727 cases, approximately 60 percent, involve both tax evasion and money laundering charges. LIFT, Pointwise Mutual Information (PMI), Normalized Pointwise Mutual Information (NPMI), and Fisher's Exact Test revealed links between specific forms of tax evasion, such as corporate pass-through schemes, payroll tax violations, and false returns, and distinct laundering mechanisms, including wire transfers, shell entities, bulk cash movements, real estate acquisitions, and cryptocurrency transactions. Overall, these findings suggest that current U.S. enforcement practice commonly treats tax evasion and money laundering as separate but interconnected offenses, supporting the view that tax evasion serves as a predicate offense in a two-step process that aligns more closely with the FATF approach than with models that equate tax evasion directly with money laundering. Collectively, this study bridges the gap between economic theory and enforcement practice. By demonstrating that U.S. prosecutors treat tax evasion and money laundering as distinct, sequential offenses, the findings refine the single-stage models proposed by Kemsley and Morgan (2022, 2024) and validate the FATF two-stage framework. Furthermore, by identifying these behaviors at the subtype level, this study provides practical benchmarks to support Anti-Money-Laundering (AML) detection and strengthen cross-border enforcement.

Keywords: Tax Evasion, Money Laundering, FATF, Expected Utility Function

Introduction

Tax evasion has long been a major fraud issue. It involves illegal income tax savings, often resulting from underreported taxes, and may or may not include schemes to conceal those savings. Money laundering, on the other hand, involves channeling illegal proceeds generated from criminal activities. In 2012, the Financial Action Task Force (FATF) added tax evasion to its list of designated predicate offenses for money laundering.

U.S. law, however, follows a traditional approach that treats tax evasion differently from money laundering. Under U.S. law, money laundering requires the source of the funds to originate from a criminal act, such as fraud or illegal arms sales. Conversely, in a tax evasion offense, the original source of income might be entirely legal, stemming from regular transactions, but later manipulated, such as by shifting profits offshore. Because U.S. courts often do not perceive unpaid taxes as "criminal proceeds," they do not normally treat tax evasion as money laundering. This position is inconsistent with the FATF framework, which designates tax evasion as a predicate offense to money laundering. The dividing line between tax evasion and money laundering can be highly blurred. Even so, treating tax evasion as money laundering can also cause practical problems. It risks mistakenly prosecuting simple, unintentional underreporting as a criminal money laundering act (Bridges and Green, 1999).

In reality, both offenses can look remarkably similar, as both involve hiding financial sources to avoid legal consequences. Yet, U.S. law and the FATF maintain inconsistent legal standings. This conflict means that because U.S. law and FATF rules do not agree on when tax evasion crosses the line into money laundering, law enforcement officials and prosecutors face confusion over when to charge someone with money laundering. In the U.S., simply hiding tax savings is not always enough to trigger a money laundering charge, whereas the FATF framework suggests it is. This inconsistency also creates hurdles for international investigations and information sharing. Countries strictly adhering to the FATF framework might treat a case as both tax evasion and money laundering, but the U.S. recognizes it solely as tax evasion. Because different definitions dictate different investigative protocols and penalties, these discrepancies can directly impact international legal cooperation and sentencing severity.

Building on this discussion, this paper analyzes IRS Criminal Investigation (IRS-CI) press releases from 2021 to June 2025 to examine fraud and money laundering cases and draw insights for the FATF framework on tax evasion. Existing literature discusses the theoretical and regulatory inconsistencies between U.S. law and the FATF guidelines. However, a gap remains in empirical assessments of how these boundaries are navigated in

practice. This study addresses this gap by shifting the focus from abstract legal doctrine to concrete enforcement actions.

Specifically, it investigates how IRS-CI prosecutions during this period clarify whether tax evasion functions as a predicate offense or as a distinct stage relative to money laundering. By analyzing these recent enforcement actions, this paper contributes to the literature in three ways. First, it maps how U.S. prosecutors operationalize the blurred boundary between legal tax avoidance, illegal tax evasion, and money laundering. Second, it offers empirical evidence to inform the ongoing international policy debate regarding the harmonization of FATF standards and domestic legal frameworks. Finally, by categorizing the specific mechanisms used to conceal tax savings versus illegal proceeds, this study helps clarify the financial crime lifecycle.

Literature Review

Financial Action Task Force (FATF) Recommendations vs U.S. Money Laundering Statutes

The Financial Action Task Force (FATF) is an independent intergovernmental organization that establishes and promotes international policies aimed at safeguarding the global financial system from money laundering, terrorist financing, and the financing of weapons of mass destruction. Its Recommendations serve as the internationally recognized benchmark for anti-money laundering (AML) and counter terrorist financing (CTF) frameworks (<https://www.fatf-gafi.org>). The organization recommends that countries apply money laundering laws to all serious offenses. It explicitly identifies tax crimes related to both direct and indirect taxes as designated predicate offenses for money laundering in the Recommendations and the accompanying glossary (FATF Recommendation, updated October 2025, pp. 12, 129).

The inclusion of tax crimes as predicate offenses evolved through the revision of the international standards. During the 2010 public consultation on the updates, the FATF proposed expanding the designated categories of predicate offenses to explicitly include tax crimes related to direct and indirect taxes (FATF, 2010). This proposal was subsequently adopted in the 2012 Recommendations, which require countries to criminalize money laundering based on a broad range of predicate offenses (FATF, 2012, Recommendation 3). The accompanying glossary expressly identifies tax crimes related to direct and indirect taxes among the designated categories of predicate offenses, establishing tax crimes as a recognized predicate offense under the international anti-money laundering framework (FATF, 2012). This shift in policy is also reflected in notes issued by the International Monetary Fund (IMF) on updates to the FATF standards, which officially include tax evasion

as a predicate offense. Furthermore, it appears in FATF guidance and typology reports describing the laundering of funds from serious tax crimes (Egmont Group of Financial Intelligence Units, 2020; International Monetary Fund, 2012).

In contrast to the FATF explicit designation, tax evasion under Title 26 of the Internal Revenue Code is not a direct predicate offense for money laundering under primary U.S. money laundering statutes, specifically 18 U.S.C. Section 1956 and Section 1957. To secure a money laundering conviction under 18 U.S.C. Section 1956, the government must prove that the financial transaction involved the proceeds of a “specified unlawful activity” (SUA). Under U.S. law, money laundering must involve illegal proceeds typically earned through SUA with certain degrees of intent. Tax evasion, conversely, does not require money to be obtained through unlawful activities. One key difference between these two offenses is the source of the funds in question. Money laundering must involve money from a specified unlawful activity, whereas tax evasion does not.

Historically, the U.S. has not treated tax evasion itself as a predicate offense for money laundering. Instead, prosecutors have often relied on fraud offenses such as wire fraud or shell company schemes when tax evasion also involves fraudulent activities. To avoid overreach and ensure that routine tax disputes or offenses are not automatically escalated to money laundering charges, the Department of Justice maintains strict authorization guidelines. As stated in Tax Division Directive Number 128:

"The Tax Division will not authorize the use of mail, wire or bank fraud charges to convert routine tax prosecutions into RICO (Racketeer Influenced and Corrupt Organizations Act) or money laundering cases. The Tax Division will authorize prosecution of tax-related RICO and money laundering offenses, however, when unusual circumstances warrant it."

Contrasting the distinct approaches of the FATF and U.S. law toward financial crimes, Kemsley, Kemsley and Morgan (2022) develop a conceptual and economic framework. They argue that tax evasion itself is a composite form of money laundering because the act of evasion simultaneously generates criminal tax savings and conceals their unlawful origin under the FATF definition. Tax evasion does not fit the traditional definition of a standard predicate offense, which requires two separate conditions: a serious crime and a distinct laundering act. The authors suggest that the act of tax evasion constitutes using an illegal scheme for illegal tax benefits, which is the serious crime, and does not require the second condition of a separate laundering step. This is because the tax savings are already concealed as an unlawful benefit the moment a taxpayer files a false tax return. Their later article (Kemsley, Kemsley and Morgan, 2024) on the tax evasion of lawful income extends this

reasoning, contending that evasion on legitimate income still meets the conditions for money laundering once tax crimes are recognized as predicates.

On the other hand, earlier legal work warns against collapsing these boundaries. Bridges and Green (1999) analyze laundering and tax crimes, noting that merging the two raises concerns about overcriminalization, double punishment, and the requirement of a distinct laundering step in some legal systems. Maugeri (2018) also argues that the inclusion of tax evasion among the predicate offenses of money laundering raises significant rule of law concerns, particularly regarding whether proceeds from tax evasion can be considered launderable criminal profits. She highlights challenges related to the principles of double jeopardy and proportionality when criminalizing the self-laundering of tax savings.

