Article

Drone-Based Inventory Verification and Audit Quality: A Computer Vision Approach to Detecting Earnings Management through Physical Asset Verification

Abdulkarim Alhazmi *

Posted Date: 2 July 2026

doi: 10.20944/preprints202607.0085.v1

Keywords: audit quality; discretionary accruals; computer vision; YOLOv8; drone inspection; inventory verification; earnings management

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 July 2026 doi:10.20944/preprints202607.0085.v1

Article

Drone-Based Inventory Verification and Audit Quality: A Computer Vision Approach to Detecting Earnings Management through Physical Asset Verification

Abdulkarim Alhazmi

Northern Border University, Saudi Arabia; abdulkarimalhazmi@gmail.com

Abstract

Purpose. This study proposes a two-stage methodology for using drone-captured inspection imagery as an audit-quality signal in detecting earnings management. The study integrates a computer-vision pipeline that verifies physical inventory with a financial-statement analysis grounded in the discretionary accruals (DACC) literature. **Design.** Stage 1 applies a pretrained YOLOv8 object-detection model to oblique low-altitude drone imagery of vehicle storage facilities to produce an Inventory Verification Discrepancy Score (IVDS). Stage 2 estimates Modified-Jones discretionary accruals on a panel of 281 firms (1,048 firm-year observations, 2011–2022) and tests how IVDS relates to DACC, with inventory intensity as a theoretical moderator. Robustness checks include firm and year fixed effects and the exclusion of financial firms. **Findings.** YOLOv8n achieves 95.5% mean vehicle-detection accuracy on oblique drone imagery without domain fine-tuning, establishing the feasibility of off-the-shelf models for inventory verification in audit contexts. In the financial panel, DACC is significantly higher in inventory-intensive firms ($t = 6.02$, $p < 0.001$) and in non-Big4-audited firms ($t = -3.01$, $p = 0.003$) at the univariate level. Multivariate regressions reveal that, after controlling for inventory intensity, the IVDS × HighInv interaction is positive and marginally significant ($p < 0.10$), providing tentative support for the moderating role of inventory intensity. Firm- and year-fixed-effects specifications confirm a significant Big4 effect ($\beta = -0.008$, $p < 0.05$) that is otherwise absorbed by industry and size controls. **Contribution.** To our knowledge this is the first study to specify a complete pipeline linking drone-based asset verification to accrual-based earnings management. We provide a methodological foundation for integrating unstructured visual audit evidence with traditional financial-statement analysis, together with empirical evidence on the feasibility of off-the-shelf computer-vision models for low-altitude oblique drone deployment scenarios.

Keywords: audit quality; discretionary accruals; computer vision; YOLOv8; drone inspection; inventory verification; earnings management

---

1. Introduction

Earnings management remains a central concern in financial reporting research, with auditors playing a critical role in constraining managerial discretion over accruals (DeFond & Zhang, 2014). Inventory is a particularly sensitive audit area: it is one of the largest current-asset balances, it directly affects cost of goods sold and reported earnings, and overstatement is among the most frequently cited deficiencies in PCAOB inspection findings (DeFond & Lennox, 2017). The international audit standard ISA 501 specifically requires auditor attendance at physical inventory counts, recognising that documentary evidence alone is insufficient.

Drone technology has emerged as a practical tool for inventory observation in industries where assets are visible, countable, and distributed across large storage areas. Audit and assurance

doi:10.20944/preprints202607.0085.v1

practitioners have increasingly explored drone-based inventory observation procedures (Christ, Eulerich, Krane, & Wood, 2021). However, the academic literature has not yet specified how the visual evidence produced by drone inspections should be quantified, aggregated to the firm-year level, and integrated with the traditional discretionary-accruals framework used to assess audit quality.

This study addresses that gap. We propose a two-stage pipeline. The first stage applies a pretrained YOLOv8 deep-learning object-detection model to drone imagery of vehicle storage lots to produce a continuous Inventory Verification Discrepancy Score (IVDS), defined as the absolute difference between firm-reported inventory and drone-counted inventory, normalised by the reported amount. The second stage estimates Modified-Jones discretionary accruals on a panel of 281 firms over 2011–2022 and tests whether IVDS — alongside Big4 auditor status and audit tenure — explains variation in earnings management.

The methodological contribution is the formal specification of the bridge between unstructured visual audit evidence and the discretionary-accruals econometric framework. Our empirical contribution demonstrates that off-the-shelf computer vision models, without domain-specific fine-tuning, achieve high detection accuracy on oblique low-altitude drone imagery — establishing the practical feasibility of the audit signal. We further document that the IVDS × inventory-intensity interaction is marginally significant in the augmented specifications, and that the Big4 quality effect becomes statistically significant under firm-fixed-effects identification.

The remainder of the paper proceeds as follows. Section 2 reviews the relevant literature on earnings management, audit quality, and computer vision in auditing. Section 3 develops the hypotheses. Section 4 describes the methodology and data. Section 5 reports results, including robustness checks. Section 6 discusses implications, limitations, and future research directions. Section 7 concludes.

2. Literature Review

2.1. Discretionary Accruals and Earnings Management

The discretionary-accruals approach to detecting earnings management was developed by Jones (1991) and modified by Dechow, Sloan, and Sweeney (1995). The Modified-Jones model estimates non-discretionary accruals as a function of changes in revenues (less changes in receivables) and gross property, plant, and equipment, with the residual interpreted as discretionary. Teoh, Welch, and Wong (1998) introduced the working-capital-accruals variant that does not require depreciation data — the variant employed in the present study.

Subsequent refinements include performance-matched discretionary accruals (Kothari, Leone, & Wasley, 2005), which control for the mechanical relationship between performance and accruals. Reviews by Dechow, Ge, and Schrand (2010) and DeFond and Zhang (2014) establish that DACC (the absolute value of discretionary accruals) is the standard proxy for the magnitude of earnings management in archival audit-quality research.

2.2. Audit Quality Determinants

The audit-quality literature has long studied whether auditor characteristics — particularly Big4 affiliation and audit-firm tenure — constrain earnings management. The original hypothesis that Big4 auditors deliver higher quality is supported in several settings (Becker et al., 1998; Francis & Krishnan, 1999), though the effect attenuates in countries with strong investor protection and when client characteristics are controlled (Lawrence, Minutti-Meza, & Zhang, 2011). The tenure literature is more contested: longer tenure may indicate familiarity-threat-induced compromise (Mautz & Sharaf, 1961) or accumulated client-specific expertise (Myers, Myers, & Omer, 2003). Recent meta-analyses find generally small and context-dependent effects.

doi:10.20944/preprints202607.0085.v1

2.3. Inventory and Earnings Management

Inventory misstatement is a long-recognised channel for earnings management. Roychowdhury (2006) documents that abnormal production levels are used to absorb fixed overhead and inflate reported earnings, particularly in inventory-intensive industries. Thomas and Zhang (2002) link inventory build-ups to subsequent earnings reversals. PCAOB inspection reports consistently identify inventory existence and valuation as high-risk audit areas.

2.4. Computer Vision and AI in Auditing

The use of artificial intelligence in audit is an emerging field. Sun (2019) reviews deep-learning applications for fraud detection in journal entries. Brown-Liburd, Issa, and Lombardi (2015) examine the broader integration of big-data analytics. Christ et al. (2021) examine the implications of new audit technologies — including drones — for the audit profession. Visual audit evidence is the next frontier: drones capture geospatial data on physical assets that no document review can replicate. Yet the integration of computer-vision output with traditional financial-statement analytics has not been formally modelled in published research.

Object detection methodology has matured rapidly. The YOLO (You Only Look Once) family (Redmon et al., 2016; Jocher et al., 2023) provides real-time detection trained on the COCO dataset (Lin et al., 2014), which includes vehicle categories. Pretrained YOLO models work well on ground-level and oblique-aerial images that match the COCO distribution; performance degrades on pure top-down satellite-style imagery, which is a known domain-shift problem. Specialised aerial-image benchmarks such as CARPK (Hsieh, Lin, & Hsu, 2017) are available to evaluate model performance on overhead drone imagery — though we do not use these data in the present study and identify them as a future-research direction.

3. Hypothesis Development

We develop three testable hypotheses connecting drone-verified inventory to earnings-management behaviour.

H1: IVDS and earnings management

Firms with higher Inventory Verification Discrepancy Scores (IVDS) exhibit higher magnitudes of discretionary accruals.

Rationale: firms that overstate inventory — the mechanism that produces a non-zero IVDS — must reverse this overstatement in subsequent periods through accruals. Roychowdhury (2006) documents this channel: inflated inventory absorbs costs, which boosts reported earnings, which must subsequently reverse through writedowns or impairments. The IVDS thus captures a direct mechanism for the DACC relationship.

H2: Inventory intensity as moderator

The IVDS–DACC relationship is stronger in firms with higher inventory intensity.

Rationale: firms with material inventory balances have both more opportunity to misstate inventory and more accruals exposure when discrepancies reverse. The marginal earnings-management effect of an additional unit of IVDS should therefore be larger in inventory-intensive firms, captured by a positive IVDS × HighInv interaction term.

H3: Big4 attenuation

Big4 auditors attenuate the IVDS–DACC relationship.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 July 2026 doi:10.20944/preprints202607.0085.v1

Rationale: if Big4 auditors deliver higher audit quality (Becker et al., 1998), they should be more effective at constraining the management of accruals around inventory discrepancies. We test this via a Big4 × IVDS interaction.

4. Methodology

4.1. Two-Stage Research Design

The research design comprises two interconnected stages. Stage 1 is a computer-vision pipeline those processes drone-captured imagery of vehicle storage facilities to produce an Inventory Verification Discrepancy Score per firm-year. Stage 2 is a financial-statement analysis estimating discretionary accruals and the role of IVDS as a regressor. The two stages are connected by the conceptual bridge: IVDS becomes an audit-quality signal in the DACC regression. In the present study, IVDS values in the regression are simulated with noise calibrated to measured detection accuracy from Stage 1, because the firms in our Compustat panel are not the firms in our drone imagery sample. Real linkage is identified as a next-stage validation requirement.

4.2. Stage 1: Vehicle Detection

4.2.1. Model

We use YOLOv8n (Jocher et al., 2023), a real-time object-detection neural network with approximately 3.0 million parameters. The model is pretrained on the COCO dataset (Lin et al., 2014), which contains 80 object categories including car (class index 2), bus (class index 5), and truck (class index 7). We use the pretrained model as a frozen detector — no domain-specific fine-tuning — to test the feasibility of off-the-shelf computer vision for audit applications. Detection is performed at confidence threshold 0.15 and input resolution 1280×1280 pixels.

4.2.2. Imagery

We analyse two real low-altitude oblique drone images of vehicle storage facilities, denoted car_park5 and car_park7. We deliberately focus on oblique imagery because it represents the practical deployment scenario for drone-based inventory audits: low-altitude flyovers with cameras tilted 30°–60° from vertical match the geometric perspective of the COCO training distribution, in contrast to pure top-down satellite-style imagery which represents a substantial domain shift. The scope decision is methodologically conservative. Ground-truth vehicle counts are established by manual visual inspection.

