Review of Accounting Studies
https://doi.org/10.1007/s11142-019-09525-9

Entropy-balanced accruals

Jeff L. McMullin[1] · Bryce Schonberger[2]

Published online: 10 January 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract
This study assesses whether the accrual-generating process is adequately described by a linear model with respect to a range of underlying determinants examined by prior literature. We document substantial departures from linearity across the distributions of accrual determinants, including measures of size, performance, and growth. To incorporate non-linear relations, we employ a recently developed multivariate matching approach (entropy balancing) to adjust for determinants in place of relying on a linear model. Entropy balancing identifies weights for the control sample to equalize the distribution of determinants across treatment and control samples. In simulations drawing random samples from deciles where a linear model displays poor fit, we find that entropy balancing significantly improves accrual model specification by reducing coefficient bias relative to linear and propensity-score matched models. Consistent with entropy balancing retaining sufficient power, we find that its estimates detect seeded accrual manipulations and explain variation in accruals around equity issuances.

Keywords Abnormal accruals · Nonlinearity · Entropy balancing · Propensity-score matching · Covariate balance · Initial public offerings · Seasoned equity offerings

We benefitted from numerous conversations with Charles Wasley and comments and suggestions from Eric Allen, Daniel Beneish, Matthew Cobabe, Ilia Dichev, Scott Dyreng, Leslie Hodder, Michelle Hutchens, Mat McCubbins, Brian Miller, Miguel Minutti-Meza, Spencer Pierce, Richard Price, Darren Roulstone, Joseph H. Schroeder, Richard Sloan (editor), Anup Srivastava (discussant), Karen Ton, Brady Twedt, Jim Wahlen, Teri Yohn, Jerry Zimmerman, an anonymous referee, and seminar participants at the American Accounting Association's Annual Meeting, Brigham Young University's Accounting Research Symposium, Indiana University, and the University of Rochester. The authors appreciate financial support from the Deloitte Foundation, the Simon Business School, and the Kelley School of Business.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11142-019-09525-9) contains supplementary material, which is available to authorized users.

✉ Bryce Schonberger
bryce.schonberger@simon.rochester.edu

Jeff L. McMullin
jemcmull@indiana.edu

[1] Kelley School of Business, Indiana University, Bloomington, IN 47405, USA

[2] Simon Business School, University of Rochester, Rochester, NY 14627, USA

JEL M4

1 Introduction

A large literature relies on linear models of the accrual-generating process to isolate "abnormal accruals," hypothesized to be the result of managerial manipulation. (See Dechow et al. 2010 for a review.) While recent studies measure abnormal accruals by matching firms on a limited set of determinants, such as performance (Kothari et al. 2005), size (Ecker et al. 2013), and size, age, and growth (Armstrong et al. 2015), these studies assume linear relations between accruals and *unmatched* determinants. In this study, we seek to examine two research questions. First, is the accrual-generating process adequately described by a linear model with respect to a broad set of fundamental determinants? Second, do multivariate matching approaches that allow for nonlinearities in the accrual-generating process significantly improve the specification and power of tests for accrual-based earnings management? Importantly, multivariate matching allows us to avoid ad hoc partitions of variables into matched and linear controls.

Beginning first with our investigation into accrual linearity, univariate plots of accruals for portfolios formed on the basis of a variety of accrual determinants (e.g., growth, capital intensity, performance, size, and net financing) reveal substantial departures from a simple linear model (see Fig. 1). For example, a linear model for sales growth, which enters as a linear control in the widely used Jones (1991) model, overestimates (underestimates) both total and working capital accruals in the lowest (second and third) deciles of the distribution.[1] To assess the importance of departures from linearity for the specification of linear accrual models, we conduct a series of simulations, drawing independent random samples from deciles with pronounced departures from linearity for each of our determinants. The results show substantial misspecification for estimates of linear models, with rejection rates for the null hypothesis of zero abnormal accruals exceeding 20% in several deciles.

Given pronounced nonlinear relations between accruals and a broad set of determinants, we turn to identifying a workable solution for conducting well-specified and powerful abnormal accrual tests. To do this, we evaluate abnormal accrual estimates formed using multivariate matching via entropy balancing (Hainmueller 2012), propensity-score matching (Rosenbaum and Rubin 1983), and propensity-score weighting (Freedman and Berk 2008). These methods are designed to systematically address covariate imbalance (i.e., differences in observables on one or more distributional moments) in multiple accrual determinants *without* relying on a first-stage linear regression, thereby relaxing the requirement to specify a functional form between independent and dependent variables (Rosenbaum and Rubin 1983; Shipman et al. 2017).[2] In particular, these approaches allow us to avoid assuming linearity for the *full* set of determinants we examine.

[1] Jones's (1991) discretionary accrual study is cited more than 8100 times, according to Google Scholar, as of December 2018.

[2] In addition, multivariate matching avoids the two-step estimation concerns raised by Chen et al. (2018) for studies using residuals from a first-stage linear model in a second-stage regression. These authors document that this two-step approach results in biased estimates (inflation, attenuation, or sign change) that lead to type I and type II errors.

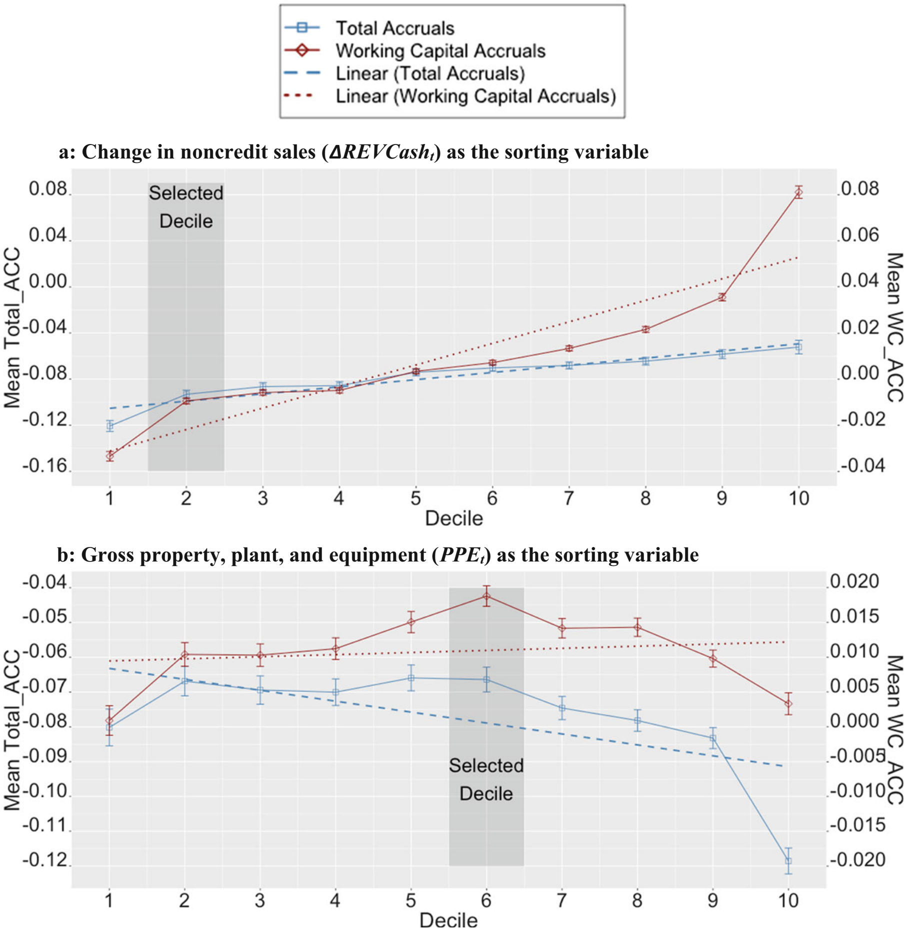
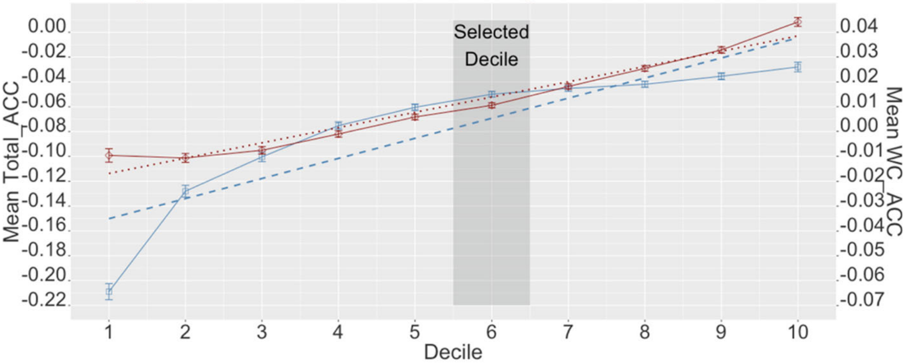


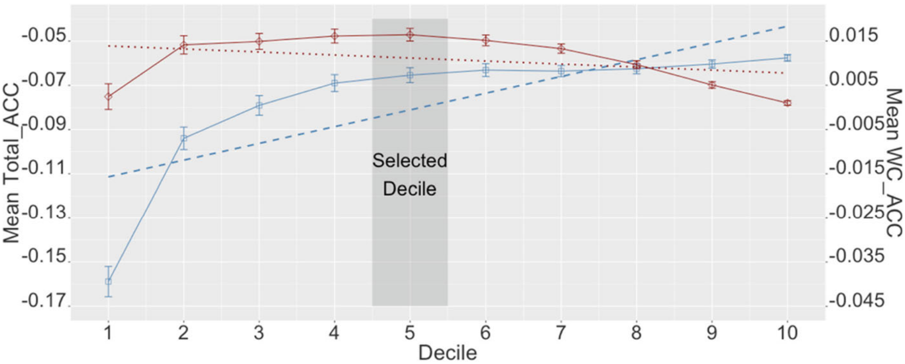
Fig. 1 Plots for assessing departures from linearity in the relation between accruals and underlying covariates: portfolio approach. This figure plots mean total (*ACC*, left vertical axis) and working capital (*WC_ACC*, right vertical axis) accruals for 10 portfolios (deciles) separately sorted each year on the following accrual determinants (on the horizontal axis of each figure): change in noncredit sales ($\Delta REVCash_t$); gross property, plant, and equipment (PPE_t); lagged return on assets (ROA_{t-1}); the natural log of prior-year book value of total assets (log_AT_{t-1}); prior-year earnings-to-price (EP_{t-1}); prior-year book-to-market (BM_{t-1}); and net financing raised (New_FIN_t). Each panel presents error bars indicating the 95% confidence interval for *ACC* and *WC_ACC*. Shaded gray areas in each figure indicate selected deciles with pronounced departures from a linear relation between either (or both) *ACC* or *WC_ACC* and the underlying determinant. These selected deciles are included in the simulations examining type I errors (Tables 2–3) and type II errors (Table 4) for each accrual model. See Appendix Table 7 for detailed variable definitions

While entropy-balancing and propensity-score approaches each target covariate balance across treatment and control samples, they differ in how they assign weights to the control sample. (See Section 3 for details.) Briefly, propensity-score matching estimates a first-stage treatment model and then matches treatment observations to control observations on the resulting propensity score, assigning a weight of either one (matched) or zero (excluded) to each control observation. In contrast, entropy balancing identifies continuous weights for all control sample observations to equalize the

c: Prior-year return on assets (*ROA_{t-1}*) as the sorting variable



d: Natural logarithm of prior-year book value of total assets (*log_AT_{t-1}*) as the sorting variable



e: Prior-year earnings-to-price ratio (*EP_{t-1}*) as the sorting variable

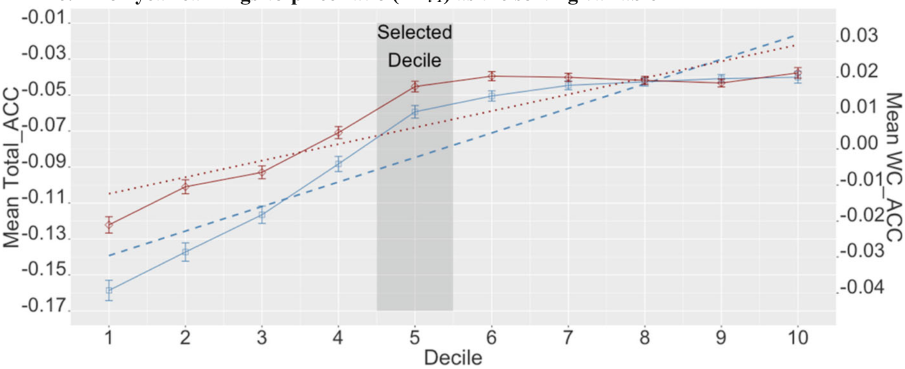


Fig. 1 (continued)

treatment and control sample distribution moments (e.g., means, variances, and skewness) for all included covariates while staying as close as possible to equal weighted. This weighting approach ensures balance for all covariates without the need for researcher adjustment of a propensity model.

