Open Journal of Accounting, 2026, 15(4), 301-325
https://www.scirp.org/journal/ojacct
ISSN Online: 2169-3412
ISSN Print: 2169-3404

Determinants of Capital Structure, Cash Holdings, and Earnings Quality among ASX-Listed Firms: Cross-Sectional Evidence for FY2025

Ibrahim Majed M. Hablayn

Monash University, Melbourne, Australia
Email: Ihab0003@student.monash.edu

How to cite this paper: Hablayn, I. M. M. (2026). Determinants of Capital Structure, Cash Holdings, and Earnings Quality among ASX-Listed Firms: Cross-Sectional Evidence for FY2025. *Open Journal of Accounting, 15,* 301-325.
https://doi.org/10.4236/ojacct.2026.154015

Received: June 26, 2026
Accepted: August 23, 2026
Published: August 26, 2026

Abstract

This study examines the firm-level determinants of capital structure, cash holdings, profitability, and the consistency between accruals and operating cash flows for a sample of 113 firms listed on the Australian Securities Exchange (ASX), using audited financial-statement data for fiscal years 2024 and 2025. Working from a standardized panel of eleven accounting ratios, we estimate four ordinary least squares (OLS) models with heteroskedasticity-robust standard errors and subject each to an extensive robustness programme (winsorization, sample screens, alternative covariance estimators, firm-type fixed effects, a pooled panel with firm-clustered standard errors, and influence diagnostics). We find strong and stable evidence that firm size is positively associated with leverage, consistent with trade-off theory, and qualified evidence that profitability is negatively associated with leverage, consistent with the pecking-order view; the latter result is sensitive to the treatment of outliers but strengthens under sample expansion, fixed effects, and influence-trimming. Larger firms hold less cash relative to assets, while the relation between leverage and cash holdings is statistically insignificant. Accruals are strongly negatively related to operating cash flow, in line with the established accruals-cash-flow relation. Because the accrual measure is constructed as net profit minus operating cash flow, we interpret this model as an accrual-cash-flow consistency check rather than a standalone test of earnings quality; an alternative specification omitting the mechanical regressor confirms the negative relation. The determinants of contemporaneous return on assets are weakly identified in this cross-section. The findings are broadly consistent with the international capital-structure literature while highlighting the limitations of a small, single-year, accounting-only sample. We discuss implications and outline extensions using market data.

Keywords

Capital Structure, Pecking Order, Trade-off Theory, Cash Holdings, Accruals, Earnings Quality, ASX, Australia

1. Introduction

How firms choose their mix of debt and equity, how much liquidity they retain, and how reliably their reported earnings map into cash flows are among the most durable questions in corporate finance and financial accounting. Despite decades of research, the empirical determinants of these choices remain contested and context-dependent, and most large-sample evidence is drawn from the United States. Evidence from other institutional settings—for example, on how auditor tenure and audit-firm type shape the quality of financial reports on the Saudi Stock Exchange (Alhazmi et al., 2024)—reinforces the view that these relations are conditioned by local market structure. The Australian market offers a distinctive setting: it combines a dividend-imputation tax system that alters the relative tax treatment of debt and equity (Twite, 2001) with a market capitalization that is heavily concentrated in financial institutions, for which conventional leverage and liquidity ratios are not comparable to those of industrial firms.

This paper provides cross-sectional evidence on four related questions for ASX-listed firms in fiscal year 2025. First, do the classic determinants of leverage—profitability and size—operate in the directions predicted by the trade-off and pecking-order theories? Second, what firm characteristics explain corporate cash holdings? Third, how closely do accruals track operating cash flow—a consistency property linking reported earnings to cash generation? Fourth, can simple firm characteristics explain contemporaneous profitability? We address these questions using a standardized panel of accounting ratios constructed from audited income statements, balance sheets, and cash-flow statements for 113 firms.

Our approach is deliberately transparent about data quality and robustness. Because the underlying statements are heterogeneous and several firms are financial institutions or report negative book equity, we conduct explicit data-integrity checks, apply documented sample screens, winsorize ratios to limit the influence of extreme observations, and report each result across a battery of alternative specifications. The contribution is therefore threefold: we assemble and document a clean cross-section of ASX accounting variables; we test a coherent set of theory-derived hypotheses on that cross-section; and we are candid about the statistical fragilities—small sample, single year, and absence of market data—that condition the inferences.

We preview the main results. Size is positively and very robustly associated with leverage; profitability is negatively associated with leverage in most specifications, though the baseline estimate is marginal and outlier-sensitive. Larger firms hold proportionally less cash. Accruals are strongly negatively related to operating cash flow. The determinants of return on assets are weakly identified. These patterns align with the dominant findings of the international literature while underscoring the caution warranted by the sample.

The remainder of the paper is organized as follows. Section 2 reviews the relevant literature and develops the hypotheses. Section 3 describes the data, variable measurement, data-quality checks, and empirical models. Section 4 reports the results. Section 5 discusses their interpretation, and Section 6 concludes.

2. Literature Review and Hypotheses

2.1. Capital Structure: Trade-Off and Pecking-Order Theories

The modern theory of capital structure originates with Modigliani and Miller (1958), whose irrelevance proposition established the frictionless benchmark against which subsequent theories introduce taxes, distress costs, and information asymmetry. The trade-off theory holds that firms balance the tax advantages of debt against expected costs of financial distress, implying an interior optimal leverage ratio (Kraus & Litzenberger, 1973; Myers, 1984). Because larger, more diversified firms face lower expected distress costs and enjoy better access to debt markets, the trade-off view predicts a positive association between firm size and leverage—one of the most reliably documented regularities in the empirical literature (Rajan & Zingales, 1995; Titman & Wessels, 1988; Frank & Goyal, 2009).

The pecking-order theory, by contrast, emphasizes information asymmetry between managers and outside investors (Myers, 1984; Myers & Majluf, 1984). Firms are predicted to prefer internal funds to external finance and debt to equity; consequently, more profitable firms—which generate more retained earnings—are expected to carry less debt. The negative profitability-leverage relation is among the most consistent empirical findings across markets (Rajan & Zingales, 1995; Fama & French, 2002; Frank & Goyal, 2009), and is frequently interpreted as evidence favouring the pecking order over a strict static trade-off (Harris & Raviv, 1991). Accordingly, we test the two leverage predictions separately:

H1 (Pecking order). Profitability is negatively associated with leverage.

H2 (Trade-off, size). Firm size is positively associated with leverage.

2.2. Cash Holdings

The determinants of corporate liquidity have been studied extensively since Opler, Pinkowitz, Stulz, and Williamson (1999), who show that cash holdings reflect a trade-off between the precautionary benefits of liquidity and its opportunity cost. Larger firms, which enjoy superior access to external capital and greater diversification of cash-flow shocks, are predicted to hold proportionally less cash (Opler et al., 1999; Bates, Kahle, & Stulz, 2009). Leverage is also expected to be negatively associated with cash, both because debt capacity substitutes for precautionary balances and because cash can be viewed as negative debt (Bates et al., 2009; Almeida, Campello, & Weisbach, 2004). We therefore test:

H3 (Cash holdings). Leverage and firm size are each negatively associated with cash holdings.

2.3. Accruals and the Accrual-Cash-Flow Relation

Accounting earnings differ from cash flows through accruals, and the properties of accruals are central to the measurement of earnings quality (Dechow, 1994; Dechow, Ge, & Schrand, 2010), and accrual-based measures remain the standard gauge of reporting quality across institutional settings, including recent evidence on artificial-intelligence adoption and financial reporting quality on emerging exchanges (Alhazmi et al., 2025b). A well-established empirical regularity is the strong negative contemporaneous relation between accruals and operating cash flow, which reflects the smoothing role of accruals and the timing differences they capture (Dechow, 1994; Dechow & Dichev, 2002). Sloan (1996) further shows that the accrual and cash-flow components of earnings have different persistence, motivating attention to their relation. A pronounced negative accrual-cash-flow association is consistent with normal accrual behaviour rather than systematic manipulation. Because our accrual measure is constructed residually as net profit minus operating cash flow scaled by total assets (**Table 1**), the hypothesis below functions as a consistency check on the accrual-cash-flow relation rather than as a standalone test of earnings quality, which would require a modelled discretionary-accrual benchmark (Jones, 1991; Kothari, Leone, & Wasley, 2005). We test:

H4 (Accrual-cash-flow consistency). Accruals are negatively associated with operating cash flow.