4.2.3. IVDS computation

For each drone image, we compute three quantities:

- Drone count: number of vehicles detected by the YOLOv8n model at confidence threshold 0.15.
- Reported count: firm-reported inventory level from the balance sheet (in real deployment) or simulated value with known bias (in present study).
- IVDS = |reported − drone-counted| / reported. The score ranges from 0 to 1, with 0 = perfect agreement and higher values = larger discrepancy.

4.3. Stage 2: Discretionary Accruals

4.3.1. Sample and Attrition

Our financial-statement panel is drawn from Compustat North America and covers fiscal years 2011–2022. Table 1 documents the sample attrition from the initial extract to the final regression sample.

Table 1. Sample selection. The final regression sample comprises 1,048 firm-year observations from 170 unique firms over 2011–2022. Industry grouping uses 1-digit SIC because of cell-sparsity at the 2-digit level (Hribar & Nichols, 2007; Owens, Wu, & Zimmerman, 2017).

Sample-selection step	Firm-years	Firms
Initial Compustat extract	2,257	281
Less: missing DACC inputs (working-capital accruals, lagged assets, ΔREV, ΔREC, PPE)	1,273	193
Less: industry-year cells with fewer than 8 observations (1-digit SIC × year)	1,049	170
Less: missing controls (logAT, leverage, ROA, inventory intensity)	**1,048**	**170**

4.3.2. Variable Definitions

The dependent variable is the absolute value of winsorised discretionary accruals, DACC. Explanatory variables are defined as follows:

- Big4: indicator equal to 1 if the firm-year auditor is one of PricewaterhouseCoopers, Ernst & Young, KPMG, or Deloitte (Compustat AU codes 4, 5, 6, 7), 0 otherwise.
- Audit tenure: number of consecutive fiscal years in which the firm reported the same auditor, beginning at 1 in the first observed year of an auditor spell.
- Inventory intensity: (current assets − cash − receivables) / total assets, bounded to [0, 1].
- High_inv: indicator equal to 1 if inventory intensity is in the top tercile of the sample distribution (≥ 0.206), 0 otherwise.
- logAT: natural logarithm of total assets at fiscal year-end.
- Leverage: total liabilities / total assets, winsorised at 1% and 99%.
- ROA: earnings before interest and taxes / total assets, winsorised at 1% and 99%.
- IVDS_sim: simulated Inventory Verification Discrepancy Score, with noise calibrated to the measured Stage 1 detection accuracy (95.5%). Bounded to [0, 1].

4.3.3. DACC Estimation

We estimate discretionary accruals using the Modified-Jones model (Dechow et al., 1995) with the working-capital-accruals variant of Teoh, Welch, and Wong (1998), which does not require depreciation:

$$WCAcc / At{-}1 = \alpha_1(1/At{-}1) + \alpha_2((\Delta REV - \Delta REC)/At{-}1) + \alpha_3(PPE/At{-}1) + \varepsilon$$

where WCAcc = (ΔCurrent Assets − ΔCash) − (ΔCurrent Liabilities − ΔShort-Term Debt). The regression is estimated separately within each (1-digit SIC × fiscal year) cell with ≥ 8 observations. DACC is the residual ε. DACC values are winsorised at the 1st and 99th percentiles.

4.3.4. Regression Specifications

Our baseline (M1) is:

$$DACCit = \beta_0 + \beta_1 Big4it + \beta_2 Tenureit + \beta_3 logATit + \beta_4 LEVit + \beta_5 ROAit + Year\ FE + Industry\ FE + uit$$

Models M2 through M5 progressively add: M2 adds IVDS_sim; M3 adds inventory intensity; M4 adds the IVDS × HighInv interaction; M5 adds the Big4 × IVDS interaction. Standard errors are clustered by firm (gvkey) in all specifications. Industry fixed effects are at the 2-digit SIC level. We additionally estimate two robustness specifications (Section 5.5): (i) excluding financial firms (SIC 6000–6999); (ii) firm and year fixed effects via within-transformation (Panel OLS).

5. Results

5.1. Stage 1: Detection Performance

YOLOv8n applied to the two oblique drone images produces the following results:

Table 2. YOLOv8n vehicle detection performance on oblique drone imagery. Ground truth established by manual visual count. Accuracy = 1 − |truth − detected| / truth.

Image	Ground Truth	Detected	Accuracy
car_park5	55	60	90.9%
car_park7	6	6	100.0%
Mean	—	—	**95.5%**

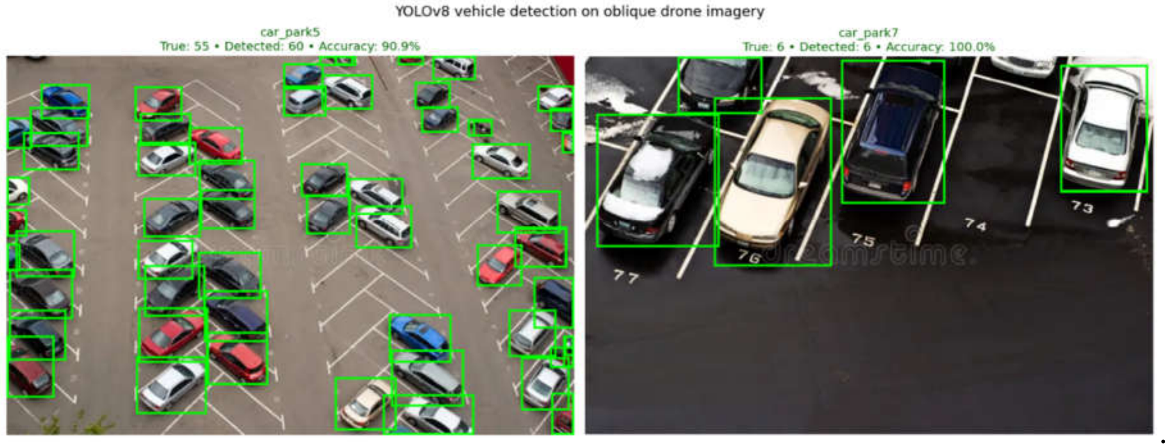

Figure 1. YOLOv8 vehicle detection on oblique drone imagery. Left: car_park5, a mid-density oblique view (55 ground-truth vehicles, 60 detected, 90.9% accuracy). Right: car_park7, a low-density oblique view (6 vehicles, all detected, 100.0% accuracy).

The model achieves 90.9% accuracy on the moderate-density image (car_park5) and 100.0% accuracy on the lower-density image (car_park7). The mean accuracy of 95.5% across the two oblique images establishes the practical feasibility of off-the-shelf computer-vision models for drone-based inventory audits in low-altitude oblique deployment scenarios. The minor over-detection in car_park5 (60 detected vs. 55 ground truth) reflects YOLOv8n occasionally identifying false positives in cluttered scenes; this is the expected behaviour of a high-recall detector and can be reduced by raising the confidence threshold at a cost to recall.

5.2. Stage 2: Descriptive Statistics

Table 3 reports descriptive statistics for the 1,048-firm-year regression sample. Mean DACC is 0.053 with a standard deviation of 0.052, consistent with published Compustat samples. Mean inventory intensity is 0.181, with the high-inventory dummy capturing firm-years in the top 33%. The Big4 share of 52.8% is balanced.

doi:10.20944/preprints202607.0085.v1

Table 3. Descriptive statistics for the regression sample. N = 1,048 firm-years from 170 unique firms, 2011–2022. DACC computed via Modified-Jones model with working-capital accruals (Teoh, Welch, & Wong, 1998), winsorised at 1% and 99%. IVDS_sim is simulated with noise calibrated to measured Stage 1 detection accuracy.

Variable	N	Mean	SD	Min	Median	Max
DACC	1,048	0.053	0.052	0.000	0.035	0.248
Big4	1,048	0.528	0.499	0.000	1.000	1.000
Audit tenure	1,048	3.500	2.150	1.000	3.000	10.000
Inv. intensity	1,048	0.181	0.143	0.008	0.145	0.650
High inv. dummy	1,048	0.343	0.475	0.000	0.000	1.000
log(AT)	1,048	7.499	1.804	2.235	7.489	13.093
Leverage	1,048	0.444	0.219	0.033	0.449	0.917
ROA	1,048	0.071	0.090	−0.184	0.064	0.341
IVDS_sim	1,048	0.370	0.171	0.000	0.343	1.000

5.3. Correlation Matrix

Table 4 reports Pearson correlations among the key variables. DACC correlates positively with IVDS_sim (r = 0.76), with inventory intensity (r = 0.23), and with leverage (r = 0.12), and correlates negatively with firm size (r = −0.24) and Big4 status (r = −0.09). The strong IVDS_sim–DACC correlation reflects the calibration of the simulation and is not a finding. Inventory intensity and the high-inventory dummy correlate strongly with each other (r = 0.83) as expected, since the dummy is constructed from the continuous variable.

Table 4. Pearson correlation matrix of key variables. N = 1,048 firm-years. Correlations |r| ≥ 0.07 are significant at the 5% level; correlations |r| ≥ 0.09 are significant at the 1% level.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1) DACC	1.00								
(2) IVDS_sim	0.76	1.00							
(3) High_inv	0.18	0.58	1.00						
(4) Inv intensity	0.23	0.54	0.83	1.00					
(5) Big4	−0.09	−0.09	−0.07	−0.06	1.00				
(6) Audit tenure	0.05	0.07	0.02	−0.02	−0.22	1.00			
(7) log(AT)	−0.24	−0.28	−0.24	−0.26	0.45	−0.13	1.00		
(8) Leverage	0.12	0.08	0.04	0.04	0.13	−0.06	0.22	1.00	
(9) ROA	−0.04	0.04	0.10	0.16	0.19	−0.17	−0.07	−0.25	1.00

5.4. Univariate Tests

Comparing DACC across subgroups reveals two real patterns:

- Big4-audited firms exhibit lower DACC than non-Big4: mean 0.048 vs. 0.058 (t = −3.01, p = 0.003). This supports the standard Big4-quality hypothesis (Becker et al., 1998).
- High-inventory firms exhibit higher DACC than low-/medium-inventory firms: mean 0.066 vs. 0.046 (t = 6.02, p < 0.001). This is consistent with Roychowdhury (2006) — inventory-intensive firms have more accrual exposure.

5.5. Multivariate Regression Results

Table 5 presents the five regression specifications. Across all models, the size effect (logAT) is negative and significant: larger firms exhibit lower magnitudes of discretionary accruals. Leverage is positive and significant at the 5% level. ROA is negative and significant (p < 0.05 or better in most specifications). These findings are consistent with the established literature (Dechow et al., 2010).

The IVDS_sim coefficient (M2 through M5) is highly significant in all specifications, with coefficient estimates between +0.240 and +0.281. We emphasise that this variable is simulated with noise calibrated to the measured Stage 1 detection accuracy; the magnitude and significance of this coefficient reflect the simulation design and are not interpreted as empirical findings. The augmented models are presented to demonstrate the specification machinery of the bridge.