To compare the relative performance of multivariate-matched normal accrual estimates, we re-run our simulations, using multivariate matching to construct estimates of normal accruals. Consistent with entropy balancing addressing nonlinear relations with

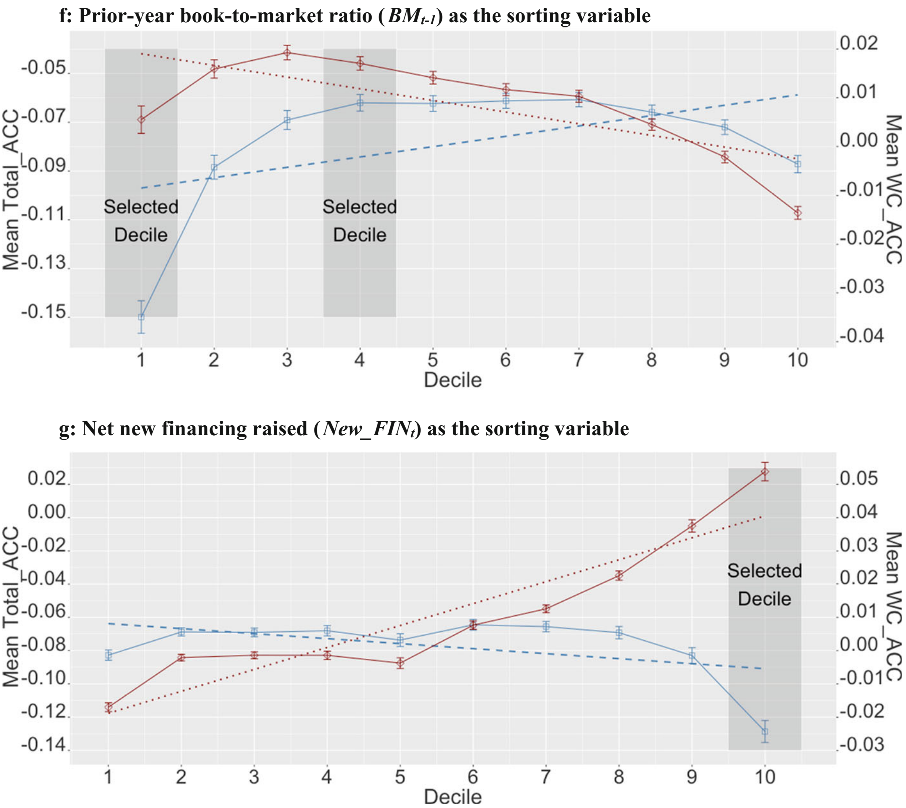


Fig. 1 (continued)

underlying determinants, rejection rates for the null hypothesis of no earnings management for both total and working capital accruals never exceed 11%, resulting in the majority of the deciles we examine being characterized by a well-specified test (i.e., a low type I error rate). Turning to the other two approaches, while these methods generally improve test specification relative to ordinary least squares models, their improvement falls short of that achieved by entropy balancing. In particular, propensity-score matching estimates are well specified in fewer than half of all portfolios, with maximum rejection rates exceeding 20% for both total and working capital accruals in some decile portfolios.

To explain how and why entropy balancing achieves superior specification, we compare mean coefficient estimates and standard errors from each estimation method. We find that entropy balancing substantially reduces absolute bias in coefficient estimates (by more than 50%), relative to linear models in the random samples we consider. In contrast, standard error estimates show relatively little difference across methods. More importantly, we find that entropy balancing markedly improves covariate balance on both the first and second moments, relative to ordinary-least-squares and propensity-score approaches, consistent with a reduction in bias due to better matching.

While the simulations favor entropy balancing in terms of lower type I errors, we also evaluate each method's power to detect earnings management to address any concern that multivariate matching sacrifices test power. Using the same random samples drawn for our specification tests, we seed accrual manipulations for firms in a pseudo-treatment sample and then measure the incidence of type II errors (false negatives). For seeded levels of 1% and 2% of total assets, we find that entropy-balanced accrual estimates correctly detect working capital (total) accrual manipulations in 33.1% (49.0%) and 70.4% (81.9%) of our random samples, respectively. These findings suggest that entropy balancing retains sufficient power while achieving proper test specification.

Given findings from our simulations, we next assess the advantages of our entropy-balancing approach by re-investigating abnormal accruals around initial public offerings (IPOs). This is a useful setting for at least three reasons. First, our simulations indicate a pronounced departure from linearity in working capital accruals as well as linear model misspecification for firms in the top decile of new financing. Second, evidence on earnings management around equity issuances is mixed (see Section 2). Finally, this setting allows us to incorporate tests for accrual reversals (Dechow et al. 2012). If a firm manages earnings via accruals around an IPO, we expect to observe elevated accruals in the year of the offering followed by a reversal in the next year. We find entropy-balanced abnormal accruals indicate *insignificant* abnormal accruals and reversals in the years surrounding an IPO, consistent with stringent reporting requirements and investor scrutiny of IPO firms limiting earnings management (Ball and Shivakumar 2008). In contrast, results based on propensity-score-matching and ordinary-least-squares approaches show *significant* abnormal working capital accruals in the years surrounding an IPO, leading to an inference of upward earnings management for these firms. Additional tests examining Accounting and Auditing Enforcement Releases (AAERs) show that entropy-balanced estimates retain sufficient power to detect accrual manipulations in these firms, indicating that insignificant results in our IPO tests using entropy balancing are unlikely due to a low power test.

Entropy balancing is not without drawbacks. Broadly, any matching estimator works only if the researcher matches on the correct set of determinants. Additionally, matching estimators are estimates of local average treatment effects rather than average treatment effects (Imbens 2010). More specific to entropy balancing is a concern with assigning large weights to a handful of control observations, resulting in estimates that are sensitive to small changes in the composition of the control sample. We assess this issue in two ways. First, we propose and compute a *Weight Ratio* that measures the relative percentage of control sample observations receiving above equal weight, where higher ratios indicate a larger pool of entropy balancing-matched observations. *Weight Ratios* throughout our tests show that entropy balancing assigns weight to more observations than one-to-one matching. Second, we use regression leverages to see whether entropy balancing assigns large weights to outlying observations. We find that its models display maximum leverages that are above (below) those for ordinary least squares when the treatment comes from an extreme (middle) decile. These diagnostic tests highlight a trade-off in using entropy balancing for treatment samples that are drawn from the extremes of a covariate distribution: achieving greater covariate balance may come at the cost of assigning larger weights to a subset of observations. As a result, while entropy balancing displays clear benefits in terms of test specification, we urge future researchers to assess the extent to which it assigns extreme weights to a subset of control observations, to potential outlying observations, or both to achieve covariate balance.

Our study has at least three implications for future research. First, our evidence suggests that linear models of accrual determinants are misspecified when examining treatments that are correlated with (and thereby reflect) the presence of widespread nonlinearities. Second, simulations suggest that entropy balancing offers the best approach for incorporating multiple accrual determinants when measuring abnormal accruals in terms of both the specification and power for detecting accrual manipulations at plausible magnitudes. While no single model will result in a well-specified, powerful test in each and every research setting, entropy balancing is useful in mitigating type I errors by reducing coefficient bias in the presence of nonlinear relations, allowing researchers to draw more reliable inferences regarding abnormal accruals. Third, we demonstrate the flexibility of entropy balancing in incorporating accrual determinants. This will allow future researchers to easily adapt the set of covariates to capture unique features of their research setting or sample. On this latter point, entropy balancing is easy to implement in Stata (Hainmueller and Xu 2013) and requires relatively few discretionary choices by the researcher, both features which should facilitate (when necessary) the replication of prior results.[3] Given entropy balancing's benefits and ease of implementation, we recommend future researchers use it in studies of abnormal accruals, if not in their main tests, at least in sensitivity tests.

We also contribute to accounting literature more broadly by providing a detailed discussion of the properties (and benefits and drawbacks) of entropy balancing along with systematic evidence on its usefulness, relative to propensity-score approaches. On this point, the benefits are not limited solely to the accruals setting but rather apply to other settings, such as abnormal audit fees, where outcome variables display a nonlinear relation with underlying controls. Consistent with entropy balancing's benefits, a number of contemporaneous studies in accounting are beginning to employ the technique (e.g., Shroff et al. 2017; Bonsall and Miller 2017; Wilde 2017), often citing the evidence we present here.

2 Related literature

A substantial body of research examines abnormal accruals. Jones (1991) provides the workhorse linear model for estimating a normal accrual, using controls for gross property, plant, and equipment (PPE), changes in revenue, and the inverse of total assets. (See Dechow et al. 2010 for a review of this accrual literature.) Subsequent research expands this set of accrual determinants to include industry (DeFond and Jiambalvo 1994), performance (Kothari et al. 2005), size (Ecker et al. 2013), and age (Armstrong et al. 2015), among others. It also proposes a range of methods for estimating normal accruals, including firm-specific regressions (Jones 1991), cross-sectional regressions within industry (DeFond and Jiambalvo 1994; Dechow et al. 1995) or size portfolios (Ecker et al. 2013) and one-to-one matching on performance (DeFond and Subramanyam 1998; Kothari et al. 2005). More recently, Armstrong et al. (2015) employ propensity-score matching to control for size, age, and growth. This

[3] Consistent with the ease of implementing entropy balancing, we provide an online appendix with detailed Stata code and examples.

matched-sample design addresses systematic differences across treatment and control groups while relaxing the assumption of a linear relation for the matched covariates. However, studies that match on a subset of covariates continue to adjust for unmatched determinants via a first-stage linear determinants model to measure abnormal accruals (via a regression residual).

Despite widespread use of linear first-stage models to estimate abnormal accruals, several studies note that estimates are frequently misspecified. For example, Dechow et al. (1995) find that abnormal accrual estimates from linear models are significantly misspecified in subsamples of firms with extreme performance and that these estimates lack power in detecting earnings management at plausible levels. Relatedly, Owens et al. (2017) show that first-stage linear models for normal accruals display poor fit in the presence of business model shocks. Consistent with these issues, Chen et al. (2018) provide evidence that a two-step approach to estimating abnormal accruals results in biased estimates (inflation, attenuation, or sign change) that lead to type I and type II errors. Research also points to issues with matching approaches. Ayers et al. (2006) and Keung and Shih (2014) find that one-to-one matched accrual estimates are biased toward zero, resulting in low power tests (elevated type II errors).

Recently, studies model the time-series properties of accruals to avoid specifying cross-sectional determinants. Dechow and Dichev (2002) argue that working capital accruals unrelated to near-term cash flows represent likely accrual estimation errors. Building on this work, Dechow et al. (2012) show that simultaneously testing for the original accrual and its subsequent reversal increases the power of earnings management tests.[4]

3 Research design

We employ multivariate matching to estimate abnormal accruals by comparing accruals across treatment and matched control samples. By reducing (or eliminating) observable differences across treatment and control groups, multivariate matching increases the credibility of using the matched control sample's average accrual to estimate the treatment sample's normal accrual. We examine two matching approaches: entropy balancing (Hainmueller 2012) and propensity-score methods, including propensity-score matching (e.g., Rosenbaum and Rubin 1983) and propensity-score weighting (e.g., Freedman and Berk 2008).

3.1 Estimating abnormal accruals using propensity-score approaches

Propensity-score methods first estimate the likelihood of being treated using a Logit or Probit model, with resulting fitted values serving as propensity scores. Propensity-score matching identifies control sample observations with near identical propensity scores as a match for each treatment observation. Using the matched observations, we estimate

[4] Similarly, Nikolaev (2017) and Bloomfield et al. (2017) model accruals using a general method of moments approach and by modeling the extent to which accrual innovations map into future cash flows, respectively.

the difference in accruals between the treated and propensity-matched control sample via a model of the following form.

$$Accruals = \alpha + \beta_1 Treat + \Gamma X + \epsilon, \tag{1}$$

where *Treat* is an indicator set to one for all observations in the treatment sample and to zero for matched control observations, and X is a vector of accrual determinants.[5] Observing a β_1 that is significantly different from zero indicates an abnormal level of accruals in the treatment sample, relative to the propensity-score-matched control sample.

In lieu of matching, estimated propensity scores can be used to assign continuous weights to the control sample (rather than integer weights via propensity-score matching). Freedman and Berk (2008) use estimated propensity scores (*PS*) to implement propensity-score weighting by assigning a weight of 1/*PS* (1/[1-*PS*]) for observations in the treated (control) sample. Estimating abnormal accruals via this approach then requires running a weighted ordinary least squares version of Eq. (1).

3.2 Estimating abnormal accruals using entropy balancing

Like propensity-score approaches, the purpose of entropy balancing is to eliminate differences in observable covariates across treatment and control samples. However, the approach differs in how it achieves covariate balance.[6] To use entropy balancing, a researcher specifies a set of covariates (e.g., accrual determinants) to be balanced, the balance conditions, and a tolerance level. Balance conditions are the moments (mean, variance, skewness, or a combination of these) of covariate distributions that should be equalized across treated and weighted control samples. The tolerance level determines the minimum degree of covariate balance that must be achieved before the entropy-balancing program ceases adjusting control sample weights, analogous to a caliper width for propensity-score matching. In the extreme, a tolerance level of zero will result in identical distribution moments for the treated and weighted control samples. More covariates, balance conditions, and restrictive tolerance levels require a control sample with sufficient distributional overlap for entropy balancing to converge. Because research on accruals tends to employ relatively large pools of control observations, convergence should rarely present an issue even with very restrictive tolerances.

Given these inputs, the entropy-balancing algorithm uses an iterative process to identify a weight for each observation in the control sample, such that the weighted control sample meets the balance conditions within the specified tolerance level. Entropy balancing seeks to stay as close to equal weight as possible and does not assign negative weights. Iterations continue until either the balance conditions are met,

[5] We estimate this difference in a multiple regression to control for any remaining difference after matching, as this approach is shown to produce the least biased estimate (Rubin 1973).

[6] Hainmueller (2012) notes that entropy balancing is a generalization of propensity-score weighting (Hirano et al. 2003). In addition, it is in essence a large-sample version of the synthetic controls method developed by Abadie and Gardeazabal (2003) and Abadie et al. (2010). Synthetic controls are designed to examine case studies by comparing a single treated case with a weighted sample of control cases (e.g., examining the effect of a cigarette tax on cigarette consumption in California or the effect of terrorism on a specific region's economic output).

or it is determined that convergence is impossible, due to the control sample's lack of covariate overlap with the treated sample.[7] Once balance conditions are met, the iterative process ceases, and control sample weights are retained. We test for abnormal accruals by estimating the difference in accruals between the treated sample and the weighted control sample via a weighted regression version of Eq. (1). Treated observations have a weight of one and control sample observations have the weight identified by the entropy-balancing algorithm.

3.3 Advantages and limitations of entropy balancing

Although propensity-score matching and entropy balancing both adjust for multiple accrual determinants, the latter has two primary conceptual advantages.[8] First, entropy balancing ensures that higher-order moments of covariate distributions are nearly identical across treated and control samples, while propensity-score matching does not. In fact, propensity-score matching generally does not ensure that means for each individual covariate are balanced, focusing instead on balancing the estimated propensity score. Ensuring covariate balance increases the plausibility that any differences in accruals we document are driven by the treatment rather than correlated differences in determinants (i.e., unconfoundedness; Rosenbaum and Rubin 1983). In addition, estimates from balanced samples are less sensitive to research design choices. (See chapter 15 of Imbens and Rubin 2015 for a discussion of this point.) In research on accruals, the ability of entropy balancing to ensure balance on variance should prove important in light of evidence that accruals are systematically related to operating and cash flow volatility (Dechow and Dichev 2002).