2.4. Determinants of Profitability

Finally, we examine whether basic firm characteristics explain contemporaneous profitability. Firms generating stronger operating cash flow and larger firms benefiting from scale economies are expected to report higher returns on assets, whereas firms undertaking rapid asset expansion may experience lower contemporaneous returns, a pattern related to the asset-growth effect documented by Cooper, Gulen, and Schill (2008). We test:

H5 (Profitability). Return on assets is positively associated with operating cash flow and firm size, and negatively associated with asset growth.

3. Data and Methodology

3.1. Sample

The sample comprises 113 firms listed on the ASX, observed across the 2024 and 2025 fiscal years to yield a total of 226 firm-year observations. The underlying variables are derived from audited consolidated income statements, balance sheets, and statements of cash flows. These financial statements were obtained directly from the ASX Market Announcements Platform, from which the necessary accounting aggregates were extracted. The 113 firms represent all ASX-listed entities for which complete, machine-readable financial statements were available for both FY2024 and FY2025 at the date of data collection. Accordingly, the sample constitutes a specific subset of ASX-listed entities rather than a full-population screen, meaning empirical inferences apply directly to this sample rather than to the exchange as a whole.

While the broader panel is noted, the primary analysis focuses on a cross-section of the 2025 fiscal year. Because several accounting ratios are undefined or non-comparable for financial institutions, and because negative book equity distorts equity-scaled metrics, these entities were removed to establish the baseline sample. Of the initial 113 firms, 28 are financial institutions (comprising 19 banks and 9 insurers), 3 report negative book equity, and 7 report their financials in a foreign currency. This currency variation only impacts the absolute size variable, as the computed financial ratios remain scale- and currency-free. Excluding these observations leaves a baseline sample of 75 non-financial, positive-equity firms, with individual regression models employing between 63 and 70 observations after accounting for listwise deletion of missing values.

3.2. Variable Measurement

Table 1. Variable definitions.

Variable	Definition
size_ln_assets	Natural log of total assets (thousands of the firm's reporting currency)
ROA	Net profit/total assets
ROE	Net profit/total equity
leverage	Total liabilities/total assets
equity_ratio	Total equity/total assets (=1 − leverage)
ocf_to_assets	Operating cash flow/total assets
cash_to_assets	Cash and equivalents/total assets
accruals	(Net profit − operating cash flow)/total assets
asset_growth	Total assets(t)/total assets (t − 1) − 1
eff_tax_rate	−Tax expense/pre-tax profit (defined for pre-tax profit > 0)

Continuous ratios are winsorized at the 1st and 99th percentiles for estimation; winsorized variables are denoted with a _w suffix.

Table 1 defines the variables. All ratios are computed from standardized accounting aggregates; numerator and denominator share each firm's reporting unit, so ratios are unaffected by currency or scale. Firm size is the natural logarithm of total assets expressed in thousands of each firm's reporting currency. Seven firms report in a currency other than the Australian dollar (three in USD and one each in SGD, NZD, PGK, and CNY), of which four enter the baseline sample; their total assets were not converted to AUD before the logarithm was taken [confirm—if assets were instead converted to AUD, state the exchange-rate convention used, e.g., the RBA spot rate at each firm's fiscal year-end]. Because size enters in logarithms, a currency difference amounts to an additive firm-specific shift in the regressor and is therefore a potential source of measurement error confined to the size variable. As a robustness check, re-estimating the two size-sensitive models after excluding the foreign-currency reporters leaves the size coefficients essentially unchanged (M1: $\beta = 0.090$, $p < 0.001$, N = 61; M2: $\beta = -0.044$, $p = 0.001$, N = 65).

3.3. Data Quality

Several integrity checks precede estimation. The panel contains no duplicate firm-year keys and is balanced across the two years (113 firms per year). The accounting identity linking the equity ratio and leverage holds exactly: equity_ratio + leverage has a mean of 1.000 with a maximum absolute deviation of 0.000, confirming internal consistency. Missing-data rates vary by variable: the effective tax rate is missing for 39.8% of FY2025 firms (it is undefined when pre-tax profit is non-positive, which is common among loss-making firms), accruals for 15.0%, return on equity for 14.2%, and return on assets for 12.4%, while size is missing for only 2.7%. Five firms report effective tax rates outside the plausible −0.5 to 0.75 range and are flagged. Distributions of ROA, ROE, and accruals are highly right-skewed (skewness exceeding 7.5 and excess kurtosis exceeding 60 in **Table 2**), driven principally by one micro-capitalization firm that recorded a large one-off gain on the disposal of subsidiaries against a small remaining asset base; winsorization is applied to mitigate the influence of such observations, and medians are emphasized alongside means.

Estimation proceeds by listwise deletion: each regression uses the baseline firms with complete data on that model's variables, which is why sample sizes vary between 63 and 70. Firms omitted in this way do not differ materially from retained firms on size: mean log assets of omitted and retained firms differ by at most 0.85 (11.45 versus 12.30, in the ROA model), and Welch two-sample tests do not reject equality of means in any of the four models (all $p \geq 0.11$). Firm type cannot drive the omissions, because the baseline sample is entirely non-financial. Omission is instead driven principally by missing profitability and accrual variables—typically loss-making firms with incomplete disclosures—and by asset growth, which is undefined where prior-year totals are unavailable.

3.4. Empirical Models

We estimate four OLS regressions, one per hypothesis, on the baseline sample using winsorized variables and heteroskedasticity-consistent (HC3) standard errors (White, 1980):

- **M1 (Leverage):** leverage = $\beta0 + \beta1 \cdot$ROA + $\beta2 \cdot$size + $\beta3 \cdot$asset_growth + $\beta4 \cdot$cash_to_assets + ε—tests H1 ($\beta1 < 0$) and H2 ($\beta2 > 0$).
- **M2 (Cash):** cash_to_assets = $\beta0 + \beta1 \cdot$leverage + $\beta2 \cdot$size + $\beta3 \cdot$ocf_to_assets + $\beta4 \cdot$asset_growth + ε—tests H3 ($\beta1 < 0$, $\beta2 < 0$).

- **M3 (Accruals):** accruals = β0 + β1·ocf_to_assets + β2·size + β3·ROA + ε—tests H4 (β1 < 0). Because accruals ≡ ROA − ocf_to_assets by construction, including both regressors brings M3 close to an accounting identity; Section 4.3 therefore also reports an alternative specification (M3′) that omits ROA.
- **M4 (ROA):** ROA = β0 + β1·size + β2·ocf_to_assets + β3·asset_growth + β4·leverage + ε—tests H5 (β2 > 0, β1 > 0, β3 < 0).

Leverage and the equity ratio are perfect linear complements and are never entered jointly. Robustness checks include re-estimation with raw (non-winsorized) variables, expansion to the full FY2025 sample, alternative covariance estimators (HC1, HC3, classical), firm-type fixed effects, a pooled two-year specification with standard errors clustered by firm (Petersen, 2009), and influence diagnostics based on Cook's distance (Cook, 1977).

All four models identify conditional associations rather than causal effects. Leverage, cash holdings, and profitability are measured contemporaneously and are plausibly jointly determined—for example, profitable firms may choose lower leverage while leverage simultaneously affects measured profitability—and the cross-section offers no instruments or timing structure with which to resolve this simultaneity. Coefficients are therefore interpreted throughout as partial correlations that are consistent with, but not dispositive of, the underlying theories. The following section presents the results, which are detailed in Appendix.