Behrendt (2020) explains and compares how tax evasion is treated in German and U.S. law. In Germany, tax evasion is explicitly listed as a predicate offense under the criminal code, including situations involving saved tax expenditures and undeclared assets. In contrast, U.S. law typically does not label tax evasion directly as a predicate crime for money laundering. Instead, U.S. prosecutors often reframe tax evasion schemes as related offenses, such as mail or wire fraud, to fit within the money laundering statutes.

Economic Models

Economic and enforcement related research uses expected utility models and enforcement data to understand how offenders choose between "evasion only" and "evasion plus laundering" strategies and how authorities respond. Allingham Sandmo (1972) developed a classic expected utility model of tax evasion that treats the taxpayer as choosing how much income to report under the risk of audits and fines. Building on this model of tax evasion under risk, Kemsley et al. introduce concealment costs and laundering choices into the utility function. This demonstrates when additional layering, wiring, or the use of shell entities becomes rational for taxpayers seeking to protect unlawful tax savings. Empirical work on IRS criminal investigation activities and IMF analysis of Anti Money Laundering (AML) shows that leveraging AML tools, including reporting obligations, financial intelligence, and freezing powers, can strengthen tax enforcement, enhance the detection of complex tax crimes, and improve deterrence. Financial Intelligence Units (FIU) typology reports further indicate that in practice, tax related money laundering often involves real estate purchases, luxury goods, shell companies, and crypto transactions.

Money laundering, according to the FATF definition, requires two distinct conditions. First, it requires a predicate offense that produces unlawful proceeds. Second, it requires a separate or identifiable act to conceal or disguise the origins of those unlawful proceeds. Kemsley, Kemsley and

Morgan (2022) use an economic model to support the proposition that tax evasion itself is money laundering. They argue that treating tax evasion as a standard predicate offense as defined by the FATF can be misleading and creates legal confusion. In their analysis of the Anderson case, they examine whether the defendant's use of a complex scheme to shift profits to a tax haven could, by itself, constitute money laundering (United States v. Anderson, 2005).

In Kemsley, Kemsley and Morgan's study (2022), the utility function of money laundering and concealment costs is expressed as follows (p. 597):

$$E[U] = qU(y-f_0) + (1-q)pzU(y-f_0-f_{ml}-tc) + (1-q)p(1-z)U(y-f_{ml}-tc) + (1-q)(1-p)U(y-tc)$$

Where:

y is the unlawful proceeds from a crime predicate to money laundering;

f₀ is the punishment for the predicate crime;

q is the probability of being detected and punished for the predicate crime;

tc is the cost or effort required to launder the unlawful proceeds;

f_{ml} is the punishment for money laundering;

p is the probability of being detected and punished for money laundering; and

z is the conditional probability that if money laundering is detected and punished, the predicate crime will then also be detected and punished.

The authors reconciled the above model to the following tax evasion model:

$$E[U] = (1-p)U(W-\theta X-tc) + pU(W-\theta X-\pi(W-X-tc))$$

Where:

W is the true taxable income

X is the amount of taxable income to declare

θ is the tax rate faced by the decision maker

π accounts for the normal required tax rate θ plus any penalty for undeclared taxable income

Two condition for money laundering as defined in the FATF's framework:

Condition 1 (Unlawful Predicate Proceeds): $u(W-X) > 0$

Condition 2 (Effort to Conceal or Disguise Unlawful Proceeds): $tc > 0$

The authors prove that tax evasion generates unlawful proceeds through unpaid taxes and fundamentally requires an effort to conceal or disguise those savings from authorities. Because underreporting income is impossible without a strategic effort to deceive, evading taxes always involves both illegal retention and concealment. Consequently, the concealment cost (t_c) is always positive, which satisfies the second condition of money laundering. The authors therefore conclude that tax evasion inherently constitutes money laundering because it satisfies both core conditions of the offense. In other words, tax evasion itself is money laundering.

Concealment Costs and Modified Model

The assumption of concealment costs (t_c) in the utility function proposed by Kemsley, Kemsley and Morgan is open to question for two primary reasons.

First, underreported taxes can stem from honest, incidental errors, such as erroneously claiming a tax credit rather than committing intentional evasion. If the resulting tax savings are simply absorbed into everyday household spending without any motivation or deliberate action to hide them, concealment costs are effectively zero ($t_c=0$). Second, even in scenarios of intentional underreporting, the decision to launder those illicit savings depends heavily on expected costs, which are unlikely to scale linearly as implied by Kemsley, Kemsley and Morgan's model.

For intentional tax evasion, Kemsley, Kemsley and Morgan model the evasion and concealment process as a linear function. In reality, however, hiding unlawful proceeds often requires significant effort and becomes increasingly complex as the volume of unlawful proceeds grows. Larger amounts typically require more diverse channels and highly sophisticated schemes to conceal. Consequently, these concealment costs are nonlinear rather than the constant t_c modeled in the original utility function. Thus, the concealment cost t_c can be modified to $c(W-X)$, where $W-X$ represents the amount of undeclared income. The concealment function is increasing and convex ($c' > 0$, $c'' > 0$) to reflect the rising marginal difficulty of concealment. The modified utility function is presented below:

$$E[U] = p(W-X, c(W-X), E) * U(W_{\text{caught}}) + (1-p)(W-X, c(W-X), E) * U(W_{\text{free}})$$

$$W_{\text{caught}} = W - \theta X - \pi(W-X) - c(W-X)$$

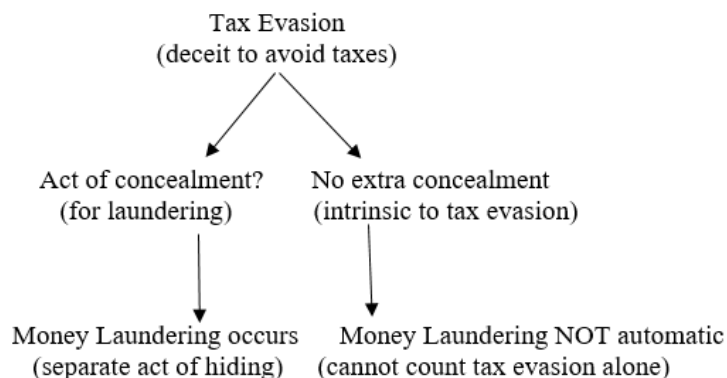
$$W_{\text{free}} = W - \theta X - c(W-X)$$

Where E is enforcement efforts, θ is the tax rate and π is penalty on undeclared taxable income

If tax evasion does not inherently constitute money laundering, such as in cases of failing to file payroll taxes, the expected utility framework in the study by Kemsley, Kemsley and Morgan (2022) can be modified by adjusting the structure of the necessary conditions. Logically, tax evasion is a necessary condition for money laundering because without unlawful proceeds, there is no money to launder. However, tax evasion itself is not a sufficient condition for money laundering because not every act of underreported taxes leads to money laundering.

Tax evasion generally involves the underpayment of taxes through unlawful savings or the misrepresentation or concealment of income within the taxpayer's own reporting system. For example, undeclared taxable income may be used to pay imminent household bills rather than laundered through additional transactions. In such cases, tax evasion alone does not constitute a sufficient condition for money laundering. Concealment costs may therefore be nonlinear, and the intent to launder may be entirely absent, as seen in failures to declare cash income.

The key question is whether the identifiable or separable act of concealment that defines money laundering is conceptually distinct from the deceit inherent in tax evasion. If the concealment is intrinsic rather than distinct, tax evasion cannot automatically satisfy the laundering requirement. Legally, the issue is whether the hiding required for money laundering differs from the deception involved in tax evasion. If concealment is merely part of the evasion itself, then tax evasion does not necessarily imply that money laundering has occurred, as U.S. law requires a distinct concealment step beyond the act of evasion.



Kemsley, Kemsley and Morgan's composite model assumes that tax evasion simultaneously generates and launders illicit proceeds. However, one can argue that individuals who underreport taxable income do not engage in money laundering when the underreporting stems from unintentional filing mistakes. Furthermore, even in intentional cases, individuals may have a

limited incentive to launder the resulting tax savings because the additional costs and risks associated with concealing those funds can outweigh the potential benefits. For unintentional incidents of underreported tax liabilities, the motive to launder the unreported income simply does not exist.

