Online Supplement A. Derivation of the distribution of observed earnings in model (1)–(3)

Let $m \in \{0,1\}$ be the earnings management decision ($m = 0$ when earnings are reported "as is" and $m = 1$ when earnings are managed to report a small profit). After integrating out unobservable $EARN^*$ and m, the probability density function of reported earnings is

$$f(EARN|X) = \int_{EARN^*} \sum_{m=0,1} f(EARN|EARN^*, m, X) \Pr(m|EARN^*, X) f^*(EARN^*|X) dEARN^* \quad (A1)$$

where $f(EARN|EARN^*, m, X)$ is the density of reported earnings conditional on pre-managed earnings $EARN^*$, the earnings management decision m, and explanatory variables X.

We simplify (A1) for three scenarios on $EARN$.

Case 1: $EARN < -K^-$ or $EARN \geq K^+$, i.e., a large loss or a large profit. The conditional density $f(EARN|EARN^*, m, X)$ is non-zero only for $EARN^* = EARN$ and $m = 0$.[1] Equation (A1) simplifies to

$$f(EARN|X) = \Pr(m = 0|EARN^* = EARN, X) f^*(EARN^* = EARN|X) \quad (A2)$$

where $\Pr(m = 0|EARN^* = EARN, X) = 1 - P(EARN^* = EARN, X) = 1$ because $EARN \notin [-K^-, 0)$. Therefore, (A2) simplifies to equation (4a) in Section 2

$$f(EARN|X) = f^*(EARN^* = EARN|X) \quad (4a)$$

Case 2: $EARN \in [-K^-, 0)$, i.e., a small loss. The conditional density $f(EARN|EARN^*, m, X)$ is non-zero only for $EARN^* = EARN$ and $m = 0$.[2] Therefore, (A1) simplifies to (A2), similar to Case 1. For $EARN \in [-K^-, 0)$, $\Pr(m = 0|EARN^* = EARN, X) = 1 - P(EARN^* = EARN, X)$. Therefore, (A2) simplifies to equation (4b) in Section 2

$$f(EARN|X) = [1 - P(EARN^* = EARN, X)] f^*(EARN^* = EARN|X) \quad (4b)$$

Case 3: $EARN \in [0, K^+)$, i.e., a small profit. The conditional density $f(EARN|EARN^*, m, X)$ is non-zero in two situations: (a) $m = 0$ and $EARN^* = EARN$, and (b) $m = 1$ and $EARN^* \in [-K^-, 0)$. Therefore, (A.1) becomes

$$f(EARN|X) = \Pr(m = 0|EARN^* = EARN, X) f^*(EARN^* = EARN|X) + \int_{-K^-}^{0} f(EARN|EARN^*, m = 1, X) \Pr(m = 1|EARN^*, X) f^*(EARN^*|X) dEARN^* \quad (A3)$$

In the first line of (A3), $\Pr(m = 0|EARN^* = EARN, X) = 1$ because $EARN \notin [-K^-, 0)$. In the second line of (A3), $f(EARN|EARN^*, m = 1, X) = g(EARN|EARN^*, X)$ per (2b) and $\Pr(m = 1|EARN^*, X) = P(EARN^*, X)$. Therefore, (A3) simplifies to equation (4c) in Section 2

$$f(EARN|X) = f^*(EARN^* = EARN|X) + G(EARN, X) \quad (4c)$$

where

$$G(EARN, X) = \int_{-K^-}^{0} g(EARN|EARN^*, X) P(EARN^*, X) f^*(EARN^*|X) dEARN^* \quad (A4)$$

[1] When $m = 1$ (i.e., earnings are managed to report a small profit), the probability of reporting a large loss or a large profit is zero. When $m = 0$ (i.e., no earnings management), reported $EARN$ must equal the pre-managed $EARN^*$. Because $f(EARN|EARN^*, m, X)$ collapses to a mass point with weight 1, it drops out of the computation.

[2] When $m = 1$, the probability of reporting a small loss is zero. When $m = 0$, reported $EARN$ must equal the pre-managed $EARN^*$.

Online Supplement B. Implementation details of maximum likelihood (ML) estimation

For each firm-year observation i, t, the log-likelihood is $\ln f(EARN_{i,t}|X_{i,t})$ as defined in (4a)–(6b), where $EARN_{i,t}$ and $X_{i,t}$ are from the data, and the coefficient vector comprises $\alpha_{0,0} \dots \alpha_{P,J}$ and $\pi_0 \dots \pi_J$. Maximum likelihood estimation finds the coefficients $\alpha_{0,0} \dots \alpha_{P,J}$ and $\pi_0 \dots \pi_J$ that maximize the total log-likelihood for the sample.[3] We code the log-likelihood as a user-defined Stata function. The *ml* command in Stata takes this function as an input and handles the numerical optimization and the computation of the standard errors.