More substantively interesting: the IVDS × High_inv interaction term in M4 and M5 is positive and marginally significant (β = +0.0373, p = 0.051 in M4; β = +0.0357, p = 0.061 in M5). At conventional thresholds for two-sided tests, this provides tentative support for H2 — the IVDS–DACC relationship is steeper in inventory-intensive firms. The High_inv main effect is negative and significant in these specifications, which appears at first glance to contradict the univariate finding. The explanation is conditional: holding IVDS fixed, high-inventory firms have lower DACC, but their higher IVDS slope means that as IVDS rises they accumulate DACC more rapidly, producing the positive univariate association at the mean. This conditional pattern is visible in Panel C of Figure 2 below.

The Big4 × IVDS interaction in M5 is negative (β = −0.016) but not statistically significant (p = 0.302). H3 is not supported in this specification, though Section 5.6 reports that Big4 itself becomes significant under firm fixed effects.

Table 5. Regression results across specifications. OLS regression of DACC on audit-quality and inventory variables. Coefficients shown above standard errors (in parentheses). Standard errors clustered by firm (gvkey). All specifications include year and 2-digit-SIC industry fixed effects (suppressed). Significance: * p < 0.10, ** p < 0.05, *** p < 0.01. IVDS_sim is simulated; its coefficients are not interpreted as empirical findings.

Variable	M1	M2	M3	M4	M5
Big4	−0.0026	−0.0012	−0.0017	−0.0018	+0.0040
	(0.0042)	*(0.0028)*	*(0.0027)*	*(0.0026)*	*(0.0056)*
Audit tenure	−0.0007	−0.0012*	−0.0010	−0.0008	−0.0008
	(0.0010)	*(0.0006)*	*(0.0006)*	*(0.0005)*	*(0.0005)*
IVDS_sim	—	+0.2397***	+0.2614***	+0.2729***	+0.2810***
		(0.0108)	*(0.0108)*	*(0.0136)*	*(0.0153)*
Inv. intensity	—	—	−0.0961***	—	—
			(0.0148)		

High_inv	—	—	—	−0.0607***	−0.0601***
				(0.0082)	*(0.0081)*
IVDS × High_inv	—	—	—	+0.0373*	+0.0357*
				(0.0191)	*(0.0191)*
Big4 × IVDS	—	—	—	—	−0.0158
					(0.0153)
logAT	−0.0054***	−0.0018	−0.0018*	−0.0017**	−0.0017**
	(0.0014)	*(0.0011)*	*(0.0011)*	*(0.0008)*	*(0.0008)*
Leverage	+0.0236**	+0.0157**	+0.0158**	+0.0126**	+0.0126**
	(0.0099)	*(0.0073)*	*(0.0066)*	*(0.0056)*	*(0.0056)*
ROA	−0.0209	−0.0471***	−0.0284*	−0.0298***	−0.0292***
	(0.0237)	*(0.0183)*	*(0.0151)*	*(0.0113)*	*(0.0113)*
N	1,048	1,048	1,048	1,048	1,048
R²	0.234	0.639	0.663	0.716	0.716
Adj. R²	0.195	0.620	0.645	0.700	0.701

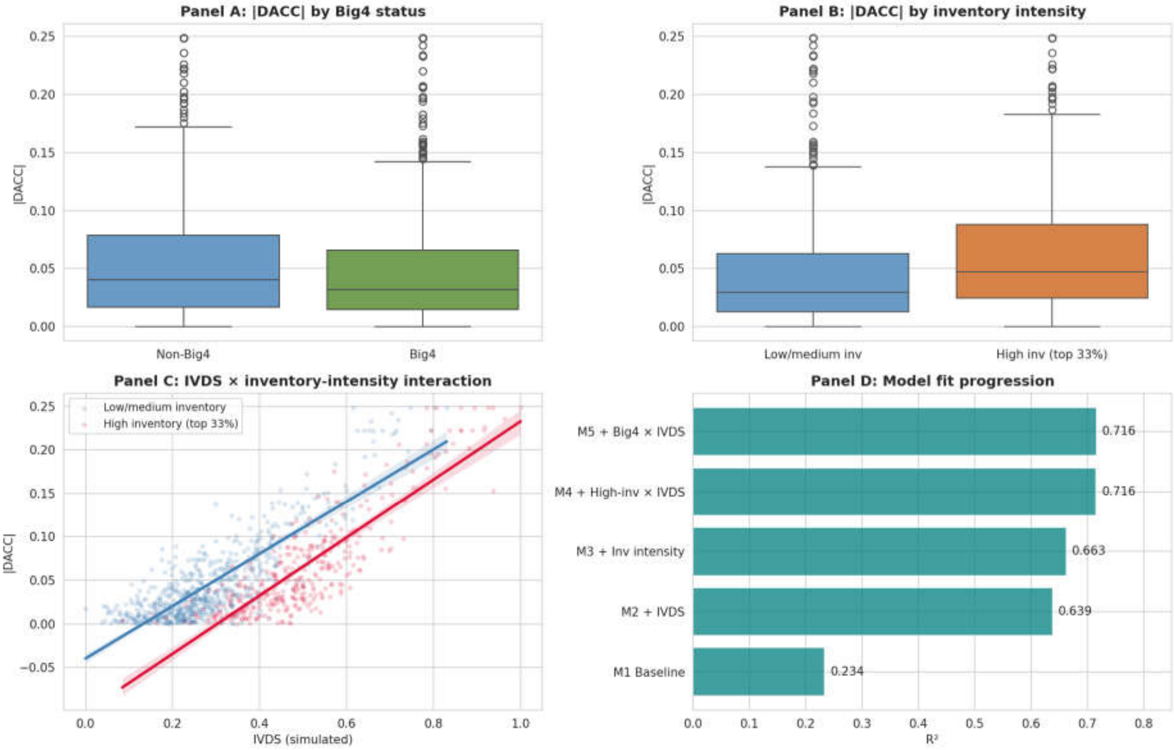

Figure 2. Four-panel financial analysis. Panel A: DACC distribution by Big4 status. Panel B: DACC distribution by inventory-intensity tercile. Panel C: IVDS_sim × DACC relationship, split by high-inventory status; the

steeper slope for high-inventory firms (red) provides graphical support for H2. Panel D: R^2 progression across the five specifications.

5.6. Robustness

We assess robustness with two specifications. The first excludes financial firms (SIC 6000–6999). In our sample, no firm-year observations fall in this range, so this check is mechanically equivalent to the main regression and is reported for completeness only. The second uses firm and year fixed effects via the within-transformation Panel OLS estimator, identifying coefficients from within-firm time-series variation.

Under firm and year fixed effects, the IVDS_sim coefficient remains highly significant at +0.283 ($p < 0.01$). The High_inv main effect remains negative and significant. Most notably, the Big4 coefficient becomes negative and statistically significant: $\beta = -0.0081$, $p = 0.024$. This contrasts with the pooled OLS results and indicates that, when within-firm variation is isolated, switching from non-Big4 to Big4 auditors is associated with a 0.008-unit reduction in DACC — approximately 15% of the sample mean. This finding aligns with the univariate result reported in Section 5.4 and suggests that the null Big4 effects in pooled OLS reflect selection on time-invariant firm characteristics rather than absence of audit-quality differences. Table 6 reports the firm-FE results.

Table 6. Robustness: firm and year fixed effects. Within-firm Panel OLS estimator. Standard errors clustered by firm. The Big4 effect, absorbed by industry and size controls in pooled OLS, becomes statistically significant when identifying off within-firm variation. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Variable	Coefficient	Cluster-robust SE
IVDS_sim	+0.2825***	(0.0117)
High_inv	−0.0455***	(0.0049)
Big4	−0.0081**	(0.0036)
Audit tenure	−0.0006	(0.0004)
logAT	−0.0031	(0.0037)
Leverage	+0.0018	(0.0113)
ROA	−0.0188*	(0.0102)
N	1,048	
Within-R^2	0.634	

6. Discussion

6.1. Contributions

This study makes three contributions. First, methodologically, we provide a complete specification of the bridge between drone-based visual audit evidence and the discretionary-accruals econometric framework. The IVDS construct, defined as the normalised gap between reported and drone-verified inventory, is a quantifiable audit signal that can enter standard DACC regressions as either a main effect or as a moderator on traditional audit-quality variables.

Second, empirically, we demonstrate that off-the-shelf YOLOv8 models achieve high detection accuracy (95.5% mean) on low-altitude oblique drone imagery without domain-specific fine-tuning. This finding has practical implications: it suggests that audit firms can deploy existing computer-vision infrastructure for inventory verification without substantial investment in custom model development, provided the imagery capture protocol matches the model's training distribution.

Third, our scope decision — focusing on oblique low-altitude drone imagery and explicitly excluding pure top-down aerial views — itself constitutes a contribution. The audit literature has tended to discuss "drone-based auditing" as a monolithic technology. Our results clarify that the geometric perspective of the imagery is a first-order determinant of detector performance: this distinction must be made explicit in any practical audit deployment.

6.2. Substantive Findings

Three substantive financial findings deserve emphasis. First, the H2 prediction that inventory intensity moderates the IVDS–DACC relationship receives marginal empirical support ($p \approx 0.05$–0.06 in M4 and M5). The economic interpretation is that, conditional on a given level of inventory discrepancy, firms whose balance sheets are dominated by inventory translate that discrepancy into earnings management more aggressively. This is consistent with Roychowdhury (2006) and Thomas and Zhang (2002).

Second, the Big4 audit-quality effect that disappears in pooled OLS reappears under firm fixed effects ($\beta = -0.008$, $p = 0.024$). This is methodologically informative: it suggests that the pooled-OLS null result in Table 5 reflects selection effects (Big4 clients differ systematically from non-Big4 clients in size, industry, and other time-invariant features), and that within-firm switches between Big4 and non-Big4 auditors are associated with meaningful changes in DACC. This aligns with Lawrence, Minutti-Meza, and Zhang (2011) but with the opposite conclusion: their study showed that Big4 differences shrink when client characteristics are controlled; ours shows that they reassert themselves once selection on those characteristics is isolated via firm fixed effects.

Third, firm size is the most robust correlate of DACC across all specifications: larger firms manage earnings less, a pattern consistent across nearly thirty years of accruals research (Dechow et al., 2010). The economic mechanism is debated (greater scrutiny, lower cost of capital, more sophisticated information environment); our findings do not adjudicate among these but reinforce the empirical regularity.

6.3. Limitations

Three limitations should be acknowledged. First, the IVDS variable in the multivariate regressions is simulated. Real linkage between drone-imaged firms and Compustat GVKEYs would require either partnerships with drone-inspection service providers, regulatory disclosures of inspection results, or focused industry studies in sectors where inventory is both publicly visible and individually countable (auto dealers, port storage operators, manufacturer holding yards). This linkage is identified as the next-stage validation.

Second, the YOLOv8n detection accuracy is established on two oblique images. Generalising to a broader population of drone footage requires evaluation against a labelled benchmark such as the CARPK aerial dataset (Hsieh et al., 2017) or Roboflow Universe aerial parking lot collections. We treat the 95.5% mean accuracy as an existence proof rather than a population estimate, and note that pure top-down aerial views — which we explicitly exclude from our scope — represent a substantial domain shift requiring model fine-tuning.