4. Results

4.1. Descriptive Statistics

Table 2 reports descriptive statistics for the baseline sample. The mean log of total assets is 12.13 (median 12.11), spanning a wide range (7.50 to 17.35) that reflects the inclusion of both micro-capitalization and large firms. Mean leverage is 0.440 (median 0.403) and the equity ratio averages 0.560, indicating that the typical non-financial firm finances a little under half of its assets with liabilities. Cash holdings average 17.7% of assets (median 11.1%). The mean return on assets (0.123) is inflated by the extreme right tail noted above; the median of 0.031 is more representative. The large skewness and kurtosis of ROA, ROE, and accruals motivate the winsorization applied in estimation.

Table 2. Descriptive statistics (baseline sample, FY2025).

Variable	N	Mean	SD	Min	25%	Median	75%	Max	Skew	Kurt.
size_ln_assets	80	12.132	2.332	7.502	10.597	12.111	13.881	17.348	−0.03	−0.75
ROA	72	0.123	1.123	−1.027	−0.003	0.031	0.091	9.277	7.80	64.55
ROE	70	0.225	1.857	−2.470	−0.001	0.097	0.150	15.155	7.70	63.04
leverage	77	0.440	0.319	0.001	0.148	0.403	0.703	0.984	0.42	−1.21
equity_ratio	77	0.560	0.319	0.016	0.297	0.597	0.852	0.999	−0.42	−1.21
ocf_to_assets	77	0.031	0.255	−1.006	−0.008	0.038	0.102	0.979	−1.04	7.53

Continued

cash_to_assets	77	0.177	0.200	0.000	0.047	0.111	0.219	0.925	1.87	3.19
accruals	70	0.111	1.241	−1.134	−0.047	−0.013	0.024	10.222	8.04	66.52
asset_growth	77	0.075	0.253	−0.778	−0.004	0.056	0.155	0.890	0.10	3.89
eff_tax_rate	47	0.026	0.590	−2.693	−0.000	0.201	0.295	0.632	−3.40	12.84

The shape of these distributions is shown in Figure 1: profitability and accruals are visibly dominated by a single observation in the extreme right tail, whereas leverage, firm size, and the equity ratio are more evenly dispersed, with leverage and the equity ratio displaying the bimodality expected of a sample combining lowly- and highly-levered firms.

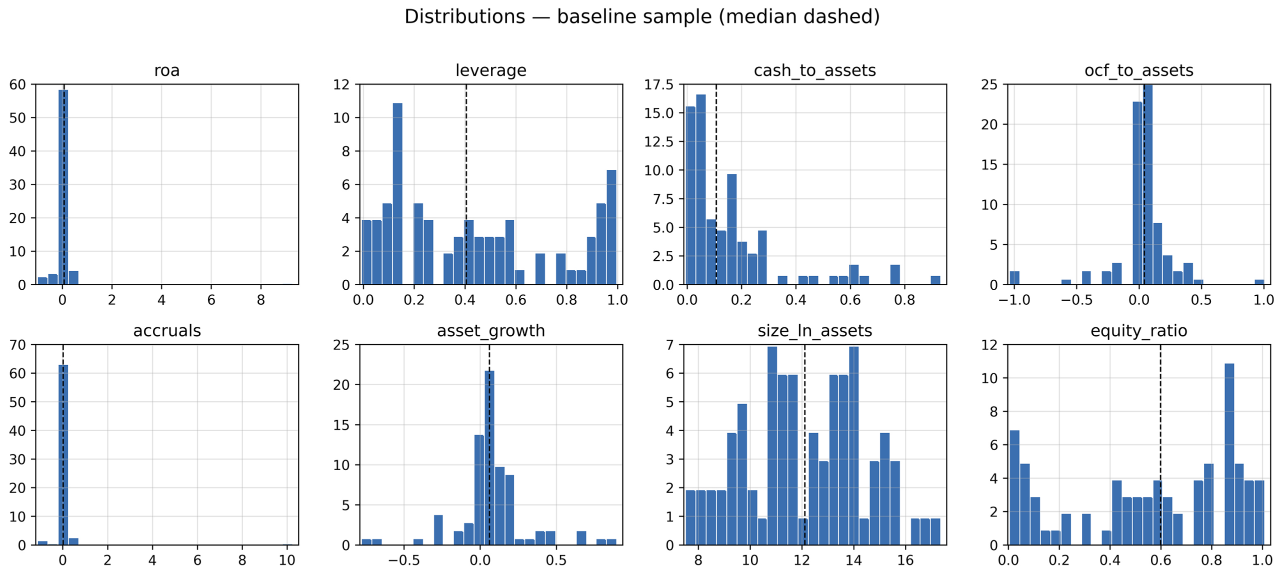

Figure 1. Distributions—baseline sample.

Table 3 reports mean characteristics by firm type over the full FY2025 sample and underscores why financial firms are screened from the baseline. Banks carry mean leverage of 0.835 and insurers 0.702, far above the 0.472 of non-financial firms, while reporting markedly lower returns on assets (0.012 and 0.054, versus 0.092). These structural differences confirm that pooling financial and non-financial firms without controls would confound the leverage and profitability relations of interest.

Table 3. Mean characteristics by firm type (full FY2025 sample).

Firm type	size_ln_assets	ROA	ROE	leverage	cash_to_assets	ocf_to_assets
Bank	16.891	0.012	0.078	0.835	0.112	0.028
Insurer	15.682	0.054	0.145	0.702	0.092	−0.090
Non-financial	12.051	0.092	0.332	0.472	0.176	0.009

4.2. Correlations and Multicollinearity

Table 4 presents the Pearson correlation matrix. Three features merit comment. First, leverage and the equity ratio are perfectly negatively correlated (−1.00) by construction, confirming the identity in Table 1 and justifying their mutual exclusion within a model. Second, ROA and accruals are very strongly correlated (0.98) because both are scaled by total assets and share net profit in the numerator; this mechanical linkage is relevant to the interpretation of M3 below. Third, the theory-relevant correlations have the expected signs: size is positively correlated with leverage (0.56) and negatively with cash holdings (−0.44), and operating cash flow is negatively correlated with accruals (−0.52). Variance inflation factors for all four models are below 1.6, well under conventional thresholds, indicating that multicollinearity does not threaten the estimates (untabulated; all VIF < 1.6).

Table 4. Pearson correlation matrix (baseline sample).

Variables	size	ROA	lev	eq	ocf	cash	accr	growth
size_ln_assets	1.00	−0.11	0.56	−0.56	0.28	−0.44	−0.16	0.26
ROA	−0.11	1.00	−0.02	0.02	−0.36	0.16	0.98	−0.31
leverage	0.56	−0.02	1.00	−1.00	−0.09	−0.26	−0.01	0.15
equity ratio	−0.56	0.02	−1.00	1.00	0.09	0.26	0.01	−0.15
ocf_to_assets	0.28	−0.36	−0.09	0.09	1.00	0.11	−0.52	0.06
cash_to_assets	−0.44	0.16	−0.26	0.26	0.11	1.00	0.15	−0.15
accruals	−0.16	0.98	−0.01	0.01	−0.52	0.15	1.00	−0.30
asset growth	0.26	−0.31	0.15	−0.15	0.06	−0.15	−0.30	1.00

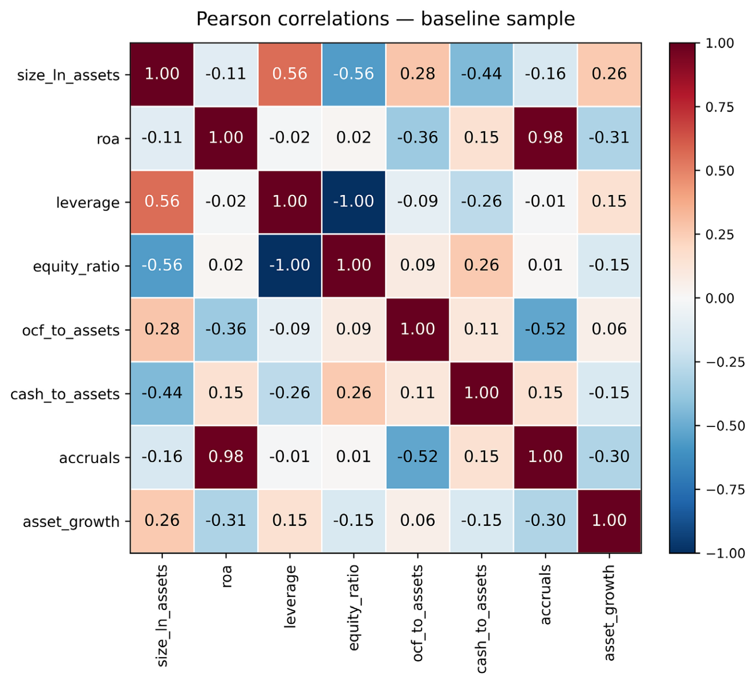

Figure 2. The Pearson correlation heat map.