We estimate the model for the subsample with earnings in a relatively narrow interval $[-R, R)$ around zero (e.g., $EARN \in [-0.04, 0.04)$ for our main definitions), as illustrated in Panel C of Figure 3. Although this approach involves selection on the dependent variable, we show that it yields the same parameter estimates as conventional maximum likelihood estimation on the full sample. The conventional ML approach solves

$$\max_{\theta} \sum_{i=1\dots N, t=1\dots T} \ln f(EARN_{i,t}|X_{i,t}) =$$

$$= \max_{\theta} \left(\sum_{i,t:\, EARN_{i,t} \in [-R,R)} \ln f(EARN_{i,t}|X_{i,t}) + \sum_{i,t:\, EARN_{i,t} \notin [-R,R)} \ln f(EARN_{i,t}|X_{i,t}) \right) \quad (B1)$$

where θ is the full parameter vector, which comprises $\alpha_{0,0} \dots \alpha_{P,J}$, $\pi_0 \dots \pi_J$, and additional parameters that determine the earnings distribution outside $[-R, R)$; $f(EARN_{i,t}|X_{i,t})$ is the probability density function of reported earnings from equations (4a)–(4c), evaluated at the parameter values θ; and $[-R, R)$ is the earnings interval used in our subsample-specific estimation.

Suppose that a researcher uses a separate subset of parameters $\theta_{outside}$ for the pre-managed earnings distribution outside the estimation interval $[-R, R)$. From equations (4a)–(4c) in Section 2, all other components of θ (i.e., the pre-managed distribution parameters $\alpha_{0,0} \dots \alpha_{P,J}$ for the interval $[-R, R)$ and the earnings management parameters $\pi_0 \dots \pi_J$) affect the likelihood only for observations with reported earnings inside $[-R, R)$. Therefore, (B1) can be rewritten as

$$\max_{\pi_0 \dots \pi_J,\, \alpha_{0,0} \dots \alpha_{P,J}} \sum_{i,t:\, EARN_{i,t} \in [-R,R)} \ln f(EARN_{i,t}|X_{i,t}) + \max_{\theta_{outside}} \sum_{i,t:\, EARN_{i,t} \notin [-R,R)} \ln f(EARN_{i,t}|X_{i,t}) \quad (B2)$$

The first maximization in this expression is equivalent to our maximum likelihood estimation procedure for the subsample with earnings in the interval $[-R, R)$. It fully determines all of the parameters that we are interested in, i.e., $\alpha_{0,0} \dots \alpha_{P,J}$ and $\pi_0 \dots \pi_J$. Therefore, our estimation of $\alpha_{0,0} \dots \alpha_{P,J}$ and $\pi_0 \dots \pi_J$ on the restricted sample is equivalent to conventional maximum likelihood estimation on the full sample with an additional parameter vector $\theta_{outside}$.

[3] For each observation i, t, we normalize the probability density function (6a) by imposing the standard restriction that the total mass $\int f(EARN|X_{i,t})dEARN$ of the conditional earnings distribution must equal 1. Without this normalization, the ML estimation procedure would artificially drive the log-likelihood to infinity by increasing the coefficients α to infinity. This issue is specific to ML estimation. In our two-stage method, the estimates minimize the distance between predicted and actual bin frequencies conditional on *X*, and thus the actual bin frequencies directly determine the scale of the density parameters, removing the need for an explicit normalization.

Online Supplement C. Implementation details of the two-stage estimation method

We omit the firm and year indexes for brevity. To prepare the data, we restrict the sample to observations with scaled earnings $EARN$ in the estimation interval (Panel C of Figure 3) and convert each of these firm-year observations into B firm-year-bin observations with a dummy dependent variable Y_b ($b = 1 \dots B$) that equals 1 if $EARN$ is in bin b and 0 otherwise. Because our estimation equations pool all relevant bins, the bin grid can have many bins in the small-loss and small-profit intervals (versus just one bin for each interval in the standard tests). The maximum number of bins in practice is restricted only by computation time and available memory.

Stage 1 uses all firm-years in the estimation sample; for each firm-year, we only include firm-year-bins outside the small-loss and small-profit intervals, as illustrated in Panel A of Figure 4.[4] We estimate a discrete version of equation (6a) for a pooled sample of these bins

$$Y_b = \alpha_0(X) + \alpha_1(X) \times z_b + \alpha_2(X) \times z_b^2 + \cdots + \alpha_P(X) \times z_b^P + \varepsilon_b \qquad (C1)$$

where the polynomial coefficients $\alpha_p(X) = \alpha_{p,0} + \alpha_{p,1}X_1 + \cdots + \alpha_{p,J}X_J$ follow (6b), $X = X_1 \dots X_J$ are the explanatory variables for the firm-year, and z_b is the midpoint of earnings bin b. Because the dependent variable Y_b is binary, the predicted value in regression (C1) captures the bin probability $Pr(Y_b = 1|X, z_b)$ conditional on explanatory variables X.[5] This probability interpretation follows the linear probability model (e.g., Wooldridge, 2002, Ch. 15), and the pooling of data for all relevant bins imposes the smooth polynomial structure (C1) across the bins.