Second, focusing almost solely on setting a tolerance level for convergence of the algorithm, entropy balancing permits less researcher discretion, relative to propensity-score matching. In particular, using entropy balancing allows researchers to avoid specifying the propensity-score model, deciding to match with or without replacement, selecting caliper distance, conducting one-to-one versus one-to-many matching, and assessing match quality. On this point, Shipman et al. (2017, p. 213) find "seemingly innocuous design choices greatly influence sample composition and estimates" of propensity-score-matching treatment effects.[9]

However, entropy balancing is not without drawbacks. In particular, it may assign large weights to a small set of control observations. This will occur in settings where the treatment sample differs markedly from control sample observations on at least one covariate. For example, treatment observations selected from the top decile of the distribution of firm size will have fewer control observations with a similar size, relative to treatment firms chosen from a middle decile (where control observations are both

[7] These latter samples will not converge when using only positive control sample weights (e.g., a treatment sample of large firms is compared to a control sample of small firms, such that there is limited or no overlap in their distributions).

[8] In addition to these advantages, Hainmueller (2012) notes that entropy balancing is in a class of matching methods called "equal percent bias reducing," which guarantee that covariate imbalance after matching will be lower than before matching, and Zhao and Percival (2016) note that entropy balancing is "doubly robust with respect to linear outcome regression and logistic propensity score regression." See work by King et al. (2011) for a discussion of propensity-score matching's shortcomings.

[9] Similarly, research on audit quality finds that estimates of Big N auditor effects are sensitive to changes in propensity-score-matching assumptions (DeFond et al. 2017).

larger and smaller). In cases where entropy balancing assigns large weights to a few control observations, slight alterations to the control sample could produce significantly different estimates of abnormal accruals (i.e., suffer from small sample problems).

We assess whether entropy balancing assigns extreme weights in two ways. First, we calculate a *Weight Ratio*, measured as the number of control observations receiving above equal weight in the entropy-balancing regression divided by the number of observations appearing in a one-to-one match without replacement. *Weight Ratios* below (above) one indicate that entropy balancing assigned large weights to a small subset of control firms (included more firms in the weighted control sample relative to a one-to-one match). Second, we conduct regression diagnostics by examining regression leverages for our entropy-balancing-weighted regression models. Regression leverages measure whether entropy balancing assigns large weights to potentially outlying observations, where the independent variable values are far from those of other observations (i.e., have high leverage). Entropy balancing may exacerbate issues with outlying observations (relative to ordinary least squares) to the extent it assigns large weights to outliers.

In addition, entropy balancing is not a panacea when it comes to accruals research. Although it ensures covariate balance for all determinants included in the matching algorithm, accurately estimating abnormal accruals requires correctly specifying the set of underlying accrual determinants. While inclusion of determinants should be guided by theory, in the current study, we elect to include determinants that are common in accruals research to examine the influence of nonlinearities on these variables. Future researchers applying entropy balancing should consider whether this set of determinants is appropriate in their setting. Consistent with this, we include covariates in our tests that are correlated with the treatment of interest (such as the amount of net new financing in our equity issuance tests). Additionally, entropy balancing's sample-level estimation of abnormal accruals cannot be applied when abnormal accruals are an independent (rather than an outcome) variable, as such settings require firm-level abnormal accruals estimates.

4 Data and sample

Our sample covers fiscal years ending between December 1987 and December 2017 to allow for the availability of cash flow statement data in Compustat. We require variables necessary to measure total and working capital accruals. To measure total accruals (ACC_t), we follow the approach of Hribar and Collins (2002) to measure the difference between net income and cash flow from operations (IBC – OANCF + XIDOC, from Compustat) scaled by lagged book value of total assets. We use this measure for total accruals for two reasons. First, it is commonly used in the literature, including in recent studies examining accruals around new equity issuances (e.g., Armstrong et al. 2015). Second, alternative measures of comprehensive accruals include accruals relating to investment and non-articulating accruals relating to acquisitions or divestments (Larson et al. 2018). These large investment accruals pose an issue with respect to the planned investment of proceeds by firms raising new capital. For working capital accruals (WC_ACC_t), we follow the approach recommended by Larson et al. (2018) to measure changes in working capital across successive balance sheets (ΔACT – ΔCHE – ΔLCT

+ ΔDLC, from Compustat) scaled by lagged total assets. We require the absolute value of both ACC_t and WC_ACC_t to be less than one to eliminate outliers and obvious errors.

Next, we require availability for the following determinants of accruals examined in prior research (see Section 2): the natural logarithm of lagged book value of total assets (log_AT_{t-1}); gross property, plant, and equipment scaled by lagged assets (PPE_t); change in cash sales scaled by lagged assets ($\Delta REVCash_t$); prior-year return on assets (ROA_{t-1}); and the amount of net debt and equity financing raised as a proportion of lagged assets (New_FIN_t). In addition, we examine a subsample of observations with lagged equity market capitalization available to compute lagged book-to-market (BM_{t-1}) and earnings-to-price (EP_{t-1}) ratios, as requiring the availability of these lagged ratios would eliminate all IPOs from our main tests. We winsorize all continuous determinants (with the exception of variables in logged form) at the top and bottom percentiles to reduce the influence of outlying observations. We provide detail for calculations and data sources for all variables in Appendix Table 7. We also require at least 10 observations in each two-digit SIC industry-year to allow for the estimation of firm-year abnormal accrual measures. After these filters, we are left with a final sample of 185,166 firm-years between 1987 and 2017.

Data on IPOs and seasoned equity offerings (SEOs) come from the SDC Platinum database. We consider IPOs and SEOs issuing common shares in the U.S. public equity market between 1987 and 2017, consistent with our full sample. Data on firms subject to AAERs comes from a hand-collected sample used by Dechow et al. (2011).[10]

Table 1 provides descriptive statistics for the final sample of firm-years. This table shows that, while total accruals (ACC_t) display a negative mean (median) of −0.080 (−0.059), consistent with these accruals omitting initial investments in noncurrent assets, working capital accruals (WC_ACC_t) display a small positive mean (median) of 0.008 (0.003). Table 1 also reports descriptive statistics for abnormal accrual estimates, using total (DAC_t) and working capital (WC_DAC_t) accruals for a modified-Jones (1991) model that includes changes in cash revenue ($\Delta REVCash_t$) and PPE_t (e.g., Dechow et al. 1995) as linear controls estimated within industry-year portfolios.[11] We use these abnormal accruals to generate performance-matched abnormal accruals ($DAC\_ROAMatch_t$, $WC\_DAC\_ROAMatch_t$), following the approach outlined by Kothari et al. (2005) to select a peer from the same industry-year with the closest value of ROA_{t-1} before subtracting the peer's abnormal accrual.

Turning to our fundamental determinants, mean ROA_{t-1} is negative for our full sample period, while a positive mean $\Delta REVCash_t$ indicates positive average sales growth. A mean (standard deviation) of 0.623 (0.513) for PPE_t indicates sizeable average (highly variable) capital intensity across firms. In addition, average new financing raised (New_FIN_t) represents 16.4% of lagged total assets. However, New_FIN_t is highly skewed, with some firms obtaining substantial amounts of new financing. Consistent with this, we find that only 1.8% (4.0%) of our firm-year observations undergo an IPO (SEO) during the year.

[10] This dataset is available through the University of California at Berkeley's Haas School of Business and is current through August 31, 2012. AAERs typically take at least three years for the SEC to file. As a result, AAERs are considered for a subsample of observations over the 1987–2009 period.

[11] In untabulated tests, we find nearly identical results using the standard Jones (1991) model that includes a control for ROA_t (Kothari et al. 2005) in place of this normal accrual model.

Table 1 Descriptive statistics for firm-year observations in the full sample (1987–2017) and AAER subsample (1987–August 2009)

Variable	Mean	St. deviation	Min	Q1	Median	Q3	Max
Full sample (1987–2017)							
ACC_t	−0.080	0.168	−1.000	−0.124	−0.059	−0.013	0.992
WC_ACC_t	0.008	0.135	−1.000	−0.032	0.003	0.043	1.000
DAC_t	0.000	0.157	−0.979	−0.042	0.015	0.066	0.996
$PMDAC_t$	0.000	0.208	−1.000	−0.082	0.000	0.081	1.000
WC_DAC_t	0.000	0.125	−0.995	−0.039	0.000	0.038	0.999
WC_PMDAC_t	0.000	0.172	−1.000	−0.063	0.000	0.063	1.000
$\Delta REVCash_t$	0.102	0.351	−0.970	−0.023	0.043	0.177	1.786
PPE_t	0.623	0.513	0.000	0.226	0.492	0.912	2.652
log_AT_{t-1}	5.009	2.553	−6.908	3.177	4.923	6.800	13.010
ROA_{t-1}	−0.086	0.347	−2.088	−0.084	0.022	0.066	0.321
New_FIN_t	0.164	0.533	−0.244	−0.019	0.003	0.095	3.632
IPO_t	0.018	0.134	0	0	0	0	1
SEO_t	0.040	0.196	0	0	0	0	1
BM_{t-1}	0.582	0.891	−3.840	0.231	0.465	0.811	4.521
EP_{t-1}	−0.140	0.640	−4.812	−0.083	0.027	0.064	0.402
AAER subsample (1987–Aug 2009)							
AAER (Y)	0.007	0.084	0	0	0	0	1
AAER (Y + 1)	0.002	0.050	0	0	0	0	1
$\Delta REVCash_t$	0.110	0.353	−0.970	−0.023	0.059	0.200	1.786
PPE_t	0.616	0.493	0.000	0.242	0.493	0.878	2.652
ROA_{t-1}	−0.070	0.320	−2.088	−0.071	0.026	0.070	0.321
B/M_t	0.534	1.188	−6.866	0.238	0.480	0.831	4.741
log_MCAP_t	5.377	2.346	−6.180	3.673	5.251	6.977	14.417
Lev_t	0.253	0.275	0.000	0.031	0.201	0.372	1.788
New_FIN_t	0.131	0.448	−0.244	−0.021	0.003	0.087	3.632
$RevCash\_Growth_t$	0.223	0.779	−1.000	−0.044	0.085	0.268	5.292
ΔROA_t	0.013	0.225	−0.746	−0.042	0.006	0.048	1.225

This table presents descriptive statistics for the full sample of 185,166 firm-year observations over the 1987–2017 sample period with data available to calculate total and working capital accruals along with base model accrual determinants in the form of changes in cash sales ($\Delta REVCash_t$), size (natural log of lagged total assets, log_AT_{t-1}), capital intensity (gross property, plant, and equipment scaled by lagged assets, PPE_t), and prior-year performance (ROA_{t-1}). We also present descriptive statistics for variables measuring the amount of net new financing (New_FIN_t) and indicator variables set to one for years in which the firm conducts an initial public (IPO_t) or seasoned equity (SEO_t) offering. Descriptive statistics for SEO_t exclude observations with an IPO in the current or prior year to distinguish IPO and SEO effects in later tests. In addition, we present descriptive statistics for 159,520 firm-year observations with lagged equity market capitalization in Compustat available to measure book-to-market (BM_{t-1}) and earnings-to-price (EP_{t-1}) ratios for use in our simulations in Tables 2–4. We also present descriptive statistics for 117,671 firm-year observations with data available to compute the F-Score (Dechow et al. 2011) and with Accounting and Auditing Enforcement Releases (AAERs) identified using the hand-collected sample of Dechow et al. (2011) updated through August 31, 2012. All accruals measures are trimmed if absolute values exceed 100% of lagged total assets to eliminate outlying observations. See Appendix Table 7 for detailed variable definitions

5 Results

5.1 Nonlinearities in accrual distributions with respect to underlying fundamentals

We begin by plotting values for total (*ACC*) and working capital (*WC_ACC*) accruals for portfolios formed on the basis of annual sorts of each of the accrual determinants in Table 1. Figure 1 presents mean ACC_t and WC_ACC_t values and their corresponding standard error bands for 10 portfolios (deciles) of each determinant along with the linear line of best fit for the mean *ACC* and *WC_ACC* values to facilitate assessing the fit of a simple linear model.

The takeaway from Fig. 1 is that both *ACC* and *WC_ACC* display substantial departures from linearity for each of the determinants examined. Further, these departures are not isolated to the extremes of each distribution. For example, mean values of *ACC* and *WC_ACC* for the $\Delta REVCash_t$ sort indicate that, while a simple linear model will underestimate accruals in the top decile for *WC_ACC*, deciles 2–3 (6–9) also display a pronounced under- (over-) estimate of *WC_ACC*. In contrast, a linear model for *ACC* displays a fairly good fit for all but the lowest deciles of $\Delta REVCash_t$. Panels B–G of Fig. 1 reveal similar (more pronounced) evidence of departures from linearity for *WC_ACC* (*ACC*) values. For each accrual determinant, we select a decile (shaded in Fig. 1) with a substantial departure from linearity to examine the specification of our normal accrual models. While we select a middle decile (between two and nine) for most determinants, we also select two extreme deciles: the bottom (top) decile for BM_{t-1} (New_FIN_t). We select these extreme deciles because the bottom decile for BM_{t-1} displays among the most extreme departures from linearity of any of the portfolio sorts for both *ACC* and *WC_ACC* and the top decile of New_FIN_t aligns with the treatment samples for our subsequent equity issuance tests. We expect matching to yield lower efficiency gains in these extreme deciles, relative to treatments drawn from middle deciles, as matching algorithms require a pool of control observations with similar determinant values (i.e., there is less distributional overlap at the tails of the distribution, increasing the likelihood of noisy matching).

5.2 Simulation results—Type I error rates

For each selected decile with a significant departure from linearity in Fig. 1, we randomly select a sample of 500 firm-years (without replacement) to use as a "pseudo treatment" sample. We combine this sample with 4500 control observations randomly chosen (without replacement) from the remaining pooled sample of firm-years without regard to the distribution of the determinant. This approach results in a treatment sample that is directly correlated with each accrual determinant along with a control group reflecting the average Compustat sample. This aligns with the observation by Kothari (2001) that earnings management studies typically examine samples of firms with unusual performance. We follow Kothari et al. (2005) and conduct 1000 independent draws of 5000 observations each (500 treated and 4500 control).