Within the context of intentional tax evasion, the Kemsley, Kemsley and Morgan framework can be modified to decouple the choice to evade from the subsequent choice to launder. By modeling tax evasion and money laundering as two distinct, sequential decisions, we can capture the unique incentives at each stage. Under this modified framework, the expected utility for an individual deciding whether to evade taxes and whether to subsequently launder the resulting proceeds is specified as follows:

$$E[U] = qU(y-f_e-f_l-t_c) + (1-q)U(y-t_c)$$

Where:

y= total post-tax income (compliance)

f_e= penalty if evasion is detected

f_l= penalty if laundering is detected (only if a laundering process is added)

t_c= concealment cost

q= probability of detection

Scenario 1- Tax Evasion Only

$$U_E = (1-p)U(W - \theta X - t_c) + pU(W - \theta X - f_e - t_c)$$

Where W-X is the amount of undeclared income and θ is tax rate.

Proposition 1: No laundering without offsetting benefit

Under risk neutrality, if the laundering stage only introduces positive concealment costs ($t_{ml} > 0$) and additional penalty ($f_l \geq 0$) without reducing the expected tax evasion sanction or detection probability, then $U_{E+L} < U_E$. Consequently, a rational agent will optimally choose tax evasion without laundering. This relationship is illustrated in Figure 1, which demonstrates that the expected utility function of tax evasion combined with money laundering lies entirely below the utility function of tax evasion alone ($U_{E+L} < U_E$).

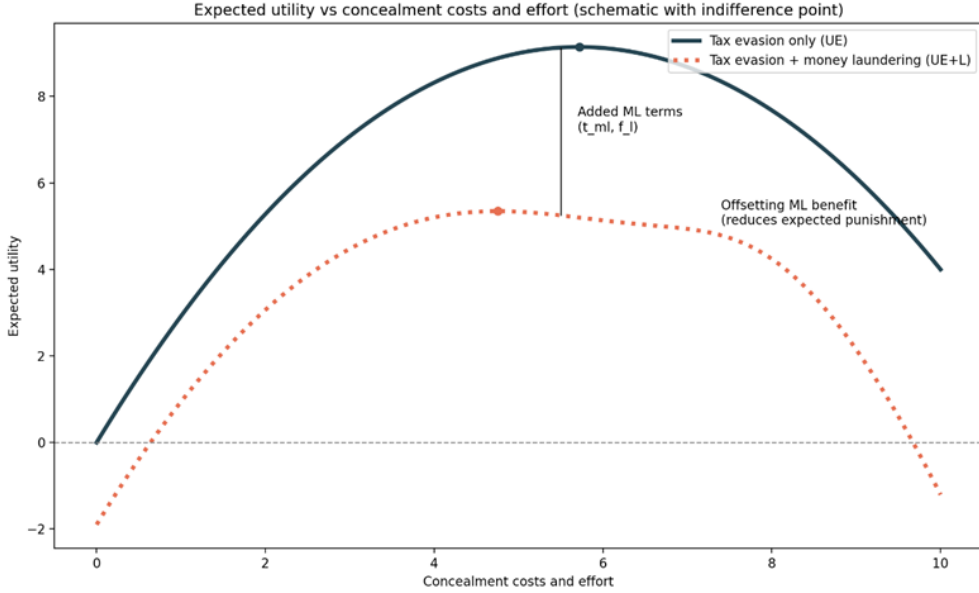


Figure 1: Expected Utilities and Concealment Costs (Tax Evasion only)

Scenario 2- Tax Evasion and Money Laundering

$$U_{E+L} = (1-p)U(W - \theta X - t_c - t_{ml}) + pU(W - \theta X - f_e - f_l - t_c - t_{ml})$$

Proposition 2: When laundering is attractive

Laundering occurs only when it is sufficiently justified by a reduction in the expected punishment for evasion. Specifically, laundering becomes a rational decision only if it reduces the expected sanction from evasion, for example, by lowering the probability of detection, p or the penalty, f_e .

Assume that t_{ml} represents the distinct incremental concealment efforts related to money laundering. when $t_{ml} > 0$, laundering involves distinct efforts and costs. If this cost does not reduce the probability of being caught, it functions purely as an extra expense. A rational agent will not invest in these unnecessary extra efforts without expecting a corresponding payoff. Laundering only makes sense when the fine from evasion is greater than the penalty from laundering ($f_e > f_l$), or when laundering reduces subsequent penalties. Otherwise, laundering is not an inherently required step.

In other words, tax evasion does not automatically trigger money laundering. The conditions for optimal concealment do not necessarily require a separate laundering process, nor is the amount of unpaid tax always significant enough to justify additional concealment efforts. Money laundering occurs only if there is an additional incentive to proceed with this step, such as avoiding regulatory scrutiny or escaping a higher penalty. Thus, tax evasion does not automatically constitute money laundering.

Condition under which a rational agent chooses to launder:

$$U_{E+L} \geq U_E$$

Figure 2 demonstrates the expected utility for scenarios of tax evasion only and tax evasion combined with money laundering. Both curves trend upward until a certain peak and then turn downward. This downward turn occurs because the concealment cost function $c(W-X)$ is increasing and convex, meaning each additional unit of hidden income is progressively more difficult and expensive to conceal. The evasion-only curve lies above the tax evasion plus laundering curve at most effort levels, since adding a distinct laundering stage introduces incremental concealment costs t_{ml} and potential laundering penalties f_l .

The second curve is lower because money laundering involves distinct efforts and costs, t_{ml} . Consequently, a rational agent will not launder unless the expected utility exceeds these added efforts, which occurs where the dotted line crosses the solid line. The concave shape of both utility curves reflects the diminishing returns that occur as the convex concealment costs increase ($c' > 0$, $c'' \leq 0$).

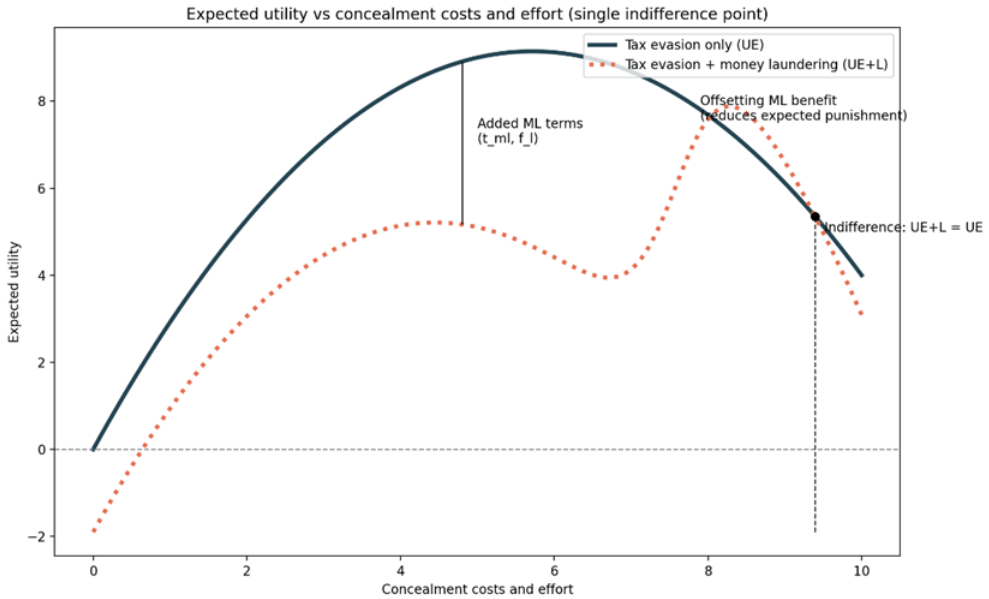


Figure 2: Expected Utilities and Level of Concealment Efforts (TE and ML, when ML is attractive)

Methodology and Analysis

To evaluate the theoretical propositions established in the modified utility framework, this section details an analysis of U.S. legal prosecutions using Internal Revenue Service Criminal Investigation (IRS CI) press releases.

The financial crime prosecutions are classified into three mutually exclusive categories: tax evasion only, money laundering only, or concurrent offenses. By mapping these real-world sentencing outcomes against our conceptual models of optimal concealment and risk-reward behavior, this study aims to derive insights into the sentencing patterns for each offense and explore how the complex legal system structures these charges.