Stage 2 uses all firm-years in the estimation sample; for each firm-year, we only include firm-year-bins inside the small-loss and small-profit intervals (Panel B of Figure 4). The regression model is

$$Y_b - \hat{Y}_b = \left(\pi_0 + \pi_1 X_1 + \cdots + \pi_J X_J\right) \times W_b + u_b \qquad (C2)$$

where $\hat{Y}_b$ is the predicted probability of bin b for the firm-year based on the pre-managed earnings distribution (C1) from stage 1, $Y_b - \hat{Y}_b$ is the deviation from the pre-managed distribution, $\pi_0 \dots \pi_J$ are the earnings management parameters, and W_b is a synthetic explanatory variable shown in Panel C of Figure 4 that embeds the earnings management definitions of Models I and II.

For the small-loss bins, W_b is defined as $-\hat{Y}_b$ in Model I and $-q(z_b) \times \hat{Y}_b$ in Model II, where $q(z_b)$ is the triangular interaction term from (5b). These definitions implement the dip in the density of small losses due to earnings management, represented by $-f^*(EARN|X)P(EARN, X)$ in equation (4b). In Model I, $P(EARN, X)$ is flat with respect to the size of the small loss $EARN$ (Panel A of Figure 3), and thus the dip is proportional to the pre-managed earnings density $f^*(EARN|X)$, approximated by $\hat{Y}_b$. In Model II, the earnings management probability is triangular with respect to $EARN$ (Panel B of Figure 3), and thus the dip is proportional to the triangular interaction term $q(z_b)$ times the pre-managed density approximated by $\hat{Y}_b$.

For the small-profit bins, W_b is defined as $-\frac{K^-}{K^+} W_{mean}$ in Model I and $-q(z_b)\frac{K^-}{K^+} W_{mean}$ in Model II, where the ratio K^-/K^+ adjusts for the relative widths of the small-loss and small-profit intervals, W_{mean} is the mean of W_b across the small-loss bins for the firm-year, and $q(z_b)$ is the triangular interaction term for small profits. These definitions implement the bump in the density

---

[4] In both stages, this bin selection is unrelated to reported earnings for the year. For example, firm-years with small (moderately large) profits are included in stage 1 (stage 2), but all of the included bin dummies equal zero.

[5] The discrete bin probability (C1) approximates the integral of the continuous density function (6a) for the bin using the function value at the bin midpoint. Therefore, the empirical coefficients α in (C1) are not entirely equivalent to the theoretical α in (6a). Because the α-s are not individually interpretable (as parts of a polynomial), and their only job is to approximate the underlying smooth distribution, we slightly abuse the notation and reuse α in (C1) for brevity.

of small profits due to earnings management, represented by $G(EARN, X)$ in equation (4c). For each firm-year, the total probability of the bump for small profits must equal the total probability of the dip for small losses (Panel C of Figure 1). In both models, this restriction is implemented through $\frac{K^-}{K^+}W_{mean}$. In Model I, $G(EARN, X)$ is flat with respect to the size of the small profit $EARN$ (Panel A of Figure 3), and thus W_b does not require any further adjustments. In Model II, $G(EARN, X)$ is triangular (Panel B of Figure 3), and thus W_b incorporates the triangular interaction term $q(z_b)$.

Standard errors of the two-stage estimates

Because the explanatory variables in the second-stage regression (C2) are constructed based on the first-stage estimates $\hat{\alpha} = (\hat{\alpha}_{0,0} \dots \hat{\alpha}_{P,J})'$ from (C1), the standard errors of $\hat{\pi} = (\hat{\pi}_0 \dots \hat{\pi}_J)'$ in the second stage should be adjusted for the first-stage estimation noise. The usual OLS standard errors (with appropriate clustering) do not incorporate this adjustment and should not be used.[6]