For each of the 1000 independent samples, we first address covariate imbalance by estimating entropy-balanced, propensity-score-matched, propensity-score-weighted, and performance-matched models. Specifically, we estimate a Logit propensity model using a set of four determinants that the literature generally employs, following the

Jones (1991) model and its modifications (log_AT_{t-1}, PPE_t, $\Delta REVCash_t$, ROA_{t-1}), along with the sorting variable used to select the 500 treatment observations (*Sort_VAR*). In cases where the *Sort_VAR* is one of the other four covariates (e.g., PPE_t) or is highly correlated with one of the covariates (e.g., EP_{t-1} and ROA_{t-1}), we include only the four baseline covariates in the Logit propensity model (and second-stage regression).[12] We conduct propensity-score matching using both one-nearest neighbor (with replacement) and five-nearest neighbor matching with a caliper of 0.03, following the recommendation by Shipman et al. (2017). For entropy balancing, we balance this same set of covariates in each iteration on three moments (mean, variance, and skewness) with a tolerance of 0.1.[13] We performance-match each of the 5000 observations in the random sample to the closest ROA_{t-1} peer within the same two-digit SIC industry-year in the full Compustat sample, without replacement. This exploits the full set of industry peers for performance-matching, consistent with the common approach in other studies.

To assess type I error rates, we estimate the following ordinary least squares (or weighted ordinary least squares in the case of matching) model for each accrual measure in each of our 1000 independent random samples.

$$[ACC\ or\ WC\_ACC]_{it} = \alpha_0 + \beta_1 TREAT_{it} + \beta_2 log\_AT_{it-1} + \beta_3 \Delta REVCash_{it} + \beta_4 PPE_{it} + \beta_5 ROA_{it-1} + \beta_6 Sort\_VAR + \varepsilon_{it}, \tag{2}$$

where *ACC* (*WC_ACC*) is the abnormal or unadjusted total (working capital) accrual variable of interest and *TREAT* is an indicator variable set to one for the 500 pseudo-treatment observations and to zero for the remaining 4500 control observations in each random sample. Remaining controls are as defined above and in detail in Appendix Table 7. This two-stage regression approach follows the typical use of signed abnormal accrual estimates in the literature. The intuition underlying this simulation is that, as long as abnormal accruals are randomly distributed with respect to the subsample of interest, the proportion of t-statistics for the *TREAT* coefficient that are significantly different from zero (i.e., the incidence of type I errors) should approach the confidence level selected, in this case, 5%.[14]

In contrast to our approach, Dechow et al. (1995) and Kothari et al. (2005) compare mean abnormal accruals to zero, using univariate tests for randomly selected samples of firms from upper and lower quartiles of their accrual determinant distributions. While both approaches generate correlations with underlying fundamentals, our approach more directly targets the departures from linearity identified in Fig. 1. In addition,

[12] In the case of the EP_{t-1} sort, we include EP_{t-1} as a control in place of ROA_{t-1}. Results are qualitatively unchanged if we include ROA_{t-1} instead.

[13] While this tolerance for entropy balancing does prevent some random samples out of the 1000 from converging (due to a lack of covariate overlap, collinearity in covariates, or both), the number of unconverged iterations never exceeds 2%.

[14] The 95% confidence interval for binomial tests, using rejection rates computed according to the test statistic from the ordinary least squares-estimated coefficient, ranges from 3.5% to 6.7% for the 1000 independent draws. Rejection rates outside of these bounds are consistent with an abnormal accrual measure that is biased for (rejection rates above 6.7%) or against (rejection rates below 3.5%) the hypothesis of no earnings management. Because samples are randomly drawn from a sample of observations that is pooled across years and industries, clustering concerns are not expected to be present within the simulated data. We verify that clustering by firm and year does not alter results.

our two-stage regression approach (rather than a univariate comparison) allows us to include the *same set* of controls in the second stage for estimates using multivariate matching and linear abnormal accruals measures.

Panels A and B of Table 2 report rejection rates for the null hypothesis that the β_1 coefficient from Eq. (2) is zero for each of our selected accrual determinants (representing the columns in Table 2). Table 2 shows substantial misspecification for ordinary-least-squares estimates, with rejection rates for the test that $\beta_1 = 0$ exceeding 20% in several instances. For example, results for total accruals (*ACC*) in Table 2 show rejection rates exceeding 20% in Panel A (B) for the test $\beta_1 < 0$ ($\beta_1 > 0$) for samples drawn from the lowest decile for BM_{t-1} (middle decile for ROA_{t-1}). When we move to results for entropy-balanced estimates, we find that rejection rates never exceed 11% across all deciles chosen for both *WC_ACC* and *ACC* tests. Of the 32 rejection rates for these models in Panels A and B of Table 2 (16 each for *WC_ACC* and *ACC*), we find that 22 of 32 indicate well-specified models (within the bands suggested for a 95% confidence interval). In contrast, only four of 32 tests for ordinary-least-squares models are well specified.

Moving to results in Table 2 for performance-matched (*PMDAC*) and five-nearest neighbor propensity-score-matching (*PSM [5NN]*) models, we observe that, while these alternative approaches generally improve test specification, relative to ordinary-least-squares models, the improvement falls short of that achieved using entropy balancing. For example, we observe that rejection rates exceed 20% in 2 out of 32 tests in Table 2 for both *PMDAC* and *PSM* estimates. Additionally, only four (12) of 32 tests fall within the bounds indicating a well-specified test for *PMDAC* (*PSM [5NN]*) models. Broadly, these results indicate excess type I errors for linear estimation methods in samples drawn from deciles with departures from linearity and that entropy balancing maintains a significant advantage in reducing type I errors for these samples, relative to other approaches.

Panel C of Table 2 explores whether the improved specification for entropy-balanced models in Panels A and B can be traced to an advantage in reducing coefficient bias or due to less efficient standard error estimates (which would bias against a significant test statistic by reducing power). For ease of presentation, Panel C summarizes mean coefficient bias and standard errors across all the middle versus extreme deciles in Panels A and B. Beginning with ordinary-least-squares estimates, we find a mean absolute coefficient bias for the β_1 coefficient on *TREAT* of 0.42% (0.22%) as a proportion of lagged total assets for *ACC* (*WC_ACC*) models in the middle deciles. Consistent with less extreme departures from linearity for middle deciles, absolute bias is larger for these models in our two extreme deciles for BM_{t-1} and New_FIN_t, with a mean bias of 0.80% (0.45%) for *ACC* (*WC_ACC*) models. We find that entropy balancing substantially reduces the absolute bias in coefficient estimates (by more than 50%), relative to linear models, in both sets of portfolios, with a mean absolute coefficient bias of 0.1% (0.1%–0.3%) in the middle (extreme) decile sorts. In line with results from Panels A and B of Table 2, Panel C shows that while performance matching and propensity-score matching tend to improve on ordinary-least-squares estimates, both *PMDAC* and *PSM* models display only a fraction of the reduction in absolute coefficient bias as entropy balancing. Interestingly, *PSWT* estimates display evidence of larger coefficient bias, relative to ordinary least squares, for the extreme deciles for *ACC* and *WC_ACC* models, suggesting that *PSWT* estimates

Table 2 Specification tests: Rejection rates for the null hypothesis that abnormal accrual estimates are zero

Panel A: Tests for (ACC Measure) < 0 (specification tests for type I errors)[a]								
Sorting variable	**ΔREVCash$_t$**	**PPE$_t$**	**ROA$_{t-1}$**	**AT$_{t-1}$**	**EP$_{t-1}$**	**BM$_{t-1}$**	**BM$_{t-1}$**	**New_FIN$_t$**
Decile	**2**	**6**	**6**	**5**	**5**	**4**	**1**	**10**
Total accruals (ACC):								
EBACC[b]	4.1%	**7.2%**	3.6%	6.5%	4.4%	6.0%	**10.0%**	4.7%
OLS	**15.7%**	***0.7%***	***0.0%***	4.4%	***0.8%***	***2.1%***	**51.3%**	**12.0%**
MJDA	**17.5%**	***1.4%***	***0.0%***	***1.1%***	***0.1%***	***0.3%***	**51.2%**	**8.7%**
PMDAC	**10.8%**	***1.5%***	***0.7%***	***2.4%***	**7.3%**	***2.4%***	**28.3%**	**8.8%**
PSM (1 NN)	5.7%	***2.8%***	***0.2%***	3.7%	3.6%	5.2%	**14.8%**	3.9%
PSM (5 NN)	5.1%	***2.3%***	***0.0%***	3.8%	***2.7%***	4.5%	**17.5%**	3.9%
PSWT	**10.0%**	***0.9%***	***0.0%***	6.0%	***1.7%***	***2.9%***	**34.6%**	5.8%
Working capital accruals (WC_ACC):								
EBACC[b]	3.8%	4.9%	4.3%	***2.7%***	***2.9%***	5.5%	5.4%	***2.9%***
OLS	***2.5%***	***0.3%***	***1.5%***	***0.7%***	***2.3%***	***1.0%***	**16.2%**	***2.9%***
MJDA	***2.7%***	***1.3%***	***1.8%***	***0.3%***	***0.7%***	***0.9%***	**14.1%**	***2.0%***
PMDAC	***2.3%***	***1.0%***	***1.0%***	***0.7%***	***2.3%***	5.1%	**8.9%**	***1.6%***
PSM (1 NN)	3.8%	***1.0%***	6.3%	***1.5%***	***2.7%***	6.2%	4.5%	***1.8%***
PSM (5 NN)	***2.1%***	***0.9%***	**8.5%**	***1.1%***	3.5%	5.4%	5.2%	***1.4%***
PSWT	***3.0%***	***0.6%***	4.6%	***2.3%***	4.8%	***2.3%***	4.1%	***1.0%***
Panel B: Tests for (ACC Measure) > 0 (specification tests for type I errors)[a]								
Sorting variable	**ΔREVCash$_t$**	**PPE$_t$**	**ROA$_{t-1}$**	**AT$_{t-1}$**	**EP$_{t-1}$**	**BM$_{t-1}$**	**BM$_{t-1}$**	**New_FIN$_t$**
Decile	**2**	**6**	**6**	**5**	**5**	**4**	**1**	**10**
Total accruals (ACC):								
EBACC[b]	6.3%	***3.3%***	5.7%	4.3%	5.7%	***3.4%***	***2.9%***	4.4%
OLS	***1.3%***	**13.0%**	**39.4%**	***3.4%***	**13.8%**	3.6%	***1.3%***	**13.9%**
DAC	***1.1%***	**8.3%**	4.0%	**8.3%**	**42.3%**	**10.5%**	***0.9%***	**16.9%**
PMDAC	***2.8%***	**8.5%**	***1.1%***	**8.6%**	***0.7%***	6.1%	***3.0%***	**14.3%**
PSM (1 NN)	6.0%	**9.9%**	**29.4%**	6.5%	6.2%	4.8%	***1.9%***	**7.6%**
PSM (5 NN)	5.8%	**11.1%**	**43.1%**	**6.8%**	**8.9%**	3.9%	***1.1%***	6.6%
PSWT	***2.2%***	**16.4%**	**56.6%**	**7.3%**	**14.5%**	**9.0%**	***0.2%***	**7.4%**
Working capital accruals (WC_ACC):								
EBACC[b]	6.3%	6.5%	6.4%	**10.5%**	**7.7%**	4.9%	6.5%	5.8%
OLS	5.4%	**26.9%**	***1.0%***	**8.5%**	5.4%	**7.2%**	**13.4%**	**35.3%**
WC_DAC	3.9%	**14.2%**	***0.8%***	**12.3%**	**9.2%**	**9.1%**	**14.6%**	**38.4%**
WC_PMDAC	**10.2%**	**16.0%**	***0.9%***	**11.8%**	3.8%	3.6%	**15.2%**	**38.3%**
PSM (1 NN)	**8.1%**	**17.9%**	***3.4%***	**12.9%**	6.0%	4.6%	5.6%	**11.0%**
PSM (5 NN)	**7.8%**	**24.3%**	***3.0%***	**16.5%**	5.9%	***3.2%***	5.0%	**12.6%**
PSWT	6.3%	**31.9%**	4.5%	**14.5%**	5.0%	**10.3%**	**8.2%**	**25.8%**

Panel C: Mean coefficient bias and standard errors from specification tests				
Statistic	**Mean coefficient bias**		**Mean Standard Errors**	
Deciles	**2–9**	**1 and 10**	**2–9**	**1 and 10**
Total accruals (ACC):				
EBACC[b]	0.10%	0.30%	0.007	0.018
OLS	0.42%	0.80%	0.007	0.009
DAC	0.47%	0.90%	0.007	0.009
PMDAC	0.27%	0.60%	0.010	0.012
PSM (1 NN)	0.23%	0.75%	0.009	0.021

Table 2 (continued)

PSM (5 NN)	0.23%	0.70%	0.007	0.017
PSWT	0.42%	0.85%	0.007	0.014
Working capital accruals (WC_ACC):				
EBACC[b]	0.08%	0.10%	0.006	0.016
OLS	0.22%	0.45%	0.006	0.007
WC_DAC	0.20%	0.45%	0.006	0.007
WC_PMDAC	0.23%	0.70%	0.008	0.010
PSM (1 NN)	0.23%	0.35%	0.007	0.018
PSM (5 NN)	0.22%	0.40%	0.006	0.015
PSWT	0.22%	0.75%	0.006	0.013

Table 2 compares type I error rates for alternative abnormal accrual measures. The table reports results of 1000 independent random draws of 5000 firm-years each, without replacement. Of the 5000 firm-years selected in each draw, 500 observations are randomly selected as treatment observations from selected deciles with pronounced departures from a linear relation between accruals and the underlying sorting variable (identified from a review of Fig. 1). We then select 4500 control observations from the remaining pooled firm-year sample of 159,520 observations with data available for accrual determinants including book-to-market (BM_{t-1}) and earnings-to-price (EP_{t-1}) ratios. Panels A and B present results of significance tests from regressions of each accrual measure on an indicator for the 500 treatment observations (*TREAT*), along with controls for firm size (log_AT_{t-1}), capital intensity (PPE_t), changes in noncredit sales ($\Delta REVCash_t$), prior-year performance (ROA_{t-1} or EP_{t-1} for the *EP* sorts), and the financial sorting variable used to select treatment observations (*Sort_Var*) in cases where the *Sort_Var* is not one of the other control variables. Panel A (B) presents the percentage of t-statistics from the treatment indicator that are rejected at the 5% one-tailed level as significantly below (above) zero. Rejection rates are calculated for the following dependent variables: entropy-balanced *ACC* or *WC_ACC* (*EBACC*), where the control sample of 4500 firms is weighted to match the distribution of the 500 treatment firms; unadjusted *ACC* or *WC_ACC* (*OLS*); modified-Jones-model discretionary accruals (*DAC*, *WC_DAC*); performance-matched discretionary accruals (*PMDAC*, *WC_PMDAC*); propensity-score matched *ACC* and *WC_ACC* (*PSM*), using one-nearest neighbor (*1 NN*) and five-nearest neighbors (*5 NN*); and propensity-score weighted *ACC* and *WC_ACC* (*PSWT*), where the weight for the 500 treated (4500 control) observations is 1/Prop. Score (1/[1-Prop. Score]). Appendix Table 7 provides details of variable measurement and data sources

