Online Appendix:

Does CEO Inside Debt Really Improve Financial Reporting Quality?

Stefano Cascino
London School of Economics

Máté Széles
Tilburg University

David Veenman
University of Amsterdam

Contents

OA1. Details on compensation variable definitions

This section provides details on the construction of the inside debt variables. Because the inside debt variables rely on estimates of CEO equity holding values and sensitivities, we first describe our estimation of these inputs. The value of equity ownership is determined by the sum of the values of shares and options held. We first determine the value of a CEO's share ownership by multiplying the total number of vested and unvested shares held, including restricted stock, by the end-of-fiscal-year stock price (Compustat: `PRCC_F`). The variable for the number of shares held is taken from Execucomp's data item `SHROWN_EXCL_OPTS` in the Annual Compensation table, similar to Coles et al. (2006).[1]

Next, we determine the value of stock option holdings using the Black-Scholes formula for European call options, modified by Merton to account for dividends (e.g., Core and Guay, 2002). Option holdings are obtained from the Execucomp Outstanding Equity Awards data, which contain information on individual tranches of vested and unvested options held by executives. For each tranche of options with similar characteristics (e.g., exercise price and expiration date), we compute the value of a call option (C) as follows:

$$C = Se^{(-dT)}N(d_1) - Xe^{(-rT)}N(d_2), \tag{1}$$

where:[2]

S = the firm's stock price at fiscal year-end (Compustat: `PRCC_F`);

d = the natural logarithm of (one plus) the expected dividend yield over the maturity of the stock option, measured using the average realized dividend yield for the firm over the years $t-2$ through t, where the dividend yield is defined as the ex-date fiscal year dividends per share (Compustat: `DVPSX_F`) divided by the end-of-fiscal-year stock price (Compustat: `PRCC_F`);

[1] See also the programming code provided at https://sites.temple.edu/lnaveen/data/.

[2] Following the programming code of Coles et al. (2006), we winsorize the dividend yield (d) and stock return volatility (σ) at the 5^{th} and 95^{th} percentiles of the pooled-sample distribution.

T = the expected time-to-maturity for the options in years, measured as the difference between the option expiration date (Execucomp: `EXDATE`) and the current fiscal year-end date (Compustat: `DATADATE`) in years, multiplied by 0.7 to account for early exercise (Lee et al., 2018);

N = the cumulative standard normal probability distribution function;

X = the option exercise price (Execucomp: `EXPRIC`);

r = the natural logarithm of (one plus) the 7-year U.S. treasury rate (Lee et al., 2018);[3]

$$d_1 = \frac{\ln(\frac{S}{X}) + T(r - d + \frac{\sigma^2}{2}))}{\sigma\sqrt{T}};$$

σ = the annualized (i.e., multiplied by $\sqrt{12}$) standard deviation of monthly stock returns, calculated based on data from CRSP over the 60-month period ending in the firm's fiscal year-end month, requiring a minimum of 12 months;

$d_2 = d_1 - \sigma\sqrt{T}$.

We next use the value estimates per option (C) to determine the total value of each tranche of options, by multiplying the total number of vested and unvested options in the tranche (Execucomp: `OPTS_UNEX_EXER` and `OPTS_UNEX_UNEXER`, respectively) by C. We then determine the total value of options held by CEO in the firm-year by aggregating the values obtained for the individual tranches. Combined with the estimated value of share ownership, this procedure gives us an estimate of the total value of CEO equity ownership (*CEO Equity*).

To evaluate the validity of our estimation of *CEO Equity*, we obtain data on the same construct shared by Professor Lalitha Naveen for the extended sample of the Coles et al. (2006) study. The (untabulated) Spearman correlation between our measure and the external measure is very high at 0.9947. For 50 (75) percent of the sample, our estimates are within 1.5 (5.4) percent of the estimates in the external data. The closeness of the estimates reassures us that the procedures described above are valid and consistent with the literature.

[3]See https://home.treasury.gov/interest-rates-data-csv-archive.

Returning to the Black-Scholes option valuation formula presented above, we obtain the sensitivity of an option's value to a change in share price ("delta") as $\Delta = e^{-dT}N(d1)$. To compute the sensitivity of the CEO's *total* equity portfolio to a change in share price, we first set Δ equal to 1 for each share held. Next, for each tranche of securities (options or shares), we multiply Δ by the number of securities $\times$ 0.01 $\times$ S to obtain the sensitivity to *a one percent change in share price* following Core and Guay (2002). Finally, to determine the overall sensitivity of the equity portfolio to a one percent change in share price, we aggregate the resulting values across the different tranches of the securities held by the CEO in a firm-year (*Delta*).

We obtain the sensitivity of an option's value to changes in volatility ("vega") in a similar way. An option's vega is computed as $V = e^{-dT}N'(d1)S\sqrt{T}$, where N' captures the standard normal probability density function. The sensitivity of the equity portfolio to a 0.01 change in stock return volatility ($Vega$) is then determined by multiplying V by the number of options $\times$ 0.01 for each tranche of options. We then aggregate the values across the different tranches in the portfolio (note that vega is assumed to be zero for shares).

The inside debt variables are constructed as follows. First, from Execucomp's Annual Compensation table, we obtain variables `PENSION_VALUE_TOT` and `DEFER_BALANCE_TOT`. The sum of these variables captures the CEO's total value of inside debt holdings for the firm-year ($CEO\ IDH$). Variable $InsideDebtDum$ is an indicator variable set equal to 1 when this sum is nonzero, and 0 otherwise. Next, we construct variable $InsideDebtRatio$, which captures the ratio of the value of inside debt ownership to the sum of the values of debt and equity ownership of the CEO (where the latter is captured by variable $CEO\ Equity$ defined above):

$$InsideDebtRatio = \frac{CEO\ IDH}{CEO\ IDH + CEO\ Equity}. \tag{2}$$

The third and fourth variables we study in the paper are based on the CEO relative leverage ratio, which is defined as the ratio of the personal leverage of the CEO to the leverage of the firm. The CEO's personal leverage is computed as the ratio of the value of inside debt holdings ($CEO\ IDH$) to the value of equity holdings ($CEO\ Equity$). Firm leverage is defined as the ratio of debt ($Firm\ debt$, the sum of Compustat data items `DLTT` and `DLC`) to the market value of equity ($Firm\ equity$, `PRCC_F` × `CSHO`). Because of the use of $Firm\ debt$ in the denominator, this variable is defined only for firms with nonzero debt. Combined, we obtain the CEO relative leverage ratio similar to prior research (e.g., Cassell et al., 2012; Anantharaman et al., 2014; Dhole et al., 2016; Chi et al., 2017):

$$RelativeLev = \frac{CEO\ IDH/CEO\ Equity}{Firm\ debt/Firm\ equity}. \tag{3}$$

Based on this definition, we create an indicator variable ($RelativeLevDum$) set equal to 1 when the ratio exceeds one, and 0 otherwise (He, 2015; Chi et al., 2017). In addition, to capture more variation in the magnitude of the ratio, we also transform the continuous $RelativeLev$ variable into a rank variable. Specifically, we decile-rank the positive values of $RelativeLev$ in the sample. For cases where $RelativeLev = 0$, we set the rank to zero. Finally, we scale the rank variable by 10 to obtain values between 0 and 1. The benefit of this rank transformation is that the $RelativeLev$ is extremely right-skewed, even after winsorization or log transformation (e.g., Campbell et al., 2016).

Although we do not use this variable in the paper, we also examine an additional inside debt variable in our replication of the Dhole et al. (2016) study. In addition to the relative *leverage* ratio, Dhole et al. (2016) examine the relative *incentive* ratio of the CEO based on Wei and Yermack (2011) (see also Campbell et al., 2016). Conceptually, this measure captures how sensitive the value of the CEO's inside debt versus equity ownership is to a change in firm value, *relative to* the sensitivity of the firm's debt versus equity to the same change in value. Similar to Dhole et al. (2016), this variable is measured as follows:[4]

$$RelativeInc = \frac{CEO\ IDH/CEO\ delta}{Firm\ debt/Firm\ delta}. \tag{4}$$

Dhole et al. (2016) label this variable as *CEO relative incentive ratio* in their tests. CEO

[4] Wei and Yermack (2011) explain that, conceptually, the values of $CEO\ IDH$ and $Firm\ debt$ should be replaced by their sensitivities (i.e., the "delta" of (inside) debt). Because these sensitivities are difficult to measure, they suggest using $CEO\ IDH$ and $Firm\ debt$ as a simplification.

inside debt holdings, CEO equity delta, and firm debt are defined as above. "Firm delta" is a measure of the sensitivity of the value of the *firm's* total equity (shares and options outstanding) to a dollar change in share price. For shares, we take the total number of shares outstanding in Compustat (`CSHO`) with an assumed delta of 1. For options, we take the total number of employee stock options outstanding from Compustat (`OPTOSEY`) and compute delta for these options using the Black-Scholes formula. As inputs, we use the average exercise price of outstanding employee options (`OPTPRCBY`), the end-of-year stock price (`PRCC_F`), an assumed remaining useful life of four years (Wei and Yermack, 2011), and the same firm-year-level estimates of the risk-free rate, dividend payout rate, and volatility as we use to compute the CEO's delta.

OA2. Additional test: effects of (not) controlling for broad-based pension plans

Recall from Table 5 in the paper that our six operating environment characteristics play a key role in explaining the relation between inside debt variables and accrual-based financial reporting quality variables. While the relation is significantly negative in most regression specifications when these variables are not controlled for, it becomes insignificant after including the variables as controls.

Although we do not have a specific prediction on the impact of this variable, in this section we explore the effect of additionally excluding the indicator variable for broad-based employee pension plans from the regressions. The reasons for doing so are that (a) consistent with Cadman and Vincent (2015), Table 3 of the paper suggests a strong relation with inside debt, (b) the (untabulated) correlation between the $PensionPlan$ indicator variable and our combined $OE\ factor$ variable is -0.2824, and (c) both He (2015) and Dhole et al. (2016) do not control for this firm characteristic. The correlation result is consistent with firms being more likely to provide broad-based employee pension plans in less volatile and uncertain operating environments.

In Table OA-1, we present an analysis similar to Table 5 in the paper. The key difference is that we now additionally exclude the $PensionPlan$ indicator variable from the regressions because of its simultaneous correlation with the dependent variables and independent (test) variables. For comparison purposes, Panel A of Table OA-1 is a copy of Panel B of Table 5 in the paper. The results reported in Panel B of Table OA-1 are striking: after additionally excluding the pension-plan variable from the regressions, the negative relation between inside debt and inverse financial reporting quality variables becomes statistically significant in almost all specifications. Thus, in addition to the operating environment variables we identify in the paper, the usage of broad-based pension plans by firms is another characteristic of firms that (a) is associated with a less volatile and uncertain operating environment and (b) induces a negative relation between inside debt variables and accrual-based measures of inverse financial reporting quality.

OA3. Other dependent variables: severe restatements and meeting/beating earnings forecasts

In the paper, we examine restatement incidence as an alternative measure of (inverse) financial reporting quality. Because the data we use capture a wide range of smaller and larger restatements, we also focus on a subset of more severe restatements related to fraud and/or SEC investigations, as identified by the Audit Analytics data. In Table OA-2, we find that the four inside debt variables have no significant relation with severe restatement incidence—both before and after controlling for other factors.

Both He (2015) and Dhole et al. (2016) also examine firms' propensity to meet or (just) beat analyst earnings forecasts as an alternative, non-accrual-based, measure of financial reporting quality. In Table OA-3, we similarly examine the relation between our four inside debt variables and two meet/beat indicator variables. Inconsistent with the results of He (2015) and Dhole et al. (2016), but consistent with the broader conclusion of our study, we find no evidence to support a link between inside debt and financial reporting quality based on these alternative outcome variables.

OA4. Accrual results using quantile regression

The prediction that prior studies test by focusing on absolute discretionary accruals is that, conditional on determinants of accruals, firms with (higher) inside debt are associated with less dispersed distributions of unexplained ("discretionary") accruals. However, because the dispersion in discretionary accrual estimates is a function of the measurement error in first-step estimations, unsigned accrual-model regression residuals are systematically correlated with variables related to this measurement error (Hribar and Nichols, 2007; Owens et al., 2017). This is at least partly the reason why our operating environment variables display such significant explanatory power for absolute discretionary accruals and why they help to explain the univariate relation between absolute discretionary accruals and inside debt variables.

To more directly test the prediction regarding the dispersion of unexplained accruals, we follow Chen et al. (2024) by using quantile regressions for different quantiles of the conditional distribution of accruals. Specifically, we estimate the following model using quantile regression:

$$\begin{aligned} TACC_{it} =& \beta_0 + \beta_1(\Delta Rev_{it} - \Delta Rec_{it}) + \beta_2 PPE_{it} + \gamma I_{it} \\ &+ \boldsymbol{\theta Interactions} + \boldsymbol{\delta Controls} + \varepsilon_{it}, \end{aligned} \tag{5}$$

where $(\Delta Rev_{it} - \Delta Rec_{it})$ and PPE_{it} are the same variables used in the paper, I_{it} is the inside debt test variable, $\boldsymbol{Interactions}$ is a vector of year indicator variables and their interactions with the $(\Delta Rev_{it} - \Delta Rec_{it})$ and PPE_{it} variables, and $\boldsymbol{Controls}$ is the full set of control variables employed in the paper. We estimate this model using quantile regression for quantiles 0.1 through 0.9 with increments of 0.1 similar to Chen et al. (2024). If indeed (higher) inside debt curbs earnings management and leads to higher quality financial reporting, we should observe a positive coefficient (γ) on the inside debt test variable for the lowest conditional quantiles of accruals (e.g., $q = 0.1$ and $q = 0.2$) and a negative value for γ for the highest quantiles (e.g., $q = 0.8$ and $q = 0.9$).

For consistency with He (2015), we tabulate results based on the relative leverage dummy variable ($RelativeLevDum$) as the inside debt test variable in Table OA-4. Inferences are similar for the other inside debt test variables. The results reveal that, inconsistent with the above prediction, the lowest regression quantiles are associated with lower, rather than higher, conditional accruals. At the same time, we also find no evidence to suggest that the highest quantiles are associated with significantly lower conditional accruals. By contrast, when examining the other variables, we can see that firms with more volatile cash flows (high σCFO have more negative (positive) accruals for the lowest (highest) conditional quantiles, which is consistent with firms having more dispersed accruals in more volatile operating environments.

Overall, using a research design that is arguably more appropriate for the specific prediction tested by prior research (Chen et al., 2024), we find no evidence that suggests that (higher) inside debt is associated with smaller unexplained ("discretionary") accruals. These results are consistent with the conclusions drawn in the main paper.

OA5. Replication of He (2015)

In this section, we present the results from our attempt at replicating the main sample and tests of He (2015). We follow the sample selection and research design steps described in the original study as closely as possible. Doing so, we obtain a sample of 5,522 firm-year observations with discretionary accrual data, very close to, but slightly less than, the 5,596 firm-years reported by He (2015). Because some of the steps in the sample selection procedure are not clearly described in the paper (e.g., outlier treatment or specification of the accrual model), we contacted the author, who kindly clarified some of the questions we had on the sample selection procedure and research design choices. While our sample and results display some differences from those reported by He (2015), we should emphasize that it remains unclear what is driving these differences. Historical Compustat data are frequently adjusted (Lyle et al., 2024), which essentially precludes us from performing an exact replication. That said, our analyses do help in assessing the robustness of the original results.

5.1. *Notes on research design choices*

Some of the key design choices we make are the following. First, we follow He (2015) by estimating the cross-sectional modified Jones model for each two-digit SIC industry-year group. We do so for all firm-years with available data in the intersection of the Compustat and CRSP databases with at least 20 observations per industry-year group. As disclosed privately by the author, the accrual-model regressions include both the traditional asset-scaled intercept and a regular unscaled intercept, and the input variables are winsorized at the 1^{st} and 99^{th} percentiles. Because the estimations are performed by industry-year group, we winsorize each cross-section by year. Discretionary accruals (DA) are obtained as the residuals from the industry-year regressions, and we use the absolute values ($|DA|$) to measure (inverse) financial reporting quality.

Second, we do not exclude specific industries such as financial firms or utilities, because the original study does not appear to do so. Following this choice, we obtain a sample size that is close to that of He (2015).

Third, the He (2015) study explains it estimates the Dechow and Dichev (2002) measure of accrual quality following Francis et al. (2005). This implies that, as a first step, working capital accruals are regressed on contemporaneous, lagged, and forward cash flows, the change in revenues, and gross property plant and equipment for each industry-year group. All variables are scaled by average total assets (Francis et al., 2005) and winsorized at the 1^{st} and 99^{th} percentiles by year. Again, the estimations are performed for each two-digit SIC industry-year group with at least 20 observations. As the second step, the DD accrual quality measure is computed as the standard deviation of the five accrual-model residuals over the years $t-4$ through t, requiring data on all five years. Compared to He (2015), we lose fewer observations with these additional data requirements (our sample size drops from

5,522 to 4,778, while the sample of He (2015) drops from 5,596 to 4,238).

Fourth, we present regression results both before and after the winsorization of the continuous regression variables. He (2015) disclosed to us that only the variables used as inputs to the accrual model estimations were winsorized, while the main regression variables were not. Since the regression variables were not winsorized, it is not entirely clear whether key control variables function effectively in controlling for confounding effects. Given the unpredictability of extreme outlying observations and our inability to perfectly match the sample of He (2015), the lack of an outlier treatment for the regression variables in the original study makes it more difficult for us to reconcile any differences in results we may find.