IRS Criminal Investigation Case Analysis

To construct the dataset, tax fraud and financial crime cases from 2021 to June 2025 were compiled from the official IRS Criminal Investigation (IRS-CI) press releases archive. A custom Python script was chosen for web scraping and data preparation due to Python's strengths in rapid prototyping and its extensive data science library ecosystem; however, its technical limitations were carefully considered during the analytical process. As an interpreted, dynamically typed language, Python exhibits slower execution speeds, higher memory consumption, and potential runtime errors compared to compiled languages (Gavrilova, 2023). To mitigate these performance and security challenges during the text extraction phase, memory-optimized data structures and vectorized operations within the Pandas library were implemented to ensure efficient, stable processing of the IRS-CI corpus. The resulting unstructured textual data was cleaned, structured, and aggregated into a comprehensive corpus. This final dataset contains structured variables for the publication date, pre-processed full body text, and source URL.

Tax Evasion Cases

Cases were categorized into three mutually exclusive groups using a rule-based classification framework applied to the case narratives and offense indicators. Each case was assessed across two binary variables: the presence of a tax evasion-related offense and the presence of a money laundering-related offense.

A tax evasion indicator was assigned when the case title, narrative, or extracted offense subtype identified conduct such as tax evasion, filing false tax returns, failure to file, payroll or employment tax violations, unreported income, improper deductions, or other forms of federal tax fraud. A money laundering indicator was assigned when the case included evidence of laundering-related activities, such as money laundering charges, structuring or other violations of the Bank Secrecy Act, bulk cash smuggling, funnel account transactions, the use of shell or nominee entities, unlicensed money transmitting or hawala operations, cryptocurrency-based concealment, or other methods used to disguise the proceeds of criminal activity. Using these two indicators, cases were classified into three groups: (1) records with tax evasion indicators but no money laundering indicators, (2) records with money

laundering indicators but no tax evasion indicators, and (3) records containing both tax evasion and money laundering indicators.

A total of 5,123 prosecution cases were extracted from the website. Out of these 5,123 cases, 4,534 cases were identified as involving either money laundering or tax evasion. Using Python, the cases were then sorted into money laundering only, tax evasion only, and both tax evasion and money laundering. A list of keywords used for classification is provided in the Appendix.

A second validation step was applied to the tax evasion only (TE only) group to ensure that it contained only cases involving tax evasion without evidence of money laundering charges. During this review, each TE only case was examined for laundering-related indicators in the case narrative. This included explicit references to money laundering, laundering, or structuring, as well as broader concealment mechanisms such as shell companies, nominee entities, front companies, cryptocurrency transactions, offshore accounts, funnel accounts, money mules, hawala networks, and Bank Secrecy Act or anti-money laundering (BSA AML) related language. Cases containing any of these indicators were reclassified from the TE-only group to the combined tax evasion and money laundering category. This additional validation step improved classification accuracy by identifying cases in which the narrative indicated a distinct money laundering stage that was not captured by the initial binary coding process. As a third validation step, a manual spot check reading was conducted to verify that cases were accurately classified into the three groups.

After the validation process, the final dataset contains 631 cases charged with tax evasion only, 1,176 cases charged with money laundering only, and 2,727 cases charged with both tax evasion and money laundering.

As illustrated in Table 1, tax evasion prosecutions without concurrent money laundering charges are heavily clustered within two primary categories: False Return False Deductions (251 cases) and Payroll Employment Tax (189 cases). Combined, these two categories account for approximately 70 percent (440 out of 631) of the entire sample. The median fraud volume for these dominant categories resides in the low single-digit millions (\$2.0 million and \$3.0 million, respectively). Conversely, highly specialized and digitally complex subtypes, such as Crypto Virtual Currency (\$33.0 million) and Structuring CTR Evasion (\$22.5 million), exhibit dramatically larger median fraud scales despite their lower overall case counts. Judicial sentencing severity varies substantially across categories; the most severe prison sentences are concentrated in Preparer Driven violations (155.25 months) and Crypto Virtual Currency crimes (133.08 months), whereas Structuring and Failure to File offenses yield the lowest average incarceration

times (35 to 36 months). Figure 3 summarizes the tax evasion by crime types, respective median penalty amount, and average judicial sentence.

This case distribution supports the theoretical framework's proposition regarding sequential decision making and non-linear concealment costs. In mainstream, high-frequency categories like Payroll Employment Tax or Unreported Cash Income, the illicit proceeds are often immediately absorbed into routine operational expenses or everyday consumer spending. Because the intent to hide remains intrinsic to the tax filing omission itself, the incremental concealment cost (t_{ml}) of establishing an independent laundering scheme is absent, resulting in a rational choice of Tax Evasion Only ($U_{E+L} < U_E$).

In contrast, the median fraud scale observed in Crypto/Virtual Currency (\$33.0 million) highlights where the model's convex concealment cost function ($c'' > 0$) originates. Safely absorbing or utilizing tens of millions of dollars in virtual assets introduces compounding regulatory and tracing difficulties. When the scale of the fraud is this massive, the taxpayer faces an inflection point: they must either accept an exceptionally high probability of detection (p) due to the sheer size of the undeclared wealth, or incur the separate, distinct operational costs of money laundering to actively reduce that detection risk. The presence of these massive dollar amounts in the Evasion Only category suggests that these specific actors chose to accept the heightened detection risk of standalone evasion rather than navigate or pay for a secondary laundering infrastructure.

Table 1: Tax Evasion Only Cases, Fraud Amount and Sentence

Crime types (Tax Evasion Only Charges)	Number of cases	Median fraud amount (thousand)	Avg sentence (months)
False Return/False Deductions	251	\$ 2,000	75.09
Payroll/Employment Tax	189	\$ 3,000	82.05
Unreported Income/Cash Business	46	\$ 810	68.62
General - Other Specific	42	\$ 2,268	99.85
Failure to File	32	\$ 1,000	36.08
Corporate/Pass-Through Schemes	22	\$ 1,375	86.4
Preparer Driven	17	\$ 600	155.25
Crypto / Virtual Currency	15	\$ 33,000	133.08
False Deductions / Credits	11	\$ 13,000	65.25
Structuring / CTR Evasion	6	\$ 22,500	35
Total	631		

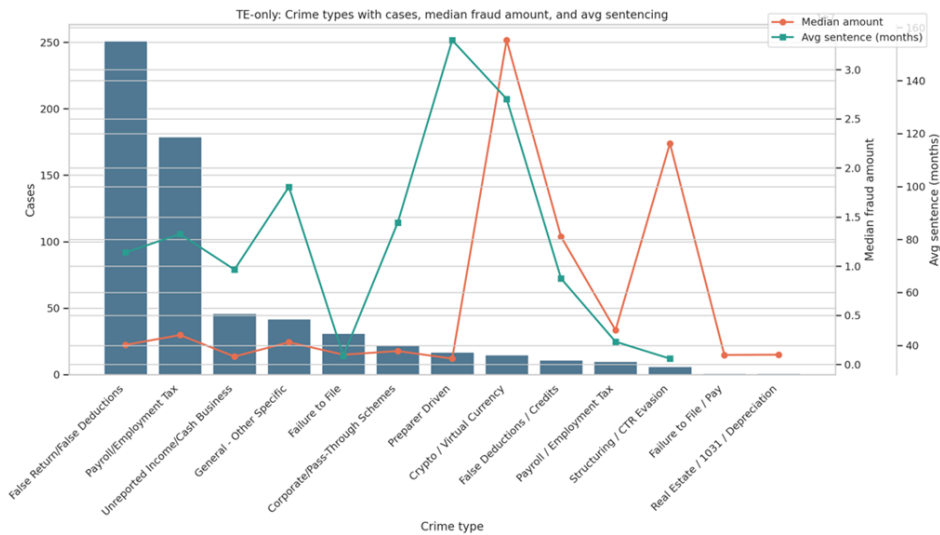


Figure 3: Tax Evasion by Crime Types

Money Laundering Cases

There is a total of 1,176 money laundering cases with various predicate offenses, excluding tax evasion. General money laundering cases were classified using a rule-based text analysis approach. Each case title and text case description were converted to a preprocessed text format in lowercase and evaluated for keywords associated with categories such as drug trafficking, wire fraud, COVID-19 relief loan fraud including the Paycheck Protection Program (PPP), Economic Injury Disaster Loans (EIDL), and Unemployment Insurance (UI), securities and investment fraud, and health care fraud and social engineering schemes including business email compromise (BEC), romance scams and phishing.