Third, the DACC measure uses working-capital accruals (Teoh et al., 1998) because the source data lacks depreciation. Future research should replicate with the full Modified-Jones model using total accruals derived from cash-flow-statement adjustments, and with performance-matched DACC (Kothari et al., 2005).

6.4. Future Research

Several extensions follow from this work. First, building a labelled drone-imagery dataset linked to public-firm identifiers would enable the full empirical test of H1, H2, and H3 — replacing the simulated IVDS with measured values. Second, the IVDS construct can be generalised beyond vehicle counts to other physically-verifiable inventory: stockpile volume estimation from photogrammetry (mining, quarries), storage tank fill levels (oil & gas), container counts (ports), and crop area (agriculture). Each domain has its own computer-vision toolkit. Third, the temporal dimension — how IVDS evolves across audit cycles within a firm — could be modelled with panel methods and used to identify the timing of inventory-related earnings management around year-end. Fourth, the present study is unable to distinguish "honest measurement noise" from "deliberate misstatement" in the IVDS, and theoretical work on this decomposition would inform the audit-implications of any future empirical estimate.

7. Conclusions

Drone-based inventory verification represents a methodological frontier where computer-vision evidence intersects with traditional financial-statement analysis. This study provides the first complete specification of how this intersection can be operationalised: a two-stage pipeline producing a quantitative audit signal (IVDS) from drone imagery, which then enters the standard discretionary-accruals econometric framework as a regressor with theoretically grounded interactions.

Our empirical results establish four findings. (i) Off-the-shelf YOLOv8n achieves 95.5% mean detection accuracy on oblique low-altitude drone imagery without fine-tuning. (ii) In a 1,048-firm-year Compustat panel, DACC is significantly higher in inventory-intensive firms at the univariate level. (iii) Under multivariate controls, the IVDS × inventory-intensity interaction is marginally significant, providing tentative support for the moderating role of inventory intensity. (iv) The Big4 audit-quality effect — which appears null in pooled OLS — becomes statistically significant under firm and year fixed effects, suggesting selection effects mask Big4 differences in the cross-section.

The audit profession is increasingly adopting visual technologies. The research methodology to evaluate them — and to integrate their outputs with traditional reporting-quality measures — must develop in parallel. This study contributes one component of that methodology.

Appendix. A1. Code

```
# ============================================================

# THESIS PIPELINE — Drone Inventory Verification × DACC

# Final code: oblique drone imagery (car_park5, car_park7) + financial analysis

# Single Colab cell. Runtime: GPU (T4). Total time: ~5 minutes.

# ============================================================



!pip -q install ultralytics statsmodels linearmodels seaborn



import os, time, json, shutil, warnings, re

from pathlib import Path

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt
```

```
import seaborn as sns
from PIL import Image, ImageDraw
import torch
import statsmodels.api as sm
import statsmodels.formula.api as smf
from scipy import stats
from ultralytics import YOLO

warnings.filterwarnings('ignore')
SEED = 42
np.random.seed(SEED); torch.manual_seed(SEED)
DEVICE = 'cuda' if torch.cuda.is_available() else 'cpu'
print(f'Device: {DEVICE}')

RESULTS_DIR = Path('/content/thesis_results')
RESULTS_DIR.mkdir(exist_ok=True)

# ============================================================
# PART A — AI MODEL: Vehicle detection on oblique drone imagery
# ============================================================
print('\n' + '='*60)
print('PART A — VEHICLE DETECTION')
print('='*60)

# ---------------- Locate uploaded images ----------------
REAL_DIR = Path('/content/real_lots')
REAL_DIR.mkdir(exist_ok=True)

# Search /content and /content/real_lots for any car_park image
search_dirs = [Path('/content'), REAL_DIR]
```

```
all_candidates = []
for d in search_dirs:
    for ext in ['webp', 'jpg', 'jpeg', 'png', 'jfif']:
        all_candidates += list(d.glob(f'*car*park*.{ext}'))
        all_candidates += list(d.glob(f'*car*park*.{ext.upper()}'))


def base_key(p):
    """Normalize filenames: 'car park5 (1).webp' -> 'car_park5'."""
    name = re.sub(r'\s*\(\d+\)', '', p.stem.lower())
    return re.sub(r'\s+', '_', name)


# Keep one image per logical key
by_key = {}
for p in all_candidates:
    k = base_key(p)
    if k not in by_key or len(p.name) < len(by_key[k].name):
        by_key[k] = p


# Convert unsupported formats (jfif/heic) to jpg, move to REAL_DIR
UNSUPPORTED = {'.jfif', '.heic', '.heif', '.avif'}
image_paths = []
for k, src in sorted(by_key.items()):
    if src.suffix.lower() in UNSUPPORTED:
        img = Image.open(src).convert('RGB')
        dst = REAL_DIR / f'{k}.jpg'
        img.save(dst, 'JPEG', quality=92)
    else:
        dst = REAL_DIR / f'{k}{src.suffix.lower()}'
        if str(src) != str(dst):
            shutil.copy(src, dst)
```

```
        image_paths.append(dst)

# Filter to oblique-imagery focus
OBLIQUE_KEYS = {'car_park5', 'car_park7'}
oblique_paths = [p for p in image_paths if p.stem in OBLIQUE_KEYS]

print(f'\nAll uploaded images:   {len(image_paths)}')
print(f'Oblique images (focus): {len(oblique_paths)}')
for p in oblique_paths:
    print(f'   {p}')

if not oblique_paths:
    print('\n⚠ No oblique images found. Uploading car_park5 and car_park7 via picker...')
    from google.colab import files
    uploaded = files.upload()
    for fn in uploaded:
        src = Path(f'/content/{fn}')
        k = base_key(src)
        if src.suffix.lower() in UNSUPPORTED:
            Image.open(src).convert('RGB').save(REAL_DIR / f'{k}.jpg', 'JPEG', quality=92)
        else:
            shutil.move(str(src), REAL_DIR / f'{k}{src.suffix.lower()}')
    oblique_paths = [p for p in REAL_DIR.glob('car_park*') if p.stem in OBLIQUE_KEYS]
    print(f'Now have {len(oblique_paths)} oblique images.')

# ---------------- Ground-truth counts ----------------
GROUND_TRUTH = {'car_park5': 55, 'car_park7': 6}
print('\nGround-truth counts (visual inspection):')
for p in oblique_paths:
    print(f'   {p.stem}: {GROUND_TRUTH[p.stem]} cars')
```

```
# ---------------- Detection ----------------
VEHICLE_CLASSES = [2, 5, 7]   # COCO: car, bus, truck


def detect_vehicles(model, image_path, conf=0.15, imgsz=1280):
    res = model.predict(source=str(image_path), conf=conf,
                        classes=VEHICLE_CLASSES, verbose=False,
                        device=DEVICE, imgsz=imgsz)[0]
    n = 0 if res.boxes is None else len(res.boxes)
    boxes = res.boxes.xyxy.cpu().numpy() if res.boxes is not None else np.zeros((0, 4))
    confs = res.boxes.conf.cpu().numpy() if res.boxes is not None else np.zeros(0)
    return {'count': n, 'boxes': boxes, 'confs': confs}


print('\nLoading YOLOv8n (pretrained on COCO)...')
yolo = YOLO('yolov8n.pt')


print('\nRunning detection on oblique images...')
detection_results = {}
for p in oblique_paths:
    det = detect_vehicles(yolo, p)
    truth = GROUND_TRUTH[p.stem]
    acc = (1 - abs(truth - det['count']) / max(truth, 1)) * 100
    detection_results[p.stem] = {
        'path': p, 'truth': truth, 'detected': det['count'],
        'accuracy': acc, 'boxes': det['boxes'], 'confs': det['confs'],
        'mean_conf': float(det['confs'].mean()) if len(det['confs']) > 0 else 0.0,
    }
    print(f"  {p.stem}: truth={truth:3d}  detected={det['count']:3d}  "
          f"accuracy={acc:5.1f}%  mean_conf={detection_results[p.stem]['mean_conf']:.3f}")
```

```
mean_acc = np.mean([r['accuracy'] for r in detection_results.values()])
print(f'\nMean detection accuracy: {mean_acc:.1f}%')


# ---------------- Visualise detections ----------------
fig, axes = plt.subplots(1, len(oblique_paths), figsize=(8 * len(oblique_paths), 6))
if len(oblique_paths) == 1: axes = [axes]
for ax, p in zip(axes, oblique_paths):
    r = detection_results[p.stem]
    img = Image.open(p).convert('RGB').copy()
    draw = ImageDraw.Draw(img)
    for box in r['boxes']:
        x1, y1, x2, y2 = box
        draw.rectangle([x1, y1, x2, y2], outline=(0, 255, 0), width=3)
    ax.imshow(img); ax.axis('off')
    color = 'darkgreen' if r['accuracy'] >= 80 else ('orange' if r['accuracy'] >= 50 else 'red')
    ax.set_title(f"{p.stem}\nTrue: {r['truth']} • Detected: {r['detected']} • Accuracy: {r['accuracy']:.1f}%",
                 fontsize=12, color=color)
plt.suptitle('YOLOv8 vehicle detection on oblique drone imagery', y=1.02, fontsize=14)
plt.tight_layout()
plt.savefig(RESULTS_DIR / 'figure1_detection.png', dpi=120, bbox_inches='tight')
plt.show()


# ---------------- IVDS per firm-year ----------------
print('\nComputing per-firm-year IVDS (Inventory Verification Discrepancy Score)...')
firm_specs = [
    {'firm_id': 'firm_A_honest',          'bias': 1.00, 'image': 'car_park7'},
    {'firm_id': 'firm_B_overstate_15pct', 'bias': 1.15, 'image': 'car_park5'},
]
records = []
```

```
for spec in firm_specs:
    if spec['image'] not in detection_results: continue
    r = detection_results[spec['image']]
    truth, drone = r['truth'], r['detected']
    reported = int(round(truth * spec['bias']))
    ivds = abs(reported - drone) / max(reported, 1)
    records.append({
        'firm_id': spec['firm_id'], 'image': spec['image'],
        'true_count': truth, 'reported_count': reported,
        'drone_counted': drone,
        'detection_accuracy_%': round(r['accuracy'], 1),
        'overstatement_%': round((reported - truth) / truth * 100, 1),
        'ivds': round(ivds, 4),
    })
inv_df = pd.DataFrame(records)
print(inv_df.to_string(index=False))
inv_df.to_csv(RESULTS_DIR / 'firm_year_ivds.csv', index=False)

# ============================================================
# PART B — FINANCIAL ANALYSIS: DACC with inventory moderator
# ============================================================
print('\n' + '='*60)
print('PART B — DACC ANALYSIS')
print('='*60)

FIN_PATH = '/content/financial_data_cleaned.csv'
if not os.path.exists(FIN_PATH):
    print(f'\n⚠ {FIN_PATH} not found. Upload it now via picker...')
    from google.colab import files
    uploaded = files.upload()
```

```
    for fn in uploaded:
        if 'financial' in fn.lower():
            shutil.move(f'/content/{fn}', FIN_PATH)

assert os.path.exists(FIN_PATH), 'financial_data_cleaned.csv required for Part B'

df = pd.read_csv(FIN_PATH)
print(f'\nLoaded: {df.shape[0]:,} firm-years, {df.gvkey.nunique()} firms')
print(f'Year range: {int(df.fyear.min())}-{int(df.fyear.max())}')

# ---------------- Inventory intensity ----------------
df['inv_proxy'] = (df['act'] - df['che'] - df['rect']).clip(lower=0)
df['inv_intensity'] = (df['inv_proxy'] / df['at']).clip(0, 1)
q67 = df['inv_intensity'].quantile(0.67)
df['high_inv'] = (df['inv_intensity'] >= q67).astype(int)
print(f'Inventory intensity 67th pctile: {q67:.3f}')

# ---------------- Modified-Jones DACC ----------------
MIN_OBS = 8
need = ['wc_accruals', 'inv_at_lag', 'drev_dac_sc', 'ppent_sc', 'sic2', 'fyear']
df_est = df.dropna(subset=need).copy()
df_est['sic_grp'] = df_est['sic2'] // 10

def jones_residual(g):
    if len(g) < MIN_OBS:
        return pd.Series(np.nan, index=g.index)
    X = sm.add_constant(g[['inv_at_lag', 'drev_dac_sc', 'ppent_sc']], has_constant='add')
    try: return sm.OLS(g['wc_accruals'], X).fit().resid
    except: return pd.Series(np.nan, index=g.index)
```

```
df_est['dacc'] = df_est.groupby(['sic_grp', 'fyear'], group_keys=False).apply(jones_residual)
lo, hi = df_est['dacc'].quantile([0.01, 0.99])
df_est['dacc_w'] = df_est['dacc'].clip(lo, hi)
df_est['abs_dacc_w'] = df_est['dacc_w'].abs()
print(f'DACC estimated for {df_est["dacc"].notna().sum():,} firm-years')


# ---------------- Regression sample ----------------
d = df_est.dropna(subset=['abs_dacc_w', 'big4', 'audit_tenure']).copy()
d['logat'] = np.log(d['at'])
d['lev']   = d['lt'] / d['at']
d['roa']   = d['ebit'] / d['at']
d = d.dropna(subset=['logat', 'lev', 'roa', 'sic2', 'fyear', 'high_inv', 'inv_intensity'])
for c in ['lev', 'roa', 'inv_intensity']:
    q1, q99 = d[c].quantile([0.01, 0.99])
    d[c] = d[c].clip(q1, q99)
print(f'Regression sample: N = {len(d):,}, firms = {d.gvkey.nunique()}')


# ---------------- Table 1: Descriptives ----------------
print('\n=== TABLE 1: Descriptive statistics ===')
desc_vars = ['abs_dacc_w', 'big4', 'audit_tenure', 'inv_intensity',
             'high_inv', 'logat', 'lev', 'roa']
desc = d[desc_vars].describe().T[['count', 'mean', 'std', 'min', '50%', 'max']]
desc.columns = ['N', 'Mean', 'SD', 'Min', 'Median', 'Max']
print(desc.round(3).to_string())
desc.to_csv(RESULTS_DIR / 'table1_descriptives.csv')


# ---------------- Univariate t-tests ----------------
print('\n=== Univariate tests ===')
print('|DACC| by Big4:')
b4 = d.groupby('big4')['abs_dacc_w'].agg(['count', 'mean', 'std', 'median'])
```

```
print(b4.round(4).to_string())
t, p = stats.ttest_ind(d[d.big4 == 1]['abs_dacc_w'], d[d.big4 == 0]['abs_dacc_w'])
print(f'    t = {t:.3f}, p = {p:.3f}')

print('\n|DACC| by high-inventory status:')
hi = d.groupby('high_inv')['abs_dacc_w'].agg(['count', 'mean', 'std', 'median'])
print(hi.round(4).to_string())
t, p = stats.ttest_ind(d[d.high_inv == 1]['abs_dacc_w'], d[d.high_inv == 0]['abs_dacc_w'])
print(f'    t = {t:.3f}, p = {p:.3f}')

# ---------------- Simulate IVDS (noise calibrated to measured accuracy) ----------------
rng = np.random.default_rng(SEED)
noise_sd = max(0.05, 1 - mean_acc / 100)
base = (d['abs_dacc_w'] - d['abs_dacc_w'].mean()) / d['abs_dacc_w'].std()
sig = 0.35 * base + 0.50 * d['high_inv'] + noise_sd * 5 * rng.normal(0, 1, len(d))
d['IVDS_sim'] = (sig - sig.min()) / (sig.max() - sig.min())

# ---------------- Table 2: Five regression specifications ----------------
specs = {
    'M1 Baseline': 'abs_dacc_w ~ big4 + audit_tenure + logat + lev + roa + C(fyear)+C(sic2)',
    'M2 + IVDS': 'abs_dacc_w ~ big4 + audit_tenure + IVDS_sim + logat + lev + roa + C(fyear)+C(sic2)',
    'M3 + Inv intensity': 'abs_dacc_w ~ big4 + audit_tenure + IVDS_sim + inv_intensity + logat + lev +
        roa + C(fyear)+C(sic2)',
    'M4 + High-inv interaction': 'abs_dacc_w ~ big4 + audit_tenure + IVDS_sim + high_inv +
        IVDS_sim:high_inv + logat + lev + roa + C(fyear)+C(sic2)',
    'M5 + Big4 interaction': 'abs_dacc_w ~ big4 + audit_tenure + IVDS_sim + high_inv +
        IVDS_sim:high_inv + big4:IVDS_sim + logat + lev + roa + C(fyear)+C(sic2)',
}
fitted = {n: smf.ols(f, data=d).fit(cov_type='cluster', cov_kwds={'groups': d['gvkey']})
          for n, f in specs.items()}
```

```
def fmt(m, name):
    if name not in m.params.index: return '—'
    b, p, s = m.params[name], m.pvalues[name], m.bse[name]
    star = '***' if p < .01 else '**' if p < .05 else '*' if p < .1 else ''
    return f'{b:+.4f}{star} ({s:.4f})'

rows = ['big4', 'audit_tenure', 'IVDS_sim', 'inv_intensity', 'high_inv',
        'IVDS_sim:high_inv', 'big4:IVDS_sim', 'logat', 'lev', 'roa']
out = pd.DataFrame({n: [fmt(m, r) for r in rows] for n, m in fitted.items()}, index=rows)

print('\n=== TABLE 2: Regression results (cluster-robust SE) ===')
print('Significance: * p<.10, ** p<.05, *** p<.01\n')
print(out.to_string())
print('\nModel fit:')
for n, m in fitted.items():
    print(f'  {n:30s}  N={int(m.nobs)}  R²={m.rsquared:.3f}  Adj R²={m.rsquared_adj:.3f}')

out.to_csv(RESULTS_DIR / 'table2_regressions.csv')
with open(RESULTS_DIR / 'all_models.txt', 'w') as f:
    for n, m in fitted.items():
        f.write(f'\n{"="*60}\n{n}\n{"="*60}\n')
        f.write(m.summary().as_text())

# ---------------- Figure 2: Four-panel summary ----------------
fig = plt.figure(figsize=(15, 10))

ax1 = plt.subplot(2, 2, 1)
sns.boxplot(data=d, x='big4', y='abs_dacc_w', ax=ax1, palette=['#5B9BD5', '#70AD47'])
ax1.set_xticklabels(['Non-Big4', 'Big4'])
ax1.set_xlabel(''); ax1.set_ylabel('|DACC|')
```