Fifth, we measure the main inside debt variable based on the CEO relative leverage ratio, defined exactly as in our study and as described by He (2015). When the CEO relative leverage ratio exceeds the value of one, the variable $InsiDebt$ is set equal to 1 and 0 otherwise (He, 2015; Chi et al., 2017). Sixth, following He (2015), we include all firm-years with zero inside debt but exclude firms with zero debt, because the calculation of the relative leverage ratio requires nonzero debt as input in the denominator. Finally, following He (2015) we obtain analyst data from the IBES summary history files, director data from ISS (formerly RiskMetrics), restatement data from Audit Analytics, institutional ownership data from Thomson 13-F data, and all remaining control variables from Compustat and Execucomp.

5.2. Descriptive statistics

Table OA-5 presents descriptive statistics similar to Table 1 of He (2015) for our sample. As the table suggests, despite our relatively close replication of the sample size, differences remain in some of the descriptive statistics.[5]

The mean and median values of absolute discretionary accruals ($|DA|$) in our sample are

[5] In contrast to He (2015), who tabulates the largest possible sample size for each variable, we focus on a sample that is restricted by the availability of discretionary accrual data. This choice does not change inferences, but it makes it easier in later tests to winsorize the continuous variables on a constant sample (recall that He (2015) does not winsorize regression variables).

0.0449 and 0.0305, respectively, higher than the values of 0.0247 and 0.0115 reported by He (2015). We are unable to explain or reconcile this difference, but the values reported by He (2015) are small compared to those typically found in the literature (e.g., Bergstresser and Philippon, 2006; Hribar and Nichols, 2007; Armstrong et al., 2013; Owens et al., 2017).

By construction, the mean residual from a regression should be zero. However, our mean value of 0.0033 for signed discretionary accruals (DA) is nonzero because the accrual models were initially estimated using the CRSP/Compustat universe of firm-years, while our final sample focuses on the Execomp universe of S&P 1500 firms. We also examine a discretionary accrual measure that is obtained after controlling for performance (net income scaled by lagged total assets) in the accrual model. The mean, median, and standard deviation of these absolute discretionary accruals ($|DA^{ROA}|$) are smaller than for $|DA|$, because more of the variation in accruals is explained by the model when performance is included in the estimations.

The mean value of 0.0377 for DD is much smaller than the value of 0.1228 reported by He (2015), despite the fact that our approach relies on the same procedure described by Francis et al. (2005). Although estimated for a different sample and time period, our mean and median values of 0.0377 and 0.0273, respectively, are much closer to the values of 0.0442 and 0.0313 reported by Francis et al. (2005, Table 1). We further find a slightly higher frequency of restatements than He (2015) (mean Res of 0.1018 versus 0.0788) and a substantially higher frequency of firm-years that exactly meet or just beat the analyst median consensus forecast by a maximum of one percent (mean $AnaSur$ of 0.2173 versus 0.0948).

Focusing on the key test variable that captures the CEO relative leverage, $InsiDebt$, we find that the CEO relative leverage exceeds the value of one in 34 percent of firm-years. This frequency is higher than the 21.6 percent reported by He (2015). However, our frequency is very close to the 32 percent reported by Chi et al. (2017).

For the control variables, we highlight the following. First, our estimates of independent director share ownership are substantially higher despite the use of the same input data source (ISS/RiskMetrics). The median director ownership as a percentage of shares outstanding is five times higher in our sample (0.0035 versus 0.0007). Second, our estimates of the natural logarithm of (one plus) vega are also much higher than those of He (2015). He (2015) does not report the procedure used to calculate vega, which makes it difficult to infer what drives this large difference. Third, He (2015) reports a mean for the $Big4$ indicator variable of 0.0970, which is counterintuitively low. Instead, our descriptive statistics suggest that this value is more consistent with the typical frequency with which firms hire a *non*-Big-4 auditor. In our descriptives, we find that 89.9 (10.1) percent of firm-years in the sample have a Big-4 (non-Big-4) auditor.

Fourth, the mean and median values of ROA in our sample are close to those reported by He (2015). There is an important difference, however, in the standard deviations. In our sample, the standard deviation of ROA is 0.1005, which is close to the value of 0.12 reported by Armstrong et al. (2013). By contrast, He (2015) reports a standard deviation of 0.7375, which suggests that the variable contains substantially more extreme values than in our sample. Given these extreme values, the ROA variable is potentially less effective as a control variable in the regressions.

Fifth, although the variable $AnaCov$ is defined as the natural logarithm of the number of analysts following a firm, its median value of 1 in He (2015, Table 1) is difficult to explain. This value suggests that the median analyst coverage equals the exponential constant $e^1 = 2.718$, which is not an integer number and therefore cannot represent the number of analysts following a firm.

Finally, the variable for the standard deviation of sales ($StdSales$) is extremely skewed in Table 1 of He (2015). Its mean of 0.3024 is much higher than the median of 0.0778, with an extreme standard deviation of 8.77. We instead find a mean (median) of 0.1245 (0.0904) with a standard deviation of 0.1323. Despite our choice to closely follow He (2015) by not winsorizing the variables before tabulating the descriptives, our variable is not as extremely skewed. This result highlights the unpredictable nature of extreme observations and, again, underscores the difficulty of exactly replicating the original study.

5.3. Replicating Tables 3 and 4 of He (2015)

Table OA-6 presents the results from testing the relation between the accrual-based measures of financial reporting quality and the $InsiDebt$ variable used by He (2015). We do so without winsorizing the continuous regression variables, similar to He (2015), although the inferences from Table OA-6 do not change when we do winsorize the continuous variables (untabulated).

Results in column (1) of Table OA-6 indicate that, when we do not include control variables in the estimation, we are able to obtain the same result as He (2015). When using absolute discretionary accruals ($|DA|$) as the dependent variable, the coefficient estimate of -0.007 on the $InsiDebt$ variable is statistically highly significant and very close to the value of -0.0080 reported in Table 3 of He (2015). In column (2), however, we find that the coefficient on $InsiDebt$ attenuates to -0.001 and becomes statistically insignificant after we include the control variables of He (2015) in the regression.

Although it is impossible to precisely pinpoint the source of this difference in results, the difference in explanatory power between our regression in column (2) and the regression reported in Table 3 of He (2015) hints at the role played by the control variables. The adjusted R^2 of He (2015) is equal to 0.024 when the control variables are included, while it is 0.255 in our column (2) regression. That is, compared to column (1), the addition of control variables increases the adjusted R^2 from 0.040 to 0.255, a substantial increase. However, the adjusted R^2 of 0.024 for the regression reported by He (2015, Table 3) suggests that much less of the variation in $|DA|$ remains unexplained by the independent variables.

When zooming in on the coefficients on key control variables, we find that the coefficient on the standard deviation of cash flows ($StdCash$) in our regression is positive and highly significant, consistent with the main analyses in our paper. However, the coefficient on this variable is statistically insignificant in Table 3 of He (2015), which suggests it does not account for the predictable variation in absolute discretionary accruals that is documented in prior research (e.g., Hribar and Nichols, 2007; Owens et al., 2017). Because the $InsiDebt$ variable is also correlated with characteristics of firms' operating environments, the limited ability of $StdCash$ to explain variation in $|DA|$ in Table 3 of He (2015), combined with the low explanatory power, hints at a potential correlated omitted variable problem where the confounding effect of the operating environment is not sufficiently controlled for.

It is also important to note that the coefficient on the ROA control variable is negative and highly significant in column (2) of Table OA-6 ($t = -9.12$). This result is inconsistent with the significantly positive coefficient reported by He (2015, column (1) of Table 3), while consistent with the results of Armstrong et al. (2013, Table 3) and Liu (2016, Table 2). We are unable to reconcile this difference, but recall from the discussion of the descriptive statistics that the ROA variable in He (2015, Table 1) has a standard deviation that is about seven times higher than the ROA variable in our sample, which suggests the variable contains more extreme values in the He (2015) study.

The surprisingly limited explanatory power of the control variables can also be inferred by contrasting the univariate comparisons in Table 2 of He (2015) with the regression results in Table 3 of He (2015). Without control variables and year-fixed effects, the difference in absolute discretionary accruals ($|DA|$) between $InsiDebt = 1$ and $InsiDebt = 0$ firm-years is -0.0092, but this reduces only slightly to -0.0080 when control variables and year fixed effects are included in the regression. Given the relation between $InsiDebt$ and the control variables, such as the operating environment variables, this small change in the difference after including control variables is surprising and inconsistent with our results. Recall that in Table OA-6, we find that the coefficient changes from a significant -0.007 to an insignificant -0.001 after the control variables are included. Unfortunately, it remains unclear why the explanatory power is so low in He (2015, Table 3) and why the coefficients on key control variables are so different from those reported by prior research.

To further assess the role of the control variables, we explore the effects of excluding control variables from our regression. In column (3) of Table OA-6, we exclude the ROA variable, which has a seemingly inconsistent coefficient in He (2015, column (1) of Table 3) and an unusually high standard deviation in He (2015, Table 1). The results reveal the effect of excluding this control variable is material. The coefficient on $InsiDebt$ becomes more negative and regains its statistical significance ($p < 0.05$). Moreover, the exclusion of this single control variable makes the adjusted R^2 drop from 0.255 to 0.146.

In column (4) of Table OA-6, we additionally exclude the two operating volatility variables ($StdCash$ and $StdSales$) and find that the coefficient on $InsiDebt$ and its significance further strengthen ($p < 0.01$).[6] The adjusted R^2 further drops to 0.094, leaving even more of the variation in the dependent variable unexplained. Combined, these results are consistent with our conjecture that it may be the limited effectiveness of key control variables (ROA and the volatility variables) in the He (2015) study that explains the difference in results between the regression in our column (2) and that reported by He (2015).

Although our main focus is on absolute discretionary accruals as an inverse measure of financial reporting quality, in untabulated tests, we also examine signed discretionary accruals as a dependent variable to be consistent with He (2015). Note that, conceptually, it is not clear that signed discretionary accruals are a useful dependent variable in a broad sample covering multiple years per firm. The reason is that if a firm manages earnings upwards in one year with positive discretionary accruals, the accruals should be expected to reverse in the next year and lead to negative discretionary accruals. In other words, even if accrual-based earnings management takes place, the average expected discretionary accruals over multiple years for a firm should still be close to zero. This is the reason why it is common to focus on unsigned discretionary accruals. Setting this conceptual issue aside, we find an insignificant relation between $InsiDebt$ and signed DA after including all control variables from Table OA-6.

[6] We exclude both variables even though the coefficient on $StdSales$ is not significant in columns (2) and (3). The coefficient $StdSales$ *is* significant when $StdCash$ is excluded as a control variable.

It is important to note that He (2015) additionally explains that the significant relation between $InsiDebt$ is also found when examining discretionary accruals that account for the effect of firm performance (Kothari et al., 2005). In Table OA-7, we therefore examine absolute discretionary accruals obtained from accrual model estimations that additionally include a control for performance (ROA).[7] The results in columns (1) and (2) suggest that when using this other measure of inverse financial reporting quality, we *are* able to find a significant negative relation between $InsiDebt$ and $|DA^{ROA}|$ even after control variables are included. Yet, we find that the inclusion of control variables has a much stronger impact on the coefficient magnitude compared to the difference between the univariate and regression results of He (2015). After including control variables, the coefficient changes from -0.007 in column (1) to -0.003 in column (2). This result also suggests that the economic magnitude of the difference in conditional average $|DA^{ROA}|$, between $InsiDebt = 1$ and $InsiDebt = 0$ firm-years, is substantially smaller than that found in the original study.

In column (3), we explore the effect of winsorizing all continuous variables in the regression at the 1$^{\text{st}}$ and 99$^{\text{th}}$ percentiles of their distributions (recall that He (2015) does not winsorize regression variables). Interestingly, the impact of the winsorization on the $InsiDebt$ coefficient appears limited. The coefficient remains -0.003, although the t-statistic drops slightly from 2.54 to 2.30. The coefficient on the key control variable $StdCash$ becomes statistically more significant (the t-statistic increases from 10.68 to 13.02), and the adjusted R^2 of the regression increases from 0.148 to 0.169 after the continuous variables are winsorized.

The results in our main paper suggest that the standard deviations of cash flows and sales are not the only confounding variables. First, although both the inside debt usage by firms and the measures of financial reporting quality vary across industries, the regressions

[7] Note that the inclusion of ROA in the first-stage accrual model is different from controlling for ROA in a second-stage regression where absolute discretionary accruals are used as the dependent variable. The first-stage inclusion helps control for the relation between ROA and the average level of accruals (Kothari et al., 2005), while the second-stage control helps control for the relation between ROA and the variance of (unexplained) accruals (Hribar and Nichols, 2007).

estimated by He (2015) do not account for industry-fixed effects. Second, the recent history of losses, idiosyncratic return volatility, and R&D intensity are control variables that additionally account for the confounding effects of the operating environment. In columns (4) and (5), we therefore also control for industry fixed effects and the operating environment factor ($OE\,factor$), respectively.[8] The results suggest that these choices cause the coefficient on $InsiDebt$ to become insignificant, which is consistent with the results in our study but inconsistent with the results of He (2015).

In Table OA-8, we attempt to replicate the results of Table 4 in He (2015) with the DD variable as an inverse financial reporting quality measure (Dechow and Dichev, 2002; Francis et al., 2005). Consistent with the results of He (2015) and our tests in Table OA-7, we find a significant negative relation between $InsiDebt$ and inverse financial reporting quality even after controlling for confounding factors. However, the coefficient on $InsiDebt$ is only marginally statistically significant in column (2) with a t-value of -1.77 (p-value of 0.077). Again, our adjusted R^2 of 0.243 is also much higher than the value of 0.057 in Table 4 of He (2015).

Next, the results in column (3) suggest that the coefficient turns insignificant after we winsorize the continuous variables at the extreme percentiles, a choice that is consistent with most empirical archival studies. Interestingly, the coefficient on $StdSales$ is now statistically significant, which is consistent with this variable better explaining variation in the dependent variable after winsorization. The significance of the coefficient on $StdCash$ also strengthens after winsorization. In other words, the control variables are more effective at capturing predictable variation in the dependent variable. As a result, we find that the coefficient on $InsiDebt$ is no longer statistically significant.

In columns (4) and (5), we either include industry fixed effects or the operating environment factor ($OE\ factor$) as an additional control variable. We find, similar to the winsorization of continuous variables, that the results of He (2015) based on the DD vari-

[8] As explained in the main paper, $OE\ factor$ is defined as the first principal component factor based on the six operating environment characteristics σCFO, $\sigma Sales$, $Idioshock2$, ROA, $Loss\%$, and $R\&D$.

able are sensitive to the inclusion of additional operating environment characteristics in the model.

In summary, we conclude that some of the results of He (2015) using accrual-based measures of financial reporting quality can be replicated. However, we also find that important design choices, such as the winsorization of regression variables and/or the inclusion of additional variables that intend to capture firms' operating environments, render the significant relations between financial reporting quality and inside debt insignificant. These results support the general conclusions in our paper that tests of the relation between inside debt variables and accrual-based measures of financial reporting quality are sensitive to confounding effects.

5.4. Restatement and meet/beat tests

As part of the main tests, He (2015) also examines the relation between CEO relative leverage and two non-accrual-based measures of financial reporting quality: the incidence of restatements and firms' propensity to meet or just beat analyst earnings forecasts. We present our attempt at replicating the tests of Tables 5 and 6 of the He (2015) study in our Table OA-9. The results in the first column suggest that the coefficient on $InsiDebt$ is insignificant when restatement incidence (Res) is used as the dependent variable. In the second column, we find a significantly positive, instead of negative, coefficient on $InsiDebt$ when firms' propensity to meet or just beat analyst earnings forecasts ($AnaSur$) is used as the dependent variable. Both results are inconsistent with those reported by He (2015).[9]

The differences in results may be explained by differences in variable construction. Although He (2015) mentions the use of Audit Analytics as the source of restatement data, there are several design choices a researcher can make when using these data. We impose minimal restrictions on the data, only requiring restatements to be accounting-related, not

[9]Note that He (2015) estimates the regressions for these tests on larger samples that are not restricted by the availability of discretionary accrual data. Our inferences do not change when we similarly rely on larger samples for these tests.

caused by clerical errors, and be associated with an overstatement of performance. The frequency of restatements of 10.2 percent in our sample of firm-years (see Table OA-5) is somewhat higher than the frequency of 7.9 percent in Table 1 of He (2015).

To assess the sensitivity of the results to a different definition of restatements, we also examine the more severe restatements that Audit Analytics identifies as being related to fraud or an SEC investigation. In our sample, 1.1 percent of firm-years is associated with such a severe restatement, which is much lower than the frequency reported by He (2015) or our Table OA-5. Based on this narrower definition of restatements, the results in column (3) of Table OA-9 suggest we now find evidence that supports the conclusion of He (2015) that CEO relative leverage *is* negatively associated with restatement incidence.[10] This result provides some evidence in support of the predicted link between inside debt and financial reporting quality. However, we are able to reproduce this result only after changing the definition of the restatement variable into a narrower classification that captures more severe restatements, which differs from the definition that He (2015) relies upon.

Finally, we also examine a different construction of the meet/beat variable. Most studies define this variable as an indicator variable based on whether or not the company exactly meets (i.e., an earnings surprise of 0¢ per share) or just beats the analyst consensus earnings forecast by 1¢ per share (e.g., Cheng and Warfield, 2005; Lim and Tan, 2008; Cohen et al., 2008; Bissessur and Veenman, 2016). He (2015) defines this variable differently based on firms exactly meeting or just beating the consensus forecast by "one percent." In column (4), we change the calculation of the meet/beat variable to the more common definition based on cents per share and still find a (marginally significant) positive coefficient on the $InsiDebt$ variable.

To summarize, our results from replicating the non-accrual-based tests of He (2015) suggest that we are unable to corroborate the conclusions drawn from Tables 5 and 6 of He (2015). We *do* find a significantly lower incidence of severe restatements for firms where the

[10]Untabulated tests suggest this result is robust to additionally including our operating environment factor as a control variable in the estimation.