Because individual cases could involve more than one predicate offense, each case was assigned one primary classification using a predefined hierarchy based on substantive specificity. Specific predicate-offense categories were prioritized over broad charging statutes. For example, cases involving PPP or EIDL fraud were classified as COVID-19 relief loan fraud rather than wire fraud, even when wire fraud was also charged. Similarly, cases involving drug proceeds were classified as drugs or narcotics when drug trafficking was explicitly identified. All additional applicable offenses were retained as secondary labels. Using this approach, drugs or narcotics, wire fraud, and COVID-19 relief loan fraud accounted for approximately 75% of the 1,176 money-laundering cases.

Interestingly, a standard median sentence of 240 months, or 20 years, applies consistently across almost all high-volume categories, including health care and securities fraud, despite varying financial impacts. Conversely, categories like corruption or bribery and gambling represent the lower thresholds of the data, carrying minimal financial impacts and the shortest median jail sentences. The summary of crime types and their visualization is presented in Table 2 and Figure 4. The size of each bubble in Figure 4 represents the number of cases associated with a particular predicate offense that leads to money laundering.

Table 2: Table 1 Money Laundering Only Cases, Fraud Amount and Sentence

Crime Types	Number of Cases	Median fraud amount (thousand)	Median sentence (month)
Drugs/Narcotics	505	10,000	240
Wire fraud	198	5,500	240
COVID/relief loan fraud (PPP/EIDL/UI)	180	20,000	240
Securities/Investment fraud	76	42,534	240
Social engineering (BEC/Romance/Phishing)	68	31,500	240
Health-care fraud (Medicare/Medicaid/Durable Medical Equipment- DME)	43	48,000	240
Other generic fraud	38	29,000	106.5
Mail fraud	16	20,392	204
Wildlife/Environmental	11	8,946	240
Bank fraud	10	2,728	121
Benefits (non-UI)	5	52,000	72
Identity theft	4	6,000	162
Charity/Nonprofit	3	300,000	582
Corruption/Bribery	3	10	24
Cybercrime	3	250	36
Elder fraud	3	3,000	48
Human Trafficking/Smuggling	3	77,000	137
Procurement/Contracting	3	20,300	240
Gambling	2	1,900	12
Counterfeit goods	1	750	24
Mortgage/Real estate	1	66,000	126
total	1176		

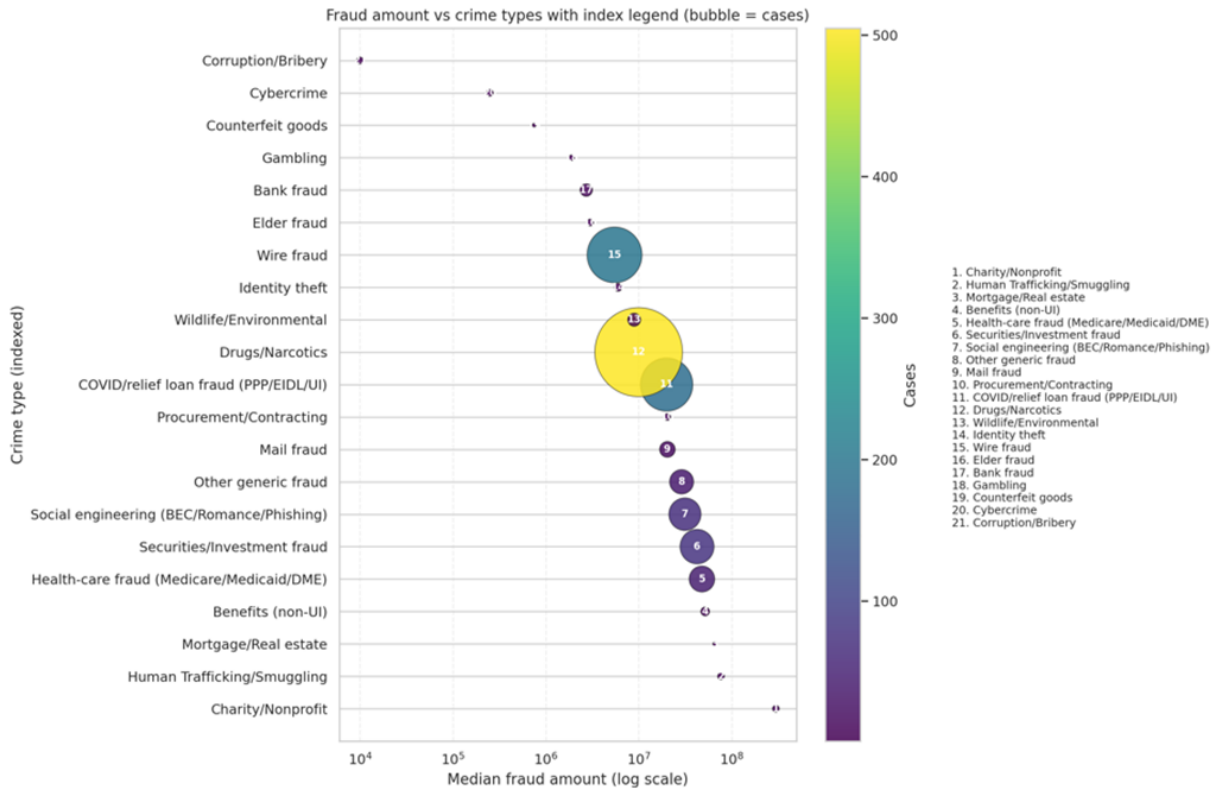


Figure 4: Money Laundering- Crime Types and Fraud Amount

Tax Evasion and Money Laundering Cases

A total of 2,272 tax evasion and money laundering cases were analyzed. To evaluate the tax evasion that associates with different forms of money laundering, LIFT and pointwise mutual information (PMI) scores are used to measure the strength of the relationship for each different combination of tax evasion and money laundering pairing. A value of 1 for the LIFT measure indicates independence. Values greater than 1 indicate that two events co-occur more often than expected by chance, which represents a positive association, and values below 1 indicate a negative association.

PMI is the log-scaled format of LIFT, with positive values indicating positive association and negative values indicating negative association. Pointwise Mutual Information (PMI) is a measure of association that quantifies how much more frequently two events co-occur than would be expected under statistical independence (Church and Hanks, 1990). Because PMI emphasizes unexpected co-occurrences rather than simple frequency, it has been commonly applied in text mining, natural language processing, and network analysis to identify meaningful relationships among entities.

However, a well-known limitation of PMI is its bias toward rare events. Infrequent event pairs can receive disproportionately high PMI values even when supported by only a few observations, making the measure sensitive to sparse data and potentially overstating weak or spurious associations (Church and Hanks, 1990; Damani, 2013). To address this limitation, Bouma (2009) proposed Normalized Pointwise Mutual Information (NPMI), which rescales PMI to a bounded range between -1 and 1, improving interpretability and reducing the influence of rare event inflation.

Researchers have also emphasized that measures of association should be evaluated together with statistical significance rather than relying solely on effect size. Damani (2013) suggested that NPMI should be complemented with significance testing to distinguish meaningful associations from chance co-occurrences. Accordingly, in addition to NPMI, this study also includes Fisher's Exact Test to evaluate the significance of each tax evasion and money laundering subtype pairing.

Since numerous subtype pairs are tested simultaneously, p-values are adjusted using the Benjamini-Hochberg False Discovery Rate (FDR) procedure to control for multiple comparisons (Benjamini and Hochberg, 1995). In addition, minimum support thresholds based on observed and expected frequencies are applied to reduce unstable estimates arising from sparse data. The final association analysis combines LIFT, NPMI, Fisher's Exact Test, and FDR-adjusted p values to identify statistically significant and substantively meaningful subtype relationships, which mitigates the rare event bias associated with raw PMI. Table 3 summarizes the top 15 pairings based on NPMI and the related Fisher's p value and FDR-adjusted p values. Figure 5 visualizes the top 15 pairings based on NPMI. Detailed descriptions of LIFT, PMI, NPMI, Fisher's exact test, and the FDR-adjusted p-value are provided in Appendix 2.

The analysis identifies several statistically significant positive pairings after Benjamini–Hochberg False Discovery Rate (BH FDR) correction, indicating that these co-occurrences are highly unlikely to be due to chance. The strongest relative associations are other or broadly defined entity-related tax-evasion conduct (*entity_misuse_misc*) paired with unspecified or general money-laundering activity (*general:general_misc*; LIFT = 6.20) and the willful failure to file required tax returns (*failure_to_file*) paired with unspecified laundering (*general:general_misc*; LIFT = 3.65). However, both findings are based on a relatively small number of pairs (e.g., 10 cases for the former) and involve a residual laundering category (*general:general_misc*) where no specific mechanism was described in the record, necessitating a cautious interpretation.