Figure 2 visualizes the same correlation structure as a heat map, making the two mechanical relationships—the perfect negative association between leverage and the equity ratio, and the near-perfect positive association between return on assets and accruals—immediately apparent, alongside the theory-relevant size-leverage (positive) and size-cash (negative) gradients.

4.3. Regression Results and Hypothesis Tests

Table 5 reports the four baseline regressions and Table 6 summarizes the hypothesis tests.

Table 5. Baseline OLS regressions (FY2025, winsorized variables, HC3 standard errors).

Regressor	M1: Leverage	M2: Cash	M3: Accruals	M4: ROA
Intercept	−0.713*** (0.199)	0.708*** (0.162)	−0.098 (0.098)	−0.385 (0.313)
ROA_w	−0.340* (0.186)		0.703*** (0.180)	
size_ln_assets	0.091*** (0.014)	−0.044*** (0.013)	0.007 (0.007)	0.041 (0.026)
asset_growth_w	0.046 (0.183)	−0.059 (0.105)		0.050 (0.318)
cash_to_assets_w	0.136 (0.161)			
leverage_w		0.018 (0.075)		−0.235* (0.132)
ocf_to_assets_w		0.180 (0.205)	−0.669*** (0.214)	0.217 (0.566)
R^2	0.417	0.261	0.916	0.193
Adjusted R^2	0.377	0.214	0.912	0.138
N	63	68	70	64

Standard errors in parentheses. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Table 6. Hypothesis test summary (baseline estimates).

Hypothesis	Predicted sign	Coefficient	*p*-value	Outcome
H1: ROA → Leverage	−	−0.340	0.067	Supported (10%)
H2: Size → Leverage	+	0.091	0.000	Supported (1%)
H3a: Leverage → Cash	−	0.018	0.804	Not supported
H3b: Size → Cash	−	−0.044	0.001	Supported (1%)
H4: OCF → Accruals	−	−0.669	0.002	Supported (1%)
H5a: OCF → ROA	+	0.217	0.701	Not supported
H5b: Size → ROA	+	0.041	0.116	Not supported
H5c: Asset growth → ROA	−	0.050	0.874	Not supported

Capital structure (M1). Consistent with H2, firm size is strongly and positively associated with leverage ($\beta = 0.091$, $p < 0.001$; Table 5, Table 6). Consistent with H1, profitability enters negatively ($\beta = -0.340$, $p = 0.067$), supporting the pecking-order prediction at the 10% level, although the baseline estimate is marginal and—as Section 4.4 shows—sensitive to outlier treatment. The model explains a substantial share of cross-firm variation in leverage (adjusted $R^2 = 0.377$).

Cash holdings (M2). The size prediction in H3 is supported: larger firms hold significantly less cash relative to assets ($\beta = -0.044$, $p = 0.001$; Table 5). The leverage prediction is not supported—the coefficient is small, positive, and statistically insignificant ($\beta = 0.018$, $p = 0.804$)—so H3 is only partially confirmed.

Accruals (M3). Consistent with H4, accruals are strongly negatively related to operating cash flow ($\beta = -0.669$, $p = 0.002$; Table 5). The model's very high explanatory power (adjusted $R^2 = 0.912$) and the positive loading on ROA ($\beta = 0.703$) are, however, largely mechanical: because accruals equal ROA minus ocf_to_assets by construction, M3 with both regressors approximates an accounting identity, and its R^2 should not be read as evidence on earnings quality. To provide a non-degenerate estimate, we re-estimate the model omitting ROA (M3′: accruals regressed on operating cash flow and size). The negative accrual-cash-flow relation survives on this basis, with the expected loss of mechanical fit ($\beta = -0.375$, $p = 0.025$, adjusted $R^2 = 0.224$, N = 70). H4 is therefore supported, understood as an accrual-cash-flow consistency check rather than a standalone earnings-quality test, as discussed in Section 5.

Profitability (M4). H5 receives no support in the baseline. None of operating cash flow ($\beta = 0.217$, $p = 0.701$), size ($\beta = 0.041$, $p = 0.116$), or asset growth ($\beta = 0.050$, $p = 0.874$) is significant, and asset growth carries the opposite sign to that predicted (Table 6). Leverage enters negatively and marginally ($\beta = -0.235$, $p = 0.076$). The low adjusted R^2 (0.138) indicates that the included characteristics explain little contemporaneous variation in return on assets in this cross-section.

4.4. Robustness and Diagnostics

Table 7 collects the robustness evidence for the central leverage relations. Three patterns emerge. First, the size-leverage relation (H2) is highly stable: the coefficient remains positive and significant at the 1% level across raw variables ($\beta = 0.077$, $p = 0.005$), the full sample, firm-type fixed effects ($\beta = 0.07$, $p < 0.001$), and the pooled clustered specification ($\beta = 0.083$, $p < 0.001$). Second, the profitability-leverage relation (H1), while marginal in the baseline, strengthens under several checks: it becomes more negative in the full sample ($\beta = -0.657$), under firm-type fixed effects ($\beta = -0.63$, $p = 0.02$), in the pooled two-year model with firm-clustered standard errors ($\beta = -0.326$, $p = 0.002$), and after removing six high-influence observations identified by Cook's distance ($\beta = -0.552$). It is significant under HC1 and classical standard errors ($p = 0.013$) and marginal only under the conservative HC3 estimator ($p = 0.067$). The exception is the raw-variable specification, in which the coefficient is essentially zero ($\beta = 0.020$), confirming that the baseline result is driven away from zero by winsorization of an extreme micro-cap observation. Third, the accrual-cash-flow (H4) and size-cash (H3b) relations are stable across the baseline and full samples.

Residual diagnostics for the leverage model are reassuring. The Breusch-Pagan test does not reject homoskedasticity ($p = 0.637$; Breusch & Pagan, 1979), and the

Shapiro-Wilk test does not reject normality of residuals (p = 0.592). The HC3 standard errors used throughout are therefore a conservative choice rather than a necessity, and the inferences are not an artefact of obvious specification failures. Influence analysis identifies six observations exceeding the 4/n Cook's-distance threshold (AFA, CI1, SOR, PNI, ARC, and INV), the most extreme being the micro-cap firm noted earlier; their removal strengthens rather than overturns the profitability-leverage relation.

Table 7. Robustness checks for the leverage model (key coefficients).

Specification	ROA → Leverage	Size → Leverage	N
Baseline (winsorized, HC3)	−0.340*	0.091***	63
Raw (non-winsorized)	0.020	0.077*** (p = 0.005)	63
Full FY2025 sample	−0.657	—	84
HC1 standard errors	−0.340 (p = 0.013)	—	63
Classical standard errors	−0.340 (p = 0.013)	—	63
Firm-type fixed effects (full)	−0.63** (p = 0.02)	0.07***	84
Pooled FY2024-25, firm-clustered SE	−0.326*** (p = 0.002)	0.083***	127
Cook's-distance trimmed	−0.552	—	57

*p < 0.10, **p < 0.05, ***p < 0.01. Cells marked "—" not separately tabulated.