CEO has a relative leverage ratio above one ($InsiDebt = 1$), which is the only evidence we find that suggests inside debt is associated with higher quality financial reporting. Results for accrual-based measures of financial reporting quality suggest that the relation documented by He (2015) is sensitive to the effective inclusion of the control variables. Overall, we therefore argue that there is no conclusive evidence of a systematic relation between CEO inside debt ownership and firms' financial reporting quality.

5.5. *Instrumental variable results*

Recall that He (2015) finds statistically and economically significant associations between $InsiDebt$ and all measures of financial reporting quality. To address the endogenous nature of firms' inside debt use, He (2015) also presents results from an instrumental variables (IV) approach. Using this approach, all significant relations found in the earlier analyses remain statistically significant with the same sign (Table 7). Performing the same tests is perhaps less obvious for our replication because we are unable to replicate many of the significant relations reported by He (2015) in the first place. However, we do focus on the effects of using the IV approach on three specifications where we find significant results: column (2) of Table OA-7, column (2) of Table OA-8, and column (3) of Table OA-9.

Similar to other studies (e.g., Anantharaman et al., 2014; Cassell et al., 2012; Dhole et al., 2016), He (2015) uses state marginal tax rates as instruments under the assumptions that (a) deferred compensation (such as inside debt) is more attractive for CEOs when tax rates are higher, and (b) state-level tax rates do not directly influence financial reporting quality outcomes. These tax rates are available from the TAXSIM model by state and year (Feenberg and Coutts, 1993).[11]

Anantharaman et al. (2014) use three state-level tax rates: (1) the maximum rate for wages, (2) the maximum rate for long-term capital gains, and (3) the maximum mortgage subsidy rate. He (2015) relies only on (1) and (3) as instruments, which makes sense because

[11] See `https://taxsim.nber.org/state-rates/`.

the correlation between (1) and (2) is 0.9879 in our sample. Because the subsidy rate is presented in the data as a negative tax rate, we multiply this by -1. Doing so, we obtain an instrument that better aligns with the prediction from prior research (e.g., Anantharaman et al., 2014) that a higher mortgage subsidy rate makes inside debt less attractive as a compensation form.

When matching firm-years to state-level data, we rely on firms' historical headquarter data from 8-K filings shared by Gao et al. (2021), instead of the header Compustat data that can lead to significant measurement error (Jennings et al., 2024).[12] Because we are not able to match every firm-year in the sample to a headquarter state, the sample size drops slightly from 5,522 to 5,444 firm-year observations.

Another notable observation is that He (2015) appears to implement the IV procedure using two separate steps instead of the more classical two-stage least squares (2SLS) procedure. That is, He (2015) estimates the first step regression that explains $InsiDebt$ based on state tax rates and control variables using probit. Although not explicitly disclosed, we assume the predicted values from this first-stage probit estimation are then transformed into the variable included in the second-step regression, which is either estimated with OLS for continuous dependent variables or probit for indicator dependent variables. He (2015) reports that the standard errors for the second-step regression are obtained using a bootstrap procedure, although it is unclear whether this approach is correct given the need to account for the uncertainty in the first and second stages *combined* (Chen et al., 2023). By contrast, we follow the more standard 2SLS approach (Cassell et al., 2012; Anantharaman et al., 2014), combined with a cluster-robust bootstrap procedure that encompasses both the first- and second-step estimations.

In the first column of Table OA-10, we first follow He (2015) by estimating a probit regression with $InsiDebt$ as the dependent variable and the two IVs ($StateWage$ and $StateMortgage$) together with all control variables. Consistent with He (2015) and Anan-

[12] See `https://mingze-gao.com/posts/firm-historical-headquarter-state-from-10k/`.

tharaman et al. (2014), the coefficient on $StateWage$ is significantly positive, while the coefficient on $StateMortgage$ is significantly negative. The magnitudes of the coefficients obtained using the probit estimation are close to those reported by He (2015). In column (2), we obtain similar significant relations of $InsiDebt$ with the IVs when using OLS as an estimator—which is the standard first-step estimation performed in 2SLS IV estimations.

Larcker and Rusticus (2010) demonstrate the importance of testing for the relevance of the IVs in the first stage estimation. He (2015), therefore, reports the partial F-statistic for the instruments in the first stage. As Larcker and Rusticus (2010) explain, the partial F-statistic is the F-statistic that is obtained from testing whether the coefficients on the IVs in the first stage are jointly zero (after controlling for the influence of the control variables). This F-statistic can then be compared with critical values from Stock et al. (2002) to evaluate whether the instruments are sufficiently relevant in explaining variation in the endogenous variable.

The first column in Table 7 of He (2015) suggests the partial F-statistic of the first-stage estimation is equal to 40.67, which is much higher than the critical value of 11.59 when using two instruments (Larcker and Rusticus, 2010). Surprisingly, this F-statistic is obtained after a probit estimation, which is unusual because a coefficient test after a probit or logit estimation should provide a χ^2-statistic instead of an F-statistic. Moreover, we find that even for our full model estimation in column (2) of Table OA-10, the total regression F-statistic is 20.31 (untabulated), which is much lower than the *partial* F-statistic that He (2015) reports for the combined significance of the two instruments.[13]

When we test the joint significance of the $StateWage$ and $StateMortgage$ variables in column (1) of Table OA-10, we obtain a χ^2-statistic equal to 7.06 with a p-value of 0.029. In column (2), we obtain a partial F-statistic of 3.83 with a p-value of 0.022. Although both statistics suggest statistical significance at the 0.05 level, the results also suggest that the instruments are weak: the partial F-statistic of 3.83 is well below the critical value of 11.59

[13]When we do not adjust the standard errors for clustering by firm, the total regression F-statistic equals 40.28 (untabulated).

when using two instruments, suggesting these are weak instruments in this setting. It is unclear why the partial F-statistics in Table 7 of He (2015) are so much higher than both the partial F-statistic and the *total* regression F-statistic we find.

Setting aside the weak instrument problem, we still use the instruments in a 2SLS estimation for the three specifications mentioned earlier. The results in columns (3) through (5) of Table OA-10 reveal that each of the previously significant negative relations between $InsiDebt$ and the (inverse) measures of financial reporting quality becomes statistically insignificant with the IV approach. Because of the weak instrument problem described above, we find it difficult to attribute these results to the endogenous nature of inside debt usage per se. The use of weak instruments can lead to significantly biased coefficients in the second stage, and it is unclear in which direction this bias goes (Larcker and Rusticus, 2010).

That said, our replication of the IV approach used by He (2015) suggests that the state-level marginal tax rates may not be providing instruments as strong as presented in the original study. Our replication suggests that the instruments are weak in this setting, which makes it unclear how well the IV approach is able to alleviate biases induced by the endogenous nature of firms' inside debt usage. This result further underpins the significance of the omitted variable problem that our paper illustrates.

OA6. Replication of Dhole et al. (2016)

The following sections present our attempt to replicate the main sample and tests of Dhole et al. (2016) as closely as possible. This study has its own unique sample selection criteria and research design choices that we follow based on the descriptions presented in the paper.

6.1. *Notes on research design choices*

6.1.1. *Variable measurement*

Some of the design choices require some additional discussion. First, in addition to the relative *leverage* ratio, Dhole et al. (2016) examine the relative *incentive* ratio of the CEO (Wei and Yermack, 2011). Conceptually, this measure captures how sensitive the value of the CEO's inside debt versus equity ownership is to a change in firm value, *relative to* the sensitivity of the firm's debt versus equity to the same change in value. This variable, labeled *CEO relative incentive ratio* by Dhole et al. (2016), is defined in more detail in Section OA-1 (see variable $RelativeInc$).

Second, the *CEO relative leverage ratio* variable used by Dhole et al. (2016) is the same variable as we construct in the paper and the variable underlying the $InsideDebt$ indicator variable used by He (2015). However, a key difference is that Dhole et al. (2016) do not transform this measure to mitigate extreme skewness and, instead, focus on the effect of the continuous variable. The same holds for *CEO relative incentive ratio*, which is neither transformed into an indicator, nor a rank variable. As we will see, these continuous variables are extremely skewed. This issue is recognized and addressed by a log transformation in prior research (e.g., Cassell et al., 2012; Anantharaman et al., 2014; Campbell et al., 2016; Chi et al., 2017).[14] Dhole et al. (2016) do not mention the skewed nature of the variables or the use of such a transformation.

Third, Dhole et al. (2016) use a measure of financial reporting quality based on the absolute value of accrual model residuals similar to He (2015) and our main study. The key difference is that they estimate the accrual model without an intercept (i.e., they only include an asset-scaled intercept). In addition, they do not disclose the minimum number of firm-year observations required per industry-year group. Because this choice is unknown, we follow the common choice of requiring at least 20 observations per industry-year group,

[14] Campbell et al. (2016) additionally note that the variables remain skewed even after the log transformation. They therefore combine the log transformation with median regression estimation, which is a robust alternative to OLS that improves estimation precision when the dependent variable is skewed or fat-tailed (e.g., Gassen and Veenman, 2024).

similar to He (2015) and our paper. Although also not disclosed, we further assume Dhole et al. (2016) winsorize the accrual model inputs before estimation following common practice. Because the estimations are performed by industry-year, we winsorize the input variables at the 1^{st} and 99^{th} percentiles of their annual distributions.

Fourth, Dhole et al. (2016) compute measures of real earnings management (REM) using a procedure that is similarly based on a first-step estimation by industry-year group. Although not the focus of our analysis, we follow their estimation because the REM measure is used as a control variable in the main tests. Again, we assume they require at least 20 firm-years per industry-year group.

Fifth, although not explicitly explained in the research design section, we assume from their industry distribution in Panel C of Table 1 that Dhole et al. (2016) exclude both financial firms (SIC codes 6000–6999) and utilities from their sample (SIC codes 4900–4949). Their Table 1 refers to "nonfinancial firms in Execucomp," although the exclusion of utility firms is not mentioned. Yet, their industry distribution does not show two-digit SIC group 49. We find that, when we do not impose the restriction of excluding firms with SIC codes 4900–4949, we have 404 observations for utility firms. This number clearly exceeds the threshold that Dhole et al. (2016) use to tabulate the frequencies per individual two-digit industry group, which makes us conclude that they did exclude utility firms. Following this design choice, we also obtain a sample size that is close to theirs. Our sample comprises 4,864 firm-year observations while that of Dhole et al. (2016) contains 4,845 firm-years over the period 2006–2010.

Finally, in addition to including year-fixed effects similar to He (2015), Dhole et al. (2016) also include (untabulated) industry-fixed effects in their regressions. This design choice is important and helps mitigate omitted variable problems, given the systematic variation in inside debt usage and financial reporting quality measures across industries.

6.1.2. Standard error clustering

Dhole et al. (2016) cluster the standard errors of the regressions by two dimensions: firm and year. This choice can be helpful because it is often desirable with panel data to cluster by both the cross-sectional and the time dimensions (Gow et al., 2010). Doing so helps reduce the risk of obtaining understated standard errors and inflated t-statistics when the data are dependent over time and across firms. However, in this setting, the smallest clustering dimension (time) has only five unique clusters (i.e., five years from 2006–2010). It is well known that having very few clusters can lead to understated standard errors and biased inferences in situations where it is important to cluster the standard errors on that dimension (Bertrand et al., 2004; MacKinnon et al., 2023).

The impact of clustering by a dimension that is too small is unclear in this setting. First, it is unclear whether there is a cross-sectional dependence in the regression errors that should be accounted for when clustering the standard errors by year. If clustering by the time dimension is not required, the standard errors will just be noisy but not biased when clustering by that dimension with a small number of clusters. Second, even if clustering by time is required, and the standard errors *are* understated because the clustering-adjustment is performed using too few clusters, it is still possible that clustering at only the first (i.e., firm) dimension would lead to even less conservative standard error estimates. In other words, the choice by Dhole et al. (2016) to cluster by both firm and year can still provide more reliable inferences than when the standard errors were clustered by firm only.

More practically, the use of two-way clustering with one small clustering dimension affects the degrees of freedom we should use to evaluate the t-statistics. Here, with five unique clusters in the time dimension, the common rule of thumb suggests that the degrees of freedom are equal to four, i.e., the number of clusters in the smallest clustering dimension minus one ($G_{min} - 1$, see e.g., Cameron and Miller, 2015). As a result, with four degrees of freedom, the critical value of the t-statistic with a two-tailed significance level of 0.05 is equal to 2.776 instead of the critical value of 1.96 used for large samples. This degrees-of-freedom adjustment is standard in the most recent versions of statistical software (e.g., with `reghdfe` in Stata), but it was less common in earlier applications following the publication of Petersen (2009) and Gow et al. (2010).[15]

In the regression tables of Dhole et al. (2016), we observe that the determination of p-values and the significance of coefficient estimates at the bright-line cutoffs 0.01, 0.05, and 0.10 do not appear to be based on the degrees-of-freedom adjustment discussed above. For example, the coefficient of 0.0080 on *Ln(Cash pay)* in column (3) of Table 4 in Dhole et al. (2016) has a t-value of 1.76. The asterisk shown with the coefficient suggests it is statistically significantly different from zero at the 0.10 level. With a critical t-value of 1.645 for a two-sided test and a significance level of 0.10, this is correct. But with only four degrees of freedom, this critical value rises to 2.132. Accordingly, we assume Dhole et al. (2016) relied on statistical software that did not make the degrees-of-freedom adjustments for clustered standard errors. Because we *do* rely on this adjustment, our interpretation of statistical significance based on the two-way clustered standard errors will slightly differ from the interpretation of Dhole et al. (2016).

Even though the overall implications of our replication do not depend on this specific research design choice, we believe it is indeed important to clarify this issue. To make the issue more salient, we tabulate the p-values for the coefficients on the main test variables in addition to the t-statistics and significance levels. We will also see in Section OA-6.5 that the choice of whether and how to cluster the standard errors can directly affect the interpretation of tests that evaluate the strength of an instrumental variable.

6.2. Descriptive statistics

Table OA-11 presents descriptive statistics similar to Table 2 of Dhole et al. (2016). Despite the closeness in sample size, we find several differences with the descriptive statistics

[15] For example, in the years following publication of Petersen (2009) and Gow et al. (2010), it was common for researchers using Stata to rely on the cluster2 program (see `https://www.kellogg.northwestern.edu/faculty/petersen/htm/papers/se/se_programming.htm`). This program does not apply the degrees-of-freedom adjustment before calculating and reporting p-values.

of Dhole et al. (2016) that are worth mentioning. Similar to our attempt at replicating the sample and tests of He (2015) in the previous sections, we are unable to fully explain these differences given that we closely followed the sample selection criteria and research design choices of Dhole et al. (2016). As Lyle et al. (2024) explain, the frequent adjustment of Compustat data can make it virtually impossible to exactly replicate prior studies.

The (winsorized) mean inside debt holdings by CEOs in our sample equal $4.8 million, compared to $3.2 million reported by Dhole et al. (2016). Both the deferred compensation and pension components of inside debt have higher means in our sample.[16] We also find that the other compensation variables have higher values in our sample compared to the descriptives presented by Dhole et al. (2016). Given the closeness in sample size and our attempts at following the sample selection and research design choices as closely as possible, we are unable to explain these differences. Interestingly, our mean (standard deviation) of inside debt holdings for CEOs of $4.77 mln ($10.188) is very close to the $4.82 mln ($10.191) reported by Chi et al. (2017) for their sample.

The two primary test variables, *CEO relative leverage ratio* and *CEO relative incentive ratio* are extremely right-skewed. For example, *CEO relative leverage ratio* has a mean of 3.4787 while the median equals 0.1742 and the 75$^{\text{th}}$ percentile value equals 1.2447. Note that this mean is obtained *after* winsorizing the variable, which suggests that the variable still contains extreme values. Indeed, the (untabulated) value of the winsorized 99$^{\text{th}}$ tail equals 146.7. Figure OA-1 visualizes the extreme skewness in the variables. The descriptive statistics of Dhole et al. (2016) suggest that their variables are also highly skewed but somewhat less extreme than in our sample. For example, their mean of 1.2539 for *CEO relative leverage ratio* is close to the 75$^{\text{th}}$ percentile value of 1.1643, while our sample mean of 3.4787 is substantially larger than the 75$^{\text{th}}$ percentile value of 1.2447.[17]

[16] In their footnote 11, Dhole et al. (2016) note that 39 percent of their observations have zero inside debt. In our sample, we find that 36 percent of observations have zero inside debt (untabulated).

[17] Interestingly, we find that, when we winsorize the upper tail of the *CEO relative leverage ratio* variable at the 95$^{\text{th}}$ percentile instead of the 99$^{\text{th}}$ percentile, our mean equals 1.3357. This value is much closer to the mean of 1.2539 reported by Dhole et al. (2016).

It is not clear what drives this difference in the extreme right tail of the variable distributions. However, it is important to note that the CEO equity holdings (and sensitivity) components of the relative leverage and incentive ratios require a range of calculations, assumptions, and inputs that could lead to differences. To assess the sensitivity of our estimates to alternative calculations of the CEO equity components of the ratios, we obtain data from Professor Lalitha Naveen, who generously shares estimates of CEO equity ownership values and deltas on her website.[18] In untabulated tests, we find for our replication sample that our estimates of CEO equity ownership (delta) value have a Spearman correlation of 0.9968 (0.9978) with the variable available in the Naveen data. When applying Naveen's `firm_related_wealth` variable to calculate *CEO relative leverage ratio* instead of our estimates, we find a distribution that is extremely close to that displayed in Table OA-11. For example, the mean and 75$^{\text{th}}$ percentile values are equal to 3.3788 and 1.2375, respectively, and the maximum value is 137.8 after winsorization. Similarly, the values of *CEO relative incentive ratio* are close to those displayed in Table OA-11 when we apply Naveen's CEO equity delta estimates instead of our own calculations.