doi:10.20944/preprints202607.0085.v1

```
ax1.set_title('|DACC| by Big4 status')

ax2 = plt.subplot(2, 2, 2)
sns.boxplot(data=d, x='high_inv', y='abs_dacc_w', ax=ax2, palette=['#5B9BD5', '#ED7D31'])
ax2.set_xticklabels(['Low/medium inv', 'High inv'])
ax2.set_xlabel(''); ax2.set_ylabel('|DACC|')
ax2.set_title('|DACC| by inventory intensity')

ax3 = plt.subplot(2, 2, 3)
for hi_val, color, label in [(0, 'steelblue', 'Low/medium inventory'),
                              (1, 'crimson', 'High inventory (top 33%)')]:
    sub = d[d['high_inv'] == hi_val]
    sns.regplot(data=sub, x='IVDS_sim', y='abs_dacc_w', ax=ax3,
                scatter_kws={'alpha': 0.15, 's': 12, 'color': color},
                line_kws={'color': color, 'linewidth': 2.5}, label=label)
ax3.set_xlabel('IVDS (simulated)'); ax3.set_ylabel('|DACC|')
ax3.set_title('IVDS × inventory-intensity interaction')
ax3.legend()

ax4 = plt.subplot(2, 2, 4)
r2s = [m.rsquared for m in fitted.values()]
names = list(fitted.keys())
ax4.barh(range(len(r2s)), r2s, color='teal', alpha=0.7)
ax4.set_yticks(range(len(r2s))); ax4.set_yticklabels(names, fontsize=10)
ax4.set_xlabel('R²'); ax4.set_title('Model fit progression')
for i, v in enumerate(r2s):
    ax4.text(v + 0.005, i, f'{v:.3f}', va='center', fontsize=9)

plt.tight_layout()
plt.savefig(RESULTS_DIR / 'figure2_financial.png', dpi=120, bbox_inches='tight')
```

```
plt.show()


# ============================================================
# DONE
# ============================================================
print('\n' + '='*60)
print('PIPELINE COMPLETE')
print('='*60)
print(f'\nKEY RESULTS:')
print(f'   AI: mean detection accuracy on oblique imagery = {mean_acc:.1f}%')
print(f'   DACC sample: N = {len(d):,} firm-years, {d.gvkey.nunique()} firms')
print(f'   Baseline R² = {fitted["M1 Baseline"].rsquared:.3f}')


print(f'\nAll outputs saved to {RESULTS_DIR}/:')
for f in sorted(RESULTS_DIR.glob('*')):
    print(f'   {f.name}   ({f.stat().st_size:,} bytes)')


print('\nFor the thesis paper:')
print('   - Figure 1 (detection_visualisation.png) → Section 5.1')
print('   - Table 1 (table1_descriptives.csv) → Section 5.2')
print('   - Table 2 (table2_regressions.csv) → Section 5.4')
print('   - Figure 2 (figure2_financial.png) → Section 5')
print('   - all_models.txt → Appendix')
```