Using the method of moments representation of OLS (e.g., Wooldridge, 2002, Ch. 14), the regression estimates $\hat{\alpha}$ and $\hat{\pi}$ in the two stages (C1) and (C2) are defined by the moment conditions

$$\bar{h}(\hat{\alpha}) = \frac{1}{N} \sum_{\substack{i,t,b: \\ b \notin small\ loss/profit\ bins}} \left(Y_{i,t,b} - \hat{\alpha}' P_{i,t,b}\right) P_{i,t,b} = 0 \qquad (C3)$$

$$\bar{g}(\hat{\alpha}, \hat{\pi}) = \frac{1}{N} \sum_{\substack{i,t,b: \\ b \in small\ loss/profit\ bins}} \left(Y_{i,t,b} - \hat{Y}_{i,t,b}(\hat{\alpha}) - \hat{\pi}' Q_{i,t,b}(\hat{\alpha})\right) Q_{i,t,b}(\hat{\alpha}) = 0 \qquad (C4)$$

where $P_{i,t,b}$ is the full vector of explanatory variables in stage 1 (i.e., $1, z_b, z_b^2 \dots z_b^P$ and its interactions with $X_{i,t,1} \dots X_{i,t,J}$), $\hat{Y}_{i,t,b}(\hat{\alpha})$ is the predicted value from stage 1, and $Q_{i,t,b}(\hat{\alpha})$ is the full vector of explanatory variables in stage 2 (i.e., $W_{i,t,b}$ and its interactions with $X_{i,t,1} \dots X_{i,t,J}$). The estimation noise in $\hat{\alpha}$ affects the second-stage standard errors through both $\hat{Y}_{i,t,b}(\hat{\alpha})$ and $Q_{i,t,b}(\hat{\alpha})$ in (C4).

The Taylor expansion of (C3) and (C4) around the true values α^* and π^* is

$$\begin{bmatrix} \bar{h}(\hat{\alpha}) \\ \bar{g}(\hat{\alpha}, \hat{\pi}) \end{bmatrix} \approx \begin{bmatrix} \bar{h}(\alpha^*) \\ \bar{g}(\alpha^*, \pi^*) \end{bmatrix} + \begin{bmatrix} \nabla_\alpha \bar{h} & 0 \\ \nabla_\alpha \bar{g} & \nabla_\pi \bar{g} \end{bmatrix} \begin{bmatrix} \hat{\alpha} - \alpha^* \\ \hat{\pi} - \pi^* \end{bmatrix} \qquad (C5)$$

After combining (C5) with (C3) and (C4), we have

$$\begin{bmatrix} \hat{\alpha} - \alpha^* \\ \hat{\pi} - \pi^* \end{bmatrix} \approx -\begin{bmatrix} \nabla_\alpha \bar{h} & 0 \\ \nabla_\alpha \bar{g} & \nabla_\pi \bar{g} \end{bmatrix}^{-1} \begin{bmatrix} \bar{h}(\alpha^*) \\ \bar{g}(\alpha^*, \pi^*) \end{bmatrix} \qquad (C6)$$

From (C6), the asymptotic covariance matrix of the estimates $\hat{\alpha}$ and $\hat{\pi}$ is

$$Cov\left(\sqrt{N} \begin{bmatrix} \hat{\alpha} \\ \hat{\pi} \end{bmatrix}\right) = \Gamma' \Omega \Gamma \qquad (C7a)$$

where

$$\Gamma = \left(-plim \begin{bmatrix} \nabla_\alpha \bar{h} & 0 \\ \nabla_\alpha \bar{g} & \nabla_\pi \bar{g} \end{bmatrix}\right)^{-1} \qquad (C7b)$$

$$\Omega = Cov\left(\sqrt{N} \begin{bmatrix} \bar{h}(\alpha^*) \\ \bar{g}(\alpha^*, \pi^*) \end{bmatrix}\right) \qquad (C7c)$$

[6] In untabulated simulations, conventional single-stage clustered standard errors are biased downward slightly, as expected, which leads to moderate over-rejection in hypothesis tests. Our adjustment resolves this bias. Because the bin dummies are mutually exclusive, the bin-level observations are correlated within each firm-year. Therefore, our adjustment must be combined with clustering to address the within-firm-year correlation.

Because $\hat{\alpha}$ and $\hat{\pi}$ converge in probability to the (unknown) true values α^* and π^*, the matrices Γ and Ω can be evaluated at $\hat{\alpha}$ and $\hat{\pi}$ instead of α^* and π^*. The covariance matrix Ω of the moment conditions is clustered as needed. Our Stata command incorporates these computations.

Online Supplement D. Type-I error and power simulations for basic distribution discontinuity tests without explanatory variables

We generate 1,000 artificial samples of pre-managed earnings $EARN^*$ on the estimation interval $[-0.04, 0.04)$, using the earnings distribution parameters from the main empirical specification (column 4 in Panel A of Table 2), for sample size $N = 5{,}000$ (less than 1/6 of our main sample) and $N = 30{,}000$ (slightly less than the main sample). In Type-I error simulations, the null hypothesis of no earnings management is true, and therefore we do not manage simulated earnings. In test power simulations, we convert some of the small losses into small profits per Model II with the true earnings management probability $\pi_0^{true} = 0.025$ and 0.05 (i.e., 2.5% or 5% of small losses are managed on average). For each simulated sample, we test for earnings discontinuity using Burgstahler and Dichev (1997) standardized difference tests, the main two-stage version of our method, and the ML version of our method as an asymptotically efficient benchmark (Wooldridge, 2002).