In contrast, a highly robust and specific finding is the pairing of false tax returns, fabricated deductions, or improper tax credits

(*false_return/false_deductions_credits*) with laundering enabled by financial, business, professional, or intermediary services (*services_enabled_general*). This combination exhibits substantial volume (64 pairs), a strong LIFT of 2.43 (measuring how much more frequently a pair occurs than expected under independence), and an extremely small adjusted q -value ($q < 0.0001$). Additionally, corporate or pass-through entity tax schemes (*corporate/pass-through_schemes*) paired with the electronic movement of funds via wire or bank-account transfers (*wires_transfers*) are statistically significant and represent the largest volume in the dataset (1,132 pairs). However, its modest LIFT (1.13) and low Normalized Pointwise Mutual Information (NPMI = 0.138) suggest a broad, common operational baseline rather than a highly distinctive typological signature. In contrast, pairings with a LIFT near or below 1.00, such as payroll or employment-tax evasion (*payroll_tax_evasion/employment_tax*) paired with wire transfers (*wires_transfers*; $q = 0.9162$) or with professional intermediary services (*services_enabled_general*) show no statistically significant evidence of association, indicating their co-occurrence is not meaningfully different from what would be expected by chance. (Detailed definitions of these tax-evasion and money-laundering typology pairings are provided in Appendix 3)

Table 3: Top 15 TE x ML paring with LIFT and PMI Measures

Pairing	N of Pairs	LIFT	PMI	NPMI	Two-sided Fisher p-value	BH FDR q-value
entity misuse misc x general:general misc	10	6.2	2.63	0.325	<0.0001	<0.0001
failure to file x general misc	27	3.65	1.87	0.28	<0.0001	<0.0001
unreported income/cash business x general misc	19	2.44	1.29	0.18	0.0002	0.0009
false return/false deductions credits x services enabled general	64	2.43	1.28	0.237	<0.0001	<0.0001
false return/false deductions credits x general:crypto virtual	14	1.97	0.98	0.128	0.0094	0.0288
payroll tax evasion/employment tax x general:general misc	51	1.81	0.86	0.149	<0.0001	0.0001
corporate/pass-through schemes x bulk cash/funnel	22	1.72	0.78	0.112	0.0001	0.0007
corporate/pass-through schemes x money transmitting/hawala	31	1.57	0.65	0.101	0.0002	0.0009
false return/false deductions credits x general misc	61	1.54	0.63	0.114	0.0002	0.0009
failure to file x structuring/bsa	17	1.3	0.38	0.052	0.2280	0.3630
corporate/pass-through schemes x structuring/bsa	213	1.13	0.18	0.049	0.0050	0.0171
corporate/pass-through schemes x wires transfers	1132	1.13	0.18	0.138	<0.0001	<0.0001
payroll tax evasion/employment tax x wires transfers	263	0.99	-0.01	-0.003	0.8582	0.9162
unreported income/cash business x structuring/bsa	13	0.95	-0.08	-0.01	0.1012	0.1007
payroll tax evasion/employment tax x services enabled general	17	0.91	-0.14	-0.019	0.7981	0.8623

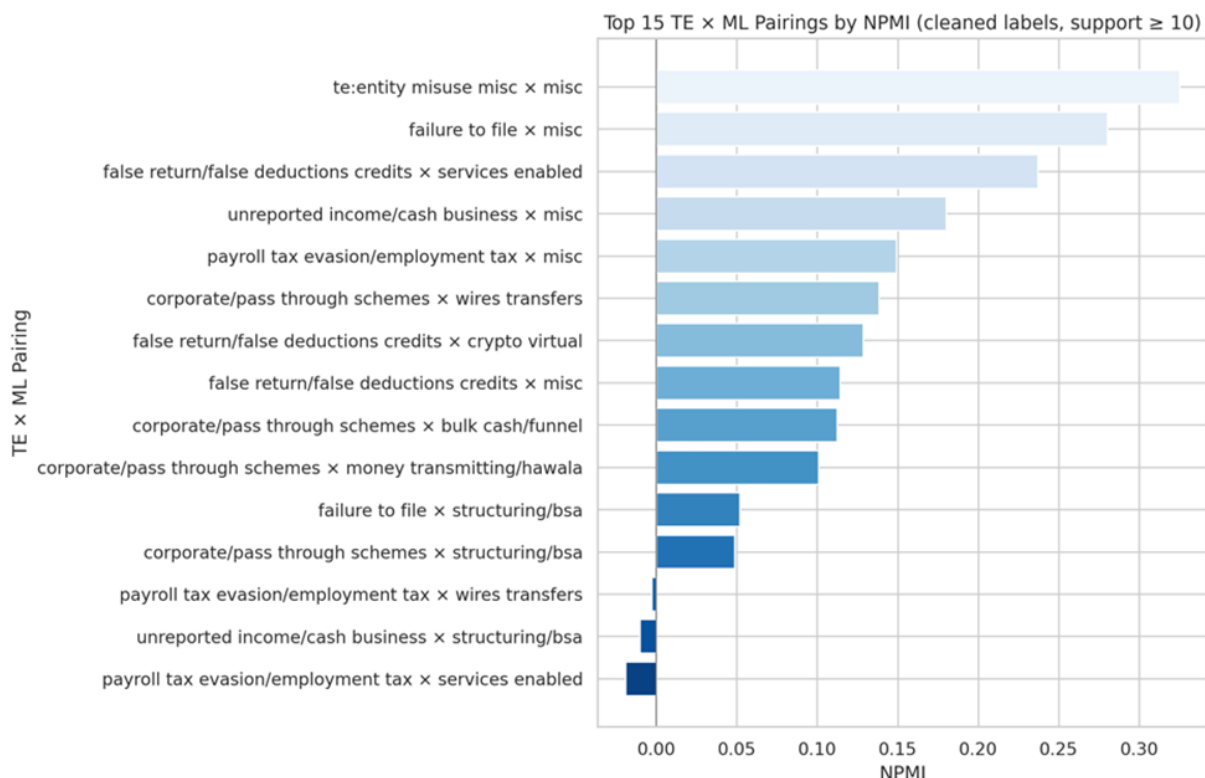


Figure 5: Top 15 TE x ML pairing with LIFT and NPMI Measures

Discussion and Conclusion

The results of the case analysis of the 4,534 IRS Criminal Investigation cases examined between 2021 and mid 2025 show that approximately 60 percent (2,727 cases) involved charges of both tax evasion and money laundering. Compared to tax evasion only and money laundering only, joint tax evasion and money laundering charges account for the majority of indicted cases. This also indicates that tax evasion is a significant predicate of money laundering. This is consistent with the fact that FATF added tax evasion as a designated offense predicate to money laundering. It also suggests that current US enforcement practice commonly treats these as separate but connected offenses.

The findings of this study both complement and qualify the theoretical position advanced by Kemsley, Kemsley and Morgan (2022, 2024). On one hand, a subset of tax evasion cases, such as small-scale household underreporting used immediately for routine consumption, involves no additional wiring, transferring, or entity structuring. This resembles a single-

stage structure, in which the act of filing a false or underreported return simultaneously generates and conceals the illegal tax savings.

On the other hand, IRS CI case evidence reveals a substantial number of prosecutions in which tax evasion is followed by clearly separable concealment behaviors, including structured cash movements, the use of shell entities, and crypto asset transactions. In these cases, prosecution press release documents explicitly state charges for two distinct offenses: tax evasion and money laundering. This aligns more closely with the FATF two-stage predicate offense framework, which requires both the generation of unlawful proceeds and a distinct laundering step. The high percentage of joint charges for tax evasion and money laundering, along with the significant LIFT and NPMI scores for specific evasion laundering pairings, reinforce this sequential relationship. This indicates that money laundering typically follows tax evasion as an additional layer of financial disguise rather than being intrinsic to it.

Overall, the results suggest that money laundering methods are not uniformly associated with tax evasion schemes but instead exhibit distinct subtype-specific patterns. Some rare subtype pairings demonstrated the strongest relative associations with high LIFT and NPMI scores. Conversely, more common pairings, such as corporate or pass-through entity schemes paired with wire transfers, combined moderate association strength with high case frequency, indicating operationally important laundering pathways. In contrast, several subtype pairs showed little or no meaningful association, highlighting that certain laundering methods are uncommon for specific tax evasion schemes. These findings underscore the value of subtype-level analysis for improving risk assessment and supporting more targeted investigative and enforcement efforts.