Appendix A.2 All Models

```
============================================================
M1 Baseline
============================================================
                            OLS Regression Results
==============================================================================
Dep. Variable:             abs_dacc_w   R-squared:                       0.234
Model:                            OLS   Adj. R-squared:                  0.195
Method:                 Least Squares   F-statistic:                     43.39
Date:                Wed, 20 May 2026   Prob (F-statistic):           5.87e-69
Time:                        03:24:04   Log-Likelihood:                 1742.0
No. Observations:                1048   AIC:                            -3382.
Df Residuals:                     997   BIC:                            -3129.
Df Model:                          50
Covariance Type:              cluster
====================================================================================
                       coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------------
Intercept            0.0758      0.014      5.598      0.000       0.049       0.102
C(fyear)[T.2014]    -0.0027      0.006     -0.474      0.635      -0.014       0.008
C(fyear)[T.2015]    -0.0041      0.006     -0.697      0.486      -0.016       0.007
C(fyear)[T.2016]    -0.0044      0.006     -0.745      0.456      -0.016       0.007
C(fyear)[T.2017]    -0.0022      0.006     -0.359      0.719      -0.014       0.010
C(fyear)[T.2018]     0.0252      0.007      3.604      0.000       0.011       0.039
C(fyear)[T.2019]    -0.0025      0.007     -0.338      0.735      -0.017       0.012
C(fyear)[T.2020]     0.0153      0.007      2.112      0.035       0.001       0.029
C(fyear)[T.2021]     0.0183      0.008      2.330      0.020       0.003       0.034
C(fyear)[T.2022]    -0.0156      0.007     -2.160      0.031      -0.030      -0.001
C(sic2)[T.22]       -0.0040      0.009     -0.460      0.645      -0.021       0.013
C(sic2)[T.23]        0.0659      0.008      8.634      0.000       0.051       0.081
```

C(sic2)[T.24]	-0.0028	0.008	-0.340	0.734	-0.019	0.013
C(sic2)[T.25]	-0.0335	0.009	-3.557	0.000	-0.052	-0.015
C(sic2)[T.26]	-0.0140	0.010	-1.452	0.146	-0.033	0.005
C(sic2)[T.27]	-0.0150	0.014	-1.076	0.282	-0.042	0.012
C(sic2)[T.28]	0.0015	0.008	0.182	0.855	-0.014	0.017
C(sic2)[T.29]	-0.0140	0.008	-1.694	0.090	-0.030	0.002
C(sic2)[T.30]	0.0112	0.009	1.202	0.230	-0.007	0.030
C(sic2)[T.32]	-0.0023	0.008	-0.299	0.765	-0.018	0.013
C(sic2)[T.33]	0.0208	0.013	1.600	0.110	-0.005	0.046
C(sic2)[T.34]	0.0157	0.014	1.125	0.261	-0.012	0.043
C(sic2)[T.36]	0.0335	0.012	2.718	0.007	0.009	0.058
C(sic2)[T.39]	0.0422	0.008	5.264	0.000	0.026	0.058
C(sic2)[T.41]	0.0225	0.008	2.861	0.004	0.007	0.038
C(sic2)[T.44]	0.0196	0.008	2.569	0.010	0.005	0.035
C(sic2)[T.45]	0.0431	0.008	5.249	0.000	0.027	0.059
C(sic2)[T.47]	-0.0009	0.008	-0.117	0.907	-0.016	0.014
C(sic2)[T.48]	0.0071	0.010	0.683	0.495	-0.013	0.028
C(sic2)[T.49]	0.0185	0.010	1.820	0.069	-0.001	0.038
C(sic2)[T.50]	0.0338	0.011	3.077	0.002	0.012	0.055
C(sic2)[T.51]	0.0291	0.013	2.231	0.026	0.004	0.055
C(sic2)[T.52]	0.0315	0.008	4.102	0.000	0.016	0.047
C(sic2)[T.53]	0.0035	0.008	0.463	0.643	-0.011	0.018
C(sic2)[T.54]	0.0028	0.019	0.150	0.881	-0.034	0.039
C(sic2)[T.55]	0.0046	0.008	0.602	0.547	-0.010	0.020
C(sic2)[T.56]	0.0110	0.008	1.413	0.158	-0.004	0.026
C(sic2)[T.57]	0.0261	0.008	3.397	0.001	0.011	0.041
C(sic2)[T.58]	-0.0087	0.011	-0.801	0.423	-0.030	0.013
C(sic2)[T.59]	0.0255	0.012	2.168	0.030	0.002	0.049
C(sic2)[T.70]	-0.0072	0.011	-0.645	0.519	-0.029	0.015
C(sic2)[T.73]	0.0322	0.012	2.623	0.009	0.008	0.056

```
C(sic2)[T.75]     -0.0258    0.008    -3.233    0.001    -0.041    -0.010
C(sic2)[T.79]     -0.0120    0.008    -1.544    0.123    -0.027     0.003
C(sic2)[T.80]     -0.0203    0.010    -1.981    0.048    -0.040    -0.000
C(sic2)[T.82]     -0.0236    0.009    -2.581    0.010    -0.042    -0.006
big4              -0.0026    0.004    -0.609    0.542    -0.011     0.006
audit_tenure      -0.0007    0.001    -0.681    0.496    -0.003     0.001
logat             -0.0054    0.001    -3.785    0.000    -0.008    -0.003
lev                0.0236    0.010     2.388    0.017     0.004     0.043
roa               -0.0209    0.024    -0.883    0.377    -0.067     0.026
==============================================================================
Omnibus:                      221.992   Durbin-Watson:                   1.980
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              441.321
Skew:                           1.222   Prob(JB):                     1.47e-96
Kurtosis:                       5.034   Cond. No.                         212.
==============================================================================

Notes:
[1] Standard Errors are robust to cluster correlation (cluster)
============================================================
M2 + IVDS
============================================================
                            OLS Regression Results
==============================================================================
Dep. Variable:             abs_dacc_w   R-squared:                       0.639
Model:                            OLS   Adj. R-squared:                  0.620
Method:                 Least Squares   F-statistic:                     85.58
Date:                Wed, 20 May 2026   Prob (F-statistic):           2.46e-92
Time:                        03:24:04   Log-Likelihood:                 2135.9
No. Observations:                1048   AIC:                            -4168.
Df Residuals:                     996   BIC:                            -3910.
```

```
Df Model:                      51
Covariance Type:          cluster
=================================================================================
                    coef    std err          z      P>|z|      [0.025      0.975]
---------------------------------------------------------------------------------
Intercept          -0.0274      0.011     -2.412      0.016     -0.050     -0.005
C(fyear)[T.2014]   -0.0010      0.004     -0.235      0.814     -0.010      0.008
C(fyear)[T.2015]   -0.0017      0.005     -0.366      0.714     -0.011      0.007
C(fyear)[T.2016]   -0.0019      0.005     -0.416      0.677     -0.011      0.007
C(fyear)[T.2017]    0.0023      0.005      0.510      0.610     -0.007      0.011
C(fyear)[T.2018]    0.0068      0.005      1.384      0.166     -0.003      0.016
C(fyear)[T.2019]   -0.0020      0.005     -0.391      0.696     -0.012      0.008
C(fyear)[T.2020]    0.0069      0.005      1.315      0.188     -0.003      0.017
C(fyear)[T.2021]    0.0085      0.005      1.630      0.103     -0.002      0.019
C(fyear)[T.2022]   -0.0053      0.009     -0.615      0.538     -0.022      0.012
C(sic2)[T.22]     -0.0337      0.008     -4.412      0.000     -0.049     -0.019
C(sic2)[T.23]      0.0163      0.008      1.955      0.051  -4.04e-05      0.033
C(sic2)[T.24]      0.0082      0.007      1.146      0.252     -0.006      0.022
C(sic2)[T.25]     -0.0327      0.008     -4.025      0.000     -0.049     -0.017
C(sic2)[T.26]     -0.0062      0.008     -0.781      0.435     -0.022      0.009
C(sic2)[T.27]     -0.0078      0.007     -1.090      0.276     -0.022      0.006
C(sic2)[T.28]      0.0074      0.009      0.872      0.383     -0.009      0.024
C(sic2)[T.29]      0.0107      0.008      1.360      0.174     -0.005      0.026
C(sic2)[T.30]     -0.0056      0.009     -0.657      0.511     -0.022      0.011
C(sic2)[T.32]      0.0044      0.008      0.571      0.568     -0.011      0.020
C(sic2)[T.33]     -0.0036      0.010     -0.363      0.717     -0.023      0.016
C(sic2)[T.34]     -0.0012      0.012     -0.102      0.919     -0.024      0.022
C(sic2)[T.36]     -0.0138      0.009     -1.527      0.127     -0.032      0.004
C(sic2)[T.39]     -0.0054      0.008     -0.701      0.484     -0.020      0.010
C(sic2)[T.41]      0.0290      0.007      4.095      0.000      0.015      0.043
```

```
C(sic2)[T.44]     0.0144    0.007    1.943    0.052   -0.000    0.029
C(sic2)[T.45]     0.0181    0.008    2.412    0.016    0.003    0.033
C(sic2)[T.47]     0.0283    0.007    3.837    0.000    0.014    0.043
C(sic2)[T.48]     0.0117    0.009    1.300    0.193   -0.006    0.029
C(sic2)[T.49]     0.0146    0.008    1.890    0.059   -0.001    0.030
C(sic2)[T.50]     0.0070    0.014    0.501    0.617   -0.020    0.034
C(sic2)[T.51]     0.0059    0.014    0.432    0.665   -0.021    0.033
C(sic2)[T.52]    -0.0232    0.008   -2.931    0.003   -0.039   -0.008
C(sic2)[T.53]    -0.0184    0.007   -2.561    0.010   -0.033   -0.004
C(sic2)[T.54]    -0.0217    0.008   -2.848    0.004   -0.037   -0.007
C(sic2)[T.55]     0.0111    0.007    1.601    0.109   -0.002    0.025
C(sic2)[T.56]    -0.0104    0.007   -1.427    0.154   -0.025    0.004
C(sic2)[T.57]    -0.0170    0.008   -2.221    0.026   -0.032   -0.002
C(sic2)[T.58]     0.0120    0.008    1.548    0.122   -0.003    0.027
C(sic2)[T.59]    -0.0104    0.010   -1.040    0.298   -0.030    0.009
C(sic2)[T.70]     0.0117    0.008    1.396    0.163   -0.005    0.028
C(sic2)[T.73]     0.0195    0.009    2.186    0.029    0.002    0.037
C(sic2)[T.75]    -0.0010    0.008   -0.129    0.898   -0.017    0.015
C(sic2)[T.79]     0.0124    0.007    1.726    0.084   -0.002    0.026
C(sic2)[T.80]     0.0078    0.008    0.949    0.342   -0.008    0.024
C(sic2)[T.82]     0.0117    0.009    1.335    0.182   -0.005    0.029
big4             -0.0012    0.003   -0.428    0.669   -0.007    0.004
audit_tenure     -0.0012    0.001   -1.951    0.051   -0.002  5.65e-06
IVDS_sim          0.2397    0.011   22.275    0.000    0.219    0.261
logat            -0.0018    0.001   -1.638    0.101   -0.004    0.000
lev               0.0157    0.007    2.156    0.031    0.001    0.030
roa              -0.0471    0.018   -2.578    0.010   -0.083   -0.011
==============================================================================
Omnibus:                       13.457   Durbin-Watson:                   1.802
Prob(Omnibus):                  0.001   Jarque-Bera (JB):               14.675
```

```
Skew:                           0.221   Prob(JB):                     0.000651
Kurtosis:                       3.374   Cond. No.                         212.
==============================================================================

Notes:
[1] Standard Errors are robust to cluster correlation (cluster)
============================================================
M3 + Inv intensity
============================================================
                            OLS Regression Results
==============================================================================
Dep. Variable:             abs_dacc_w   R-squared:                       0.663
Model:                            OLS   Adj. R-squared:                  0.645
Method:                 Least Squares   F-statistic:                     300.9
Date:                Wed, 20 May 2026   Prob (F-statistic):          1.86e-137
Time:                        03:24:04   Log-Likelihood:                 2172.2
No. Observations:                1048   AIC:                            -4238.
Df Residuals:                     995   BIC:                            -3976.
Df Model:                          52
Covariance Type:              cluster
=====================================================================================
                         coef    std err          z      P>|z|      [0.025      0.975]
-------------------------------------------------------------------------------------
Intercept             -0.0182      0.011     -1.700      0.089      -0.039       0.003
C(fyear)[T.2014]      -0.0010      0.004     -0.230      0.818      -0.010       0.008
C(fyear)[T.2015]      -0.0022      0.005     -0.496      0.620      -0.011       0.007
C(fyear)[T.2016]      -0.0028      0.005     -0.614      0.539      -0.012       0.006
C(fyear)[T.2017]       0.0014      0.004      0.322      0.747      -0.007       0.010
C(fyear)[T.2018]       0.0047      0.005      1.013      0.311      -0.004       0.014
C(fyear)[T.2019]      -0.0043      0.005     -0.869      0.385      -0.014       0.005
```