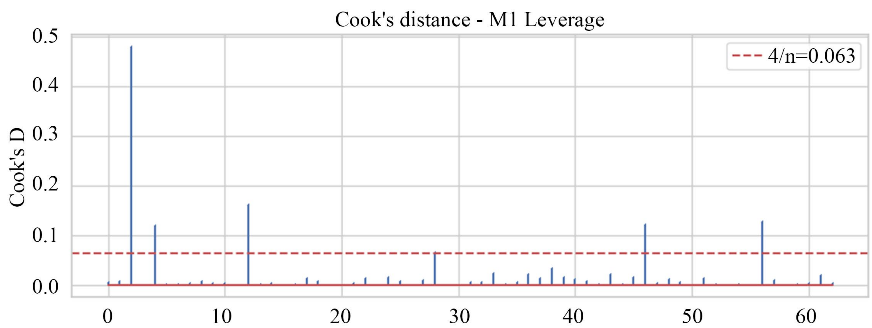

Figure 3. Residual diagnostics for the baseline leverage model (M1). Left: normal quantile-quantile plot of residuals (Shapiro-Wilk p = 0.592). Right: residuals versus fitted values (Breusch-Pagan p = 0.637). Neither plot indicates material departures from the OLS assumptions of normality and homoskedasticity.

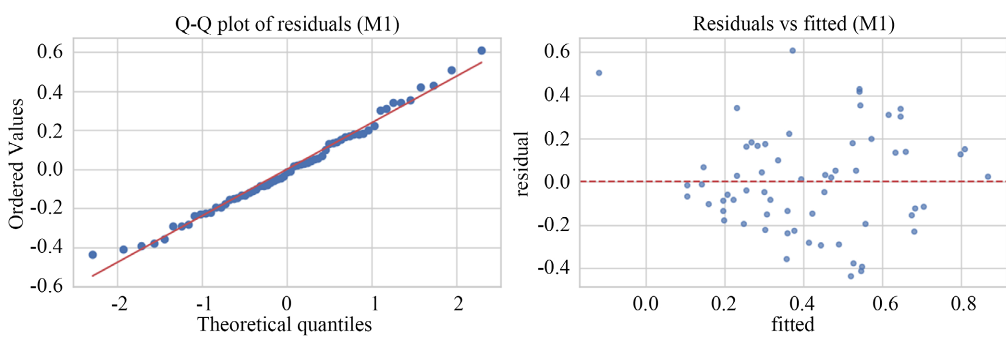

Figure 4. the Q-Q plot + residuals-vs-fitted pair.

Figure 3 plots the Cook's-distance values, in which the six influential points stand out clearly above the 4/n threshold, and **Figure 4** presents the corresponding residual diagnostics, which display only mild departures in the tails and no systematic funnelling—visually confirming the Breusch-Pagan and Shapiro-Wilk results reported above.

5. Discussion

The evidence paints a coherent, if partial, picture. The most robust finding—the positive size-leverage relation (H2)—replicates one of the best-established regularities in the capital-structure literature (Rajan & Zingales, 1995; Frank & Goyal, 2009) and is consistent with the trade-off prediction that larger, more diversified firms have greater debt capacity and lower expected distress costs. Its stability across every specification in **Table 7** gives confidence that it is not an artefact of sample composition or estimator choice.

The profitability-leverage relation (H1) is more nuanced. The negative coefficient predicted by the pecking-order theory (Myers, 1984; Myers & Majluf, 1984; Fama & French, 2002) is recovered in the baseline and strengthens under sample expansion, fixed effects, clustering, and influence-trimming, but it collapses when variables are left un-winsorized. This sensitivity is instructive rather than fatal: it reflects the presence of a small number of extreme micro-cap observations whose unwinsorized ratios dominate the un-trimmed estimate, and the fact that the relation reasserts itself once those observations are tempered or removed is consistent with a genuine underlying negative association. The result should nonetheless be reported as qualified, and it illustrates the importance of outlier treatment in small samples (Cook, 1977).

For cash holdings (H3), the size effect is clear—larger ASX firms hold proportionally less cash, consistent with superior access to external finance and diversification of cash-flow shocks (Opler et al., 1999; Bates et al., 2009)—but the leverage effect is absent. One interpretation is that, in this cross-section, both leverage and cash are strongly related to size, so that conditioning on size absorbs much of the leverage-cash association; the negative raw correlation between leverage and cash (−0.26 in **Table 4**) is consistent with this reading. The result may also reflect the limited power of an 82-firm cross-section to detect a second-order effect.

The strong negative accrual-cash-flow relation (H4) aligns with the foundational accounting evidence (Dechow, 1994; Dechow & Dichev, 2002; Sloan, 1996). We caution, however, that because accruals are measured as net profit minus operating cash flow scaled by assets, a negative loading on operating cash flow and a positive loading on profitability are partly definitional. The high R^2 of M3 should therefore be read as a consistency check confirming that the constructed measures behave as the accounting identity requires, rather than as an independent behavioural test of earnings management. The alternative specification omitting ROA (M3′, Section 4.3) confirms that a genuine, non-degenerate negative accrual-cash-flow association remains, though with far more modest explanatory power. A genuine test of discretionary accruals would require a modelled “normal” accrual benchmark, such as the Jones (1991) model or its performance-matched variant (Kothari, Leone, & Wasley, 2005), which the present data do not support.

The weak profitability results (H5) are the least satisfying but perhaps the most honest. With a single cross-sectional year, a noisy and heavily-tailed dependent variable, and only four regressors, the model has limited power to detect the expected associations; the insignificant and wrong-signed asset-growth coefficient should not be read as contradicting the asset-growth literature (Cooper et al., 2008), which concerns subsequent stock returns rather than contemporaneous accounting profitability. The marginal negative leverage-profitability coefficient is consistent with the reverse of H1—mechanically, more profitable firms can be less levered—and reinforces the endogenous, jointly-determined nature of these variables, which a single-equation cross-section cannot disentangle.

Two cross-cutting themes deserve emphasis. First, the structural heterogeneity of the ASX—documented in **Table 3**—means that the inclusion or exclusion of financial firms materially affects estimated magnitudes, as the baseline-versus-full comparisons in **Table 7** show. Researchers using ASX accounting data should treat the financial/non-financial split as a first-order design decision. Second, the sensitivity of the pecking-order result to outlier handling is a reminder that, in small samples of accounting ratios, winsorization and influence diagnostics are not cosmetic but central to inference.

6. Conclusion

This paper assembled a standardized cross-section of financial variables for 113 ASX-listed firms to test a series of theory-driven hypotheses concerning corporate leverage, cash holdings, accruals, and profitability. The empirical findings are broadly aligned with the international corporate finance literature. Specifically, firm size is robustly and positively associated with leverage, while profitability exhibits a negative relationship with leverage that is consistent with pecking order theory (subject to outlier treatment). Additionally, larger firms tend to hold lower cash reserves, and total accruals move strongly opposite to operating cash flows. Because this inverse accrual-cash-flow relationship persists even when the specification is purged of its mechanical components, it is interpreted here as a fundamental accounting consistency check rather than direct evidence of earnings quality. Conversely, the determinants of contemporaneous profitability remain weakly identified within this framework.

The primary contributions of this study lie in the rigorous construction and documentation of a clean ASX accounting cross-section, the transparent testing of coherent corporate finance hypotheses, and an explicit accounting of the robustness and data-quality constraints that condition the empirical results.

Naturally, several limitations bound the inferences that can be drawn:

- Sample and Horizon: The analysis rests on a modest sample across a single primary fiscal year, which limits statistical power and precludes the use of firm fixed-effects estimation to control for unobserved time-invariant heterogeneity.

- Measurement Risk: Relying on automated extraction from heterogeneous financial statements introduces potential measurement error, even though the data extraction protocol was validated against a manual sub-sample.
- Data Scope: The dataset relies exclusively on accounting aggregates, preventing the evaluation of market leverage, market-to-book ratios, Tobin's Q, or the asset-growth return anomaly.
- Statistical and Methodological Adjustments: The mechanical nature of the accrual-cash-flow relation restricts a substantive earnings-quality interpretation. Furthermore, because eight distinct hypothesis tests were conducted, applying a family-wise error rate correction would weaken the statistical significance of the more marginal results, notably hypothesis H_1 in the baseline model.

Several valuable extensions follow directly from this work. Merging market share-price data would permit the construction of market-based leverage and valuation metrics, facilitating a proper test of the asset-growth effect on equity returns. Lengthening the data panel would enable the deployment of firm fixed effects and dynamic specifications, while adopting a finer industry classification would allow for precise sector controls beyond the broad financial and non-financial dichotomy.