Table D1 presents the simulated rejection rates. Columns 1 and 4 reflect Type-I error. The left standardized difference test in column 4 has a slightly elevated Type-I error of 6.6%, while all other Type-I errors for all tests are in line with the nominal level.[7]

Consistent with Burgstahler and Chuk's (2014) simulations, the standardized difference test successfully detects earnings management. For example, when $\pi_0^{true} = 0.025$, earnings management affects just 13 observations on average out of $N = 5{,}000$ and 80 out of $N = 30{,}000$.[8] The rejection rates based on the left (right) standardized difference are 22.0% and 59.3% (15.9% and 42.4%) for $N = 5{,}000$ and 30,000, respectively. When $\pi_0^{true} = 0.05$, the rejection rates are 44.3% and 97.7% (36.5% and 92.3%), respectively.

Across all test power scenarios in Table D1, the rejection rate in the left standardized difference test is considerably higher than that in the right standardized difference test. This asymmetry is an artifact of the linear interpolation used in these tests. Because the simulated distribution of pre-managed earnings is convex, following the empirical estimates shown in Figure 7, the linear interpolation overstates the missing earnings density below zero and understates the excess earnings density above zero. These interpolation biases cause systematic over-rejection in the left standardized difference test and systematic under-rejection in the right standardized difference test in this simulation. Therefore, the reported rejection rates in the left (right) standardized difference test should be interpreted as an upper (lower) bound on the actual statistical power of these tests.[9]

Our main two-stage tests improve these rejection rates by 1.3–29.3 percentage points, as represented by the grey bars in Figure D1.[10] For example, for $\pi_0^{true} = 0.05$ and $N = 5{,}000$ in

[7] When the rejection rate is 5%, the total number of rejections in 1,000 simulations is a binomial random variable with $n = 1{,}000$ and $p = 0.05$. The corresponding 95% confidence interval for the rejection rate is $[3.6\%, 6.4\%]$.

[8] Small pre-managed losses are approximately 10.7% of the N earnings observations in the interval $[-0.04, 0.04)$. Therefore, when $\pi_0^{true} = 0.025$, the expected number of managed small losses per sample is approximately $0.107 \times 0.025 \times 5{,}000 = 13.38$ for $N = 5{,}000$ and $0.107 \times 0.025 \times 30{,}000 = 80.25$ for $N = 30{,}000$.

[9] Our method addresses this source of bias by directly modeling the curvature of the distribution through the polynomial approximation. Further, even if a researcher negligently uses a bad approximation (including the unreasonable examples shown in Figure 5), our method is less sensitive to bad approximation quality than the standard tests because the approximation biases below and above zero partly offset each other in our combined test statistic.

[10] Because the rejection rates in the left standardized difference test in these simulations are biased upward due to interpolation bias, the reported improvement over this test in Panel A of Figure D1 should be viewed as a lower bound on the actual improvement in test power. Similarly, because the rejection rates in the right standardized difference test are biased downward, the reported improvement in Panel B should be viewed as an upper bound on the actual improvement.

column 3, the rejection rates improve from 36.5–44.3% in the standardized difference tests to 52.2–57.5% in our method. Notably, the improvement is larger when we use a finer bin grid (the darker grey bars in Figure D1), and it gradually approaches the upper bound of potential test performance, approximated by the ML results (the black bars). For example, when the bin width is 0.005, the power improvement for our two-stage tests is on average 78% as large as that for ML; when the bin width is reduced to 0.0025 and 0.001, the ratio increases to 92% and 98%, respectively.[11] Thus, our main method performs almost as well as ML without any of ML's numerical complications.

In untabulated simulations, we examine earnings management scenarios with asymmetric small-loss and small-profit intervals ($K^{-} = 0.01$, $K^{+} = 0.005$; and $K^{-} = 0.005$, $K^{+} = 0.01$). Even when our main method uses incorrect interval definitions $K^{-} = K^{+} = 0.01$ in estimation, it outperforms the standardized difference tests in most cases. With correct asymmetric interval definitions, it dominates in all cases.

Next, we assess to what extent the test power improvement in our method can be attributed to (1) our polynomial approximation, which is more accurate than the standard linear interpolation, and (2) our overidentifying restriction from Panel C of Figure 1, which combines information from excess small profits and missing small losses. To isolate the first channel, we replace the polynomial approximation with a linear approximation (option `degree(1)` in our Stata command), and we keep the earnings management specification of Model II. We use narrower estimation intervals $[-0.015, 0.015)$ and $[-0.02, 0.02)$ for these tests because linear interpolation over the main estimation interval $[-0.04, 0.04)$ in untabulated exploratory simulations is highly vulnerable to distribution non-linearities and often has excessive Type-I errors. To isolate the second channel, we relax our model's overidentifying restriction on small losses and small profits by replacing the earnings management probability parameter π_0 with two separate probability parameters: one for the missing small losses and another one for the excess small profits. We use an F-test to assess the joint significance of these two parameters, and we keep the cubic polynomial approximation and all other empirical settings from our main Model II estimates. To further explore the role of model restrictions, we next remove all of Model II's earnings management structure and estimate eight unconstrained parameters for the individual small-loss and small-profit bins following Chetty et al. (2011). We jointly test these eight parameters using an F-test. For brevity, we present the results only for bin width 0.0025 (the results are robust to alternative bin width definitions).