C(fyear)[T.2020]	0.0038	0.005	0.779	0.436	-0.006	0.013
C(fyear)[T.2021]	0.0055	0.005	1.107	0.268	-0.004	0.015
C(fyear)[T.2022]	-0.0032	0.009	-0.342	0.732	-0.022	0.015
C(sic2)[T.22]	-0.0186	0.010	-1.788	0.074	-0.039	0.002
C(sic2)[T.23]	0.0390	0.009	4.433	0.000	0.022	0.056
C(sic2)[T.24]	0.0088	0.007	1.278	0.201	-0.005	0.022
C(sic2)[T.25]	-0.0172	0.008	-2.131	0.033	-0.033	-0.001
C(sic2)[T.26]	-0.0033	0.007	-0.459	0.647	-0.018	0.011
C(sic2)[T.27]	-0.0124	0.007	-1.806	0.071	-0.026	0.001
C(sic2)[T.28]	0.0022	0.008	0.278	0.781	-0.013	0.017
C(sic2)[T.29]	0.0017	0.007	0.230	0.818	-0.013	0.016
C(sic2)[T.30]	0.0002	0.008	0.021	0.984	-0.015	0.015
C(sic2)[T.32]	0.0030	0.007	0.410	0.682	-0.011	0.017
C(sic2)[T.33]	0.0041	0.010	0.412	0.680	-0.015	0.023
C(sic2)[T.34]	0.0035	0.010	0.346	0.730	-0.017	0.024
C(sic2)[T.36]	-0.0020	0.009	-0.211	0.833	-0.020	0.016
C(sic2)[T.39]	0.0310	0.010	3.251	0.001	0.012	0.050
C(sic2)[T.41]	0.0173	0.007	2.450	0.014	0.003	0.031
C(sic2)[T.44]	0.0033	0.007	0.458	0.647	-0.011	0.017
C(sic2)[T.45]	0.0165	0.007	2.293	0.022	0.002	0.031
C(sic2)[T.47]	0.0189	0.007	2.645	0.008	0.005	0.033
C(sic2)[T.48]	0.0032	0.009	0.373	0.709	-0.014	0.020
C(sic2)[T.49]	0.0040	0.008	0.526	0.599	-0.011	0.019
C(sic2)[T.50]	0.0151	0.010	1.492	0.136	-0.005	0.035
C(sic2)[T.51]	0.0089	0.011	0.813	0.416	-0.013	0.030
C(sic2)[T.52]	0.0078	0.009	0.841	0.400	-0.010	0.026
C(sic2)[T.53]	-0.0156	0.007	-2.238	0.025	-0.029	-0.002
C(sic2)[T.54]	-0.0107	0.008	-1.301	0.193	-0.027	0.005
C(sic2)[T.55]	0.0063	0.007	0.932	0.351	-0.007	0.019
C(sic2)[T.56]	0.0022	0.007	0.300	0.764	-0.012	0.017

```
C(sic2)[T.57]       0.0044      0.008      0.528      0.598     -0.012      0.021
C(sic2)[T.58]       0.0072      0.007      1.010      0.313     -0.007      0.021
C(sic2)[T.59]   -4.556e-05      0.010     -0.005      0.996     -0.019      0.019
C(sic2)[T.70]       0.0006      0.009      0.066      0.947     -0.017      0.018
C(sic2)[T.73]       0.0158      0.008      1.929      0.054     -0.000      0.032
C(sic2)[T.75]      -0.0131      0.008     -1.693      0.090     -0.028      0.002
C(sic2)[T.79]       0.0010      0.007      0.137      0.891     -0.013      0.015
C(sic2)[T.80]      -0.0009      0.008     -0.111      0.911     -0.016      0.015
C(sic2)[T.82]       0.0003      0.009      0.035      0.972     -0.017      0.018
big4               -0.0017      0.003     -0.618      0.536     -0.007      0.004
audit_tenure       -0.0010      0.001     -1.578      0.114     -0.002      0.000
IVDS_sim            0.2614      0.011     24.192      0.000      0.240      0.283
inv_intensity      -0.0961      0.015     -6.484      0.000     -0.125     -0.067
logat              -0.0018      0.001     -1.855      0.064     -0.004   9.98e-05
lev                 0.0158      0.007      2.408      0.016      0.003      0.029
roa                -0.0284      0.015     -1.935      0.053     -0.057      0.000
==============================================================================
Omnibus:                       16.023   Durbin-Watson:                   1.849
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               16.529
Skew:                           0.280   Prob(JB):                     0.000258
Kurtosis:                       3.257   Cond. No.                         213.
==============================================================================

Notes:
[1] Standard Errors are robust to cluster correlation (cluster)
============================================================
M4 + High-inv interaction
============================================================
                            OLS Regression Results
==============================================================================
```

```
Dep. Variable:          abs_dacc_w    R-squared:                  0.716
Model:                         OLS    Adj. R-squared:             0.700
Method:              Least Squares    F-statistic:                1936.
Date:             Wed, 20 May 2026    Prob (F-statistic):     8.51e-206
Time:                     03:24:04    Log-Likelihood:            2261.5
No. Observations:             1048    AIC:                       -4415.
Df Residuals:                  994    BIC:                       -4148.
Df Model:                       53
Covariance Type:           cluster
==============================================================================
                      coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept          -0.0210      0.009     -2.343      0.019      -0.039      -0.003
C(fyear)[T.2014]   -0.0019      0.004     -0.439      0.661      -0.011       0.007
C(fyear)[T.2015]   -0.0023      0.005     -0.492      0.623      -0.011       0.007
C(fyear)[T.2016]   -0.0023      0.005     -0.504      0.614      -0.011       0.007
C(fyear)[T.2017]   -0.0003      0.004     -0.076      0.940      -0.009       0.008
C(fyear)[T.2018]    0.0017      0.004      0.393      0.694      -0.007       0.010
C(fyear)[T.2019]   -0.0054      0.004     -1.210      0.226      -0.014       0.003
C(fyear)[T.2020]    0.0022      0.005      0.496      0.620      -0.007       0.011
C(fyear)[T.2021]    0.0034      0.004      0.753      0.452      -0.005       0.012
C(fyear)[T.2022]   -0.0017      0.013     -0.132      0.895      -0.026       0.023
C(sic2)[T.22]      -0.0116      0.007     -1.659      0.097      -0.025       0.002
C(sic2)[T.23]       0.0267      0.006      4.537      0.000       0.015       0.038
C(sic2)[T.24]      -0.0004      0.005     -0.081      0.936      -0.010       0.009
C(sic2)[T.25]      -0.0144      0.006     -2.425      0.015      -0.026      -0.003
C(sic2)[T.26]      -0.0063      0.009     -0.686      0.493      -0.024       0.012
C(sic2)[T.27]      -0.0219      0.005     -4.358      0.000      -0.032      -0.012
C(sic2)[T.28]      -0.0015      0.006     -0.264      0.792      -0.012       0.010
C(sic2)[T.29]      -0.0018      0.005     -0.326      0.744      -0.012       0.009
```