Policy and Practical Implications

The findings of this study have important implications for tax enforcement policy, regulatory frameworks, and future research. First, with respect to legal classification and sentencing policy, the analysis of prosecution cases involving both tax evasion and money laundering charges demonstrates that tax evasion frequently functions as an underlying offense associated with broader financial concealment activities. Therefore, legal frameworks should recognize tax evasion as a significant predicate offense but still maintain clear distinctions between tax violations that involve only the underreporting or nonpayment of tax obligations and those that involve additional money laundering conduct, such as the concealment, layering, transfer, or integration of illicit proceeds. Such differentiation is important to ensure proportional enforcement, avoid the overcriminalization of tax

offenses, and allow prosecutors and courts to appropriately assess the severity and complexity of financial crimes.

Second, the findings highlight the importance of regulatory coordination and international cooperation. Greater alignment between U.S. prosecutorial practices and international standards, particularly the Financial Action Task Force framework recognizing tax crimes as potential predicate offenses for money laundering, may enhance consistency in cross-border investigations, mutual legal assistance, financial intelligence sharing, and asset recovery efforts. As financial crimes increasingly involve multiple jurisdictions, coordinated approaches among tax authorities, law enforcement agencies, and financial intelligence units are essential for identifying complex schemes that combine tax evasion with illicit financial flows.

Third, the results provide important implications for financial institutions and regulatory oversight. Although the prosecution cases examined in this study reveal a strong association between tax evasion and money laundering, this relationship does not suggest that all tax evasion should be treated as money laundering. Rather, the distinguishing factor is the presence of additional behaviors intended to disguise the origin, ownership, movement, or use of illegally obtained funds. This distinction is particularly relevant for anti-money laundering (AML) compliance programs. Financial institutions can improve risk assessment and transaction monitoring by focusing on specific indicators of concealment and laundering rather than treating all tax-related offenses as equivalent risks.

Future Research

As an exploratory analysis, this study identifies co-occurrence patterns rather than causal relationships. Its findings are constrained by a reliance on publicly available case narratives, automated or rule-based subtype coding, and broad residual categories (such as general money laundering). To address these limitations, future research should prioritize developing a validated, multi-dimensional coding framework that clearly distinguishes tax-evasion conduct, predicate-offense proceeds, and laundering mechanisms. Doing so would enhance the legal precision of association results, mitigate classification-driven biases, and establish a more robust foundation for mapping specific tax-evasion typologies to distinct money-laundering methods.

Additionally, future work could expand on this analysis of IRS Criminal Investigation (IRS-CI) cases by comparing observed prosecution patterns with typologies documented in Financial Intelligence Unit (FIU) reports. These reports highlight prevalent tax-related money laundering channels, including real estate transactions, luxury asset acquisitions, shell companies, offshore structures, and cryptocurrency activities (Financial

Intelligence Analysis Unit, 2021). Researchers might also leverage machine learning, network analysis, and cross-jurisdictional comparisons to uncover emerging risk indicators and complex interactions among tax evasion methods, laundering techniques, and enforcement outcomes.

Conflict of Interest: The author reported no conflict of interest.

Data Availability: All data are included in the content of the paper.

Funding Statement: The author did not obtain any funding for this research.

References:

1. Allingham, M.G., and Sandmo, A. (1972). *Income Tax Evasion: A Theoretical Analysis*. *Journal of Public Economics*, 1 (3-4), 323-338.
2. Behrendt, P. (2020). *Tax evasion as the predicate offense of money laundering under German and US law*. *Zeitschrift für Internationale Strafrechtsdogmatik*, 4, 196–209. https://www.zis-online.com/dat/artikel/2020_4_1360.pdf
3. Benjamini, Y., & Hochberg, Y. (1995). *Controlling the false discovery rate: A practical and powerful approach to multiple testing*. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), 289–300.
4. Bouma, G. (2009). *Normalized (Pointwise) Mutual Information in Collocation Extraction*. In *Proceedings of the Biennial GSCL Conference 2009* (pp. 31–40)
5. Bridges, M. and Green, P. (1999), “*Tax evasion and money laundering – an open and shut case?*”, *Journal of Money Laundering Control*, Vol. 3, pp. 51-56.
6. Church, K. W., & Hanks, P. (1990). *Word association norms, mutual information, and lexicography*. *Computational Linguistics*, 16(1), 22–29.
7. Damani, O. P. (2013). *Improving Pointwise Mutual Information by Incorporating Significant Co-occurrence*. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing* (pp. 163–169).
8. Egmont Group of Financial Intelligence Units. (2020, July). *Money laundering of serious tax crimes: Enhancing financial intelligence units’ detection capacities and fostering information exchange* [Public bulletin]. https://egmontgroup.org/wp-content/uploads/2021/09/2020_Public_Bulletin_Money_Laundering_of_Serious_Tax_Crimes.pdf

9. Financial Intelligence Analysis Unit. (2021, November). *Typologies, red flags and indicators of tax-related money laundering*. FIAU. <https://fiaumalta.org/app/uploads/2021/11/FIAU-Factsheet-Typologies-Red-Flags-Indicators-of-Tax-Related-ML.pdf>
10. International Monetary Fund. (2012, July 17). *Revisions to the Financial Action Task Force (FATF) standard: Information note to the Executive Board*. <https://www.imf.org/external/np/pp/eng/2012/071712a.pdf>
11. Kemsley, D., Kemsley, S. A., & Morgan, F. T. (2022). *Tax evasion and money laundering: a complete framework*. *Journal of Financial Crime*, 29(2), 589-602.
12. Deen Kemsley, Sean A. Kemsley, Frank T. Morgan; *Tax evasion on lawful income: is it a form of money laundering?* *Journal of Financial Crime* 11 January 2024; 31 (1): 33–43. <https://doi.org/10.1108/JFC-11-2022-0268>
13. FATF, Money Laundering: <https://www.fatf-gafi.org/en/pages/frequently-asked-questions.html#tabs-36503a8663-item-6ff811783c-tab>
14. Financial Action Task Force Recommendation: <https://www.fatf-gafi.org/en/publications/Fatfrecommendations/Fatf-recommendations.html>
15. Financial Action Task Force Recommendation (Updated October 2025): www.fatf-gafi.org/content/dam/fatf-gafi/recommendations/FATF%20Recommendations%202012.pdf?utm_source=chatgpt.com
16. Gavrilova, Y. (2023, October 31). Pros and cons of Python programming language. <https://serokell.io/blog/python-pros-and-cons>
17. IRS Criminal Investigation press releases: <https://www.irs.gov/compliance/criminal-investigation/criminal-investigation-press-releases>
18. Maugeri, A. M. (2018). *Self-laundering of the proceeds of tax evasion in comparative law: Between effectiveness and safeguards: Between effectiveness and safeguards*. *New Journal of European Criminal Law*, 9(1), 83-108. <https://doi.org/10.1177/2032284418757385>
19. US Department of Justice, Tax Division Directive No. 128. https://www.justice.gov/archives/usam/tax-resource-manual-14-tax-division-directive-no-128?utm_source=chatgpt.com
20. 18 U.S.C. §§ 1956 and 1957. https://www.law.cornell.edu/uscode/text/18/1956?utm_source=chatgpt.com

21. United States v. Anderson, 1:05-cr-00056 (D.D.C. 2005).
<https://www.justice.gov/archive/opa/pr/2008/November/us-vs-anderson.pdf>

Appendix 1: Keyword lists used for classification

Tax Evasion Indicators

Tax evasion, tax fraud, false tax return, false return, false deductions, inflated deductions, fabricated expenses, sham losses (fabricated tax loss created through transactions that have no genuine economic substance), bogus credits, false schedule c, false 1099, false w-2, overstated expenses, unreported income, underreport, understate, skim (intentional diversion of cash receipts before they are recorded in a business's accounting records, cash business), off the books, failure to file, fail to file, did not file, non filer, nonfiler, employment tax, payroll tax, withholding, trust fund, 941(federal payroll tax form), 940 (federal payroll tax form), nominee (a person or entity that holds legal title to assets or income on behalf of the true owner in order to conceal the owner's identity or avoid tax collection), conceal income, offshore, foreign account, fbar, Swiss accounts, s corporation, partnership, llc (limited liability company), llp (limited liability partnership), k-1(Form 1065 for partnership), pass-through, entity.