Table D2 presents the simulation results. When we replace the polynomial approximation with a linear interpolation in a narrower estimation interval, there is only a slight decrease in test power (0.4–1.1 percentage points and 2.3–3.5 percentage points for the larger and smaller estimation intervals in Table D2, respectively, compared to our main Model II tests[12]). In contrast, when we relax the overidentifying restriction that the total probability of excess small profits must equal the total probability of missing small losses (test A in Table D2), there is a considerably larger decrease in test power (5.0–8.6 percentage points). When we estimate eight unconstrained parameters for the individual small-loss and small-profit bins following Chetty et al. (2011), thus entirely

[11] Bin width = 0.001 can increase the estimation time considerably because it increases the number of firm-year-bin observations in estimation, and the computational burden of major estimation steps such as sorting is non-linear with respect to data size.

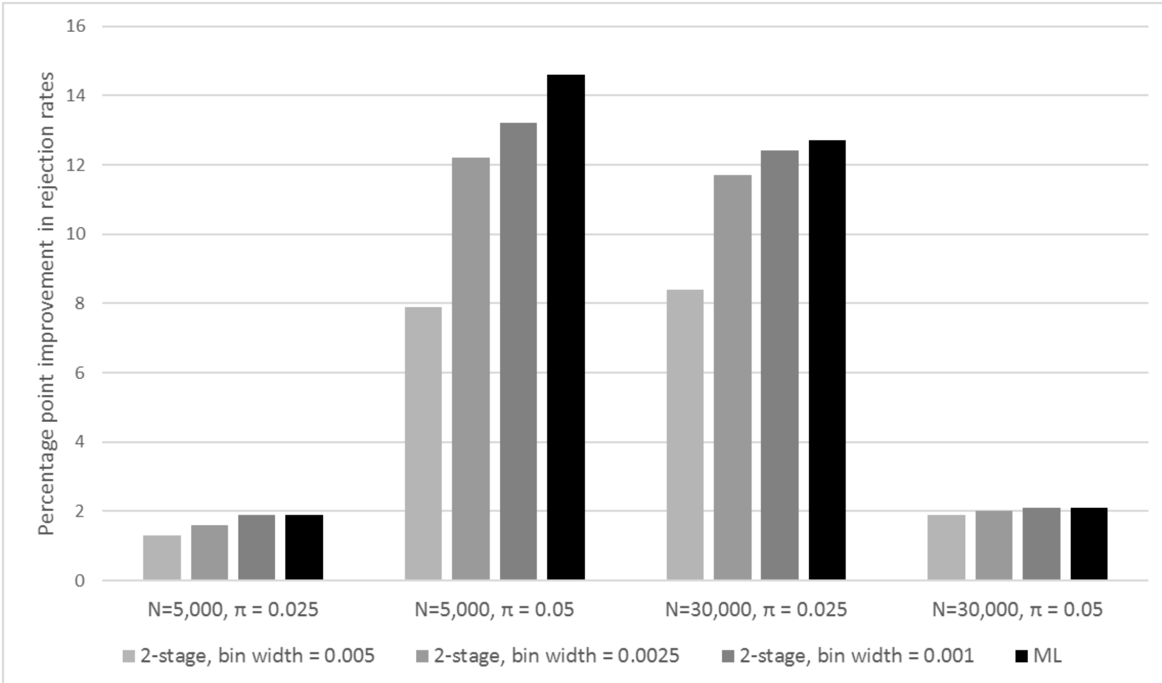
[12] Everywhere in this paragraph, we exclude scenario (6) from the computation of percentage point differences. The rejection rates in all tests for scenario (6) are very close to the upper bound of 100% because both the sample size and the effect size are large. Therefore, the measured differences in statistical power across tests are predictably miniscule even when one test performs much better than another.

dismantling the model structure in the small-loss and small-profit intervals (test B in Table D2), there is an even larger decrease in test power (7.5–30.7 percentage points). Further, in most cases this approach performs worse than the standardized difference tests because, in addition to not using the overidentifying restriction on missing small losses and excess small profits, it carves up the small-loss and small-profit intervals into multiple unconstrained estimates.