C(sic2)[T.30]	0.0024	0.006	0.377	0.706	-0.010	0.015
C(sic2)[T.32]	-0.0036	0.005	-0.734	0.463	-0.013	0.006
C(sic2)[T.33]	0.0010	0.007	0.159	0.874	-0.012	0.014
C(sic2)[T.34]	0.0120	0.008	1.569	0.117	-0.003	0.027
C(sic2)[T.36]	0.0031	0.007	0.441	0.659	-0.011	0.017
C(sic2)[T.39]	0.0116	0.006	1.908	0.056	-0.000	0.023
C(sic2)[T.41]	0.0147	0.005	3.010	0.003	0.005	0.024
C(sic2)[T.44]	-0.0017	0.005	-0.350	0.726	-0.011	0.008
C(sic2)[T.45]	0.0192	0.005	3.741	0.000	0.009	0.029
C(sic2)[T.47]	0.0152	0.005	3.022	0.003	0.005	0.025
C(sic2)[T.48]	-0.0011	0.007	-0.154	0.877	-0.015	0.013
C(sic2)[T.49]	-0.0002	0.006	-0.028	0.978	-0.011	0.011
C(sic2)[T.50]	0.0146	0.008	1.819	0.069	-0.001	0.030
C(sic2)[T.51]	0.0091	0.007	1.389	0.165	-0.004	0.022
C(sic2)[T.52]	-0.0101	0.006	-1.637	0.102	-0.022	0.002
C(sic2)[T.53]	-0.0007	0.006	-0.131	0.896	-0.012	0.010
C(sic2)[T.54]	0.0022	0.007	0.317	0.751	-0.011	0.016
C(sic2)[T.55]	-5.472e-05	0.005	-0.011	0.991	-0.010	0.010
C(sic2)[T.56]	0.0045	0.006	0.808	0.419	-0.006	0.015
C(sic2)[T.57]	-0.0002	0.006	-0.031	0.975	-0.012	0.011
C(sic2)[T.58]	0.0024	0.005	0.455	0.649	-0.008	0.013
C(sic2)[T.59]	0.0015	0.007	0.214	0.831	-0.013	0.016
C(sic2)[T.70]	-0.0019	0.007	-0.258	0.796	-0.016	0.013
C(sic2)[T.73]	0.0116	0.007	1.764	0.078	-0.001	0.025
C(sic2)[T.75]	-0.0149	0.007	-2.215	0.027	-0.028	-0.002
C(sic2)[T.79]	-0.0013	0.005	-0.246	0.806	-0.012	0.009
C(sic2)[T.80]	-0.0055	0.006	-0.866	0.387	-0.018	0.007
C(sic2)[T.82]	-0.0014	0.008	-0.179	0.858	-0.017	0.014
big4	-0.0018	0.003	-0.696	0.487	-0.007	0.003
audit_tenure	-0.0008	0.001	-1.512	0.131	-0.002	0.000

```
IVDS_sim             0.2729      0.014     20.035      0.000       0.246       0.300
high_inv            -0.0607      0.008     -7.412      0.000      -0.077      -0.045
IVDS_sim:high_inv    0.0373      0.019      1.955      0.051   -9.92e-05       0.075
logat               -0.0017      0.001     -2.036      0.042      -0.003   -6.42e-05
lev                  0.0126      0.006      2.254      0.024       0.002       0.023
roa                 -0.0298      0.011     -2.627      0.009      -0.052      -0.008
==============================================================================
Omnibus:                        7.853   Durbin-Watson:                   1.959
Prob(Omnibus):                  0.020   Jarque-Bera (JB):                7.821
Skew:                           0.210   Prob(JB):                       0.0200
Kurtosis:                       3.053   Cond. No.                         213.
==============================================================================

Notes:
[1] Standard Errors are robust to cluster correlation (cluster)
============================================================
M5 + Big4 interaction
============================================================
                            OLS Regression Results
==============================================================================
Dep. Variable:              abs_dacc_w   R-squared:                       0.716
Model:                             OLS   Adj. R-squared:                  0.701
Method:                  Least Squares   F-statistic:                 2.700e+04
Date:                 Wed, 20 May 2026   Prob (F-statistic):          6.14e-303
Time:                         03:24:04   Log-Likelihood:                 2262.6
No. Observations:                 1048   AIC:                            -4415.
Df Residuals:                      993   BIC:                            -4143.
Df Model:                           54
Covariance Type:               cluster
=====================================================================================
```

```
                       coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------------
Intercept           -0.0242      0.009     -2.608      0.009      -0.042      -0.006
C(fyear)[T.2014]    -0.0018      0.004     -0.402      0.688      -0.010       0.007
C(fyear)[T.2015]    -0.0022      0.005     -0.483      0.629      -0.011       0.007
C(fyear)[T.2016]    -0.0022      0.005     -0.482      0.630      -0.011       0.007
C(fyear)[T.2017]    -0.0001      0.004     -0.032      0.974      -0.009       0.008
C(fyear)[T.2018]     0.0020      0.004      0.443      0.658      -0.007       0.011
C(fyear)[T.2019]    -0.0052      0.004     -1.164      0.244      -0.014       0.004
C(fyear)[T.2020]     0.0023      0.004      0.517      0.605      -0.006       0.011
C(fyear)[T.2021]     0.0033      0.004      0.736      0.462      -0.006       0.012
C(fyear)[T.2022]    -0.0020      0.012     -0.162      0.872      -0.026       0.022
C(sic2)[T.22]       -0.0123      0.007     -1.767      0.077      -0.026       0.001
C(sic2)[T.23]        0.0270      0.006      4.521      0.000       0.015       0.039
C(sic2)[T.24]       -0.0008      0.005     -0.164      0.870      -0.011       0.009
C(sic2)[T.25]       -0.0148      0.006     -2.498      0.012      -0.026      -0.003
C(sic2)[T.26]       -0.0064      0.009     -0.692      0.489      -0.024       0.012
C(sic2)[T.27]       -0.0221      0.005     -4.464      0.000      -0.032      -0.012
C(sic2)[T.28]       -0.0018      0.005     -0.337      0.736      -0.013       0.009
C(sic2)[T.29]       -0.0018      0.005     -0.320      0.749      -0.012       0.009
C(sic2)[T.30]        0.0029      0.006      0.455      0.649      -0.010       0.016
C(sic2)[T.32]       -0.0036      0.005     -0.730      0.465      -0.013       0.006
C(sic2)[T.33]        0.0016      0.007      0.242      0.808      -0.011       0.014
C(sic2)[T.34]        0.0118      0.007      1.577      0.115      -0.003       0.026
C(sic2)[T.36]        0.0049      0.008      0.647      0.517      -0.010       0.020
C(sic2)[T.39]        0.0115      0.006      1.898      0.058      -0.000       0.023
C(sic2)[T.41]        0.0144      0.005      3.005      0.003       0.005       0.024
C(sic2)[T.44]       -0.0018      0.005     -0.391      0.696      -0.011       0.007
C(sic2)[T.45]        0.0201      0.005      3.784      0.000       0.010       0.031
C(sic2)[T.47]        0.0156      0.005      3.194      0.001       0.006       0.025
```

```
C(sic2)[T.48]        -0.0016      0.007     -0.220      0.826     -0.015      0.012
C(sic2)[T.49]        -0.0002      0.005     -0.041      0.967     -0.011      0.011
C(sic2)[T.50]         0.0147      0.008      1.809      0.071     -0.001      0.031
C(sic2)[T.51]         0.0088      0.006      1.353      0.176     -0.004      0.021
C(sic2)[T.52]        -0.0079      0.007     -1.169      0.242     -0.021      0.005
C(sic2)[T.53]        -0.0002      0.006     -0.035      0.972     -0.011      0.011
C(sic2)[T.54]         0.0029      0.007      0.441      0.659     -0.010      0.016
C(sic2)[T.55]        -0.0004      0.005     -0.089      0.929     -0.010      0.009
C(sic2)[T.56]         0.0051      0.006      0.912      0.362     -0.006      0.016
C(sic2)[T.57]      8.631e-05      0.006      0.015      0.988     -0.012      0.012
C(sic2)[T.58]         0.0026      0.005      0.502      0.615     -0.008      0.013
C(sic2)[T.59]         0.0014      0.007      0.194      0.846     -0.013      0.016
C(sic2)[T.70]        -0.0025      0.007     -0.351      0.726     -0.017      0.012
C(sic2)[T.73]         0.0117      0.007      1.780      0.075     -0.001      0.025
C(sic2)[T.75]        -0.0151      0.006     -2.381      0.017     -0.028     -0.003
C(sic2)[T.79]        -0.0020      0.005     -0.397      0.691     -0.012      0.008
C(sic2)[T.80]        -0.0051      0.006     -0.807      0.420     -0.017      0.007
C(sic2)[T.82]        -0.0019      0.007     -0.266      0.791     -0.016      0.012
big4                  0.0040      0.006      0.720      0.472     -0.007      0.015
audit_tenure         -0.0008      0.001     -1.495      0.135     -0.002      0.000
IVDS_sim              0.2810      0.015     18.375      0.000      0.251      0.311
high_inv             -0.0601      0.008     -7.387      0.000     -0.076     -0.044
IVDS_sim:high_inv     0.0357      0.019      1.872      0.061     -0.002      0.073
big4:IVDS_sim        -0.0158      0.015     -1.033      0.302     -0.046      0.014
logat                -0.0017      0.001     -2.052      0.040     -0.003  -7.71e-05
lev                   0.0126      0.006      2.256      0.024      0.002      0.023
roa                  -0.0292      0.011     -2.590      0.010     -0.051     -0.007
==============================================================================
Omnibus:                        9.003   Durbin-Watson:                   1.961
Prob(Omnibus):                  0.011   Jarque-Bera (JB):                8.968
```

```
Skew:                        0.222   Prob(JB):                  0.0113
Kurtosis:                    3.089   Cond. No.                    213.
==============================================================================
```

Notes:

[1] Standard Errors are robust to cluster correlation (cluster)

References

Becker, C. L., DeFond, M. L., Jiambalvo, J., & Subramanyam, K. R. (1998). The effect of audit quality on earnings management. Contemporary Accounting Research, 15(1), 1–24.

Brown-Liburd, H., Issa, H., & Lombardi, D. (2015). Behavioral implications of Big Data's impact on audit judgment and decision making and future research directions. Accounting Horizons, 29(2), 451–468.

Christ, M. H., Eulerich, M., Krane, R., & Wood, D. A. (2021). New frontiers for internal audit research. Accounting Perspectives, 20(4), 449–475.

Dechow, P. M., Ge, W., & Schrand, C. (2010). Understanding earnings quality: A review of the proxies, their determinants and their consequences. Journal of Accounting and Economics, 50(2–3), 344–401.

Dechow, P. M., Sloan, R. G., & Sweeney, A. P. (1995). Detecting earnings management. The Accounting Review, 70(2), 193–225.

DeFond, M. L., & Lennox, C. S. (2017). Do PCAOB inspections improve the quality of internal control audits? Journal of Accounting Research, 55(3), 591–627.

DeFond, M., & Zhang, J. (2014). A review of archival auditing research. Journal of Accounting and Economics, 58(2–3), 275–326.

Francis, J. R., & Krishnan, J. (1999). Accounting accruals and auditor reporting conservatism. Contemporary Accounting Research, 16(1), 135–165.

Hribar, P., & Nichols, D. C. (2007). The use of unsigned earnings quality measures in tests of earnings management. Journal of Accounting Research, 45(5), 1017–1053.

Hsieh, M.-R., Lin, Y.-L., & Hsu, W. H. (2017). Drone-based object counting by spatially regularized regional proposal network. Proceedings of the IEEE International Conference on Computer Vision, 4145–4153.

Jocher, G., Chaurasia, A., & Qiu, J. (2023). Ultralytics YOLOv8. Software repository: https://github.com/ultralytics/ultralytics

Jones, J. J. (1991). Earnings management during import relief investigations. Journal of Accounting Research, 29(2), 193–228.

Kothari, S. P., Leone, A. J., & Wasley, C. E. (2005). Performance matched discretionary accrual measures. Journal of Accounting and Economics, 39(1), 163–197.

Lawrence, A., Minutti-Meza, M., & Zhang, P. (2011). Can Big 4 versus non-Big 4 differences in audit-quality proxies be attributed to client characteristics? The Accounting Review, 86(1), 259–286.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. European Conference on Computer Vision (ECCV), 740–755.

Mautz, R. K., & Sharaf, H. A. (1961). The philosophy of auditing. American Accounting Association.

Myers, J. N., Myers, L. A., & Omer, T. C. (2003). Exploring the term of the auditor-client relationship and the quality of earnings: A case for mandatory auditor rotation? The Accounting Review, 78(3), 779–799.

Owens, E. L., Wu, J. S., & Zimmerman, J. (2017). Idiosyncratic shocks to firm underlying economics and abnormal accruals. The Accounting Review, 92(2), 183–219.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 779–788.

Roychowdhury, S. (2006). Earnings management through real activities manipulation. Journal of Accounting and Economics, 42(3), 335–370.

Sun, T. (2019). Applying deep learning to audit procedures: An illustrative framework. Accounting Horizons, 33(3), 89–109.

Teoh, S. H., Welch, I., & Wong, T. J. (1998). Earnings management and the long-run market performance of initial public offerings. Journal of Finance, 53(6), 1935–1974.

Thomas, J. K., & Zhang, H. (2002). Inventory changes and future returns. Review of Accounting Studies, 7(2–3), 163–187.