For their dependent variables capturing earnings management, Dhole et al. (2016) first present descriptives for their measure of earnings smoothing, $CORRK$. To construct this variable, they first compute discretionary accruals based on the residuals from industry-year regressions. Second, they use data from years $t-4$ through t for each firm-year to derive the correlation between the change in discretionary accruals and the change in earnings before discretionary accruals ($CORR$). Because a more negative correlation between these variables suggests more earnings smoothing, they multiply the correlation measure by -1. Finally, because of the skewness in this correlation measure, they transform it to a percentile rank that is rescaled to values between 0 and 1. We follow this approach.

Despite the use of a rank transformation to eliminate skewness in the correlation variable, the $CORRK$ variable that Dhole et al. (2016) present in their Table 2 is still skewed. The

[18]See `https://sites.temple.edu/lnaveen/data/`.

median value is 0.8437, while the mean is 0.6373. For a variable that captures a percentile rank and that is rescaled to values between 0 and 1, one would expect the mean and median to be close to 0.5. This is what we find in our Table OA-11 when we present the descriptives for $CORRK$. However, when we tabulate the descriptives for the negative of the underlying correlation variable $CORR$, we find values that are much closer to those presented by Dhole et al. (2016) and that reflect the same skewness. The median (mean) value of $-CORR$ equals 0.7927 (0.5440). Hence, given the study's use of $CORRK$ as a dependent variable in the regressions, it is not entirely clear whether this variable captures the ranked variable or the (negative of the) underlying correlation estimates.

Mean (median) absolute discretionary accruals ($|DA|$) are equal to 0.0523 (0.0396), which is lower than the value of 0.0865 (0.0765) reported in Dhole et al. (2016, Table 1). However, our median estimate of 0.0396 is substantially closer to the 0.047 reported by Owens et al. (2017, Table 1) for the absolute value of Jones-model discretionary accruals. Our descriptives for the absolute values of real earnings management measures suggest that the values of $|RCFO|$ are also somewhat smaller in our sample, while the values of $|RPROD|$ and $|RDISX|$ are close to those reported by Dhole et al. (2016, Table 1).

The variable $|RMPROXY|$, labeled $|REM\ PROXY|$ in Table 2 of Dhole et al. (2016), differs substantially in our sample. Dhole et al. (2016, p. 525) explain this variable is the absolute value of $RMPROXY$, which is "the sum of the three standardized variables $RCFO$, $RPROD$, and $RDISX$." While the original work of Cohen et al. (2008) does not describe this either, we assume that "standardized" refers to the standard practice of subtracting the sample mean from the variable and scaling by the sample standard deviation. Also, because Cohen et al. (2008, p. 766) predict that *"firms that manage earnings upward are likely to have one or all of these: unusually low cash flow from operations, and/or unusually low discretionary expenses, and/or unusually high production costs,"* we take the negative of $RPROD$ when computing the sum of the three standardized variables:

$$RMPROXY = \frac{(RCFO - \mu_{RCFO})}{\sigma_{RCFO}} - \frac{(RPROD - \mu_{RPROD})}{\sigma_{RPROD}} + \frac{(RDISX - \mu_{RDISX})}{\sigma_{RDISX}}. \quad (6)$$

Based on this calculation, lower values of $RMPROXY$ imply earnings are managed upwards more by means of real activities. Dhole et al. (2016) use the absolute values of $RMPROXY$ and its three components in their empirical tests.

Oddly, the descriptive statistics in Table 2 of Dhole et al. (2016) suggest that the absolute value of $RMPROXY$ (their variable $|REM\ PROXY|$) has a mean of zero and contains negative values. We therefore assume they tabulate the signed value of $REM\ PROXY$ rather than the absolute value. Indeed, when we tabulate descriptives for our signed variable $RMPROXY$, the mean and standard deviation of -0.0089 and 2.2002 are close to the values of 0.0000 and 2.3569 reported for $|REM\ PROXY|$ by Dhole et al. (2016). Instead, our values for $|RMPROXY|$ are, by construction, strictly positive with a right-skewed distribution.

Focusing on the remaining variables, we find that our sample firms have somewhat higher frequencies of meeting or beating analyst expectations, although the differences are not large. Our median sample firm is slightly smaller ($1.77 versus $2.04 billion in total assets), but has substantially higher analyst coverage. In our sample, median analyst coverage equals 8 analysts, while Dhole et al. (2016) report a median of only 5 analysts. Given that we similarly focus on the number of analysts in the consensus annual earnings forecast for the year, we are unable to explain what drives this substantial difference. At the same time, the values of the descriptives for the other control variables all seem close to those reported by Dhole et al. (2016).

To summarize, although we attempt to follow the sample selection and design choices of Dhole et al. (2016) as closely as possible, the descriptive statistics for some of the variables in our sample do not match those of Dhole et al. (2016). Because we are unable to derive and explain the sources of these differences, we conclude that any differences we may find in our regression analyses with those of Dhole et al. (2016) are difficult to fully reconcile. That said, our analysis is still helpful in assessing the general robustness of the original results in an attempt to replicate the paper as closely as possible.

6.3. *Replicating Table 4 of Dhole et al. (2016)*

Table OA-12 presents results from regressing two earnings management measures, $CORRK$ and $|DA|$, on the CEO relative leverage and incentive ratios and control variables. The coefficients on the two test variables *CEO relative leverage ratio* and *CEO relative incentive ratio* are close to zero and statistically insignificant in all four columns. These results are inconsistent with the results presented in Table 4 of Dhole et al. (2016). In untabulated tests, we find that the coefficients remain close to zero and statistically insignificant even when we exclude the control variables from the estimation.

At the same time, when we focus on key control variables, we do find coefficients that are in line with those reported by Dhole et al. (2016). For example, $CORRK$ and $|DA|$ are significantly positively related to $|RMPROXY|$. The standard deviation of earnings ($Evol$) is strongly negatively correlated with $CORRK$, which is consistent with earnings being less volatile in the presence of more earnings smoothing (i.e., a stronger negative correlation between discretionary accruals and earnings before discretionary accruals).[19] Consistent with Hribar and Nichols (2007), more volatile operations (higher $Evol$) are associated with larger absolute discretionary accruals.

We conjecture that the difference in results between our Table OA-12 and Table 4 of Dhole et al. (2016) may be explained by the distributions of the test variables *CEO relative leverage ratio* and *CEO relative incentive ratio*. Because our construction of these variables follows the descriptions provided by Dhole et al. (2016) and others in the inside debt literature (e.g., Cassell et al., 2012), we expect that differences in variable definitions or input variables are unlikely to explain the differences. Instead, the extreme values in our test variables, which appear more extreme than those presented by Dhole et al. (2016) in their Table 2 descriptives, could play a role. Extreme values of an independent variable, even after winsorization, could lead to a situation where a few observations are so influential (i.e., they have high "leverage")

[19]In their Table 4, Dhole et al. (2016) report coefficients on the $Evol$ variable that are very close to ours for the $CORRK$ regressions, except for the fact that their coefficients are positive instead of negative. Because the t-statistics they report for these coefficients *are* negative, we assume that the missing negative sign on the coefficients is a presentation error.

that the coefficient estimates become biased and/or imprecise (Leone et al., 2019; Gassen and Veenman, 2024). To address this concern, we follow prior studies (He, 2015; Chi et al., 2017) by transforming the continuous inside debt variables. Specifically, we transform the two test variables to indicator variables that are set equal to 1 when the ratio exceeds the value of one, and 0 otherwise. Table OA-13 presents the results from replacing the continuous inside debt variables with these indicator variables.

The results from applying these variable transformations are noteworthy. First, recall that the $CORRK$ variable is defined as the percentile rank of the *negative* of the correlation between discretionary accruals and earnings before discretionary accruals. In other words, higher values of $CORRK$ suggest a more negative correlation. For both inside debt variables, the results in columns (1) and (2) of Table OA-13 suggest that firms whose CEOs have higher relative leverage and incentive ratios are associated with *more* earnings smoothing. That is, the coefficients on the inside debt variables are positive, suggesting inside debt is associated with a stronger negative correlation between discretionary accruals and earnings before discretionary accruals.[20] This result does not align with the negative coefficients found by Dhole et al. (2016) and their conclusion that inside debt ownership is associated with *less* earnings smoothing, as captured by the $CORRK$ variable.

Our results in columns (3) and (4) for the absolute discretionary accrual variables, however, *are* consistent with the results of Dhole et al. (2016). Based on the indicator variables capturing the inside debt ratios, we find that greater inside debt ownership and incentives are associated with smaller absolute discretionary accruals. The negative coefficients are statistically significant at the 0.05 level (p-values of 0.049 and 0.042, respectively). The significant coefficient of -0.004 in column (3) suggests that firms whose CEOs have a relative leverage ratio exceeding one are associated with absolute discretionary accruals that are 0.4

[20]When we replace the percentile rank of -1 times the correlation ($CORRK$) with the underlying correlation itself ($CORR$) in untabulated tests, we indeed find a negative coefficient on the inside debt variables ($p < 0.05$). The negative coefficients suggest greater inside debt is associated with more earnings smoothing, as captured by a more negative correlation between discretionary accruals and earnings before discretionary accruals.

percent of total assets smaller than for other firms. Compared to the mean $|DA|$ of 0.0523 presented in Table OA-11, this coefficient is economically meaningful.

In columns (5) and (6), we also examine whether the significant negative relation between $|DA|$ and the inside debt variables is robust to additionally controlling for other factors related to the firm's operating environment. The Dhole et al. (2016) study controls for operating volatility through the *Evol* variable based on the standard deviation of past earnings, but it does not include the other variables we label as operating environment characteristics in the paper: σCFO, $\sigma Sales$, $Idioshock2$, ROA, $Loss\%$, and $R\&D$. We therefore examine the effects of controlling for the operating environment factor variable ($OE\,factor$) that we obtain from these variables. The results suggest that with this additional control variable included in the regressions, the coefficients remain negative but lose statistical significance. The p-values for the regressions in columns (5) and (6) of Table OA-13 are 0.203 and 0.222, respectively. Thus, similar to our replication of the He (2015) study, we conclude that the discretionary accruals results from the regressions reported by Dhole et al. (2016) are sensitive to the inclusion of specific control variables.

To summarize, our tests provide the following insights. First, when we follow the description of the sample selection and research design choices of Dhole et al. (2016), we obtain a sample that is very close in size. However, some of the variables have different distributional properties. Second, we are unable to replicate the results presented in Table 4 of Dhole et al. (2016) when using the (winsorized) continuous variables for the CEO relative leverage and incentive ratios. Third, with transformed (indicator) variables, we do find that greater inside debt holdings and incentives are associated with significantly smaller absolute discretionary accruals, which is more in line with the conclusions of Dhole et al. (2016). However, the t-statistics and significance levels are substantially weaker than those reported by Dhole et al. (2016), and the significance disappears after we control for additional factors. Moreover, we find a relation between inside debt and the measure of earnings smoothing that is *opposite* to that found by Dhole et al. (2016). Combined, these results suggest that we find some evidence consistent with Dhole et al. (2016) when we adjust the research design, but the overall results are mixed and not robust to additional controls.

6.4. *Replicating Tables 5 and 7 of Dhole et al. (2016)*

Dhole et al. (2016) additionally examine measures of real earnings management in relation to the inside debt variables in their Table 5. Although not the focus of our study, we do replicate their tests in Table OA-14, based on the inside debt indicator variables. The overall conclusion from this analysis is that we are unable to reproduce the negative relation of the inside debt variables with any of the real earnings management measures. In fact, results in the rightmost columns suggest that there is a (slightly significant) *positive* relation between the inside debt variables and the combined real earnings management measure ($|RMPROXY|$).[21]

We also examine the relation between the inside debt variables and firms' propensity to meet or just beat analysts' consensus earnings forecasts, following Table 7 of Dhole et al. (2016). As the results in Table OA-15 reveal, using the indicator inside debt variables (results are similar for the continuous variables), we are unable to replicate the result of a negative relation of the CEO relative leverage and incentive ratios with the frequency with which firms meet or (just) beat analyst earnings expectations. All coefficients are close to zero and statistically insignificant.[22]

Overall, the results from Tables OA-14 and OA-15 suggest that we are unable to identify a significant negative relation of the inside debt variables with the measures of real earnings management or firms' propensity to meet/beat analyst earnings forecasts. The meet/beat

[21]In untabulated tests, we also examine the continuous inside debt variables instead of the indicator variables. The coefficients on these variables are statistically insignificant in all regressions except for column (5), where the coefficient on the relative relative ratio is negative with p-value of 0.095. Besides the weak statistical significance, the magnitude of the coefficient at -0.00028 is much smaller than the -0.0155 reported by Dhole et al. (2016).

[22]We estimate these regressions using OLS. Because the dependent variable is an indicator variable, we estimate a linear probability model. The reason for this choice is that it allows us to compute the two-way clustered standard errors using reghdfe in Stata, which makes appropriate degrees-of-freedom adjustments as discussed in Section OA-6.1.2. Using logit regressions with one-way clustered standard errors by firm *or* time does not change the inferences drawn from these analyses.

results are consistent with those from our replication of the He (2015) study, as well as the results in our main study.

6.5. Instrumental variable results

Similar to He (2015), Dhole et al. (2016) use an IV approach using state-level tax rates to address the endogenous nature of firms' inside debt usage in executive compensation. Similar to Anantharaman et al. (2014), Dhole et al. (2016) rely on all three state-level tax rate variables (maximum rate on wages, capital gains, and mortgage subsidy). Comparable to our replication sample for the He (2015) study, we find a correlation of 0.9893 between the first two tax rate variables (untabulated). In a regression of (the natural log of) CEO relative leverage on the three tax rate variables, we obtain variance inflation factors of 49.39 and 47.13 (untabulated) for the first two variables, which indicates severe multicollinearity. For this reason, we do not include the second tax variable (maximum rate on capital gains) as an instrument, similar to He (2015).[23]

Other differences between the IV approaches of Dhole et al. (2016) and He (2015) are that Dhole et al. (2016) (a) use 2SLS instead of a combination of a probit estimation with a second-step OLS or probit estimation, (b) additionally include industry fixed effects in the first- and second-step estimations, and (c) indicate that the t-statistics presented in their Table 8 are based on robust standard errors clustered by firm and year. As before, these t-statistics are not evaluated against critical values for a t-distribution with four degrees of freedom. Also, as discussed below, it is not entirely clear whether the additional IV-tests that Dhole et al. (2016) report (the under- and over-identification tests) are based on standard errors with a similar cluster adjustment.

The results for the first-stage regressions in Table 8 of Dhole et al. (2016) are consistent with the results of Anantharaman et al. (2014) and our replication of He (2015). The state-level tax rates on wages are positively associated with the CEO relative leverage and incentive

[23] This design choice does not materially impact the interpretation of our replication of the Dhole et al. (2016) IV tests.

ratios, while the mortgage tax subsidy is negatively associated with the CEO relative leverage and incentive ratios. The capital gains tax rate is not significantly associated with either variable.

The t-statistics for the coefficients on the instrumental variables, however, suggest that the significance of the coefficients is not strong. For example, in the first column of Table 8 of Dhole et al. (2016), the t-statistics for the wage- and mortgage-tax rates are 2.06 and 2.17, respectively, which the study interprets as being statistically significant at the 0.05 level. As our discussion on standard error clustering in Section OA-6.1.2 suggests, however, these t-statistics should be compared to critical values derived from t-distribution with four degrees of freedom. At the 0.10 significance level with a two-tailed test, this critical value is equal to 2.132. Thus, with an appropriate degrees-of-freedom adjustment with two-way clustering by firm and year, the wages and the mortgage-tax variables would have p-values that are slightly above and below 0.10, respectively. Given the importance of testing the joint significance of the instruments using a partial F-test (Larcker and Rusticus, 2010), this result makes it unclear whether a partial F-statistic would exceed conventional critical values as well.

Dhole et al. (2016) do not report a partial F-statistic for the instruments, but they do report the p-value obtained from an under-identification test based on the Kleibergen-Paap LM statistic (Kleibergen and Paap, 2006). This statistic is obtained from a matrix-rank test of whether the instruments are sufficiently relevant. At a high level, this test is comparable to the idea of the partial F-test. Somewhat surprisingly, despite the marginally significant coefficients on the instruments in the first-stage regressions of Dhole et al. (2016, Table 8), the p-values reported for the Kleibergen-Paap LM statistic are all equal to "0.00", which suggests a strong rejection of the null hypothesis that the instruments are weak.

In their discussion, Dhole et al. (2016, p. 539) further explain that *"the Cragg and Donald (1993) F statistics for IV range between 9.39 and 9.81 in columns 1.1 to 4.1, which are higher than the Stock and Yogo (2005) critical value of 9.08 for a maximal relative bias. Based on* *these test statistics, we conclude that the IV is relevant."* However, it is important to note that the Cragg-Donald F-statistic is not valid in situations where robust or cluster-robust standard errors are invoked, as is the case here (see, for example, the help file for program `ivreg2` in Stata). Thus, combined with our discussion regarding the critical values against which the t-statistics based on two-way clustered standard errors should be evaluated, it is not clear whether the instruments are sufficiently strong in this setting.