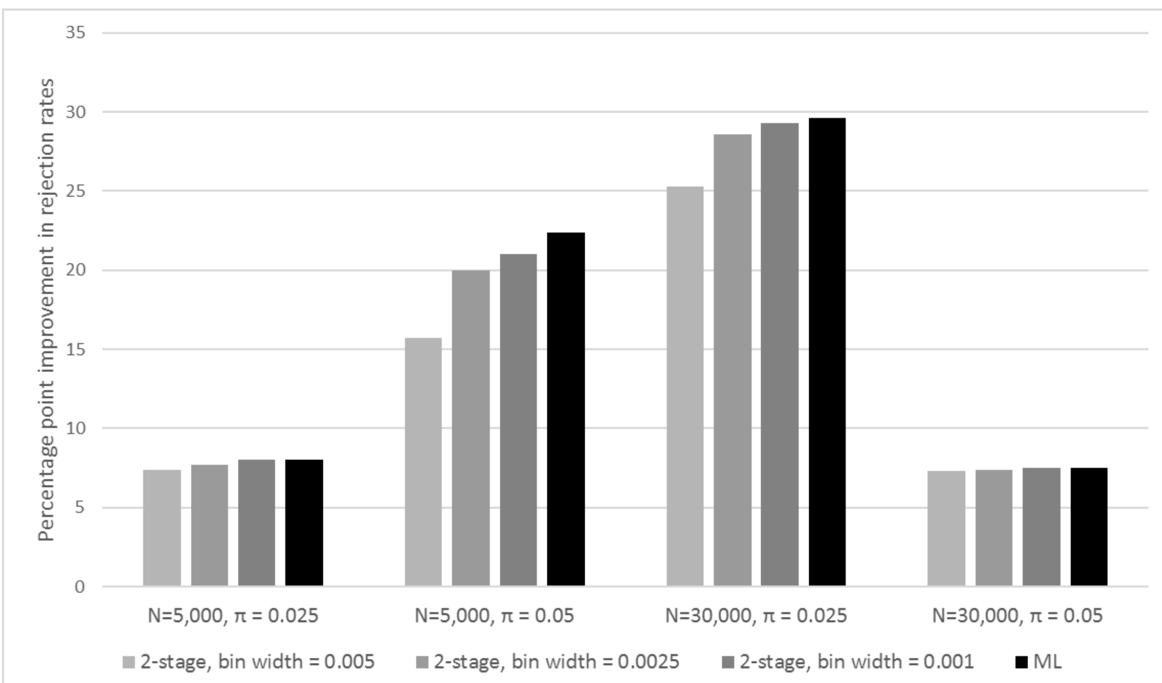
Thus, the overidentifying restriction on small losses and small profits is important for our method's statistical power, whereas the polynomial approximation is relatively less important. However, the polynomial approximation can be implemented just as easily as a linear approximation through the `degree()` option in our Stata command, and it offers slightly better test power in all scenarios in Table D2. Further, the polynomial approximation is *a priori* much more appropriate than a linear approximation in scenarios in which the discontinuity is near the peak of a distribution, especially when the distribution has high curvature (untabulated simulations confirm that the polynomial approximation improves test performance considerably in such scenarios). Therefore, we recommend that researchers use the polynomial approximation instead of the more restrictive linear interpolation.

In summary, our method offers a sizable power improvement relative to the standardized difference tests. The primary reason for this improvement is the overidentifying restriction that missing small losses must be consistent with excess small profits (Panel C of Figure 1), while the polynomial approximation plays a smaller role. Thus, even a researcher who only wants to conduct standard Burgstahler and Dichev (1997) discontinuity analysis without any explanatory variables could benefit from our method, especially when the sample size and/or effect size is small. This advantage is less important in a large sample (e.g., column 6 in Table D1), where the standard Burgstahler and Dichev (1997) tests often have sufficient power.[13]

[13] Even when the pre-managed distribution and meet-or-just-beat behavior vary with *X*, the standardized difference test can correctly detect the *existence* of distribution discontinuity. Because the pre-managed distribution conditional on *X* is smooth, the unconditional pre-managed distribution (after integrating out *X*) is also smooth. Because meet-or-just-beat behavior creates a distribution discontinuity at zero conditional on *X*, it also creates an unconditional distribution discontinuity at zero. Thus, when the conditional distribution varies with *X*, the fundamental assumptions of the standardized difference test continue to hold for the unconditional distribution. The primary advantage of our method is that it lets researchers study *multiple determinants* of distribution discontinuity; the statistical power improvement in tests of *existence* of distribution discontinuity that we document in this supplement is secondary in importance.