Furthermore, auditor-side extensions offer a natural next step. Technology-enabled audit practices, such as AI-integrated drones and big-data tools, have been shown to significantly influence external auditing performance and financial reporting quality in emerging markets (Alhazmi et al., 2025a). Linking such technological measures to the firm characteristics studied here would directly connect financial reporting outcomes to the audit production function.

While automated data extraction carries inherent measurement risks, emerging agentic architectures for autonomous asset verification and algorithmic auditor governance are actively being designed to mitigate these exact vulnerabilities (Alhazmi et al., 2026). Finally, replacing the residual accrual measure with a fully modelled discretionary-accrual benchmark would elevate the earnings-quality analysis from a baseline consistency check into a substantive empirical test. Pursued together, these iterative steps would transform the present descriptive cross-section into a more powerful, dynamic, and causally credible study of Australian corporate financial policy.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

Alhazmi, A. H. J., Islam, S. M. N., & Prokofieva, M. (2025b). The Impact of Artificial Intelligence Adoption on the Quality of Financial Reports on the Saudi Stock Exchange. *International Journal of Financial Studies, 13,* Article 21. https://doi.org/10.3390/ijfs13010021

Alhazmi, A. H. J., Islam, S. M. N., & Prokofieva, M. (2026). The CMA Agentic Platform: Autonomous Asset Verification and Algorithmic Auditor Governance. *FinTech, 5,* Article 55. https://doi.org/10.3390/fintech5020055

Alhazmi, A. H. J., Islam, S., & Prokofieva, M. (2024). The Impact of Changing External Auditors, Auditor Tenure, and Audit Firm Type on the Quality of Financial Reports on the Saudi Stock Exchange. *Journal of Risk and Financial Management, 17,* Article 407. https://doi.org/10.3390/jrfm17090407

Alhazmi, A. H. J., Islam, S., & Prokofieva, M. (2025a). The Impact of Ai-Integrated Drone Technology and Big Data on External Auditing Performance, Sustainability, and Financial Reporting Quality on the Emerging Market. *Accounting and Auditing, 1,* Article 8. https://doi.org/10.3390/accountaudit1030008

Almeida, H., Campello, M., & Weisbach, M. S. (2004). The Cash Flow Sensitivity of Cash. *The Journal of Finance, 59,* 1777-1804. https://doi.org/10.1111/j.1540-6261.2004.00679.x

Bates, T. W., Kahle, K. M., & Stulz, R. M. (2009). Why Do U.S. Firms Hold so Much More Cash than They Used to? *The Journal of Finance, 64,* 1985-2021. https://doi.org/10.1111/j.1540-6261.2009.01492.x

Breusch, T. S., & Pagan, A. R. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica, 47,* 1287-1294. https://doi.org/10.2307/1911963

Cook, R. D. (1977). Detection of Influential Observation in Linear Regression. *Technometrics, 19,* 15-18. https://doi.org/10.1080/00401706.1977.10489493

Cooper, M. J., Gulen, H., & Schill, M. J. (2008). Asset Growth and the Cross-Section of Stock Returns. *The Journal of Finance, 63,* 1609-1651. https://doi.org/10.1111/j.1540-6261.2008.01370.x

Dechow, P. M. (1994). Accounting Earnings and Cash Flows as Measures of Firm Performance: The Role of Accrual Estimation Errors. *Journal of Accounting and Economics, 18,* 3-42. https://doi.org/10.1016/0165-4101(94)90016-7

Dechow, P. M., & Dichev, I. D. (2002). The Quality of Accruals and Earnings. *The Accounting Review, 77,* 35-59. https://doi.org/10.2308/accr.2002.77.s-1.35

Dechow, P. M., Ge, W., & Schrand, C. (2010). Understanding Earnings Quality: A Review of the Proxies, Their Determinants and Their Consequences: The Role of Accounting Accruals. *Journal of Accounting and Economics, 50,* 344-401. https://doi.org/10.1016/j.jacceco.2010.09.001

Fama, E. F., & French, K. R. (2002). Testing Trade-Off and Pecking Order Predictions about Dividends and Debt. *Review of Financial Studies, 15,* 1-33. https://doi.org/10.1093/rfs/15.1.1

Frank, M. Z., & Goyal, V. K. (2009). Capital Structure Decisions: Which Factors Are Reliably Important? *Financial Management, 38,* 1-37. https://doi.org/10.1111/j.1755-053x.2009.01026.x

Harris, M., & Raviv, A. (1991). The Theory of Capital Structure. *The Journal of Finance, 46,* 297-355. https://doi.org/10.1111/j.1540-6261.1991.tb03753.x

Jones, J. J. (1991). Earnings Management during Import Relief Investigations. *Journal of Accounting Research, 29,* 193-228. https://doi.org/10.2307/2491047

Kothari, S. P., Leone, A. J., & Wasley, C. E. (2005). Performance Matched Discretionary Accrual Measures. *Journal of Accounting and Economics, 39,* 163-197. https://doi.org/10.1016/j.jacceco.2004.11.002

Kraus, A., & Litzenberger, R. H. (1973). A State-Preference Model of Optimal Financial Leverage. *The Journal of Finance, 28,* 911-922. https://doi.org/10.1111/j.1540-6261.1973.tb01415.x

Modigliani, F., & Miller, M. H. (1958). The Cost of Capital, Corporation Finance and the Theory of Investment. *American Economic Review, 48,* 261-297.

Myers, S. C. (1984). The Capital Structure Puzzle. *The Journal of Finance, 39,* 574-592. https://doi.org/10.1111/j.1540-6261.1984.tb03646.x

Myers, S. C., & Majluf, N. S. (1984). Corporate Financing and Investment Decisions When Firms Have Information That Investors Do Not Have. *Journal of Financial Economics, 13,* 187-221. https://doi.org/10.1016/0304-405x(84)90023-0

Opler, T., Pinkowitz, L., Stulz, R., & Williamson, R. (1999). The Determinants and Implications of Corporate Cash Holdings. *Journal of Financial Economics, 52,* 3-46. https://doi.org/10.1016/s0304-405x(99)00003-3

Petersen, M. A. (2009). Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches. *Review of Financial Studies, 22,* 435-480. https://doi.org/10.1093/rfs/hhn053

Rajan, R. G., & Zingales, L. (1995). What Do We Know about Capital Structure? Some Evidence from International Data. *The Journal of Finance, 50,* 1421-1460. https://doi.org/10.1111/j.1540-6261.1995.tb05184.x

Sloan, R. G. (1996). Do Stock Prices Fully Reflect Information in Accruals and Cash Flows about Future Earnings? *The Accounting Review, 71,* 289-315. https://doi.org/10.2308/tar-9608042309

Titman, S., & Wessels, R. (1988). The Determinants of Capital Structure Choice. *The Journal of Finance, 43,* 1-19. https://doi.org/10.1111/j.1540-6261.1988.tb02585.x

Twite, G. (2001). Capital Structure Choices and Taxes: Evidence from the Australian Dividend Imputation Tax System. *International Review of Finance, 2,* 217-234. https://doi.org/10.1111/1468-2443.00027

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica, 48,* 817-838. https://doi.org/10.2307/1912934