[a] Figures in bold (bold italic) signify rejection rates that significantly exceed (fall below) the 5% nominal significance level of the test, using a binomial probability at the 1% level of significance for 1000 independent draws. Rejection rates outside of this band indicate that such tests are biased against (in favor of) the null hypothesis of no earnings management within the relevant subsample

[b] Out of 1000 simulated treatment and control samples, between 0 and 20 samples (i.e., less than 2%) do not permit entropy balancing to converge for the BM_{t-1} sorts. This suggests that a handful of iterations display significant covariate imbalance, largely as a result of balancing on covariates that are correlated (i.e., including both BM_{t-1} and ROA_{t-1} in the model) in the entropy-balancing program. As a result, rejection rates are calculated assuming the number of converged iterations for these sorts rather than 1000

Panel C of Table 2 presents mean coefficient bias (as a percent of lagged total assets) and standard errors for the coefficient on the indicator variable measuring the 500 randomly selected treatment firms (*TREAT*) averaged over the 1000 independent random draws of 5000 firm-years each used to test the specification of each accrual model, where a well-specified model will display a mean coefficient estimate approaching zero. Treatment firms are selected from the shaded deciles indicated in Fig. 1 with substantial departures from linearity for each accrual determinant of interest. Mean coefficients and standard errors are calculated for models using the following dependent variables: entropy-balanced *ACC* or *WC_ACC* (*EBACC*), unadjusted *ACC* or *WC_ACC* (*OLS*), modified-Jones-model discretionary accruals (*DAC, WC_DAC*), performance-matched discretionary accruals (*PMDAC, WC_PMDAC*), propensity-score matched *ACC* and *WC_ACC* (*PSM*) using one-nearest neighbor (*1 NN*) and five-nearest neighbors (*5 NN*), and propensity-score weighted *ACC* and *WC_ACC* (*PSWT*), where the weight for the 500 treated (4500 control) observations is 1/Prop. Score (1/[1-Prop. Score]). Appendix Table 7 provides details of variable measurement and data sources tend to exacerbate bias in samples with extreme observations. In contrast to the improvement in coefficient bias, results for mean standard errors presented in the rightmost columns of Panel C show that entropy balancing displays no evidence of smaller standard errors, relative to ordinary-least-squares models. While entropy-balanced estimates do display slightly smaller standard errors relative to *PSM (1 NN)*, this advantage dissipates when moving to *PSM (5 NN)*. Overall, we conclude that entropy balancing's improved specification is due to eliminating coefficient bias stemming from underlying nonlinearities.

5.3 Simulation results—Covariate balance

In Table 3, we assess two issues. First, we examine whether entropy balancing improves the specification of accrual models by assigning extreme weights, emphasizing potential outliers in the weighted regression for accruals, or both (see Section 3.3). Second, we confirm that entropy balancing enhances covariate balance on the set of determinants included in the second-stage regression in each simulation. Table 3 presents mean *Weight Ratios*, maximum weighted regression leverages, and covariate balance for first (standardized differences) and second (variance ratios) moments of the distribution of each covariate across the 1000 random samples drawn for the decile of interest. Rather than present these balance statistics for all eight determinants/deciles selected from Fig. 1, we examine a representative extreme (decile 1 of BM_{t-1}) and middle (decile 4 of BM_{t-1}) decile for selecting the treatment sample.

Table 3 confirms that entropy balancing achieves greater covariate balance, relative to alternative approaches. Both standardized differences and variance ratios in Panels A and B show balanced covariates in our entropy-balanced models, while ordinary-least-squares and propensity-score approaches indicate imbalanced covariates in both means and variances in multiple instances. Turning to results for the *Weight Ratio,* Table 3 shows that, to achieve covariate balance, entropy balancing does emphasize a smaller number of control sample observations relative to the other approaches. Despite this, the *Weight Ratio* for entropy balancing in both sorts is above 1.0 (and exceeds the ratio for the *PSWT* approach in the lowest decile for the BM_{t-1} sort), indicating that entropy balancing continues to place weight on more control observations than a one-to-one match. Turning to regression leverages, the *Max_Lev* column shows that entropy balancing also appears to emphasize potential outlying observations to achieve covariate balance in the BM_{t-1} sort in Panel A, with a mean *Max_Lev* value of 0.068 (0.020) for entropy-balanced (ordinary-least-squares) estimates. In contrast, entropy balancing displays a lower *Max_Lev* in Panel B (for the BM_{t-1} sort using a middle decile), relative to ordinary least squares, with a value of 0.016 for the former and 0.024 for the latter. This is consistent with entropy balancing applying fewer large weights to the broader pool of control sample observations for treatments selected from the middle deciles, helping to reduce the influence of outlying observations in these settings. This suggests that, while entropy balancing has the largest potential for reducing coefficient bias for treatment samples drawn from the extremes of the distribution of a determinant (see Table 2, Panel C), it also suffers the potential for sensitivity to outlying observations in these cases. As a result, we encourage researchers to evaluate (and present) statistics similar to the *Weight Ratio* and *Max_Lev* when using entropy balancing in other research settings.

Table 3 Mean covariate balance from specification simulations

Method	Weight Ratio	Max_Lev	Standardized differences					Variance ratios				
			$\log(AT_{t-1})$	$\Delta REVCash_t$	PPE_t	ROA_{t-1}	BM_{t-1}	$\log(AT_{t-1})$	$\Delta REVCash_t$	PPE_t	ROA_{t-1}	BM_{t-1}
Panel A: Treatment firms selected from decile 1 (lowest) of *BM_{t-1}*												
EBACC	1.660	0.068	0.000	0.000	0.000	−0.001	−0.001	0.999	0.999	0.998	0.999	0.992
OLS	1.000	0.020	**−0.672**	−0.031	−0.006	**−0.735**	**−1.184**	1.198	**2.094**	**1.307**	**3.600**	**1.945**
PSM (5 NN)	2.099	0.032	**0.276**	−0.069	0.072	**0.208**	**−0.185**	**1.573**	0.984	1.013	0.931	0.848
PSWT	0.730	0.073	0.040	0.042	−0.030	−0.007	**−0.337**	1.144	**1.257**	0.966	0.865	**0.178**
Panel B: Treatment firms selected from decile 4 of *BM_{t-1}*												
EBACC	2.398	0.016	0.001	0.000	0.000	0.001	0.000	0.998	0.998	0.998	0.994	0.958
OLS	1.000	0.024	**0.175**	**0.126**	−0.062	**0.210**	**−0.402**	0.937	0.868	0.932	**0.539**	**0.006**
PSM (5 NN)	3.447	0.015	**−0.114**	−0.028	0.038	−0.087	**−0.157**	1.000	0.942	1.194	**1.428**	**0.018**
PSWT	4.138	0.088	−0.041	0.013	−0.004	**−0.129**	**−0.348**	0.988	0.815	1.097	**1.490**	**0.006**

Table 3 presents results of tests for covariate balance averaged across 1000 independent random draws of 5000 firms each: 4500 control firms selected from the pooled sample of firm-years with data available to measure total and working capital accruals along with control variables (see Table 2 for details) and 500 treatment firms selected from the lowest decile (Panel A) and the fourth decile (Panel B) of lagged book-to-market, *BM_{t-1}*, respectively. We present means across the 1000 draws for the following covariate balance statistics: the *Weight Ratio* calculated as the (number of firms with a regression weight greater than equal weight) / (number of treatment firms), the *Max_Lev* calculated as the maximum value of the regression leverage measured for the 5000 randomly selected observations in each draw multiplied by the weight assigned to each observation (i.e., one for ordinary least squares or the weight selected by the entropy-balancing, propensity-score matching, or propensity-score weighting algorithm), *Standardized differences* calculated as the difference in means between treatment and control samples divided by the standard deviation of the treatment sample for each covariate (with values near zero when the distribution for a particular covariate is more similar between treatment and control samples), and *Variance ratios* calculated as the ratio of the variance of each covariate in the treatment sample scaled by variance for the control sample. *Standardized differences* (*Variance ratios*) in bold indicate mean values outside the ± 0.1 (4/5 and 5/4) cutoffs for a balanced covariate indicated by Rubin (2001) and Austin (2011). Appendix Table 7 provides details of variable measurement and data sources

5.4 Simulation results—Type II error rates

Results in Panel C of Table 2 raise the concern that both propensity-score matching and entropy balancing sacrifice test power (i.e., increase type II errors) to improve specification (i.e., decrease type I errors) in signed abnormal accrual tests. To address this issue, our second simulation seeds accrual manipulations, following a similar approach to prior research (Dechow et al. 1995; Kothari et al. 2005). These tests follow Table 2 by seeding accrual manipulations in the treatment samples selected from the deciles with departures from linearity (see Fig. 1) to directly assess whether entropy balancing's improved test specification impairs detection rates for accrual manipulations. For each randomly selected sample, we seed accrual manipulations by adding 1% and 2% of beginning total assets to accruals for the 500 observations in the "pseudo-treatment" group. This seeded management is consistent with either a manipulation that decreases expenses, that targets credit sales to increase bottom-line income, or both (Dechow et al. 1995). Because tests control for the change in cash revenue, these manipulations will only affect current-year accruals in our models. The remaining 4500 observations from each randomly selected sample of firm-years continue to act as a control sample in each iteration.

Panel A (B) of Table 4 displays rejection rates for the null hypothesis of no accrual management by counting coefficients in Eq. (2) that are significant at a 5% one-tailed significance level for the *TREAT* indicator set to one for observations with seeded accrual manipulations at 1% (2%) of lagged total assets for total (*ACC*) and working capital (*WC_ACC*) accruals. Beginning with ordinary-least-squares models as a benchmark, Table 4 shows that detection rates for these models are largely determined by the baseline level of misspecification. For example, rejection rates in Table 2 show that these models for total accruals (*ACC*) are significantly biased in favor of (against) concluding that $\beta_1 > 0$ in the sixth decile of ROA_{t-1} (lowest decile of BM_{t-1}). Table 4, Panel A, shows that these two instances display the highest (lowest) rejection rate in the presence of seeded accrual manipulations for ordinary-least-squares models using *ACC* as the dependent variable, respectively.

Results for entropy-balanced models show that detection rates in sorts that are well specified for both ordinary-least-squares and entropy-balancing methods display similar levels of test power. For example, in the $\Delta REVCash_t$ sort, both models display rejection rates near to the 5% threshold for *WC_ACC* models (Table 2, Panels A and B), with corresponding detection rates in Table 4 showing that entropy-balanced (ordinary-least-squares) models correctly reject the null hypothesis that *WC_ACC* are zero in 52.6% (61.5%) of the iterations with 1% manipulations. In contrast, for the $\Delta REVCash_t$ portfolio sort using total accruals (*ACC*) where ordinary-least-squares models display downward biased estimates before the seeded manipulations (while entropy-balanced models are well specified), *ACC* estimates display rejection rates for entropy-balanced (ordinary-least-squares) models of 36.7% (15.4%) at a 1% manipulation level. This shows that the baseline specification of the model directly drives the detection rate. In this sense, entropy balancing's improvement in specification aids model detection rates. This inference is confirmed in the narrow dispersion of estimated coefficients for entropy-balanced models around the 1% and 2% manipulation levels for nearly every sort for both *ACC* and *WC_ACC* models, relative to *OLS*, *PMDAC*, and *PSM (5NN)* models (untabulated for brevity). In short, entropy balancing markedly lowers type I error rates by reducing coefficient bias, while displaying sufficiently low type II error rates when compared to alternatives.