Panel A: Improvement relative to the left standardized difference test



Panel B: Improvement relative to the right standardized difference test

Fig D1. The percentage point improvement in test power for our two-stage and ML estimates, relative to the standardized difference tests, based on the simulation results in Table D1

Table D1. Rejection rates in simulated distribution discontinuity tests without explanatory variables

	Simulated sample comprises 5,000 observations in the interval [-0.04, 0.04)			Simulated sample comprises 30,000 observations in the interval [-0.04, 0.04)		
	true earnings management probability π_0^{true} is					
	0%	2.5%	5%	0%	2.5%	5%
	(1)	(2)	(3)	(4)	(5)	(6)
Burgstahler and Dichev (1997) standardized difference test						
left difference	6.2	22.0	44.3	6.6	59.3	97.7
right difference	4.6	15.9	36.5	3.7	42.4	92.3
Significance test for the earnings management probability π_0 in our main Model II with $K^- = K^+ = 0.01$						
Main two-stage estimation with						
bin width = 0.005	5.2	23.3	52.2	5.3	67.7	99.6
bin width = 0.0025	5.0	23.6	56.5	5.2	71.0	99.7
bin width = 0.001	5.6	23.9	57.5	5.4	71.7	99.8
ML estimation	5.1	23.9	58.9	5.6	72.0	99.8

The table presents the rejection rates in one-tailed tests with a 5% nominal significance level in 1,000 simulated samples. For the Type-I errors in columns 1 and 4, the 95% confidence interval is 3.6% to 6.4%. The simulated distribution of pre-managed earnings follows the estimates from column 4 of Panel A in Table 2, and the simulated earnings management process follows Model II with the true earnings management probability π_0^{true} set to 0, 0.025, or 0.05. In estimation for the simulated data, the estimation interval is [-0.04, 0.04), the small-loss and small-profit interval width is 0.01, and the bin width for earnings discretization in the two-stage method varies from 0.005 to 0.001. The code fragment for the two-stage estimation is:

```
kinkyX simNI, binwidth(0.005) est_bins(8) em_bins(2) em_type(ii) degree(3) cluster(gvkey)
kinkyX simNI, binwidth(0.0025) est_bins(16) em_bins(4) em_type(ii) degree(3) cluster(gvkey)
kinkyX simNI, binwidth(0.001) est_bins(40) em_bins(10) em_type(ii) degree(3) cluster(gvkey)
```

Table D2. Rejection rates for alternative empirical models in simulated distribution discontinuity tests without explanatory variables

	Simulated sample comprises 5,000 observations in the interval [-0.04, 0.04)			Simulated sample comprises 30,000 observations in the interval [-0.04, 0.04)		
	true earnings management probability π_0^{true} is					
	0%	2.5%	5%	0%	2.5%	5%
	(1)	(2)	(3)	(4)	(5)	(6)
Main Model II tests from Table D1						
bin width = 0.0025	5.0	23.6	56.5	5.2	71.0	99.7
Replace Model II's polynomial approximation with a linear approximation on a narrower estimation interval						
[-0.015, 0.015]	4.6	21.3	53.4	4.7	67.5	99.3
[-0.02, 0.02]	5.0	22.8	56.1	4.3	69.9	99.5
Relax Model II's overidentifying restrictions on small losses and small profits						
Test A: Remove the restriction that managed small losses must become small profits, but keep all other structure of Model II	4.4	18.6	49.0	6.7	62.4	99.3
Test B: Use unrestricted bin dummies for each of the small-loss and small-profit bins, following Chetty et al. (2011)	5.6	16.1	31.3	4.8	40.3	95.6

The table presents the rejection rates in one-tailed tests with a 5% nominal significance level in 1,000 simulated samples. For the Type-I errors in columns 1 and 4, the 95% confidence interval is 3.6% to 6.4%. The simulated distribution of pre-managed earnings follows the estimates from column 4 of Panel A in Table 2, and the simulated earnings management process follows Model II with the true earnings management probability π_0^{true} set to 0, 0.025, or 0.05. In all models, the bin width for earnings discretization is 0.0025 and the small-loss and small-profit interval width is 0.01. For the tests with linear approximation, we use narrower estimation intervals [-0.015, 0.015) and [-0.02, 0.02) because linear approximation for the main estimation interval [-0.04, 0.04) in untabulated exploratory tests is highly vulnerable to distribution non-linearity. The code fragment for the linear approximation tests is `kinkyX simNI, binwidth(0.0025) est_bins(n) em_bins(4) em_type(ii) degree(1) cluster(gvkey)`, where the number of estimation bins `n` is 6 for the interval [-0.015, 0.015) and 8 for the interval [-0.02, 0.02). The tests after relaxing Model II's overidentifying restrictions on small losses and small profits in the last two rows use a cubic polynomial approximation over the estimation interval [-0.04, 0.04), similar to the main estimates, and use modified estimation code with additional earnings management parameters. The distribution discontinuity test in these scenarios is a joint *F*-test of all relevant parameters (i.e., two separate parameters for missing small losses and excess small profits in test A, and eight separate parameters for individual small-loss and small-profit bins in test B).