In Table OA-16, we present our IV tests for the Dhole et al. (2016) sample.[24] We do so for both the earnings smoothing measure ($CORRK$) and absolute discretionary accruals ($|DA|$). In columns (1) through (3), we present the 2SLS results from instrumenting the CEO relative leverage ratio, while in columns (4) through (6), we instrument the CEO relative incentive ratio. Because of the extreme skewness in these variables in our sample, we choose to follow prior research by taking the natural logarithm of the ratios in these tests. Specifically, we follow Campbell et al. (2016) by first replacing all zero values with the minimum positive value for each ratio and then taking the natural logarithm. As Campbell et al. (2016, p. 342) explain, this transformation results in a substantially less skewed and leptokurtic distribution of the relative leverage and incentive ratios. Similar to our replication of the IV tests of He (2015), we obtain cluster-robust standard errors for the second-stage regressions using a bootstrap approach that accounts for the uncertainty in both the first- and second-step regressions (Chen et al., 2023).[25]

The results in Panel A of Table OA-16 are based on standard errors clustered by firm and year. The first-stage regression results in columns (1) and (4) indicate that the IVs have the predicted positive and negative signs, and the t-statistics are close to those reported by Dhole et al. (2016). For the $StateMortgage$ IV, the t-statistics are even larger than those of Dhole et al. (2016). However, given a t-distribution with four degrees of freedom, the p-values associated with the four t-statistics reported in Panel A are 0.119, 0.070, 0.121, and

[24]Similar to our replication of the IV tests of He (2015), the sample size reduces slightly because of the additional data required to identify firms' historical headquarters.

[25]Inferences are similar when we obtain the two-way clustered standard errors from the program ivreg2 in Stata.

0.067, suggesting only marginal significance of the coefficients. The partial F-statistics for the instruments in the first-stage estimations are equal to 3.034 and 3.116 with p-values of 0.158 and 0.153, respectively. Similarly, the p-values associated with the Kleibergen-Paap LM statistic for the under-identification test are 0.232 and 0.228, respectively.

These tests suggest that, in our sample, the instruments appear to explain the variation in the endogenous variables only weakly. Also, we cannot reject the null hypothesis that the instruments are weak, as the partial F-statistics of 3.034 and 3.116 are substantially below the critical values (Larcker and Rusticus, 2010). In the second-stage regressions, we find that the instrumented endogenous variables have no significant relation with the financial reporting quality variables.[26]

Finally, we examine the importance of the clustering choice for inferences regarding instrument validity. The reason is that the value of the partial F-statistic, advocated by Larcker and Rusticus (2010), is directly affected by this choice as well. In Panel B of OA-16, we present the hypothetical situation in which a researcher would only use heteroskedasticity-robust instead of cluster-robust standard errors. In a firm-year panel dataset, this is equivalent to using one-way clustered standard errors by *firm-year* instead of two-way clustered standard errors by firm *and* year. We are not suggesting this is what the study by Dhole et al. (2016) did, but this analysis does help to illustrate the potential sensitivity of inferences drawn regarding instrument validity.

The results suggest that when using heteroskedasticity-robust standard errors, the coefficients on the IVs are all statistically significant at the 0.01 level. Both the partial F-statistics and the Kleibergen-Paap LM statistics for the under-identification test are statistically significant, although the F-statistics of 9.66 and 9.92 do not exceed the critical value of 11.59 needed to reject the null hypothesis that the instruments are weak (Larcker and Rusticus,

[26]The significance of the coefficients on the IVs differs from the somewhat stronger significance we found for our replication of the IV tests of He (2015) in Table OA-10. Untabulated tests suggest that two design choices play a role. First, we find that the significance of the instrumental variables becomes weaker when industry-fixed effects are included in the estimation. Second, we find that the additional clustering on the second (time) dimension further weakens the significance compared to the He (2015) tests.

2010). The p-values of the Kleibergen-Paap LM statistics are similar to the "0.00" values reported by Dhole et al. (2016), even though their analyses are supposed to be more consistent with those presented in Panel A of our Table OA-16. The second-stage coefficients on the instrumented variables are statistically significant, although it is unclear what to infer from these given the lack of cluster-adjustment in the standard error calculations.

Overall, we conclude that, similar to our replication of the IV tests of He (2015), the state-level tax rates do not provide instruments that are sufficiently strong to alleviate concerns over the endogenous nature of firms' inside debt usage. The weakness of the IVs makes it difficult to understand what to infer from the statistically insignificant results reported for the second-stage estimations in Panel A of Table OA-16. These insignificant results could either reflect the appropriate elimination of endogeneity bias obtained in regular regression analyses, or they could be driven by the bias induced by the use of weak instruments (Larcker and Rusticus, 2010). Besides the more specific task of replicating the IV tests of Dhole et al. (2016), our analyses also highlight the relevance of the standard error clustering choice for inferences regarding instrument validity in 2SLS IV tests.

References

Anantharaman, D., V. W. Fang, and G. Gong. 2014. Inside Debt and the Design of Corporate Debt Contracts. *Management Science* 60 (5): 1260–1280.

Armstrong, C. S., D. F. Larcker, G. Ormazabal, and D. J. Taylor. 2013. The Relation Between Equity Incentives and Misreporting: The Role of Risk-Taking Incentives. *Journal of Financial Economics* 109 (2): 327–350.

Bergstresser, D. and T. Philippon. 2006. CEO Incentives and Earnings Management. *Journal of Financial Economics* 80 (3): 511–529.

Bertrand, M., E. Duflo, and S. Mullainathan. 2004. How Much Should We Trust Differences-In-Differences Estimates? *The Quarterly Journal of Economics* 119 (1): 249–275.

Bissessur, S. W. and D. Veenman. 2016. Analyst Information Precision and Small Earnings Surprises. *Review of Accounting Studies* 21 (4): 1327–1360.

Cadman, B. and L. Vincent. 2015. The Role of Defined Benefit Pension Plans in Executive Compensation. *European Accounting Review* 24 (4): 779–800.

Cameron, A. C. and D. L. Miller. 2015. A Practitioner's Guide to Cluster-Robust Inference. *Journal of Human Resources* 50 (2): 317–372.

Campbell, T. C., N. Galpin, and S. A. Johnson. 2016. Optimal Inside Debt Compensation and the Value of Equity and Debt. *Journal of Financial Economics* 119 (2): 336–352.

Cassell, C. A., S. X. Huang, J. Manuel Sanchez, and M. D. Stuart. 2012. Seeking Safety: The Relation Between Ceo Inside Debt Holdings and the Riskiness of Firm Investment and Financial Policies. *Journal of Financial Economics* 103 (3): 588–610.

Chen, W., P. Hribar, and S. Melessa. 2023. Standard Error Biases When Using Generated Regressors in Accounting Research. *Journal of Accounting Research* 61 (2): 531–569.

Chen, W., P. Hribar, and S. Melessa. 2024. Estimating the Effects of Factors that Constrain Deviations from Expected or Normal Outcomes. *Working paper, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4572590* .

Cheng, Q. and T. D. Warfield. 2005. Equity Incentives and Earnings Management. *The Accounting Review* 80 (2): 441–476.

Chi, S., S. X. Huang, and J. M. Sanchez. 2017. CEO Inside Debt Incentives and Corporate Tax Sheltering. *Journal of Accounting Research* 55 (4): 837–876.

Cohen, D. A., A. Dey, and T. Z. Lys. 2008. Real and Accrual-Based Earnings Management in the Pre- and Post-Sarbanes-Oxley Periods. *The Accounting Review* 83 (3): 757–787.

Coles, J. L., N. D. Daniel, and L. Naveen. 2006. Managerial Incentives and Risk-Taking. *Journal of Financial Economics* 79 (2): 431–468.

Core, J. E. and W. Guay. 2002. Estimating the Value of Employee Stock Option Portfolios and Their Sensitivities to Price and Volatility. *Journal of Accounting Research* 40 (3): 613–630.

Dechow, P. M. and I. D. Dichev. 2002. The Quality of Accruals and Earnings: The Role of Accrual Estimation Errors. *The Accounting Review* 77 (4): 35–59.

Dhole, S., H. Manchiraju, and I. Suk. 2016. CEO Inside Debt and Earnings Management. *Journal of Accounting, Auditing & Finance* 31 (4): 515–550.

Feenberg, D. and E. Coutts. 1993. An Introduction to the TAXSIM Model. *Journal of Policy Analysis and Management* 12 (1): 189–194.

Francis, J., R. LaFond, P. Olsson, and K. Schipper. 2005. The Market Pricing of Accruals Quality. *Journal of Accounting and Economics* 39 (2): 295–327.

Gao, M., H. Leung, and B. Qiu. 2021. Organization Capital and Executive Performance Incentives. *Journal of Banking & Finance* 123: 106017.

Gassen, J. and D. Veenman. 2024. Estimation Precision and Robust Inference in Archival Research. *Working paper, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4975569* .

Gow, I. D., G. Ormazabal, and D. J. Taylor. 2010. Correcting for Cross-Sectional and Time-Series Dependence in Accounting Research. *The Accounting Review* 85 (2): 483.

He, G. 2015. The Effect of CEO Inside Debt Holdings on Financial Reporting Quality. *Review of Accounting Studies* 20 (1): 501–536.

Hribar, P. and D. C. Nichols. 2007. The Use of Unsigned Earnings Quality Measures in Tests of Earnings Management. *Journal of Accounting Research* 45 (5): 1017–1053.

Jennings, J., J. M. Kim, J. Lee, and D. Taylor. 2024. Measurement Error, Fixed Effects, and False Positives in Accounting Research. *Review of Accounting Studies* 29 (2): 959–995.

Kleibergen, F. and R. Paap. 2006. Generalized Reduced Rank Tests Using the Singular Value Decomposition. *Journal of Econometrics* 133 (1): 97–126.

Kothari, S. P., A. J. Leone, and C. E. Wasley. 2005. Performance Matched Discretionary Accrual Measures. *Journal of Accounting and Economics* 39 (1): 163–197.

Larcker, D. F. and T. O. Rusticus. 2010. On the Use of Instrumental Variables in Accounting Research. *Journal of Accounting and Economics* 49 (3): 186–205.

Lee, J., K. J. Murphy, P. Oh, and M. D. Vance. 2018. Inside Debt and Corporate Investment. *Working paper, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2482857* .

Leone, A. J., M. Minutti-Meza, and C. E. Wasley. 2019. Influential Observations and Inference in Accounting Research. *The Accounting Review* 94 (6): 337–364.

Lim, C.-Y. and H.-T. Tan. 2008. Non-audit Service Fees and Audit Quality: The Impact of Auditor Specialization. *Journal of Accounting Research* 46 (1): 199–246.

Liu, X. 2016. Corruption Culture and Corporate Misconduct. *Journal of Financial Economics* 122 (2): 307–327.

Lyle, M., F. Siano, and T. L. Yohn. 2024. Re-Standardized Financial Statement Data. *Working paper* .

MacKinnon, J. G., M. O. Nielsen, and M. D. Webb. 2023. Cluster-Robust Inference: A Guide to Empirical Practice. *Journal of Econometrics* 232 (2): 272–299.

Owens, E. L., J. S. Wu, and J. Zimmerman. 2017. Idiosyncratic Shocks to Firm Underlying Economics and Abnormal Accruals. *The Accounting Review* 92 (2): 183–219.

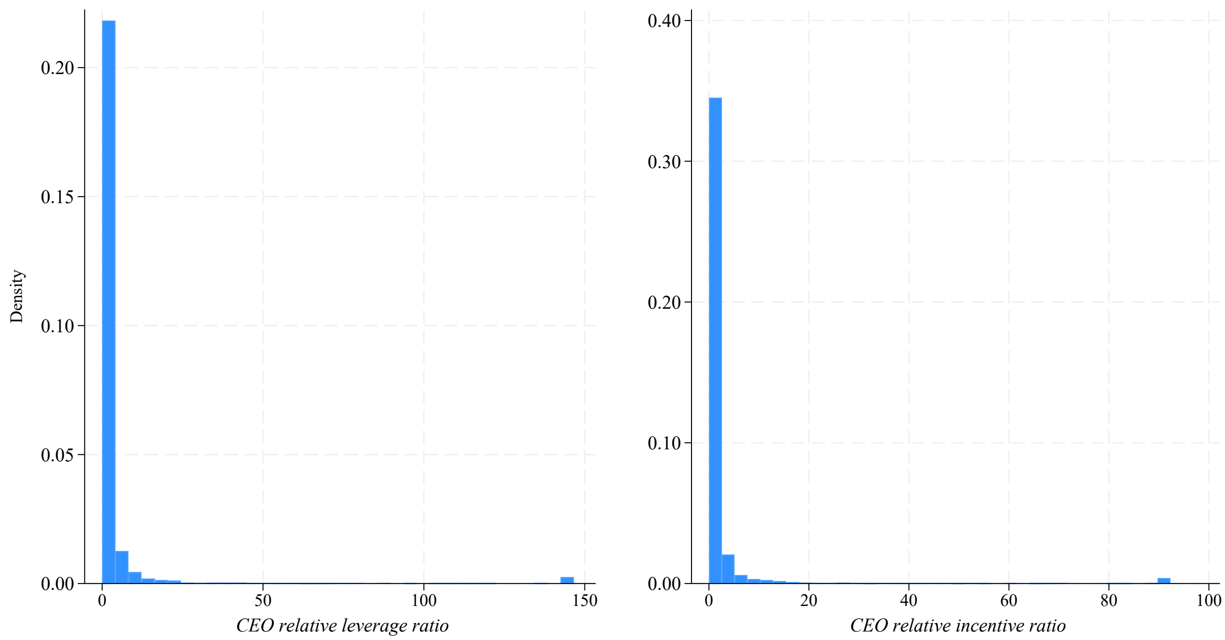
Petersen, M. A. 2009. Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches. *Review of Financial Studies* 22 (1): 435–480.

Stock, J. H., J. H. Wright, and M. Yogo. 2002. A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments. *Journal of Business & Economic Statistics* 20 (4): 518–529.

Wei, C. and D. Yermack. 2011. Investor Reactions to CEOs' Inside Debt Incentives. *The Review of Financial Studies* 24 (11): 3813–3840.

Figure OA-1

Illustration of skewness in test variables after winsorization



This figure presents histograms of the *CEO relative leverage ratio* and *CEO relative incentive ratio* for the replication of the Dhole et al. (2016) sample. The variables are defined as defined as $RelativeLev$ and $RelativeInc$ in Section OA-1 and winsorized at the 1st and 99th percentile values of their distributions.

Table OA-1

Relation between inside debt and financial reporting quality measures after controlling for (subsets) of confounding factors

Panel A: Results after excluding six operating environment characteristics from the regressions

	$\|DA\|$	$\|DA^{ROA}\|$	$\|DA^{CFO}\|$	$\|DA^{Basic}\|$	DD
InsideDebtDum	-0.004***	-0.005***	-0.004***	-0.003*	-0.007***
	(-3.58)	(-4.78)	(-3.54)	(-2.07)	(-5.30)
InsideDebtRatio	-0.003	-0.004*	-0.002	0.001	-0.008**
	(-0.70)	(-1.81)	(-0.58)	(0.28)	(-2.62)
RelativeLevDum	-0.004**	-0.002	-0.002	-0.001	-0.006***
	(-2.70)	(-1.74)	(-1.58)	(-1.02)	(-5.70)
RelativeLevDec	-0.007***	-0.005***	-0.005**	-0.004	-0.011***
	(-3.30)	(-3.48)	(-2.31)	(-1.69)	(-6.31)

Number of coefficients that are negative and significant at $p < 0.10$: 13 / 20.
Number of coefficients that are negative and significant at $p < 0.05$: 11 / 20.
Number of coefficients that are negative and significant at $p < 0.01$: 8 / 20.

Panel B: Results after excluding operating environment characteristics *and* firm pension plans indicator

	$\|DA\|$	$\|DA^{ROA}\|$	$\|DA^{CFO}\|$	$\|DA^{Basic}\|$	DD
InsideDebtDum	-0.006***	-0.007***	-0.006***	-0.004***	-0.009***
	(-4.68)	(-6.40)	(-5.00)	(-3.23)	(-6.66)
InsideDebtRatio	-0.006*	-0.009***	-0.006*	-0.002	-0.013***
	(-1.83)	(-3.83)	(-1.85)	(-0.58)	(-4.11)
RelativeLevDum	-0.005***	-0.003***	-0.003**	-0.002*	-0.007***
	(-3.75)	(-3.15)	(-2.73)	(-1.79)	(-6.76)
RelativeLevDec	-0.009***	-0.008***	-0.007***	-0.005**	-0.013***
	(-4.61)	(-5.27)	(-3.80)	(-2.88)	(-7.60)

Number of coefficients that are negative and significant at $p < 0.10$: 19 / 20.
Number of coefficients that are negative and significant at $p < 0.05$: 16 / 20.
Number of coefficients that are negative and significant at $p < 0.01$: 14 / 20.