Table 4 Detection rates (type II errors) for seeded accrual manipulations with firm performance included as a control

Sorting variable	$\Delta REVCash_t$	PPE_t	ROA_{t-1}	AT_{t-1}	EP_{t-1}	BM_{t-1}	BM_{t-1}	New_FIN_t
Decile	2	6	6	5	5	4	1	10
Panel A: Tests for (ACC Measure) > 0 with 1% seeded accrual manipulations								
Total accruals (ACC):								
EBACC[b]	36.7%	30.4%	62.6%	36.5%	41.5%	33.9%	9.2%	14.2%
OLS	15.4%	68.3%	95.6%	36.8%	68.5%	42.1%	5.8%	37.6%
DAC	14.5%	58.0%	63.0%	59.5%	91.6%	63.0%	6.0%	45.1%
PMDAC	17.8%	39.9%	19.2%	35.5%	10.5%	30.8%	11.2%	31.9%
PSM (1 NN)	24.2%	39.1%	83.0%	34.2%	37.2%	28.9%	4.6%	16.4%
PSM (5 NN)	31.3%	57.7%	95.4%	45.4%	53.4%	41.4%	4.3%	17.9%
PSWT	23.6%	66.2%	95.9%	38.1%	66.7%	45.0%	3.0%	20.9%
Working capital accruals (WC_ACC):								
EBACC[b]	52.6%	48.5%	72.8%	71.1%	57.2%	52.8%	15.4%	21.4%
OLS	61.5%	89.8%	51.6%	73.8%	61.6%	69.0%	40.1%	71.1%
WC_DAC	60.0%	78.5%	46.3%	80.0%	74.8%	74.1%	44.3%	74.3%
WC_PMDAC	48.3%	62.5%	30.2%	55.3%	37.2%	32.0%	37.4%	66.8%
PSM (1 NN)	41.4%	67.9%	45.3%	58.5%	41.0%	36.6%	13.7%	25.9%
PSM (5 NN)	57.0%	83.6%	63.9%	79.2%	56.8%	52.7%	18.2%	32.0%
PSWT	63.5%	88.2%	58.9%	73.1%	59.2%	59.0%	31.2%	51.2%
Panel B: Tests for (ACC Measure) > 0 with 2% seeded accrual manipulations								
Total accruals (ACC):								
EBACC[b]	78.9%	76.1%	98.1%	86.2%	87.3%	85.6%	19.0%	32.0%
OLS	69.8%	97.1%	100.0%	88.4%	97.4%	91.3%	22.6%	68.1%
DAC	64.5%	96.3%	98.9%	96.8%	99.6%	97.9%	22.1%	74.2%
PMDAC	64.5%	96.3%	98.9%	96.8%	99.6%	97.9%	22.1%	74.2%
PSM (1 NN)	59.6%	80.7%	99.1%	73.6%	75.4%	70.1%	12.1%	30.6%
PSM (5 NN)	75.7%	92.7%	100.0%	90.4%	92.0%	88.9%	13.2%	38.4%

Table 4 (continued)

Sorting variable	$\Delta REVCash_t$	PPE_t	ROA_{t-1}	AT_{t-1}	EP_{t-1}	BM_{t-1}	BM_{t-1}	New_FIN_t
Decile	2	6	6	5	5	4	1	10
PSWT	73.3%	96.2%	99.9%	82.3%	96.2%	82.0%	11.0%	41.4%
Working capital accruals (WC_ACC):								
EBACC[b]	95.2%	92.3%	99.8%	98.7%	96.4%	96.6%	32.1%	44.2%
OLS	98.0%	99.9%	99.2%	99.6%	98.7%	99.2%	73.8%	93.2%
WC_DAC	97.9%	99.2%	99.3%	99.9%	99.4%	99.5%	77.2%	95.3%
WC_PMDAC	97.9%	99.2%	99.3%	99.9%	99.4%	99.5%	77.2%	95.3%
PSM (1 NN)	84.8%	95.9%	95.4%	94.9%	85.0%	82.7%	31.4%	45.8%
PSM (5 NN)	95.5%	98.9%	99.3%	99.4%	96.9%	96.4%	39.9%	59.1%
PSWT	98.0%	99.5%	98.4%	98.3%	96.9%	95.0%	61.4%	76.9%

Table 4 presents detection rates from simulations that run ordinary least squares (OLS) and weighted OLS regressions to detect seeded accrual manipulations for a randomly selected set of 500 treatment firms that add 1% (Panel A) and 2% (Panel B) of beginning total assets to current-year total (ACC_t) and working capital (WC_ACC_t) accruals. For each iteration of the simulation, we run a pooled regression using 5000 observations, where 500 treatment observations selected from the decile of each sorting variable possess seeded upward earnings management ($TREAT = 1$) and 4500 firms selected from the remaining pooled sample contain no seeded accruals ($TREAT = 0$) to act as a control sample: $(ACC\ Measure) = \alpha + \beta_1 * TREAT + \Sigma\ (\beta * Controls) + \varepsilon$. Each panel reports the frequency that β_1 is positive and significant (i.e., frequency $\beta_1 > 0$) at the 5% level, using a one-tailed t-test across the 1000 random draws. Remaining control variables comprise prior-year firm performance (ROA_{t-1} or EP_{t-1} for the *EP* sorts), size (log_AT_{t-1}), capital intensity (PPE_t), changes in noncredit sales ($\Delta REVCash_t$), and the financial sorting variable used to select the treatment firms (*Sort_Var*) in cases where the *Sort_Var* is not one of the other control variables. Rejection rates are calculated for the following dependent variables: entropy-balanced *ACC* or *WC_ACC* (*EBACC*), where the control sample of 4500 firms is weighted to match the distribution of the 500 treatment firms; unadjusted *ACC* or *WC_ACC* (*OLS*); modified-Jones-model discretionary accruals (*DAC*, *WC_DAC*); performance-matched accruals (*PMDAC*, *WC_PMDAC*); propensity-score matched *ACC* and *WC_ACC* (*PSM*), using one-nearest neighbor (*1 NN*) and five-nearest neighbors (*5 NN*); and propensity-score weighted *ACC* and *WC_ACC* (*PSWT*), where the weight for the 500 treated (4500 control) observations is 1/Prop. Score (1/[1-Prop. Score]). Appendix Table 7 provides details of variable measurement and data sources

[a] Note: For the *DAC* and *PMDAC* measures (both total and working capital accrual measures), coefficients for the normal accrual model and the performance-matched peer are estimated after the seeded accrual management, which will increase current-year accruals for those firms with seeded upward management

[b] Out of 1000 simulated treatment and control samples, between 0 and 20 samples (i.e., less than 2%) do not permit entropy balancing to converge for the BM_{t-1} sorts. This suggests that a handful of iterations display significant covariate imbalance, largely as a result of balancing on covariates that are correlated (i.e., including both BM_{t-1} and ROA_{t-1} in the model) in the entropy balancing program. As a result, rejection rates are calculated assuming the number of converged iterations for these sorts rather than 1000

5.5 Empirical setting—Accrual earnings management around new equity issuances

To assess the empirical significance of our simulations, we examine accruals around new equity issuances. Studies (e.g., Loughran and Ritter 1997) note that issuers experience unusually strong operating performance, relative to non-issuers. Consistent with this, Fig. 1 shows a significant departure from a linear model for accruals for firms in the top decile of net new financing. In accounting, several studies document evidence of upward accrual management around IPOs (Aharony et al. 1993; Friedlan 1994; Teoh et al. 1998a) and SEOs (Rangan 1998; Teoh et al. 1998b; Shivakumar 2000; DuCharme et al. 2004; Cohen and Zarowin 2010). However, more recent research questions whether IPO firms' accrual processes are accurately described by accrual models. Ball and Shivakumar (2008) note that abnormal accrual estimates predict upward earnings management in excess of 100% of total assets for a number of IPOs and that coefficients for industry-year ordinary-least-squares models for a normal accrual significantly differ across samples of issuing and non-issuing firms. In line with this idea, Armstrong et al. (2015) show that abnormal accruals are attributable to the planned investment of IPO proceeds in working capital rather than to intentional misreporting.

We focus our empirical tests on comparing accruals across the forms of equity issuance (e.g., IPO, SEO, or private placement) by including the amount of net new financing (New_FIN_t) as a covariate in our matching algorithms to compare firms with similar levels of net capital raising but where the form of financing differs. This avoids combining IPO and SEO firms into one treatment sample, in line with the assumption of no hidden variation in treatment (Imbens and Rubin 2015). That is, if earnings management occurs at different rates across IPOs and SEOs, forming a treatment sample that includes both groups of firms results in a biased abnormal accrual estimate for both forms of financing. On this point, we expect regulatory scrutiny and underwriter reputation concerns to be more pronounced for IPOs, relative to other forms of raising capital, thereby limiting the scope for managing accruals. In contrast, firms pursuing SEOs should experience less scrutiny, as they are seasoned issuers. Similarly, underwriter reputation concerns are more limited, given the lower degree of adverse selection present for these issues, relative to an IPO.

In Table 5, we employ entropy balancing and propensity-score matching to weight observations in the sample of firms without IPOs (i.e., the control group), using the set of covariates we examine in our simulations along with a control for New_FIN_t. For entropy balancing, we balance on the first three moments (mean, variance, and skewness), using the default tolerance of 0.015. In implementing propensity-score matching, we employ the following Logit model to estimate propensity scores.

$$IPO(=1)_{it} = \alpha + \beta_1 New\_FIN_{it} + \beta_2 \Delta REVCash_{it} + \beta_3 PPE_{it} + \beta_4 ROA_{it-1} + \beta_5 log\_AT_{it-1} + \epsilon_{it} \tag{3}$$

Figure 2 graphically displays covariate balance before and after multivariate matching using both entropy balancing and five-nearest neighbor propensity-score matching. Before matching, we find several covariates with standardized differences

Table 5 Initial public offering (IPO) tests for accrual reversals

Method	EB		PSM (1NN)		PSM (5NN)		PMDAC	OLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Tests for abnormal accruals and reversals using total accruals (*ACC*) around IPOs								
b, IPO (t)	0.016	−0.003	0.018	0.009	0.018	0.007	0.010	0.019*
	(1.062)	(−0.192)	(1.264)	(0.883)	(1.175)	(0.721)	(0.961)	(1.741)
c, IPO (t-1)		0.008		−0.000		0.001		0.005
		(0.445)		(−0.053)		(0.133)		(0.702)
New_FIN_t		−0.017***		−0.007		−0.004		−0.007
		(−4.395)		(−1.149)		(−0.650)		(−1.430)
$\Delta REVCash_t$		0.058***		0.053***		0.049***		0.042***
		(5.476)		(7.239)		(8.730)		(9.797)
PPE_t		−0.105***		−0.086***		−0.076***		−0.040***
		(−11.382)		(−8.202)		(−9.379)		(−13.550)
ROA_{t-1}		0.141***		0.148***		0.141***		0.133***
		(12.327)		(18.978)		(23.061)		(34.211)
log_AT_{t-1}		−0.015***		−0.008***		−0.003***		0.001***
		(−6.860)		(−4.775)		(−2.797)		(3.132)
Reversal (b - c)		−0.011		0.010		0.007		0.015
χ^2 reversal test statistic		0.718		1.443		0.675		2.286
χ^2 *p*-value		0.404		0.239		0.418		0.141
Sector and Year FE?	No	Yes	No	Yes	No	Yes	No	Yes
Observations	185,166	185,166	6,796	12,537	14,110	24,886	185,166	185,166
Adjusted R^2	0.001	0.209	0.001	0.181	0.001	0.156	0.000	0.132
Panel B: Tests for abnormal accruals and reversals using working capital accruals (*WC_ACC*) around IPOs								
b, IPO (t)	0.014	0.004	0.023**	0.025***	0.020**	0.018**	0.039***	0.023***
	(1.630)	(0.524)	(2.445)	(2.969)	(2.237)	(2.513)	(5.160)	(2.949)
c, IPO (t-1)		0.005		0.007*		0.006**		0.014***
		(0.345)		(1.859)		(2.283)		(4.567)

Table 5 (continued)

Method	EB		PSM (1NN)		PSM (5NN)		PMDAC	OLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
New_FIN_t		0.006		0.014***		0.017***		0.021***
		(1.641)		(3.167)		(4.320)		(7.091)
$\Delta REVCash_t$		0.087***		0.085***		0.082***		0.076***
		(10.413)		(10.919)		(15.044)		(15.603)
PPE_t		−0.051***		−0.047***		−0.032***		−0.006***
		(−5.974)		(−6.653)		(−5.329)		(−4.104)
ROA_{t-1}		0.045***		0.038***		0.036***		0.032***
		(6.432)		(4.427)		(5.161)		(7.385)
log_AT_{t-1}		−0.016***		−0.007***		−0.004***		−0.000
		(−7.899)		(−4.902)		(−4.286)		(−1.417)
Reversal (b - c)		−0.001		0.017***		0.012*		0.010
χ^2 reversal test statistic		0.009		8.075		3.496		1.449
χ^2 p-value		0.926		0.008		0.071		0.238
Sector and Year FE?	No	Yes	No	Yes	No	Yes	No	Yes
Observations	185,166	185,166	6,796	12,537	14,110	24,886	185,166	185,166
Adjusted R^2	0.001	0.143	0.002	0.129	0.001	0.109	0.001	0.068

This table presents results of tests using firms completing an initial public offering (IPO) as the treatment variable for the full sample of firm-year observations over the 1987–2017 sample period. Panel A (B) presents regressions using total, *ACC* (working capital, *WC_ACC*), accruals as the dependent variable. Results of weighted ordinary least squares (WOLS) regressions are presented in models (1)–(2), using weights specified by the entropy balancing (*EB*) program used to achieve covariate balance. Refer to Fig. 2 for a graphical representation of covariate balance for IPOs before and after EB. The *Weight Ratio* for EB in this setting is 6.043. Results of ordinary least squares (*OLS*) regressions are presented in models (3)–(8). Models (3)–(6) use propensity-score matching (*PSM*), where models (3)–(4) employ one-nearest neighbor without replacement (*1NN*) and models (5)–(6) employ five-nearest neighbor (*5NN*) approaches to identify a matched control sample. PSM uses a Logit model and a caliper of 0.03. Model (7) uses performance matching based on the closest ROA_{t-1} peer within the same two-digit SIC industry and year (*PMDAC*) following the matching procedure recommended by Kothari et al. (2005). Chi-squared tests for the significance of accrual reversals following Dechow et al. (2012) are included below regression coefficients with corresponding *p* values computed using heteroskedasticity-consistent standard errors. See Appendix Table 7 for detailed variable definitions. Separate indicators for sector and fiscal year are included in even-numbered models. t-statistics in parentheses and corresponding *p* values are calculated using standard errors clustered by year. *, **, *** indicate *p* values that are significant at the 10%, 5%, and 1% levels, respectively

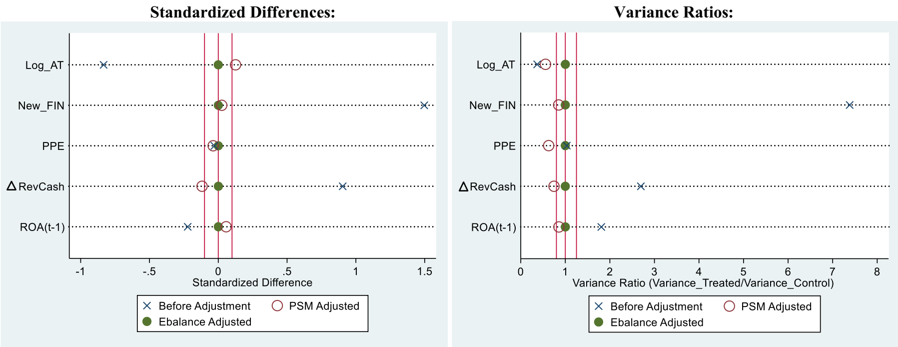


Fig. 2 Standardized differences before and after propensity-score matching (PSM) and entropy balancing (EB) using IPO firms as the treatment sample. This figure displays standardized differences and variance ratios for each covariate before and after running the five-nearest neighbor propensity-score matching (PSM) program detailed in Eq. 3 and the entropy balancing (EB) program, using firms with an initial public offering (IPO) in year t as the treatment sample. All remaining firm-year observations in the 1987–2017 sample period are used to identify the weighted control sample for EB, while PSM includes only the propensity-matched control observations from this group. Standardized differences are calculated as the difference in means between treatment and control samples divided by the standard deviation of the treatment sample for each covariate. The standardized difference will approach zero when the distribution for a particular covariate is more similar between treatment and control samples. Variance ratios for each covariate are calculated as the covariate's variance for the treatment sample scaled by variance for the control sample. Standardized differences and variance ratios outside of the vertical bands in each panel indicate differences that are significant following the guidance of Rubin (2001) and Austin (2011). See Appendix Table 7 for detailed variable definitions

and variance ratios indicating imbalance. The greatest imbalance is concentrated in New_FIN_t, with a standardized difference (variance ratio) of 1.5 (7.4), while log_AT_{t-1}, $\Delta REVCash_t$, and ROA_{t-1} also display significant imbalance on both the first and second moments. After entropy balancing, distributions of each of these variables appear indistinguishable across treated and weighted control samples, with the solid dots in Fig. 2 lying on the vertical axes, indicating balance in each panel. While propensity-score matching generally improves balance on the first moment across treatment and control samples, standardized differences remain outside of the ± 0.1 bounds for log_AT_{t-1} and $\Delta REVCash_t$. In addition, propensity-score matching actually worsens the variance ratio for PPE_t, moving this outside of the bounds suggested for a balanced covariate.