Appendix: Analyses Code on Google Colab

```
# ========================================================
# ASX FY2025 — Capital structure / cash / profitability / earnings-quality analysis
# Paste this whole block into ONE Colab cell and run.
# When prompted, upload asx_variables_panel.csv (or set PATH below).
# ========================================================

# === 0. Environment ===
try:
    import linearmodels   # noqa
except Exception:
    !pip -q install linearmodels

import numpy as np, pandas as pd
import statsmodels.api as sm
import statsmodels.formula.api as smf
from statsmodels.stats.outliers_influence import variance_inflation_factor, OLSInfluence
from statsmodels.iolib.summary2 import summary_col
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats
pd.set_option('display.width', 140); pd.set_option('display.max_columns', 60)
sns.set_theme(style='whitegrid', context='notebook')
RNG = 42
print('Environment ready.')

# === 1. Load ===
PATH = 'asx_variables_panel.csv'    # or set a Drive path, e.g. '/content/drive/MyDrive/asx_variables_panel.csv'
import os
if not os.path.exists(PATH):
    try:
        from google.colab import files
        up = files.upload()
        PATH = list(up.keys())[0]
    except Exception as e:
        raise FileNotFoundError('Upload asx_variables_panel.csv or set PATH.') from e

df = pd.read_csv(PATH)
print('Loaded:', df.shape)
```

```
df.head()

# Column inventory
RATIOS = ['size_ln_assets','roa','roa_pretax','roe','equity_ratio','leverage',
          'ocf_to_assets','accruals','cash_to_assets','eff_tax_rate','as-
set_growth']
WINS   = [c+'_w' for c in RATIOS if c+'_w' in df.columns]
FLAGS  = ['financial_firm','foreign_currency','neg_equity']
print('Ratio vars present :', [c for c in RATIOS if c in df.columns])
print('Winsorised present :', WINS)
print('Flags present      :', [c for c in FLAGS if c in df.columns])
print('Years              :', sorted(df.fiscal_year.unique()))
print('Firm types         :', df.firm_type.value_counts().to_dict())

# === 2a. Structure, dtypes, duplicates ===
print('Rows, cols:', df.shape)
print('\nDuplicate firm-year keys:', df.duplicated(['ticker','fiscal_year']).sum())
print('Unique firms:', df.ticker.nunique(), '| rows per year:')
print(df.fiscal_year.value_counts().sort_index())
print('\nDtypes:'); print(df.dtypes)

# === 2b. Missingness (FY2025) ===
f25 = df[df.fiscal_year==2025].copy()
miss = (f25[RATIOS].isna().mean()*100).round(1).sort_values(ascend-
ing=False)
print(f'Missing %% by variable (FY2025, n={len(f25)}):')
print(miss.to_string())

plt.figure(figsize=(7,4))
miss.sort_values().plot(kind='barh', color='#c44')
plt.xlabel('% missing'); plt.title('Missingness by variable (FY2025)');
plt.tight_layout(); plt.show()

# === 2c. Range / accounting sanity checks ===
checks = {
 'equity_ratio in [-1,1.05]'   : f25['equity_ratio'].between(-1,1.05),
 'leverage >= 0'               : f25['leverage'] >= 0,
 'cash_to_assets in [0,1]'     : f25['cash_to_assets'].between(0,1),
 'eff_tax_rate in [-0.5,0.75]': f25['eff_tax_rate'].between(-0.5,0.75),
}
print('Rule violations (non-missing rows):')
for name, ok in checks.items():
    bad = f25[~ok & f25[name.split()[0]].notna()]
```

```
        print(f'    {name:30} violations: {len(bad):2}    -> {list(bad.ticker)[:8]}')

    ident = (f25['equity_ratio'] + f25['leverage'])
    print(f'\nIdentity equity_ratio + leverage (should be ~1): mean={ident.mean():.3f}, max dev={(ident-1).abs().max():.3f}')

    # === 2d. Flag tallies & influential outliers ===
    print('Flag counts (FY2025):')
    for fl in FLAGS:
        print(f'    {fl:16}: {int(f25[fl].sum())} of {len(f25)}')

    print('\nMost extreme RAW values (pre-winsorising):')
    for v in ['roa','roe','accruals','asset_growth']:
        s = f25[['ticker','firm_type',v]].dropna().sort_values(v)
        print(f'\n{v}: low ->', list(zip(s.head(3).ticker, s.head(3)[v].round(2))),
              '| high ->', list(zip(s.tail(3).ticker, s.tail(3)[v].round(2))))

    # === 3. Build samples ===
    def make_sample(data, year=2025, drop_fin=True, drop_negeq=True):
        s = data[data.fiscal_year==year].copy()
        if drop_fin:    s = s[s.financial_firm==0]
        if drop_negeq: s = s[s.neg_equity==0]
        return s

    base = make_sample(df)
    full = make_sample(df, drop_fin=False, drop_negeq=False)
    print('Baseline sample (FY2025, non-fin, positive equity):', base.shape[0], 'firms')
    print('Full FY2025 sample                                    :', full.shape[0], 'firms')

    def W(v): return v+'_w' if (v+'_w') in df.columns else v
    print('\nModelling variables ->', {v: W(v) for v in ['roa','leverage','cash_to_assets','accruals','ocf_to_assets','asset_growth']})

    # === 4a. Summary table (baseline sample) ===
    desc_vars = ['size_ln_assets','roa','roe','leverage','equity_ratio',
                 'ocf_to_assets','cash_to_assets','accruals','asset_growth','eff_tax_rate']
    desc = base[desc_vars].describe(percentiles=[.25,.5,.75]).T
    desc['skew'] = base[desc_vars].skew()
    desc['kurtosis'] = base[desc_vars].kurtosis()
    desc = desc[['count','mean','std','min','25%','50%','75%','max','skew','kurto-
```

```
sis']].round(3)
    print('Descriptive statistics — baseline sample')
    print(desc)

    # === 4b. Descriptives by firm type (full FY2025 sample) ===
    by_type = full.groupby('firm_type')[['size_ln_assets','roa','roe','leverage','cash_to_assets','ocf_to_assets']].mean().round(3)
    print('\nMeans by firm type (full FY2025 sample):')
    print(by_type)

    # === 4c. Distributions ===
    plot_vars = ['roa','leverage','cash_to_assets','ocf_to_assets','accruals','asset_growth','size_ln_assets','equity_ratio']
    fig, ax = plt.subplots(2,4, figsize=(16,7))
    for a,v in zip(ax.ravel(), plot_vars):
        base[v].dropna().hist(bins=25, ax=a, color='#46c', edgecolor='white')
        a.set_title(v); a.axvline(base[v].median(), color='k', ls='--', lw=1)
    plt.suptitle('Distributions — baseline sample (median dashed)', y=1.02); plt.tight_layout(); plt.show()

    # === 5. Pearson correlation matrix + heatmap ===
    corr_vars = ['size_ln_assets','roa','leverage','equity_ratio','ocf_to_assets',
                 'cash_to_assets','accruals','asset_growth']
    corr_mat = base[corr_vars].corr()
    plt.figure(figsize=(8.5,6.5))
    sns.heatmap(corr_mat, annot=True, fmt='.2f', cmap='RdBu_r', center=0, vmin=-1, vmax=1,
                square=True, cbar_kws={'shrink':.8})
    plt.title('Pearson correlations — baseline sample'); plt.tight_layout(); plt.show()
    print(corr_mat.round(2))

    # === 6. Fit the four models (HC3 robust SE) ===
    def ols(formula, data, cov='HC3', cluster=None):
        m = smf.ols(formula, data=data)
        if cov=='cluster':
            return m.fit(cov_type='cluster', cov_kwds={'groups': data[cluster]})
        return m.fit(cov_type=cov)

    f_lev  = f"{W('leverage')} ~ {W('roa')} + size_ln_assets + {W('asset_growth')} + {W('cash_to_assets')}"