Notes: This table extends Table 5 reported in the paper. Panel A is identical to Panel B of Table 5 in the paper, where the vector of control variables contains all control variables except for the six operating environment variables: σCFO, $\sigma Sales$, $Idioshock2$, ROA, *Loss*%, and *R&D*. In Panel B, we *additionally* exclude the *PensionPlan* indicator variable from the estimations. All variables are defined in Appendix A in the paper. Continuous variables are winsorized at the 1st and 99th percentiles of their distributions. All regressions are estimated using OLS, and standard errors are clustered by firm and year. *t*-statistics are presented in parentheses below the coefficient estimates. *, **, and *** refer to statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

Table OA-2

Alternative outcome variable: Severe restatements

Panel A: Baseline association between inside debt variables and restatement incidence

	Test variable:			
	InsidedebtDum	*InsidedebtRatio*	*RelativeLevDum*	*RelativeLevDec*
	(1)	(2)	(3)	(4)
	Res_{Severe}	Res_{Severe}	Res_{Severe}	Res_{Severe}
Test variable	-0.001	0.001	-0.003	-0.004
	(-0.18)	(0.08)	(-1.44)	(-1.08)
Control variables	No	No	No	No
FE	Ind-Y	Ind-Y	Ind-Y	Ind-Y
No. obs.	19,354	19,354	16,994	16,994
Adj. R^2	0.002	0.002	0.004	0.004

Panel B: Results after controlling for other variables

	Test variable:			
	InsidedebtDum	*InsidedebtRatio*	*RelativeLevDum*	*RelativeLevDec*
	(1)	(2)	(3)	(4)
	Res_{Severe}	Res_{Severe}	Res_{Severe}	Res_{Severe}
Test variable	-0.000	-0.000	-0.002	-0.003
	(-0.13)	(-0.02)	(-0.90)	(-0.60)
Control variables	Included	Included	Included	Included
FE	Ind-Y	Ind-Y	Ind-Y	Ind-Y
No. obs.	19,354	19,354	16,994	16,994
Adj. R^2	0.005	0.005	0.006	0.006

Notes: This table presents tests of the relation between inside debt variables and severe accounting restatements before (Panel A) and after (Panel B) controlling for other factors. The tests are the same as presented in the paper, except that the dependent variable captures a more narrow category of restatements related to fraud and SEC investigations (as identified by variables `RES_FRAUD` and `RES_SEC_INVEST` in Audit Analytics). Continuous variables are winsorized at the 1st and 99th percentiles of their distributions. All regressions are estimated using OLS (i.e., as linear probability models) and include two-digit SIC industry-year fixed effects ("Ind-Y"). Standard errors are adjusted for clustering by firm and year. t-statistics are presented in parentheses below the coefficient estimates. *, **, and *** refer to statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

Table OA-3

Alternative outcome variable: Meeting/beating earnings forecasts

Panel A: Inside debt and meeting or just beating analyst earnings forecasts (MJB)

	Test variable:			
	InsidedebtDum	*InsidedebtRatio*	*RelativeLevDum*	*RelativeLevDec*
	(1)	(2)	(3)	(4)
	MJB	*MJB*	*MJB*	*MJB*
Test variable	0.003	-0.019	0.005	0.003
	(0.34)	(-0.82)	(0.56)	(0.20)
Control variables	No	No	No	No
FE	Ind-Y	Ind-Y	Ind-Y	Ind-Y
No. obs.	18,856	18,856	16,604	16,604
Adj. R^2	0.033	0.034	0.032	0.032

Panel B: Inside debt and meeting or beating analyst earnings forecasts (MBE)

	Test variable:			
	InsidedebtDum	*InsidedebtRatio*	*RelativeLevDum*	*RelativeLevDec*
	(1)	(2)	(3)	(4)
	MBE	*MBE*	*MBE*	*MBE*
Test variable	0.020*	0.011	-0.002	0.008
	(2.07)	(0.45)	(-0.14)	(0.50)
Control variables	Included	Included	Included	Included
FE	Ind-Y	Ind-Y	Ind-Y	Ind-Y
No. obs.	18,856	18,856	16,604	16,604
Adj. R^2	0.069	0.069	0.070	0.070

Notes: This table presents tests of the relation between inside debt variables and firms' propensity to meet or *just* beat ("MJB", Panel A), and meet or beat *by any amount* ("MBE", Panel B), the consensus analyst forecast. The tests are the same as those presented in Panel B of Table 7 in the paper, but with different dependent variables: MJB is an indicator variable set equal to 1 for firm-years in which reported earnings per share are equal to, or beat by one cent, the latest median consensus forecast before the annual earnings announcement, and 0 otherwise; MBE is an indicator variable set equal to 1 for firm-years in which reported earnings per share meet or beat the consensus forecast by any amount, and 0 otherwise. Analyst forecast and actual earnings data are from the IBES unadjusted summary files. Continuous variables are winsorized at the 1st and 99th percentiles of their distributions. All regressions are estimated using OLS (i.e., as linear probability models) and include two-digit SIC industry-year fixed effects ("Ind-Y"). Standard errors are adjusted for clustering by firm and year. t-statistics are presented in parentheses below the coefficient estimates. *, **, and *** refer to statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

Table OA-4

Examining discretionary accruals using quantile regression

	Regression quantile (q)								
	$q = 0.1$	$q = 0.2$	$q = 0.3$	$q = 0.4$	$q = 0.5$	$q = 0.6$	$q = 0.7$	$q = 0.8$	$q = 0.9$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	$TACC$	$TACC$	$TACC$	$TACC$	$TACC$	$TACC$	$TACC$	$TACC$	$TACC$
$RelativeLevDum$	-0.005**	-0.003	-0.003*	-0.002*	-0.002	-0.001	0.000	0.000	-0.002
	(-2.31)	(-1.59)	(-1.96)	(-1.69)	(-1.47)	(-0.55)	(0.06)	(0.03)	(-1.25)
$ln(Assets)$	0.007***	0.006***	0.006***	0.005***	0.005***	0.005***	0.005***	0.005***	0.005***
	(6.10)	(6.17)	(6.66)	(5.85)	(5.89)	(5.87)	(5.53)	(5.63)	(5.68)
BTM	0.018***	0.017***	0.017***	0.016***	0.016***	0.015***	0.013***	0.011***	0.007***
	(6.94)	(7.66)	(8.27)	(8.35)	(8.09)	(7.71)	(6.53)	(5.19)	(3.63)
$Leverage$	0.008	0.018***	0.018***	0.018***	0.015***	0.015***	0.016***	0.016***	0.015**
	(1.09)	(2.96)	(3.31)	(3.48)	(2.89)	(2.78)	(2.85)	(2.75)	(2.52)
$ln(Firm\ Age)$	0.009***	0.007***	0.007***	0.006***	0.006***	0.005***	0.005***	0.004***	0.003**
	(4.66)	(4.63)	(5.28)	(5.82)	(5.70)	(4.86)	(4.58)	(3.40)	(2.37)
$Salesgr$	0.001	0.003	-0.001	0.009**	0.004	0.004	0.001	0.003	0.015**
	(0.08)	(0.52)	(-0.13)	(2.02)	(0.91)	(0.78)	(0.10)	(0.52)	(2.14)
$Inst$	-0.002	-0.001	-0.002	-0.004	-0.004	-0.006*	-0.006*	-0.008**	-0.011***
	(-0.47)	(-0.25)	(-0.78)	(-1.42)	(-1.26)	(-1.83)	(-1.87)	(-2.10)	(-2.67)
$ln(Analysts)$	-0.015***	-0.014***	-0.014***	-0.013***	-0.013***	-0.013***	-0.012***	-0.013***	-0.014***
	(-8.47)	(-9.22)	(-10.34)	(-10.24)	(-10.36)	(-9.94)	(-9.83)	(-9.69)	(-9.13)
$Big4$	0.002	-0.001	-0.002	0.000	0.001	-0.000	-0.001	-0.001	0.000
	(0.63)	(-0.33)	(-0.79)	(0.10)	(0.25)	(-0.00)	(-0.34)	(-0.42)	(0.02)
$PensionPlan$	0.021***	0.020***	0.018***	0.016***	0.015***	0.014***	0.013***	0.012***	0.011***
	(8.32)	(9.84)	(10.37)	(10.10)	(9.67)	(8.79)	(8.05)	(7.43)	(5.85)
$ln(CashComp)$	0.002*	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
	(1.67)	(1.25)	(1.24)	(1.55)	(1.34)	(0.99)	(1.30)	(0.71)	(1.21)
$CEO\ ownership$	0.093***	0.073***	0.051**	0.041*	0.043*	0.047**	0.035*	0.025	0.040
	(4.34)	(3.61)	(2.43)	(1.77)	(1.83)	(2.10)	(1.66)	(1.05)	(1.28)
$ln(Delta)$	-0.006***	-0.006***	-0.005***	-0.004***	-0.004***	-0.004***	-0.003***	-0.003***	-0.004***
	(-5.68)	(-5.99)	(-5.92)	(-5.96)	(-5.43)	(-5.26)	(-4.29)	(-3.97)	(-4.83)
$ln(Vega)$	0.001**	0.001*	0.000	0.000	-0.000	-0.000	-0.000	-0.001*	-0.001
	(2.57)	(1.89)	(1.14)	(0.98)	(-0.33)	(-0.76)	(-1.17)	(-1.81)	(-1.61)
$ln(CEO\ Age)$	0.003	0.000	0.004	0.009	0.007	0.009	0.006	0.009	0.012
	(0.28)	(0.05)	(0.62)	(1.61)	(1.20)	(1.51)	(1.06)	(1.52)	(1.61)
$ln(CEO\ Tenure)$	0.004***	0.004***	0.003***	0.002***	0.002***	0.002*	0.001	0.001	0.001
	(3.03)	(3.54)	(3.17)	(2.72)	(2.90)	(1.82)	(1.08)	(1.13)	(1.04)
σCFO	-0.500***	-0.333***	-0.192***	-0.081**	0.049	0.180***	0.334***	0.516***	0.793***
	(-12.74)	(-9.12)	(-6.53)	(-2.43)	(1.56)	(5.48)	(8.67)	(12.98)	(19.36)
$\sigma Sales$	-0.009	0.005	0.006	0.006	-0.002	-0.007	-0.012*	-0.016**	-0.014*
	(-0.87)	(0.67)	(0.91)	(0.92)	(-0.29)	(-1.11)	(-1.80)	(-2.28)	(-1.73)
$Idioshock2$	-0.100***	-0.093***	-0.071***	-0.052**	-0.033*	-0.017	0.015	0.044**	0.050**
	(-3.62)	(-3.60)	(-3.25)	(-2.54)	(-1.69)	(-0.86)	(0.72)	(2.10)	(2.16)
ROA	0.507***	0.478***	0.448***	0.431***	0.427***	0.418***	0.404***	0.394***	0.431***
	(42.61)	(38.59)	(34.82)	(32.91)	(25.97)	(24.79)	(19.36)	(15.53)	(26.17)
$Loss\%$	0.007	0.005	0.007*	0.008**	0.011***	0.012***	0.011***	0.012***	0.016***
	(1.48)	(1.04)	(1.88)	(2.29)	(2.86)	(3.32)	(3.22)	(3.30)	(3.67)
$R\&D$	-0.101***	-0.085***	-0.079***	-0.087***	-0.081***	-0.040	-0.024	0.005	0.036
	(-3.34)	(-3.39)	(-3.35)	(-4.09)	(-3.08)	(-1.49)	(-0.93)	(0.15)	(1.28)
No. obs.	16,994	16,994	16,994	16,994	16,994	16,994	16,994	16,994	16,994
Pseudo R^2	0.415	0.339	0.286	0.243	0.207	0.180	0.166	0.172	0.221

Notes: This table presents estimations following Chen et al. (2024) using quantile regressions with total accruals ($TACC$) as dependent variable and the CEO relative ratio indicator variable (*RelativeLevDum*) and control variables as independent variables. All variables are defined in the paper. Standard errors are clustered by firm and t-statistics are presented in parentheses below the coefficient estimates. The labels *, **, and *** refer to statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

Table OA-5

Descriptive statistics for the He (2015) replication

	Mean	St. dev.	p25	Median	p75	N
Main variables						
$\lvert DA\rvert$	0.0449	0.0521	0.0139	0.0305	0.0567	5,522
DA	0.0033	0.0687	-0.0214	0.0089	0.0368	5,522
$\lvert DA^{ROA}\rvert$	0.0397	0.0414	0.0126	0.0281	0.0521	5,522
DD	0.0377	0.0345	0.0171	0.0273	0.0447	4,778
Res	0.1018	0.3024	0.0000	0.0000	0.0000	5,522
$AnaSur$	0.2173	0.4125	0.0000	0.0000	0.0000	5,480
$InsiDebt$	0.3401	0.4738	0.0000	0.0000	1.0000	5,522
Control variables						
$Insti$	0.8133	0.1831	0.7288	0.8504	0.9504	5,522
$Indpshare$	0.0145	0.0506	0.0013	0.0035	0.0084	5,522
$CEOowner$	2.0264	5.0696	0.0000	0.4820	1.8000	5,522
$Vega$	3.7469	1.8926	2.7035	4.0882	5.1293	5,522
$Big4$	0.8995	0.3007	1.0000	1.0000	1.0000	5,522
$Indp$	0.7844	0.1136	0.7143	0.8000	0.8750	5,522
$Loss$	0.0601	0.2377	0.0000	0.0000	0.0000	5,522
ROA	0.0438	0.1005	0.0228	0.0497	0.0841	5,522
$Size$	7.8771	1.4828	6.8180	7.7481	8.8056	5,522
Mb	2.8903	10.1739	1.4347	2.1406	3.3420	5,522
$AnaCov$	2.2511	0.6814	1.7918	2.3026	2.7726	5,522
$Debt$	0.1976	0.1529	0.0739	0.1826	0.2905	5,522
$SalesGrowth$	0.0871	0.2655	-0.0129	0.0708	0.1578	5,522
$StdCash$	0.0390	0.0324	0.0186	0.0305	0.0481	5,522
$StdSales$	0.1245	0.1323	0.0488	0.0904	0.1556	5,522

Notes: This table presents descriptive statistics for our replication of the He (2015) sample. Variable definitions follow the descriptions provided by He (2015) except for DA^{ROA}, which is obtained from accrual model estimations that control for ROA. See the explanations in Section OA-5.1 for more details on the sample selection and research design choices. Similar to He (2015), the continuous variables are not winsorized except for the inputs to the first-stage accrual-model estimations.

Table OA-6

Replicating Table 3 of He (2015) for absolute discretionary accruals

	(1) $\lvert DA \rvert$	(2) $\lvert DA \rvert$	(3) $\lvert DA \rvert$	(4) $\lvert DA \rvert$
$InsiDebt$	-0.007***	-0.001	-0.003**	-0.004***
	(-4.78)	(-0.69)	(-2.21)	(-2.81)
$IndpShare$		0.014	0.020	0.025
		(1.44)	(1.14)	(1.32)
$CEOowner$		-0.000	-0.000**	-0.000***
		(-0.64)	(-2.44)	(-2.67)
$Vega$		-0.000	-0.000	-0.001
		(-0.30)	(-0.80)	(-1.39)
$Big4$		-0.002	-0.003	-0.005*
		(-0.64)	(-0.99)	(-1.66)
$Indp$		-0.010	-0.007	-0.009
		(-1.50)	(-1.03)	(-1.27)
$Loss$		-0.010**	0.025***	0.036***
		(-2.28)	(4.94)	(6.88)
ROA		-0.199***		
		(-9.12)		
$Size$		-0.000	-0.001*	-0.003***
		(-0.12)	(-1.82)	(-4.25)
Mb		0.000*	0.000	0.000**
		(1.80)	(0.73)	(2.02)
$AnaCov$		0.002	0.001	0.002
		(1.25)	(0.42)	(1.14)
$Debt$		-0.025***	-0.017***	-0.028***
		(-5.21)	(-3.85)	(-5.79)
$SalesGrowth$		0.018***	0.011**	0.017***
		(2.77)	(2.30)	(3.59)
$StdCash$		0.396***	0.381***	
		(7.74)	(10.11)	
$StdSales$		0.007	0.011	
		(1.02)	(1.62)	
$Insti$		0.004	0.001	0.002
		(1.12)	(0.31)	(0.54)
Year FE	Included	Included	Included	Included
No. obs.	5,522	5,522	5,522	5,522
Adj. R^2	0.040	0.255	0.146	0.094

Notes: This table presents results from linear regressions estimated using OLS. Variables are defined as in He (2015) and Section OA-5.1. Standard errors are clustered by firm and t-statistics are presented in parentheses below the coefficient estimates. The labels *, **, and *** refer to statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

Table OA-7

Replicating Table 3 of He (2015) with alternative dependent variable

	(1) $\lvert DA^{ROA}\rvert$	(2) $\lvert DA^{ROA}\rvert$	(3) $\lvert DA^{ROA}\rvert$	(4) $\lvert DA^{ROA}\rvert$	(5) $\lvert DA^{ROA}\rvert$
$InsiDebt$	-0.007***	-0.003**	-0.003**	-0.002	-0.001
	(-5.24)	(-2.54)	(-2.30)	(-1.48)	(-1.11)
$IndpShare$		-0.007	-0.007	-0.011	-0.010
		(-0.67)	(-0.77)	(-1.19)	(-1.07)
$CEOowner$		0.000	-0.000	-0.000	0.000
		(0.65)	(-0.26)	(-0.16)	(0.31)
$Vega$		-0.000	-0.000	-0.000	-0.000
		(-0.22)	(-0.28)	(-1.17)	(-1.02)
$Big4$		0.001	-0.000	0.001	0.002
		(0.49)	(-0.01)	(0.69)	(0.81)
$Indp$		-0.005	-0.006	-0.003	-0.003
		(-0.82)	(-1.08)	(-0.58)	(-0.47)
$Loss$		-0.003	-0.005	-0.006*	-0.009***
		(-0.93)	(-1.61)	(-1.83)	(-2.70)
ROA		-0.053***	-0.075***	-0.075***	-0.051***
		(-3.76)	(-5.38)	(-5.32)	(-3.18)
$Size$		-0.000	-0.001	-0.001	-0.000
		(-0.15)	(-0.91)	(-0.98)	(-0.27)
Mb		0.000*	0.002***	0.002***	0.002***
		(1.91)	(6.42)	(5.76)	(5.47)
$AnaCov$		0.001	0.000	0.000	0.000
		(0.72)	(0.12)	(0.22)	(0.28)
$Debt$		-0.016***	-0.018***	-0.021***	-0.023***
		(-3.27)	(-4.08)	(-4.08)	(-4.30)
$SalesGrowth$		0.012**	0.016***	0.014***	0.014***
		(2.20)	(4.56)	(4.02)	(4.00)
$StdCash$		0.393***	0.373***	0.356***	0.275***
		(10.68)	(13.02)	(12.34)	(7.98)
$StdSales$		0.006	0.008	0.013*	0.004
		(1.05)	(1.20)	(1.91)	(0.63)
$Insti$		0.008**	0.007**	0.004	0.004
		(2.17)	(2.05)	(1.15)	(1.22)
$OE\ factor$					0.005***
					(3.67)
Year FE	Included	Included	Included	Included	Included
Industry FE	No	No	No	Included	Included
Winsorization	No	No	Yes	Yes	Yes
No. obs.	5,522	5,522	5,522	5,522	5,403
Adj. R^2	0.012	0.148	0.169	0.186	0.189