Given the combination of a relatively small number of treated observations (3408 IPO firms) and the significant covariate imbalance prior to matching, we assess whether entropy balancing assigns unusually large weights to a small number of control observations by calculating the *Weight Ratio* (see Section 3.3). We find a *Weight Ratio* of 6.04, consistent with entropy balancing including more than six times as many control firms as a one-to-one matching approach. This indicates limited concern that a few firms in the control sample are assigned extreme weights.

Table 5, Panel A (B) displays entropy-balancing-weighted, propensity-score-matching-weighted, and ordinary-least-squares regressions of total (working capital) accruals on an indicator for the IPO year, *b*, and an indicator for the year following an IPO, *c*, to capture accrual reversals. For the propensity-score matching tests, we follow the design of Dechow et al. (2012) to include the propensity-matched observation for each IPO firm in year t and in the subsequent year $t+1$ (if available) for both treatment and control observations to test for accrual reversals. Results in model (1) for a univariate entropy-balanced model show that accruals are insignificantly different from zero during the IPO year for both total (ACC_t) and working capital (WC_ACC_t) accruals with a b coefficient estimate in Panel A (B) of 0.016 (0.014). Turning to the multiple regression results in model (2) shows that both b and c coefficient estimates are statistically and economically insignificant for ACC and WC_ACC estimates (i.e., coefficients <0.01 in absolute value). As a result, accrual reversal tests indicate an insignificant reversal. Broadly, these results show that IPO firms' accruals resemble those of non-IPO firms after entropy balancing.

Turning to results in models (3)–(6) shows that propensity-score-matching estimates display insignificant (significantly positive) abnormal total (working capital) accruals in the year of the IPO in Panel A (B). In particular, results in Panel A show insignificant b coefficients on the IPO_t indicator that are largely in line with those for entropy-balanced estimates in models (1) and (2). In contrast, results in Panel B for working capital accruals show significantly positive b coefficients ranging between 0.018 and 0.025 in models (3)–(6) (two-tailed p values <0.05). Further, accrual reversal tests in models (4) and (6) show significant accrual reversals for IPO-year accruals using one-nearest (two-tailed p value <0.01) and five-nearest (two-tailed p value <0.10) neighbor approaches, respectively. Results for unweighted ordinary-least-squares estimates in models (7) and (8) in both Panels A and B resemble results using propensity-score matching. Performance-matched univariate estimates in model (7) show that the $PMDAC$ model indicates insignificant (significantly positive) abnormal total (working capital) accruals in the IPO year in Panel A (B), while ordinary-least-squares estimates in model (8) display coefficient estimates largely similar to those in models (4) and (6) for propensity-score matching. Despite this, reversal tests are insignificant in model (8) in both panels.

Overall, our results in Table 5 align with results from our simulations in Table 2. In particular, we observe similar results across models using total accruals, as these models are generally well specified in Table 2 for the top decile of new financing, but differing results for entropy-balanced and existing models when using working capital accruals, as existing models (including propensity-score matching) display pronounced misspecification in the form of an upward bias in WC_ACC coefficient estimates in the top decile of new financing.

In an untabulated analysis (available upon request), we also test for accrual reversals around seasoned equity offerings by assigning SEO firms as the treatment sample and remaining firms as the control group (excluding firms with an IPO in the current or prior year to avoid mechanical overlap with results in Table 5). Similar to Fig. 2, SEO firms display significant covariate imbalance in New_FIN_t, $\Delta REVCash_t$, ROA_{t-1}, and log_AT_{t-1} prior to matching, although differences are generally smaller in absolute magnitude, relative to IPOs. Also in line with the IPO setting, employing propensity-score matching for SEOs results in significant remaining imbalance in means for New_FIN_t and log_AT_{t-1}. In contrast to results for IPOs, entropy-balanced models for SEOs show evidence of significantly positive abnormal accruals in the year of the SEO that subsequently reverse, suggesting positive abnormal accruals in the SEO year of 2.3%–2.5% (1.2%–1.4%) for total (working capital) accrual models. Also in contrast to results for IPOs, propensity-score-matching and ordinary-least-

Table 6 Accrual reversal tests around Accounting and Auditing Enforcement Releases

Dependent Variable	ACC (EB)	ACC (PSM 1NN)	ACC (PSM 5NN)	ACC	DAC	PMDAC
	(1)	(2)	(3)	(4)	(5)	(6)
b, AAER (Y)	0.013**	0.019***	0.012**	0.009**	0.013***	0.011*
	(2.171)	(2.948)	(2.564)	(2.170)	(3.680)	(1.937)
c, AAER (Y + 1)	−0.053***	−0.041***	−0.048***	−0.037***	−0.038***	−0.046***
	(−4.232)	(−5.746)	(−6.772)	(−4.745)	(−6.047)	(−4.768)
$\Delta REVCash_t$	0.005	−0.012	−0.015	−0.007	−0.059***	−0.058***
	(0.315)	(−0.965)	(−1.624)	(−1.417)	(−16.769)	(−15.060)
PPE_t	−0.079***	−0.089***	−0.072***	−0.057***	−0.016***	−0.014***
	(−5.320)	(−6.830)	(−11.899)	(−21.357)	(−8.432)	(−6.152)
ROA_t	0.321***	0.354***	0.301***	0.310***	0.289***	0.151***
	(8.353)	(15.695)	(16.548)	(33.921)	(39.142)	(16.376)
New_FIN_t	0.043***	0.033*	0.027***	0.061***	0.060***	0.060***
	(3.374)	(1.994)	(3.221)	(9.908)	(11.174)	(11.962)
log_MCAP_t	−0.009***	−0.008***	−0.006***	−0.003***	−0.003***	−0.004***
	(−4.636)	(−5.630)	(−6.681)	(−10.094)	(−9.742)	(−8.459)
B/M_{t-1}	0.013***	0.009***	0.008**	0.009***	0.008***	0.008***
	(2.780)	(3.548)	(2.516)	(14.202)	(14.562)	(9.220)
Lev_t	0.083***	0.061***	0.030***	−0.010***	−0.017***	−0.011**
	(4.851)	(3.663)	(4.100)	(−4.020)	(−6.402)	(−2.591)
$RevCash\_Growth_t$	−0.014	−0.011*	0.002	−0.002	0.001	0.001
	(−1.562)	(−1.939)	(0.499)	(−1.196)	(0.490)	(0.745)
ΔROA_t	0.454***	0.478***	0.417***	0.419***	0.394***	0.376***
	(10.824)	(13.840)	(17.576)	(44.001)	(48.233)	(41.867)

Table 6 (continued)

Dependent Variable	ACC (EB)	ACC (PSM 1NN)	ACC (PSM 5NN)	ACC	DAC	PMDAC
	(1)	(2)	(3)	(4)	(5)	(6)
Reversal (b - c)	0.066***	0.060***	0.060***	0.046***	0.051***	0.057***
χ^2 reversal test statistic	19.79	44.34	57.96	32.69	51.83	35.37
χ^2 p-value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Sector and Year FE?	Yes	Yes	Yes	Yes	Yes	Yes
Observations	117,671	2,638	8,334	117,671	117,671	117,671
Adjusted R^2	0.334	0.407	0.328	0.355	0.299	0.115

This table presents results of tests using Accounting and Auditing Enforcement Releases (AAERs) as a binary treatment for likely cases of earnings management. Dependent variables are abnormal accrual measures proposed in this study and the literature. Control variables are included for underlying fundamental determinants of accruals identified in the literature and based on the F-Score prediction model for AAERs detailed by Dechow et al. (2011). Results of weighted ordinary least squares (WOLS) regressions are presented in model (1), using weights specified by the entropy-balancing (EB) program used to achieve covariate balance on means and variances for the control variables included in the table. Results of ordinary least squares (OLS) regressions are presented in models (2) through (6). Models (2) and (3) rely on propensity-score matching, using one-nearest neighbor without replacement (*1NN*) and five-nearest neighbors (*5NN*), respectively. PSM uses a Logit model to estimate the propensity score and a caliper of 0.03. Models (4)–(6) examine unadjusted total accruals (*ACC*), modified Jones (1991) abnormal accruals (*DAC*), and performance matching based on the closest ROA_{t-1} peer within the same two-digit SIC industry and year (*PMDAC*) following the matching procedure recommended by Kothari et al. (2005). Tests for the significance of accrual reversals following Dechow et al. (2012) are included below regression coefficients. These reversal tests examine whether the difference between the coefficient on an indicator for an AAER in year t (*AAER[Y]*, coefficient b) differs from the coefficient on an indicator for the year following the final year of the AAER (*AAER[Y + 1]*, coefficient c). Tests for the significance of this difference are presented using a χ^2 test statistic computed using heteroskedasticity-consistent standard errors with corresponding *p* values presented. See Appendix Table 7 for detailed variable definitions. Separate indicators for sector and fiscal year are included in all models. t-statistics in parentheses and corresponding *p* values are calculated using robust standard errors clustered by year. *, **, *** indicate *p* values that are significant at the 10%, 5%, and 1% levels, respectively

[a] Weights for the entropy-balancing program display a *Weight Ratio* of 45.05, calculated as the (number of firms with entropy-balance weights greater than equal weight) / (number of firm-years with *AAER[Y]* = 1)

squares estimates yield similar inferences to entropy-balanced models for the SEO treatment sample. Evidence from entropy-balanced models of significant positive abnormal accruals around SEOs but not IPOs is consistent with more stringent regulatory and investor scrutiny of IPO firms limiting earnings management, as argued by Ball and Shivakumar (2008). Further, our estimates that suggest modest upward accrual management for SEO firms address criticisms that existing methods indicate implausibly large manipulations around these events (Ball and Shivakumar 2008).

5.6 Additional analysis

Dechow et al. (1995, p. 202) note that SEC AAERs target a sample of firms where there are “strong a priori reasons to expect earnings management of a known sign,” since the SEC targets firms where managers likely violated GAAP (i.e., committing fraud to misstate accounts). Given the differing results between entropy-balanced and other abnormal accrual measures in the IPO setting, we examine the consistency of our approach with the other measures, using AAERs as a setting where earnings management is likely egregious. Specifically, we compare accruals for firms receiving an AAER during the year (treatment firms) to remaining firms in Compustat (control firms).

As covariates, we use the F-Score prediction model for AAERs of Dechow et al. (2011). Table 6 presents results of regressions of various accrual measures on an indicator for firms with an AAER during the year, *b*, and an indicator for the year following the final year of the AAER, *c*, to capture the accrual reversal. Results show that total accruals (ACC_t) are significantly higher for the entropy-balanced sample in model (1) during AAER years (*b* coefficient of 0.013, two-tailed p value <0.05). This is consistent with managers exercising discretion to manage earnings upward through accruals in these periods. Further, results show a significant accrual reversal in the year following an AAER (*c* coefficient of −0.053, two-tailed p value <0.01).

Table 6 also shows similar results for propensity-score matching and for linear abnormal accrual measures. We infer that, because AAERs represent a setting in which manipulations are typically egregious and violate the flexibility permitted by U.S. GAAP (Dechow et al. 2012), the magnitude of accrual manipulations are detectable across estimation approaches. Coefficient magnitudes in models (2) through (6) vary between 0.009 and 0.019 for coefficient *b* and between −0.037 and − 0.053 for coefficient *c*, indicating a similar magnitude of abnormal accruals and reversal across the models. Broadly, our results are consistent with entropy balancing improving test specification in settings with pronounced departures from linearity with respect to underlying fundamentals. In settings where accrual management is significantly egregious (e.g., AAER cases) or where the accrual-generating process is predominantly linear, entropy balancing should yield similar inferences to existing abnormal accrual measures.

5.7 Robustness checks

To confirm that our results are not sensitive to the choice of the variable used to control for firm performance, we performed all of our analysis using current-year

ROA_t as a control for performance in place of prior-year ROA_{t-1}. Results from both simulations and empirical settings using this alternative control yield similar inferences to our analysis using prior-year ROA_{t-1} (untabulated brevity, available upon request).

We also examine whether standard error estimates from entropy-balancing-weighted ordinary-least-squares regressions are biased. To do so, we calculate nonparametric bootstrapped standard errors, using 1000 iterations in tests using IPOs, SEOs and AAERs as the treatment variable of interest. Each iteration randomly selects a subset of the full sample of observations (with replacement) used for each test and re-runs entropy balancing and computes standard errors on this subsample. In results untabulated for brevity, we find that bootstrapped standard errors from these simulations are nearly identical to standard errors calculated using the weighted ordinary-least-squares regressions and that the significance of coefficients is not altered as a result.

6 Conclusion

Research relies almost exclusively on linear regression models to estimate abnormal accruals patterned on the Jones (1991) model. As a result, much of the literature implicitly assumes that the accrual-generating process is adequately described as a linear function of observable firm characteristics. In this paper, we explore whether linear estimates for abnormal accruals are misspecified, display low power, or both when treatment samples are drawn from regions with pronounced nonlinear relations with common accrual determinants. Next, we evaluate whether we can improve estimation of common accrual models if we allow for arbitrary nonlinearities in the accrual-generating process.