Money Laundering Indicators

Money laundering, launder, laundering conspiracy, structuring, structured deposits, structured withdrawals, smurfing, suspicious activity report, SARs (Suspicious activity reports which is a confidential report that financial institutions file with the U.S. government when they detect transactions that may involve money laundering, tax evasion, fraud, terrorist financing, or other financial crimes.), bank secrecy act (BSA), layering, placement, integration, shell, foreign account, unlicensed money transmitting, msb (money service business), bulk cash, funnel accounts, concealment of criminal proceeds, transaction concealment, wash trades, trade-based money laundering, shell companies used to move proceeds, wire split to avoid detection, breaking up deposits, under reporting currency, 18 u.s.c.1956, 18 u.s.c.1957 (are two of the principal federal money laundering statutes in the United States. They are part of Title 18 of the United States Code and are frequently charged in IRS Criminal Investigation (IRS-CI), FBI, and DOJ financial crime cases.), hawala (It is an informal value transfer system that allows money to be transferred between people without physically moving the money through the formal banking system. Instead, it relies on a network of trusted brokers, known as hawaladars, who settle debts among themselves.), underground banking, crypto mixing, tumbling, money mules, offshore/foreign accounts offshore.

Appendix 2: LIFT, PMI, NPMI, Fisher's Exact Test and BH FDR (Benjamini–Hochberg False Discovery Rate)

To systematically evaluate the relationships between tax evasion and money laundering subtypes, this study utilizes a suite of complementary metrics that capture the frequency, strength, standardized effect size, and statistical significance of their co-occurrence. First, Lift establishes whether a specific pair of subtypes appears together more or less frequently than random chance would predict, where values greater than 1, equal to 1, and less than 1 indicate positive association, independence, and negative association, respectively. To measure the information-theoretic strength of these relationships, Pointwise Mutual Information (PMI) compares the observed joint probability of a subtype pair against its expected baseline under independence, with higher positive values reflecting a stronger connection. Because raw PMI can be skewed by rare events, Normalized Pointwise Mutual Information (NPMI) standardizes these scores onto a strict -1 to +1 scale, where +1 signifies perfect association, 0 denotes independence, and -1 represents a complete negative association. Finally, the framework ensures statistical rigor by pairing Fisher's exact test to determine the significance of the observed co-occurrences with the Benjamini-Hochberg (BH) false discovery rate correction, which adjusts for multiple comparisons to minimize the risk of false-positive findings. The following provide definition and interpretation of each measure:

1. LIFT measures how much more often two items occur together than would be expected if they were statistically independent.

LIFT:

$$\text{Lift}(A, B) = P(A \cap B) / [P(A) \times P(B)]$$

Interpretation:

Lift = 1: A and B are independent

Lift > 1: Positive association; they co-occur more than expected

Lift < 1: Negative association; they co-occur less than expected

PMI (Pointwise Mutual Information)

PMI measures how much information the occurrence of one event provides about another, relative to what would be expected under independence.

2. PMI:

$$\text{PMI}(A, B) = \log_2 [P(A \cap B) / (P(A) \times P(B))]$$

Interpretation:

PMI = 0: Independent

PMI > 0: Positive association

PMI < 0: Negative association

Larger values indicate stronger associations.

3. NPMI (Normalized Pointwise Mutual Information):

NPMI addresses the rare-event bias of PMI by dividing the raw PMI score by the joint self-information (the upper bound of mutual information). This standardizes the metric into a fixed, easily interpretable scale. NPMI rescales PMI to a standardized range of -1 to $+1$, making association strengths comparable across subtype pairs:

$$\text{NPMI}(A, B) = \text{PMI}(A, B) / [-\log_2 [P(A \cap B)]]$$

NPMI = $+1$: Perfect association (the two subtypes always occur together).

NPMI = 0 : Statistical independence.

NPMI = -1 : Complete negative association (the two subtypes never co-occur).

Higher positive NPMI values indicate stronger positive associations, while values closer to zero indicate weaker associations and negative values indicate negative associations.

Because NPMI is normalized, it is less affected by the absolute frequency of subtype pairs than PMI, making it particularly useful for comparing the relative strength of associations across multiple tax evasion and money laundering subtype pairings.

4. Fisher's Exact Test

The Fisher's Exact Test addresses the small-sample bias of asymptotic chi-square tests by calculating the exact probability of the observed data distribution under the null hypothesis, rather than relying on large-sample approximations. This makes the metric highly robust for sparse contingency tables or rare event frequencies. The test evaluates the significance of the association between two categorical variables based on a hypergeometric distribution, yielding a bounded probability score (p-value) on a fixed, easily interpretable scale from 0 to 1 .

$p = 0$: Complete statistical dependence. The observed association between the two subtypes is highly unlikely to have occurred by chance.

$p = 1$: Perfect statistical independence. The distribution of the two subtypes matches the null hypothesis expectation exactly.

Lower p values indicate stronger, more statistically significant evidence of an association between the variables. A p value closer to 1 indicate a high probability that any observed co-occurrence is due to pure chance. Unlike NPMI, a two-sided Fisher's p value measures the significance of the association rather than its direction.

5. Summary of LIFT, PMI, NPMI, Fisher's Exact Test and BH FDR (Benjamini–Hochberg False Discovery Rate)

Measure	What it Measures	Interpretation	Typical Thresholds/Values
LIFT	How much more (or less) often a tax-evasion subtype and money-laundering subtype co-occur than expected under independence.	Effect size based on observed vs. expected frequency.	>1 = Positive association; $=1$ = Independent; <1 = Negative association
PMI (Pointwise Mutual Information)	Information-theoretic strength of association between two subtype pairings.	Higher positive values indicate stronger-than-expected co-occurrence; negative values indicate less-than-expected co-occurrence.	>0 = Positive association; 0 = Independent; <0 = Negative association; no upper bound
NPMI (Normalized PMI)	Standardized PMI that allows comparison across subtype pairs.	Normalizes PMI onto a fixed scale from -1 to $+1$.	$+1$ = Perfect association; 0 = Independent; -1 = Never co-occur
Fisher's Exact Test	Statistical significance of the association between subtype pairs.	Tests whether the observed co-occurrence differs from what would be expected by chance.	Small p-value indicates statistically significant association.

Appendix 3: Definitions of Tax-Evasion and Money-Laundering Typology Pairings

TE and ML Pairings	Definition
entity_misuse_misc x general:general_misc	Other or broadly defined entity-related tax-evasion conduct paired with money-laundering activity for which no specific mechanism was identified.
failure_to_file x general:general_misc	Willful failure to file required tax returns paired with unspecified/general money-laundering activity.
unreported_income/cash_business x general:general_misc	Concealment of income, often from cash-intensive businesses, paired with unspecified/general money-laundering activity.
false_return/false_deductions_credits x services_enabled_general	False tax returns, fabricated deductions, or improper tax credits paired with laundering enabled by financial, business, or intermediary services that are not specifically classified as hawala, crypto, or wire transfers.
false_return/false_deductions_credits x general:crypto_virtual	False tax returns, deductions, or credits paired with cryptocurrency, virtual-currency, mixing, or related digital-asset laundering activity.
payroll_tax_evasion/employment_tax x general:general_misc	Employment-tax or payroll-tax evasion paired with unspecified/general money-laundering activity.
corporate/pass-through_schemes x bulk_cash/funnel	Abuse of corporations, partnerships, LLCs, or pass-through entities paired with bulk-cash movement, cash smuggling, or funnel-account activity.
corporate/pass-through_schemes x money_transmitting/hawala	Corporate or pass-through tax schemes paired with informal value-transfer systems, unlicensed money transmission, or hawala activity.
false_return/false_deductions_credits x general:general_misc	False returns, fabricated deductions, or improper credits paired with unspecified/general money-laundering activity.
failure_to_file x structuring/bsa	Failure to file tax returns paired with structuring transactions to avoid currency-reporting requirements or other Bank Secrecy Act obligations.
corporate/pass-through_schemes x structuring/bsa	Corporate or pass-through entity schemes paired with structuring or Bank Secrecy Act-related conduct.
corporate/pass-through_schemes x wires_transfers	Corporate or pass-through entity schemes paired with wire transfers, bank-account

	transfers, or other electronic movement of funds.
payroll_tax_evasion/employment_tax x wires_transfers	Payroll or employment-tax evasion paired with wire-transfer or bank-transfer activity.
unreported_income/cash_business x structuring/bsa	Underreported income or cash-business tax evasion paired with structuring or BSA-related conduct.