    f_cash = f"{W('cash_to_assets')} ~ {W('leverage')} + size_ln_assets + {W('ocf_to_assets')} + {W('asset_growth')}"
    f_acc  = f"{W('accruals')} ~ {W('ocf_to_assets')} + size_ln_assets + {W('roa')}"
```

```
f_roa = f"{W('roa')} ~ size_ln_assets + {W('ocf_to_assets')} + {W('asset_growth')} + {W('leverage')}"

M1 = ols(f_lev,  base); M2 = ols(f_cash, base); M3 = ols(f_acc, base); M4 = ols(f_roa, base)

tbl = summary_col([M1,M2,M3,M4],
        model_names=['M1:Leverage','M2:Cash','M3:Accruals','M4:ROA'],
        stars=True, float_format='%0.3f',
        info_dict={'N':lambda x:f"{int(x.nobs)}",'R2-adj':lambda x:f"{x.rsquared_adj:.3f}"})
print(tbl)

# === 6b. Hypothesis read-out ===
def grab(model, term):
    if term in model.params.index:
        return model.params[term], model.pvalues[term]
    return (np.nan, np.nan)

rows = [
  ('H1 Pecking-order: ROA -> Leverage (exp. -)',     *grab(M1, W('roa'))),
  ('H2 Trade-off: Size -> Leverage (exp. +)',        *grab(M1, 'size_ln_assets')),
  ('H3a Cash: Leverage -> Cash (exp. -)',            *grab(M2, W('leverage'))),
  ('H3b Cash: Size -> Cash (exp. -)',                *grab(M2, 'size_ln_assets')),
  ('H4 Accruals: OCF -> Accruals (exp. -)',          *grab(M3, W('ocf_to_assets'))),
  ('H5a ROA: OCF -> ROA (exp. +)',                   *grab(M4, W('ocf_to_assets'))),
  ('H5b ROA: Size -> ROA (exp. +)',                  *grab(M4, 'size_ln_assets')),
  ('H5c ROA: Asset growth -> ROA (exp. -)',          *grab(M4, W('asset_growth'))),
]
res = pd.DataFrame(rows, columns=['Hypothesis','coef','p_value'])
res['sig'] = pd.cut(res.p_value, [0,.01,.05,.10,1], labels=['***','**','*','ns'])
res[['coef','p_value']] = res[['coef','p_value']].round(3)
print(res)

# === 7. VIF for each model's regressors ===
def vif_table(formula, data):
```

```
        X = smf.ols(formula, data=data).exog
        names = smf.ols(formula, data=data).exog_names
        v = [variance_inflation_factor(X, i) for i in range(X.shape[1])]
        return pd.DataFrame({'term':names,'VIF':np.round(v,2)}).query("term!='Intercept'")

    for name, f in [('M1 Leverage',f_lev),('M2 Cash',f_cash),('M3 Accruals',f_acc),('M4 ROA',f_roa)]:
        print(f'\n{name}'); print(vif_table(f, base.dropna(subset=[c for c in base.columns])).to_string(index=False))
    print('\nRule of thumb: VIF < 5 (strict) or < 10 (lenient) = no serious multicollinearity.')

    # === 8a. Winsorised vs RAW variables ===
    def raw_formula(f): return f.replace('_w','')
    print('M1 Leverage — winsorised vs raw key coef (ROA, Size):')
    for lbl, ff in [('winsorised',f_lev), ('raw',raw_formula(f_lev))]:
        m = ols(ff, base)
        print(f'  {lbl:11}: ROA={m.params.get(W("roa") if lbl=="winsorised" else "roa"):+.3f}'
              f'  Size={m.params["size_ln_assets"]:+.3f}  R2adj={m.rsquared_adj:.3f}  N={int(m.nobs)}')

    # === 8b. Baseline vs FULL sample ===
    print('\nKey coefficients: baseline vs full FY2025 sample')
    for name, f, key in [('M1 Leverage',f_lev,W('roa')),('M2 Cash',f_cash,W('leverage')),
                         ('M3 Accruals',f_acc,W('ocf_to_assets')),('M4 ROA',f_roa,W('ocf_to_assets'))]:
        mb, mf = ols(f, base), ols(f, full)
        print(f'  {name:12} {key:18}: baseline={mb.params.get(key,np.nan):+.3f} (N={int(mb.nobs)})'
              f'  full={mf.params.get(key,np.nan):+.3f} (N={int(mf.nobs)})')

    # === 8c. Alternative robust SEs (HC1 vs HC3) for M1 ===
    print()
    for cov in ['HC1','HC3','nonrobust']:
        m = smf.ols(f_lev, base).fit(cov_type=cov) if cov!='nonrobust' else smf.ols(f_lev, base).fit()
        se = m.bse[W('roa')]; b = m.params[W('roa')]
        print(f'  {cov:9}: ROA coef={b:+.3f}  SE={se:.3f}  t={b/se:+.2f}  p={m.pvalues[W("roa")]:.3f}')

    # === 8d. Sector (firm-type) fixed effects on the FULL sample ===
```

```
m_fe = ols(f_lev + ' + C(firm_type)', full)
print('\nM1 Leverage + firm-type FE (full sample):')
print(m_fe.summary2().tables[1].loc[[W('roa'),'size_ln_assets']].round(3))
print('R2-adj=%.3f, N=%d' % (m_fe.rsquared_adj, m_fe.nobs))

# === 8e. Pooled two-year panel with firm-clustered SEs ===
pool = df[(df.financial_firm==0) & (df.neg_equity==0)].copy()
need_pool = [W('leverage'), W('roa'), 'size_ln_assets', W('cash_to_assets')]
pool = pool.dropna(subset=need_pool).copy()
pool['firm_id'] = pool['ticker'].astype('category').cat.codes
f_lev_pool = f"{W('leverage')} ~ {W('roa')} + size_ln_assets + {W('cash_to_assets')} + C(fiscal_year)"
mp = smf.ols(f_lev_pool, pool).fit(cov_type='cluster', cov_kwds={'groups': pool['firm_id']})
print('\nPooled FY2024+FY2025, firm-clustered SEs (M1 Leverage):')
print(mp.summary2().tables[1].round(3))
print(f'N={int(mp.nobs)} firm-years, clusters={pool.firm_id.nunique()}')

# === 8f. Influence diagnostics: Cook's distance ===
need_lev = [W('leverage'), W('roa'), 'size_ln_assets', W('asset_growth'), W('cash_to_assets')]
base_cc = base.dropna(subset=need_lev).copy()
m_full = smf.ols(f_lev, base_cc).fit()
infl = OLSInfluence(m_full)
cooks = pd.Series(infl.cooks_distance[0], index=base_cc.index)
thr = 4/len(cooks)
high = cooks[cooks>thr].sort_values(ascending=False)
print(f"\nCook's distance threshold 4/n = {thr:.3f}; high-influence obs = {len(high)}")
print('Most influential:', list(base_cc.loc[high.index[:6],'ticker']))

plt.figure(figsize=(9,3.5))
plt.stem(range(len(cooks)), cooks.values, markerfmt=',')
plt.axhline(thr, color='r', ls='--', label=f'4/n={thr:.3f}'); plt.legend()
plt.title("Cook's distance - M1 Leverage"); plt.ylabel("Cook's D"); plt.tight_layout(); plt.show()

m_trim = smf.ols(f_lev, base_cc.drop(index=high.index)).fit(cov_type='HC3')
print(f"M1 ROA coef: full={m_full.params[W('roa')]:+.3f} (N={int(m_full.nobs)})"
      f" vs trimmed={m_trim.params[W('roa')]:+.3f} (N={int(m_trim.nobs)})")

# === 8g. Residual diagnostics for M1 ===
```

```
from statsmodels.stats.diagnostic import het_breuschpagan
m = smf.ols(f_lev, base).fit()
resid = m.resid
fig, ax = plt.subplots(1,2, figsize=(12,4))
stats.probplot(resid, plot=ax[0]); ax[0].set_title('Q-Q plot of residuals (M1)')
ax[1].scatter(m.fittedvalues, resid, s=18, alpha=.7); ax[1].axhline(0, color='r', ls='--')
ax[1].set_xlabel('fitted'); ax[1].set_ylabel('residual'); ax[1].set_title('Residuals vs fitted (M1)')
plt.tight_layout(); plt.show()
bp = het_breuschpagan(resid, m.model.exog)
print('Breusch-Pagan p-value = %.3f   (low => heteroskedastic => HC3 justified)' % bp[1])
print('Shapiro-Wilk normality p-value = %.3f' % stats.shapiro(resid)[1])

# === 9. Export ===
desc.to_csv('table1_descriptives.csv')
corr_mat.round(3).to_csv('table2_correlations.csv')
res.to_csv('table4_hypothesis_results.csv', index=False)
with open('table3_regressions.txt','w') as fh:
    fh.write(str(tbl))
print('\nSaved: table1_descriptives.csv, table2_correlations.csv, table3_regressions.txt, table4_hypothesis_results.csv')
try:
    from google.colab import files
    for f in ['table1_descriptives.csv','table2_correlations.csv','table3_regressions.txt','table4_hypothesis_results.csv']:
        files.download(f)
except Exception:
    pass
```