As an alternative to linear models for a normal accrual, we employ multivariate matching techniques, via propensity-score matching and entropy balancing, to control for common accrual determinants. We find that these methods improve test specification and power, relative to methods premised on a linear accrual process. Further, we find that entropy-balanced estimates for accruals outperform propensity-matched estimates. We show that using matching to control for arbitrary nonlinearities materially alters inferences in the literature on accruals surrounding new equity issuances via IPOs. Given entropy balancing's benefits and ease of use, we recommend its use to control for accrual determinants and in a wide-range of future accounting research settings. While we leave progress on identifying a comprehensive set of accrual determinants for future research, we note that entropy balancing can be easily adapted to alternative sets of determinants that fit the research setting, sample of interest, or both. Despite entropy balancing's benefits, we also point out a limitation of the approach that stems from its potential to assign large weights to a small number of control observations. We urge future researchers to assess the extent to which entropy balancing assigns extreme weights by computing a version of the weight ratio that we provide and by conducting regression diagnostics (such as weighted regression leverages) to assess the extent of any issues with extreme weights, particularly in settings with smaller samples.

Appendix A

Table 7 Variable definitions and data sources. This appendix presents variable definitions and data sources. Field names from Compustat appear in brackets below

Variable	Measurement details and data sources
Accrual measures:	
ACC	Total accruals calculated according to the cash flow statement method of Hribar and Collins (2002) as (income before extraordinary items from the cash flow statement$_t$ [*IBC*] – cash flow from operations$_t$ [*OANCF* – *XIDOC*]) / book value of total assets$_{t-1}$ [*AT*], where missing values for *XIDOC* (extraordinary items and discontinued operations cash flow) are set to zero.
DAC (WC_DAC)	Discretionary accruals calculated following Dechow et al.'s (2012) estimation of the Dechow et al. (1995) model as $ACC_{it} = \alpha + \beta_1\ PPE_{it} + \beta_2\ \Delta REVCash_{it} + \epsilon_{it}$, where variables included in the model are defined below. The *WC_DAC* measure uses working capital accruals (*WC_ACC*) in place of total accruals as the dependent variable in the ordinary-least-squares (OLS) model. This OLS regression is run by two-digit SIC industry-year (DeFond and Jiambalvo 1994) to generate coefficients that apply to all firms in the same industry-year. *DAC* (*WC_DAC*) is calculated as the residual from subtracting firm i's expected value from *ACC* (*WC_ACC*).
PMDAC (WC_PMDAC)	Discretionary accruals calculated according to the performance-matching approach of Kothari et al. (2005). Baseline discretionary accruals are calculated according to the *DAC* method for firm i. Each firm i in fiscal year t is matched to a peer firm j in the same two-digit SIC industry-year, where the peer firm is identified as the nearest match on *ROA* in prior year *t-1*. *PMDAC* (*WC_PMDAC*) is calculated as $DAC_{it} - DAC_{jt}$ ($WC\_DAC_{it} - WC\_DAC_{jt}$).
WC_ACC	Noncash working capital accruals calculated from successive balance sheets following Larson et al. (2018): (Change in current assets [*ΔACT*] – change in current liabilities [*ΔLCT*] – change in cash [*ΔCHE*] + change in short-term debt [*ΔDLC*]) / book value of total assets$_{t-1}$ [*AT*].
Independent variables and covariates:	
AAER (Y+ 1)	Indicator variable for firms subject to SEC Accounting and Auditing Enforcement Releases (AAERs). The *AAER* variable is an indicator set to one for years in which a firm is found to have manipulated accounts and to zero otherwise. The *AAER(Y+ 1)* variable is an indicator variable set to one for the year following the last year of a given AAER action and to zero otherwise. AAER firms are identified using the hand-collected sample of Dechow et al. (2011) updated through August 31, 2012. Given a typical three-year detection period for AAERs, our AAER subsample ends in August 2009. This dataset is available through the University of California at Berkeley's Haas School of Business: http://groups.haas.berkeley.edu/accounting/aaer_database/
BM	Book-to-market ratio calculated as book value of common equity$_t$ [*CEQ*] / total market-value of the firm's outstanding equity securities at fiscal year-end$_t$ [*PRCC_F* × *CSHO*].
EP	Earnings-to-price ratio calculated as income before extraordinary items$_t$ [*IB*] / total market-value of the firm's outstanding equity securities at fiscal year-end$_t$ [*PRCC_F* × *CSHO*].
IPO	Indicator variable set to one for firm-years with an initial public offering and to zero otherwise. Firms with an initial public offering are identified from SDC Platinum as conducting an initial public offering of common equity on a U.S. exchange during the fiscal year with a valid effective date for the offering.

Table 7 (continued)

Variable	Measurement details and data sources
Lev	Financial leverage calculated as (current portion of long-term debt$_t$ [*DLC*] + noncurrent portion of long-term debt$_t$ [*DLTT*]) / book value of assets$_t$ [*AT*].
log_AT	Natural log of book value of total assets$_t$ [*AT*].
log_MCAP	Natural log of market capitalization calculated as log(market value of the firm's outstanding equity securities at fiscal year-end$_t$ [*PRCC_F* × *CSHO*]).
New_FIN	New financing measured following Hoberg and Phillips (2010) as [(common and preferred stock sold [*SSTK*] - equity repurchased [*PRSTKC*]) + (long-term debt issuance [*DLTIS*] – debt retired [*DLTR*])] / book value of total assets$_{t-1}$ [*AT*].
PPE	Capital intensity calculated as gross property, plant, and equipment$_t$ [*PPEGT*] / book value of assets$_{t-1}$ [*AT*].
ΔREVCash	Change in cash sales calculated as (ΔTotal revenue$_{t-1,t}$ [*REVT*] - ΔAccounts receivable$_{t-1,t}$ [*RECTR*]) / book value of total assets$_{t-1}$ [*AT*].
REVCash_Growth	Percentage growth in cash sales calculated as $\Delta REVCash_{t-1,t}$ / $REVCash_{t-1}$, where *REVCash* = total revenue$_t$ [*REVT*] - ΔAccounts receivable$_{t-1,t}$ [*RECTR*].
ΔROA	Change in return on assets calculated as ROA_t - ROA_{t-1}, where *ROA* is defined below.
ROA_{t-1}	Prior-year return on assets calculated as income before extraordinary items$_{t-1}$ [*IB*] / book value of assets$_{t-1}$ [*AT*].
Sector FE	Sector fixed effects are measured by indicator variables set equal to one for observations falling into each of the 17 Fama-French industry classifications based on SIC codes available in Compustat. Industry classifications are available from Kenneth French's website: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html
SEO	Indicator variable set to one for firm-years with a seasoned equity offering and to zero otherwise. Firms with a seasoned equity offering are identified from SDC Platinum as conducting a secondary offering of common equity on a U.S. exchange during the fiscal year with a valid effective date for the offering.

References

Abadie, A., & Gardeazabal, J. (2003). The economic costs of conflict: A case study of the Basque Country. *American Economic Review, 93*(1), 113–132.

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association, 105*(490), 493–505.

Aharony, J., Lin, C., & Loeb, M. P. (1993). Initial public offerings, accounting choices, and earnings management*. *Contemporary Accounting Research, 10*(1), 61–81.

Armstrong, C., Foster, G., & Taylor, D. J. (2015). Abnormal accruals in newly public companies: Opportunistic misreporting or economic activity? *Management Science, 62*(5), 1316–1338.

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research, 46*(3), 399–424.

Ayers, B. C., Jiang, J., & Yeung, P. E. (2006). Discretionary accruals and earnings management: An analysis of Pseudo earnings targets. *The Accounting Review, 81*(3), 617–652.

Ball, R., & Shivakumar, L. (2008). Earnings quality at initial public offerings. *Journal of Accounting and Economics, 45*(2/3), 324–349.

Bloomfield, M., Gerakos, J., & Kovrijnykh, A. 2017. Accrual reversals and cash conversion. Chicago Booth Research Paper No. 14–29.

Bonsall, S. B., & Miller, B. P. (2017). The impact of narrative disclosure readability on bond ratings and the cost of debt. *Review of Accounting Studies, 22*(2), 608–643.

Chen, W., Hribar, P., & Melessa, S. (2018). Coefficient bias when using residuals as the dependent variable. *Journal of Accounting Research, 56*(3), 751–796.
Cohen, D., & Zarowin, P. (2010). Accrual-based and real earnings management activities around seasoned equity offerings. *Journal of Accounting and Economics, 50*(1), 2–19.
Dechow, P. M., & Dichev, I. D. (2002). The quality of accruals and earnings: The role of accrual estimation errors. *The Accounting Review, 77*(s-1), 35–59.
Dechow, P. M., Sloan, R. G., & Sweeney, A. P. (1995). Detecting earnings management. *The Accounting Review, 70*(2), 193–225.
Dechow, P., Ge, W., & Schrand, C. (2010). Understanding earnings quality: A review of the proxies, their determinants and their consequences. *Journal of Accounting and Economics, 50*(2/3), 344–401.
Dechow, P. M., Ge, W., Larson, C. R., & Sloan, R. G. (2011). Predicting material accounting misstatements. *Contemporary Accounting Research, 28*(1), 17–82.
Dechow, P. M., Hutton, A. P., Kim, J. H., & Sloan, R. G. (2012). Detecting earnings management: A new approach. *Journal of Accounting Research, 50*(2), 275–334.
DeFond, M. L., & Jiambalvo, J. (1994). Debt covenant violation and manipulation of accruals. *Journal of Accounting and Economics, 17*(1–2), 145–176.
DeFond, M. L., & Subramanyam, K. R. (1998). Auditor changes and discretionary accruals. *Journal of Accounting and Economics, 25*(1), 35–67.
DeFond, M. L., Erkens, D. H., & Zhang, J. (2017). Do client characteristics really drive the big N audit quality effect? New evidence from propensity score matching. *Management Science, 63*(11), 3531–3397.
DuCharme, L. L., Malatesta, P. H., & Sefcik, S. E. (2004). Earnings management, stock issues, and shareholder lawsuits. *Journal of Financial Economics, 71*(1), 27–49.
Ecker, F., Francis, J., Olsson, P., & Schipper, K. (2013). Estimation sample selection for discretionary accruals models. *Journal of Accounting and Economics, 56*(2), 190–211.
Freedman, D. A., & Berk, R. A. (2008). Weighting regressions by propensity scores. *Evaluation Review, 32*(4), 392–409.
Friedlan, J. M. (1994). Accounting choices of issuers of initial public offerings. *Contemporary Accounting Research, 11*(1), 1–31.
Hainmueller, J. (2012). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis, 20*(1), 25–46.
Hainmueller, J., & Xu, Y. (2013). Ebalance: A Stata package for entropy balancing. *Journal of Statistical Software, 54*(7), 1–18.
Hirano, K., Imbens, G. W., & Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica, 71*(4), 1161–1189.
Hoberg, G., & Phillips, G. (2010). Real and financial industry booms and busts. *The Journal of Finance, 65*(1), 45–86.
Hribar, P., & Collins, D. W. (2002). Errors in estimating accruals: Implications for empirical research. *Journal of Accounting Research, 40*(1), 105–134.
Imbens, G. W. (2010). Better LATE than nothing: Some comments on Deaton (2009) and Heckman and Urzua (2009). *Journal of Economic Literature, 48*(2), 399–423.
Imbens, G. W., & Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. New York: Cambridge University Press.
Jones, J. J. (1991). Earnings management during import relief investigations. *Journal of Accounting Research, 29*(2), 193–228.
Keung, E., & Shih, M. S. H. (2014). Measuring discretionary accruals: Are ROA-matched models better than the original Jones-type models? *Review of Accounting Studies, 19*(2), 736–768.
King, G., R. Nielsen, C. Coberley, J. E. Pope, & A. Wells. 2011. Comparative Effectiveness of Matching Methods for Causal Inference. Unpublished manuscript 15.
Kothari, S. P. (2001). Capital markets research in accounting. *Journal of Accounting and Economics, 31*(1), 105–231.
Kothari, S. P., Leone, A. J., & Wasley, C. E. (2005). Performance matched discretionary accrual measures. *Journal of Accounting and Economics, 39*(1), 163–197.
Larson, C. R., Sloan, R. G., & Zha Giedt, J. (2018). Defining, measuring and modeling accruals: A guide for researchers. *Review of Accounting Studies, 23*(3), 827–871.
Loughran, T., & Ritter, J. R. (1997). The operating performance of firms conducting seasoned equity offerings. *The Journal of Finance, 52*(5), 1823–1850.
Nikolaev, V. 2017. Identifying Accounting Quality. Chicago Booth Research Paper No. 14–28.
Owens, E. L., Wu, J. S., & Zimmerman, J. (2017). Idiosyncratic shocks to firm underlying economics and abnormal accruals. *The Accounting Review, 92*(2), 183–219.

Rangan, S. (1998). Earnings management and the performance of seasoned equity offerings. *Journal of Financial Economics, 50*(1), 101–122.
Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41–55.
Rubin, D. B. (1973). The use of matched sampling and regression adjustment to remove Bias in observational studies. *Biometrics, 29*(1), 185–203.
Rubin, D. B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services & Outcomes Research Methodology, 2*(3), 169–188.
Shipman, J. E., Swanquist, Q. T., & Whited, R. L. (2017). Propensity score matching in accounting research. *The Accounting Review, 92*(1), 213–244.
Shivakumar, L. (2000). Do firms mislead investors by overstating earnings before seasoned equity offerings? *Journal of Accounting and Economics, 29*(3), 339–371.
Shroff, N., Verdi, R. S., & Yost, B. P. (2017). When does the peer information environment matter? *Journal of Accounting and Economics, 64*(2–3), 183–214.
Teoh, S. H., Welch, I., & Wong, T. J. (1998a). Earnings management and the long-run market performance of initial public offerings. *The Journal of Finance, 53*(6), 1935–1974.
Teoh, S. H., Welch, I., & Wong, T. J. (1998b). Earnings management and the underperformance of seasoned equity offerings. *Journal of Financial Economics, 50*(1), 63–99.
Wilde, J. H. (2017). The deterrent effect of employee whistleblowing on firms' financial misreporting and tax aggressiveness. *The Accounting Review, 92*(5), 247–280.
Zhao, Q., & Percival, D. (2016). Entropy balancing is doubly robust. *Journal of Causal Inference, 5*(1), 1–23.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.