Notes: This table presents results from linear regressions estimated using OLS. Variables are defined as in He (2015) and Section OA-5.1. The additional variable $OE\ factor$ is the first principal component factor based on six operating environment characteristics examined in the main paper: σCFO, $\sigma Sales$, $Idioshock2$, ROA, $Loss\%$, and $R\&D$. Winsorization refers to winsorization of continuous variables at the 1st and 99th percentiles. Industry fixed effects (FE) are based on two-digit SIC codes. Standard errors are clustered by firm and t-statistics are presented in parentheses below the coefficient estimates. The labels *, **, and *** refer to statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

Table OA-8

Replicating Table 4 of He (2015)

	(1) DD	(2) DD	(3) DD	(4) DD	(5) DD
$InsiDebt$	-0.008***	-0.003*	-0.002	-0.002	-0.000
	(-4.96)	(-1.77)	(-1.55)	(-1.28)	(-0.28)
$IndpShare$		-0.006	-0.009	-0.002	-0.003
		(-0.53)	(-0.81)	(-0.20)	(-0.24)
$CEOowner$		-0.000**	-0.000**	-0.000***	-0.000
		(-2.01)	(-2.47)	(-3.15)	(-1.17)
$Vega$		0.000	0.000	-0.000	-0.000
		(1.42)	(1.33)	(-1.38)	(-0.59)
$Big4$		-0.007**	-0.007**	-0.004	-0.008**
		(-2.19)	(-2.38)	(-1.35)	(-2.46)
$Indp$		-0.001	-0.002	0.005	-0.005
		(-0.20)	(-0.28)	(0.78)	(-0.83)
$Loss$		0.009**	0.007*	0.007*	-0.005
		(1.99)	(1.71)	(1.69)	(-1.24)
ROA		-0.090***	-0.115***	-0.090***	-0.007
		(-6.98)	(-9.26)	(-7.48)	(-0.45)
$Size$		-0.002**	-0.002***	-0.001*	0.001
		(-2.30)	(-2.95)	(-1.84)	(1.13)
Mb		0.000	0.001***	0.000	0.000
		(1.15)	(3.74)	(1.04)	(1.12)
$AnaCov$		0.002	0.001	0.001	0.001
		(1.13)	(1.06)	(0.61)	(0.74)
$Debt$		0.002	-0.001	0.008	0.003
		(0.42)	(-0.20)	(1.55)	(0.67)
$SalesGrowth$		0.002	0.006**	0.001	-0.001
		(0.67)	(2.00)	(0.30)	(-0.37)
$StdCash$		0.317***	0.319***	0.280***	0.017
		(7.89)	(9.27)	(7.25)	(0.36)
$StdSales$		0.010	0.017**	0.014	-0.012
		(1.11)	(2.35)	(1.18)	(-1.26)
$Insti$		0.008**	0.007**	0.001	0.008**
		(2.15)	(2.12)	(0.26)	(2.09)
$OE\ factor$					0.021***
					(10.04)
Year FE	Included	Included	Included	Included	Included
Industry FE	No	No	No	Included	No
Winsorization	No	No	Yes	No	No
No. obs.	4,778	4,778	4,778	4,778	4,675
Adj. R^2	0.020	0.243	0.265	0.307	0.325

Notes: This table presents results from linear regressions estimated using OLS. Variables are defined as in He (2015) and Section OA-5.1. The additional variable $OE\ factor$ is defined in the notes to Table OA-7. Winsorization refers to the winsorization of continuous variables at the 1st and 99th percentiles. Industry fixed effects (FE) are based on two-digit SIC codes. Standard errors are clustered by firm and t-statistics are presented in parentheses below the coefficient estimates. The labels *, **, and *** refer to statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

Table OA-9

Replicating Tables 5 and 6 of He (2015)

	(1) Res	(2) $AnaSur$	(3) Res_{Severe}	(4) $AnaSur¢$
$InsiDebt$	-0.065	0.133***	-0.332**	0.084*
	(-0.91)	(2.85)	(-2.52)	(1.69)
$IndpShare$	0.334	0.069	2.027***	0.459
	(0.57)	(0.17)	(2.59)	(0.87)
$CEOowner$	0.002	0.008*	-0.094**	0.010**
	(0.35)	(1.80)	(-2.30)	(2.36)
$Vega$	0.002	0.031**	0.046	0.032**
	(0.09)	(2.30)	(1.04)	(2.27)
$Big4$	0.565***	-0.036	0.851**	-0.031
	(4.11)	(-0.47)	(2.24)	(-0.39)
$Indp$	-0.373	-0.298	0.924	-0.481**
	(-1.23)	(-1.47)	(1.07)	(-2.34)
$Loss$	-0.097	-0.130	0.008	0.020
	(-0.73)	(-1.11)	(0.02)	(0.17)
ROA	-0.680**	1.842***	1.295	1.129***
	(-2.35)	(3.75)	(1.06)	(3.69)
$Size$	-0.037	0.031	-0.188***	-0.098***
	(-1.03)	(1.30)	(-2.73)	(-3.86)
Mb	-0.003	0.004**	-0.003	0.002
	(-1.12)	(2.36)	(-0.85)	(1.60)
$AnaCov$	-0.159**	0.262***	-0.243**	0.386***
	(-2.50)	(5.44)	(-2.22)	(7.66)
$Debt$	0.197	0.241*	0.341	0.168
	(0.85)	(1.75)	(0.75)	(1.22)
$SalesGrowth$	-0.214*	0.046	0.092	-0.003
	(-1.70)	(0.64)	(1.00)	(-0.04)
$StdCash$	-0.074	-3.321***	2.258	-2.700***
	(-0.06)	(-3.21)	(1.21)	(-3.02)
$StdSales$	0.083	-0.170	-1.048*	-0.242
	(0.29)	(-0.76)	(-1.87)	(-0.96)
$Insti$	0.122	0.013	-0.248	-0.147
	(0.75)	(0.10)	(-0.60)	(-1.13)
Year FE	Included	Included	Included	Included
Winsorization	No	No	No	No
No. obs.	5,522	5,480	5,522	5,480
Pseudo R^2	0.0372	0.0591	0.1350	0.0339

Notes: This table presents results from probit model estimations with a binary dependent variable. Variables are defined as in He (2015) and Section OA-5.1. The additional dependent variables Res_{Severe} and $AnaSur¢$ are defined in Section OA-5.4. Standard errors are clustered by firm and z-statistics are presented in parentheses below the coefficient estimates. The labels *, **, and *** refer to statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

Table OA-10

Instrumental variables estimation for the He (2015) sample

	(1) 1st stage Probit $InsiDebt$	(2) 1st stage OLS $InsiDebt$	(3) 2nd stage OLS $\lvert DA^{ROA} \rvert$	(4) 2nd stage OLS DD	(5) 2nd stage OLS Res_{Severe}
$InsiDebt$ (predicted)			-0.026	-0.003	0.153
			(-0.73)	(-0.07)	(1.02)
$StateWage$ (IV)	0.038***	0.013***			
	(2.63)	(2.75)			
$StateMortage$ (IV)	-0.030**	-0.010**			
	(-2.26)	(-2.35)			
$IndpShare$	-0.864	-0.241	-0.011	-0.005	0.208
	(-1.31)	(-1.34)	(-0.76)	(-0.27)	(1.35)
$CEOowner$	-0.065***	-0.013***	-0.000	-0.000	0.001
	(-4.00)	(-6.42)	(-0.47)	(-0.30)	(0.70)
$Vega$	-0.019	-0.005	-0.000	0.000	0.001
	(-0.97)	(-0.81)	(-0.35)	(0.90)	(0.60)
$Big4$	0.204*	0.060*	0.003	-0.005	0.005
	(1.69)	(1.80)	(0.85)	(-1.03)	(0.43)
$Indp$	1.745***	0.540***	0.006	-0.001	-0.060
	(5.80)	(6.04)	(0.32)	(-0.03)	(-0.75)
$Loss$	-0.229*	-0.062*	-0.004	0.010*	0.012
	(-1.79)	(-1.82)	(-0.99)	(1.79)	(0.66)
ROA	1.267***	0.310***	-0.047***	-0.090***	-0.021
	(3.71)	(3.61)	(-2.59)	(-4.46)	(-0.33)
$Size$	0.183***	0.062***	0.001	-0.002	-0.013
	(5.57)	(6.10)	(0.59)	(-0.58)	(-1.22)
Mb	0.001	0.000	0.000*	0.000	-0.000
	(0.55)	(0.46)	(1.69)	(1.41)	(-0.76)
$AnaCov$	-0.267***	-0.080***	-0.001	0.001	0.007
	(-4.33)	(-4.17)	(-0.37)	(0.31)	(0.50)
$Debt$	-2.073***	-0.560***	-0.028	0.001	0.105
	(-9.42)	(-9.70)	(-1.36)	(0.05)	(1.23)
$SalesGrowth$	-0.664***	-0.127***	0.009	0.002	0.023
	(-6.10)	(-4.09)	(1.35)	(0.29)	(1.06)
$StdCash$	-1.764*	-0.493*	0.384***	0.319***	0.161
	(-1.65)	(-1.67)	(8.56)	(6.07)	(1.11)
$StdSales$	-0.125	-0.052	0.004	0.009	-0.009
	(-0.53)	(-0.79)	(0.62)	(0.87)	(-0.44)
$Insti$	-0.658***	-0.214***	0.003	0.008	0.023
	(-4.04)	(-4.00)	(0.41)	(0.75)	(0.68)
Year FE	Included	Included	Included	Included	Included
Winsorization	No	No	No	No	No
No. obs.	5,444	5,444	5,444	4,710	5,444
Pseudo / Adj. R^2	0.131	0.137	0.147	0.243	0.016
Partial χ^2 / F	7.06	3.83			
Partial χ^2 / F (p-value)	0.029	0.022			

Notes: This table presents results from two-stage instrumental variables estimations using state-level tax rates as instruments. Columns (1) and (2) present results from first-stage regression estimations using probit and OLS, respectively. Columns (3) through (5) present the second-stage results from 2SLS estimations where the first-stage estimation follows that in column (2). For the smaller sample in column (4), the first-stage estimation is performed on the reduced sample (untabulated) due to additional data requirements to calculate the DD variable. Variables are defined as in He (2015) and Section OA-5.1. The instrumental variables are defined in Section OA-5.5. The additional dependent variable Res_{Severe} is defined in Section OA-5.4. The partial χ^2-statistic (F-statistic) for the probit (OLS) estimation in the first (second) column is obtained by testing whether the coefficients on the instrumental variables $StateWage$ and $StateMortgage$ are jointly equal to zero. Standard errors are clustered by firm. For the second-stage estimations, standard errors are obtained from a cluster-robust bootstrap procedure that estimates both the first- and second-stage regressions on each bootstrap sample. For probit (OLS) estimations, z-statistics (t-statistics) are presented in parentheses below the coefficient estimates. The labels *, **, and *** refer to statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

Table OA-11

Descriptive statistics for the Dhole et al. (2016) replication

	Mean	St. dev.	p25	Median	p75
CEO compensation variables					
Inside debt holdings ($mln)	4.7710	10.1875	0.0000	0.5995	4.6674
Deferred comp. ($mln)	2.0287	5.6212	0.0000	0.1033	1.5239
Pension ($mln)	2.6198	6.0433	0.0000	0.0000	1.9317
CEO relative leverage ratio	3.4787	16.2750	0.0000	0.1742	1.2447
CEO relative incentive ratio	2.3615	10.5973	0.0000	0.1367	0.8997
Total comp. ($mln)	5.4432	5.4911	1.7811	3.6977	6.9226
Cash pay ($mln)	0.9941	0.7657	0.5977	0.8280	1.0844
Equity holdings ($mln)	55.3311	154.9033	5.5934	14.5624	39.1247
Delta ($mln)	0.6645	1.6057	0.0753	0.2046	0.5564
Vega ($mln)	0.1225	0.1924	0.0128	0.0508	0.1466
Dependent variables					
$CORRK$	0.5186	0.2738	0.2900	0.5300	0.7600
$-CORR$	0.5440	0.5462	0.3432	0.7927	0.9450
$\|DA\|$	0.0523	0.0478	0.0182	0.0396	0.0700
$\|RCFO\|$	0.0666	0.0603	0.0231	0.0504	0.0918
$\|RPROD\|$	0.1417	0.1299	0.0492	0.1049	0.1947
$\|RDISX\|$	0.1416	0.1476	0.0363	0.0935	0.1941
$RMPROXY$	-0.0089	2.2002	-1.1826	-0.0054	1.2402
$\|RMPROXY\|$	1.6392	1.5378	0.5264	1.2074	2.2496
$JustMBE$	0.1791	0.3835	0.0000	0.0000	0.0000
$LargeMBE$	0.3129	0.4637	0.0000	0.0000	1.0000
$JustMBE_{t+1}$	0.1678	0.3737	0.0000	0.0000	0.0000
Control variables					
Total assets ($bln)	6.8020	15.6375	0.6986	1.7740	5.2452
Evol	0.0586	0.0752	0.0164	0.0313	0.0682
Firm age	27.1116	16.7123	14.0000	21.0000	41.0000
Market-to-book ratio	2.7006	3.5607	1.3549	2.1176	3.3768
Leverage	0.4696	0.8970	0.0804	0.2099	0.4619
Net operating assets	0.5456	0.2092	0.4441	0.5898	0.6926
Implicit claims	0.4831	0.3713	0.2469	0.5963	0.7800
Litigation dummy	0.3100	0.4626	0.0000	0.0000	1.0000
ln(Shares)	4.4306	1.2178	3.5217	4.2087	5.1084
Big4	0.8785	0.3267	1.0000	1.0000	1.0000
Issue	0.1075	0.3098	0.0000	0.0000	0.0000
String of MBE	2.9180	1.1017	2.0000	3.0000	4.0000
Analyst following	9.7989	6.7203	5.0000	8.0000	14.0000

Notes: This table presents descriptive statistics for our replication of the Dhole et al. (2016) sample. Variable definitions follow the descriptions provided by Dhole et al. (2016). See the explanations in Section OA-6.1.1 for more details on the sample selection and research design choices. Following Dhole et al. (2016), the continuous variables are winsorized at the 1st and 99th percentiles of their distributions.

Table OA-12
Replication of Table 4 in Dhole et al. (2016)

	(1) *CORRK*	(2) *CORRK*	(3) $\|DA\|$	(4) $\|DA\|$
CEO relative leverage ratio	0.000		-0.000	
	(0.17)		(-0.51)	
	[0.872]		[0.639]	
CEO relative incentive ratio		0.000		-0.000
		(0.28)		(-0.44)
		[0.794]		[0.683]
ln(1+Cash pay)	0.004	0.004	-0.004	-0.004
	(0.35)	(0.35)	(-1.44)	(-1.45)
ln(1+Delta)	0.014**	0.014**	0.001	0.001
	(3.13)	(3.12)	(0.87)	(0.87)
ln(1+Vega)	-0.006	-0.006	-0.000	-0.000
	(-1.63)	(-1.63)	(-0.66)	(-0.68)
$\|RMPROXY\|$	0.015***	0.015***	0.006***	0.006***
	(4.96)	(4.95)	(7.88)	(7.89)
Evol	-1.536***	-1.536***	0.107**	0.107**
	(-17.32)	(-17.30)	(3.30)	(3.31)
ln(Total assets)	-0.008	-0.008	-0.005*	-0.005*
	(-0.95)	(-0.95)	(-2.69)	(-2.70)
ln(Firm age)	0.003	0.003	-0.003**	-0.003**
	(0.26)	(0.26)	(-3.14)	(-3.14)
Market-to-book ratio	0.002	0.002	0.001	0.001
	(1.16)	(1.16)	(2.05)	(2.05)
Leverage	-0.017**	-0.017**	0.001	0.001
	(-3.16)	(-3.14)	(0.88)	(0.89)
Net operating assets	-0.077*	-0.077*	-0.009	-0.009
	(-2.52)	(-2.52)	(-0.94)	(-0.94)
Implicit claims	-0.016	-0.016	0.017***	0.017***
	(-0.82)	(-0.82)	(6.75)	(6.75)
Litigation dummy	-0.055	-0.055	0.004	0.004
	(-1.99)	(-1.99)	(1.25)	(1.25)
ln(Shares)	-0.026**	-0.026**	0.005*	0.005*
	(-3.02)	(-3.02)	(2.24)	(2.25)
Big4	-0.001	-0.001	-0.002	-0.002
	(-0.04)	(-0.04)	(-1.19)	(-1.20)
Issue	0.006	0.006	0.008**	0.008**
	(0.53)	(0.53)	(3.41)	(3.41)
String of MBE	0.002	0.002	-0.001	-0.001
	(0.23)	(0.23)	(-1.72)	(-1.71)
Analyst following	0.017	0.017	0.002	0.002
	(1.81)	(1.81)	(1.65)	(1.65)
Industry and year FE	Included	Included	Included	Included
No. obs.	4,864	4,864	4,864	4,864
Adj. R^2	0.269	0.269	0.196	0.196

Notes: This table presents results from linear regressions estimated using OLS. Variables are defined as in Dhole et al. (2016) and Section OA-6.1.1. Standard errors are clustered by firm and year. p-values are obtained from `reghdfe` in Stata based on a t-distribution with four degrees of freedom (see Section OA-6.1.2). t-statistics [p-values] are presented in parentheses [brackets] below the coefficient estimates. The labels *, **, and *** refer to statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

Table OA-13

Replication of Table 4 in Dhole et al. (2016): transformed test variables

	(1) *CORRK*	(2) *CORRK*	(3) $\|DA\|$	(4) $\|DA\|$	(5) $\|DA\|$	(6) $\|DA\|$
CEO rel. leverage ratio > 1	0.036**		-0.004**		-0.002	
	(3.14)		(-2.80)		(-1.52)	
	[0.035]		[0.049]		[0.203]	
CEO rel. incentive ratio > 1		0.038**		-0.004**		-0.002
		(3.51)		(-2.96)		(-1.45)
		[0.025]		[0.042]		[0.222]
ln(1+Cash pay)	0.002	0.002	-0.004	-0.004	-0.004	-0.004
	(0.20)	(0.19)	(-1.37)	(-1.37)	(-1.45)	(-1.46)
ln(1+Delta)	0.016**	0.016**	0.000	0.000	0.001	0.001
	(3.57)	(3.52)	(0.44)	(0.53)	(1.24)	(1.34)
ln(1+Vega)	-0.006	-0.006	-0.000	-0.000	-0.000	-0.000
	(-1.83)	(-1.66)	(-0.57)	(-0.68)	(-0.60)	(-0.66)
$\|RMPROXY\|$	0.015***	0.015***	0.006***	0.006***	0.006***	0.006***
	(4.78)	(4.76)	(7.99)	(8.00)	(7.23)	(7.23)
Evol	-1.524***	-1.525***	0.105**	0.106**	0.060	0.059
	(-17.68)	(-17.60)	(3.27)	(3.26)	(1.92)	(1.93)
ln(Total assets)	-0.009	-0.008	-0.005*	-0.005*	-0.003	-0.003
	(-1.11)	(-1.05)	(-2.62)	(-2.62)	(-1.38)	(-1.38)
ln(Firm age)	-0.002	-0.001	-0.002**	-0.003**	-0.002	-0.002
	(-0.11)	(-0.08)	(-2.85)	(-2.92)	(-1.87)	(-1.89)
Market-to-book ratio	0.002	0.002	0.001	0.001	0.001*	0.001*
	(1.10)	(1.08)	(2.10)	(2.09)	(2.41)	(2.40)
Leverage	-0.014*	-0.014*	0.000	0.000	-0.001	-0.001
	(-2.63)	(-2.75)	(0.38)	(0.45)	(-1.21)	(-1.19)
Net operating assets	-0.075*	-0.075*	-0.009	-0.009	0.002	0.002
	(-2.42)	(-2.41)	(-0.97)	(-0.97)	(0.18)	(0.19)
Implicit claims	-0.014	-0.014	0.017***	0.017***	0.018***	0.018***
	(-0.75)	(-0.73)	(6.61)	(6.63)	(6.11)	(6.12)
Litigation dummy	-0.049	-0.050	0.004	0.004	0.003	0.003
	(-1.77)	(-1.80)	(1.08)	(1.12)	(0.83)	(0.85)
ln(Shares)	-0.027**	-0.027**	0.005*	0.005*	0.003	0.003
	(-3.08)	(-3.09)	(2.22)	(2.23)	(1.27)	(1.27)
Big4	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002
	(-0.09)	(-0.12)	(-1.12)	(-1.11)	(-0.92)	(-0.92)
Issue	0.008	0.008	0.008**	0.008**	0.007*	0.007*
	(0.68)	(0.70)	(3.44)	(3.45)	(2.36)	(2.36)
String of MBE	0.001	0.001	-0.001	-0.001	0.000	0.000
	(0.16)	(0.19)	(-1.53)	(-1.60)	(0.21)	(0.20)
Analyst following	0.016	0.016	0.003	0.003	0.003*	0.003*
	(1.77)	(1.75)	(1.72)	(1.70)	(2.26)	(2.25)
OE factor					0.009**	0.009**
					(4.00)	(4.04)
Industry and year FE	Included	Included	Included	Included	Included	Included
No. obs.	4,864	4,864	4,864	4,864	4,781	4,781
Adj. R^2	0.272	0.272	0.197	0.196	0.206	0.206

Notes: These tests are similar to Table OA-12, except that the test variables are transformed to indicator variables based on whether the values of the underlying variables exceed one. t-statistics [p-values] are presented in parentheses [brackets] below the coefficient estimates. The labels *, **, and *** refer to statistical significance at the 0.10, 0.05, and 0.01 levels, respectively. *OE factor* is defined in Table OA-7.

Table OA-14

Replication of Table 5 in Dhole et al. (2016)

	(1) $\lvert RCFO\rvert$	(2) $\lvert RCFO\rvert$	(3) $\lvert RPROD\rvert$	(4) $\lvert RPROD\rvert$	(5) $\lvert RDISX\rvert$	(6) $\lvert RDISX\rvert$	(7) $\lvert RMPR.\rvert$	(8) $\lvert RMPR.\rvert$
CEO RL ratio >1	0.002 (0.84) [0.446]		0.011 (1.94) [0.124]		-0.002 (-0.28) [0.793]		0.130* (2.21) [0.092]	
CEO RI ratio >1		0.004 (1.59) [0.187]		0.013 (1.94) [0.125]		0.002 (0.28) [0.793]		0.164* (2.56) [0.063]
ln(1+Cash pay)	0.002 (0.73)	0.002 (0.71)	0.007 (1.18)	0.007 (1.17)	0.002 (0.34)	0.002 (0.31)	0.032 (0.52)	0.030 (0.49)
ln(1+Delta)	0.003** (4.22)	0.003** (4.25)	0.007** (2.98)	0.007** (2.97)	0.006* (2.30)	0.006* (2.39)	0.066* (2.16)	0.067* (2.15)
ln(1+Vega)	-0.001 (-1.37)	-0.001 (-1.34)	-0.002 (-0.92)	-0.002 (-0.83)	0.000 (0.17)	0.000 (0.17)	-0.044 (-1.80)	-0.042 (-1.72)
$\lvert DA\rvert$	0.468*** (9.64)	0.468*** (9.62)	0.296*** (4.91)	0.296*** (4.90)	0.221*** (5.00)	0.222*** (5.01)	6.436*** (6.84)	6.437*** (6.83)
Evol	-0.003 (-0.13)	-0.003 (-0.12)	0.016 (0.33)	0.016 (0.32)	0.165** (3.53)	0.166** (3.53)	-0.117 (-0.18)	-0.115 (-0.17)
ln(Total assets)	-0.013*** (-5.51)	-0.013*** (-5.49)	-0.023** (-4.58)	-0.023** (-4.53)	-0.030*** (-5.73)	-0.030*** (-5.74)	-0.099 (-1.63)	-0.097 (-1.59)
ln(Firm age)	-0.003* (-2.30)	-0.004* (-2.36)	-0.001 (-0.26)	-0.001 (-0.25)	-0.004 (-0.82)	-0.005 (-0.91)	-0.048 (-0.87)	-0.049 (-0.90)
Market-to-book ratio	0.000 (0.80)	0.000 (0.78)	0.000 (0.56)	0.000 (0.55)	0.002** (3.37)	0.002** (3.35)	0.009 (1.03)	0.009 (1.02)
Leverage	-0.001 (-0.44)	-0.001 (-0.36)	0.002 (0.69)	0.002 (0.73)	0.000 (0.16)	0.001 (0.27)	0.052 (1.48)	0.054 (1.57)
Net operating assets	-0.045*** (-5.02)	-0.045*** (-5.02)	-0.111*** (-5.99)	-0.111*** (-6.01)	-0.152*** (-8.84)	-0.152*** (-8.81)	-1.499*** (-8.39)	-1.495*** (-8.39)
Implicit claims	-0.020** (-3.49)	-0.020** (-3.48)	-0.019 (-1.56)	-0.018 (-1.54)	-0.011 (-0.89)	-0.010 (-0.88)	-0.280* (-2.36)	-0.277* (-2.34)
Litigation dummy	-0.002 (-0.26)	-0.002 (-0.23)	-0.012 (-0.99)	-0.012 (-0.99)	0.010 (0.79)	0.011 (0.83)	-0.286 (-1.93)	-0.284 (-1.92)
ln(Shares)	0.011*** (4.71)	0.011*** (4.70)	0.010 (1.92)	0.010 (1.90)	0.011* (2.21)	0.011* (2.19)	0.011 (0.18)	0.009 (0.15)
Big4	-0.001 (-0.24)	-0.001 (-0.27)	0.013 (1.48)	0.013 (1.44)	0.012 (1.18)	0.012 (1.17)	0.154 (1.53)	0.151 (1.50)
Issue	0.003 (0.96)	0.003 (0.98)	0.009 (1.71)	0.009 (1.70)	0.036*** (4.92)	0.037*** (4.93)	0.151* (2.34)	0.153* (2.34)
String of MBE	-0.001 (-0.69)	-0.001 (-0.69)	0.005 (1.63)	0.005 (1.65)	0.001 (0.27)	0.001 (0.25)	0.041 (1.06)	0.041 (1.07)
Analyst following	0.007* (2.62)	0.007* (2.60)	0.011 (1.74)	0.011 (1.73)	0.017** (2.83)	0.017** (2.81)	0.035 (0.45)	0.034 (0.43)
Industry and year FE	Included	Included	Included	Included	Included	Included	Included	Included
No. obs.	4,864	4,864	4,864	4,864	4,864	4,864	4,864	4,864
Adj. R^2	0.312	0.312	0.208	0.208	0.333	0.333	0.223	0.223

Notes: This table presents results from linear regressions estimated using OLS. Variables are defined as in Dhole et al. (2016) and Section OA-6.1.1. Standard errors are clustered by firm and year. p-values are obtained from `reghdfe` in Stata based on a t-distribution with four degrees of freedom (see Section OA-6.1.2). t-statistics [p-values] are presented in parentheses [brackets] below the coefficient estimates. The labels *, **, and *** refer to statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

Table OA-15

Replication of Table 7 in Dhole et al. (2016)

	(1) *JustMBE*	(2) *JustMBE*	(3) *LargeMBE*	(4) *LargeMBE*	(5) $JustMBE_{t+1}$	(6) $JustMBE_{t+1}$
CEO RL ratio > 1	0.011		-0.002		0.025	
	(1.10)		(-0.16)		(1.79)	
	[0.334]		[0.881]		[0.148]	
CEO RI ratio > 1		-0.001		0.016		0.004
		(-0.11)		(1.04)		(0.21)
		[0.919]		[0.356]		[0.845]
ln(1+Cash pay)	-0.007	-0.006	0.025	0.024	-0.011	-0.010
	(-0.48)	(-0.45)	(1.10)	(1.06)	(-0.43)	(-0.38)
ln(1+Delta)	0.009	0.008	0.009	0.009	0.009	0.008
	(1.49)	(1.35)	(1.70)	(1.94)	(1.51)	(1.29)
ln(1+Vega)	0.006	0.006	-0.010	-0.010	0.002	0.002
	(0.88)	(0.90)	(-1.25)	(-1.24)	(0.59)	(0.73)
Evol	-0.558***	-0.562***	0.379*	0.384*	-0.279	-0.287
	(-5.04)	(-5.10)	(2.56)	(2.60)	(-1.91)	(-1.96)
ln(Total assets)	-0.097***	-0.097***	0.134***	0.134***	-0.082***	-0.081***
	(-8.21)	(-8.18)	(14.18)	(14.12)	(-5.78)	(-5.64)
ln(Firm age)	0.008	0.009	-0.019	-0.021	0.013	0.016*
	(0.84)	(1.06)	(-1.48)	(-1.73)	(2.11)	(2.55)
Market-to-book ratio	-0.001	-0.001	0.001	0.001	0.001	0.001
	(-0.65)	(-0.63)	(0.43)	(0.40)	(1.08)	(1.13)
Leverage	0.005	0.004	-0.019**	-0.017*	-0.001	-0.003
	(0.45)	(0.34)	(-2.83)	(-2.69)	(-0.12)	(-0.49)
Net operating assets	0.074*	0.073*	-0.222***	-0.221***	0.088*	0.086*
	(2.15)	(2.13)	(-7.44)	(-7.39)	(2.41)	(2.36)
Implicit claims	0.044	0.044	0.011	0.012	0.053*	0.052*
	(1.86)	(1.81)	(0.36)	(0.39)	(2.25)	(2.22)
Litigation dummy	-0.024	-0.026	0.024	0.026	-0.043	-0.046
	(-1.00)	(-1.10)	(1.15)	(1.26)	(-1.69)	(-1.85)
ln(Shares)	0.102***	0.102***	-0.153***	-0.153***	0.096***	0.096***
	(7.31)	(7.25)	(-16.60)	(-16.38)	(7.46)	(7.50)
Big4	-0.008	-0.008	0.024	0.023	-0.002	-0.002
	(-0.86)	(-0.84)	(1.35)	(1.32)	(-0.15)	(-0.11)
Issue	-0.001	-0.002	0.001	0.002	-0.022	-0.023
	(-0.09)	(-0.12)	(0.05)	(0.08)	(-1.09)	(-1.14)
String of MBE	0.017*	0.017*	0.020**	0.019**	0.022*	0.022*
	(2.24)	(2.26)	(2.80)	(2.83)	(2.42)	(2.48)
Analyst following	0.054**	0.054**	-0.043*	-0.043*	0.032	0.032
	(4.21)	(4.25)	(-2.34)	(-2.35)	(2.05)	(2.07)
Industry and year FE	Included	Included	Included	Included	Included	Included
No. obs.	4,864	4,864	4,864	4,864	4,864	4,864
Adj. R^2	0.063	0.063	0.067	0.067	0.056	0.055

Notes: This table presents results from linear regressions estimated using OLS. Variables are defined as in Dhole et al. (2016) and Section OA-6.1.1. Standard errors are clustered by firm and year. p-values are obtained from `reghdfe` in Stata based on a t-distribution with four degrees of freedom (see Section OA-6.1.2). t-statistics [p-values] are presented in parentheses [brackets] below the coefficient estimates. The labels *, **, and *** refer to statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

Table OA-16

Instrumental variables estimation for the Dhole et al. (2016) sample

Panel A: Tests using two-way clustered standard errors

	(1) 1st stage $ln(RL)$	(2) 2nd stage $CORRK$	(3) 2nd stage $\|DA\|$	(4) 1st stage $ln(RI)$	(5) 2nd stage $CORRK$	(6) 2nd stage $\|DA\|$
$StateWage$ (IV)	0.103			0.101		
	(1.98)			(1.97)		
	[0.119]			[0.121]		
$StateMortgage$ (IV)	-0.114*			-0.113*		
	(-2.46)			(-2.50)		
	[0.070]			[0.067]		
$ln(RL)$ (predicted)		-0.026	-0.006			
		(-0.15)	(-1.07)			
$ln(RI)$ (predicted)					-0.026	-0.006
					(-0.14)	(-1.10)
Control variables	Included	Included	Included	Included	Included	Included
Industry and year FE	Included	Included	Included	Included	Included	Included
No. obs.	4,766	4,766	4,766	4,766	4,766	4,766
Adj. R^2	0.369	0.271	0.195	0.360	0.271	0.195
Partial F	3.034			3.116		
Partial F (p-value)	0.158			0.153		
K-P LM stat	2.921			2.957		
K-P LM stat (p-value)	0.232			0.228		

Panel B: Tests using non-clustered robust standard errors

	(1) 1st stage $ln(RL)$	(2) 2nd stage $CORRK$	(3) 2nd stage $\|DA\|$	(4) 1st stage $ln(RI)$	(5) 2nd stage $CORRK$	(6) 2nd stage $\|DA\|$
$StateWage$ (IV)	0.103***			0.101***		
	(3.55)			(3.52)		
	[0.000]			[0.000]		
$StateMortgage$ (IV)	-0.114***			-0.113***		
	(-4.39)			(-4.45)		
	[0.000]			[0.000]		
$ln(RL)$ (predicted)		-0.026*	-0.006**			
		(-1.74)	(-1.96)			
$ln(RI)$ (predicted)					-0.026*	-0.006*
					(-1.76)	(-1.96)
Control variables	Included	Included	Included	Included	Included	Included
Industry and year FE	Included	Included	Included	Included	Included	Included
No. obs.	4,766	4,766	4,766	4,766	4,766	4,766
Adj. R^2	0.369	0.271	0.195	0.360	0.271	0.195
Partial F	9.66			9.92		
Partial F (p-value)	0.000			0.000		
K-P LM stat	19.31			19.83		
K-P LM stat (p-value)	0.000			0.000		

Notes: This table presents results from two-stage instrumental variables estimations using state-level tax rates as instruments. Columns (1) and (4) present results from first-stage regression estimations with the natural logarithm of *CEO relative leverage ratio* and *CEO relative incentive ratio*, respectively: $ln(RL)$ and $ln(RI)$. Before taking the natural logarithm, we replace zeroes with the minimum positive values of *CEO relative leverage ratio* and *CEO relative incentive ratio* following Campbell et al. (2016). The remaining columns present the second-stage results from 2SLS estimations where the first-stage estimation follows that in column (1) or (4). Variables are defined as in Dhole et al. (2016) and Section OA-6.1.1. The instrumental variables are defined in Section OA-5.5. The partial F-statistic is obtained by testing whether the coefficients on the instrumental variables $StateWage$ and $StateMortgage$ are jointly equal to zero. "K-P LM stat" refers to the Kleibergen and Paap (2006) rank LM test obtained with program `ivreg2` in Stata. Standard errors are clustered by firm and year. p-values are obtained from `reghdfe` in Stata based on a t-distribution with four degrees of freedom (see Section OA-6.1.2). For the second-stage estimations, standard errors are obtained from a cluster-robust bootstrap procedure that estimates both the first- and second-stage regressions on each bootstrap sample. t-statistics [p-values] are presented in parentheses [brackets] below the coefficient estimates. The labels *, **, and *** refer to statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.