Estimation sample selection for discretionary accruals models

Frank Ecker, Jennifer Francis *, Per Olsson, Katherine Schipper
Fuqua School of Business, Duke University, Durham, NC 27708, United States

Abstract

We examine how the criteria for choosing estimation samples affect the ability to detect discretionary accruals, using several variants of the Jones (1991) model. Researchers commonly estimate accruals models in cross-section, and define the estimation sample as all firms in the same industry. We examine whether firm size performs at least as well as industry membership as the criterion for selecting estimation samples. For U.S. data, we find estimation samples based on similarity in lagged assets perform at least as well as estimation samples based on industry membership at detecting discretionary accruals, both in simulations with seeded accruals between 2% and 100% of total assets and in tests examining restatement data and AAER data. For non-U.S. data, we find industry-based estimation samples result in significant sample attrition and estimation samples based on lagged assets perform at least as well as estimation samples based on industry membership, both in simulations and in tests examining German restatement data, with substantially less sample attrition.

1. Introduction

Using several variants and extensions of the Jones (1991) model of discretionary accruals, we examine how the selection of estimation samples affects the power of these models to detect discretionary accruals. Our research aims to provide a practical solution to the problem of substantial sample attrition when discretionary accruals models are estimated in time series (eliminating firms lacking the requisite number of time-series observations) and in industry cross-sections (eliminating firms whose industries lack the requisite number of members).

The problem we address is significant in the U.S. and acute in non-U.S. markets. The average number of firms per industry in the U.S. during 1988–2009 with the necessary data to estimate an accruals observation is 80.5 (SIC2), 21.7 (SIC3) and 13.6 (SIC4). For the 99 non-U.S. countries with data on Compustat Global over the same time period the corresponding mean values are 3.5

© 2013 Elsevier B.V. All rights reserved.

1 The accruals models require each observation to have current and one-year lagged financial data to construct the dependent and independent variables. For example, the basic Jones model requires data on lagged total assets, current total accruals, current net property, plant and equipment, and the change in sales revenues, while the Larcker and Richardson (2004) extension of the Jones model requires, additionally, current and lagged receivables, book value and market value of equity and cash from operations.

0165-4101/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jacceco.2013.07.001
imposing typical requirements—10 observations besides the event firm\(^2\) are available for an industry to be included—leads to substantial sample attrition and the complete elimination of many countries from studies of discretionary accruals. For U.S. data, the use of industry-based estimation samples eliminates 1–22% of the otherwise-available sample, depending on how industry is defined. For the 69 non-U.S. countries with at least one year with 11 firm-year observations overall, the sample loss is between 32% and 93% on average, depending on how industry is defined and the weighting of countries; requiring sufficient data to form industry-based estimation samples for a given country causes between 29 and 40 countries (out of 69) to be eliminated entirely from a study that estimates discretionary accruals using estimation samples defined by industry membership.\(^3\)

We propose and test a solution to this problem, specifically, basing estimation samples on similarity in size, not industry membership. We consider size because, as we explain in more detail later, it is an intuitively grounded alternative (to industry membership) indicator of similarity; that is, a group of larger firms is more alike than is a mixed group of larger and smaller firms. Our solution and related tests derive from two simultaneously-considered objectives: reducing sample attrition from data requirements and retaining at least the detection power obtainable from using industry membership as the estimation sample selection criterion. The objective of reducing sample attrition from data requirements suggests the criterion for choosing estimation samples should be both widely available and numerical, so large samples can be ranked on the criterion. Size-based estimation samples impose no sample loss incremental to the loss imposed by the accruals models themselves, because size-based peers\(^4\) can be defined as firms closest in size to the target firm. Since firms can always be ranked on size, there will always be a set of firms in the neighborhood of the target firm. Whether those firms are close enough for purposes of detecting discretionary accruals is the empirical question we explore.\(^5\) Evidence that firm size performs at least as well as industry membership as the criterion for selecting an estimation sample for a discretionary accruals model means researchers can substantially expand sample sizes without loss of detection power.

To meet the second (detection power) objective the criterion should also result in estimation samples with reasonably similar properties. Our use of firm size as measured by lagged total assets as an estimation sample criterion is based on previous research showing firms of similar size are also similar with regard to factors associated with accruals, such as growth, complexity and monitoring. Relative to smaller firms, larger firms are likely to be older (so more stable, with lower growth rates), to have more segments (so more complex) and to be more closely monitored (larger analyst following, more regulatory oversight, greater likelihood of Big-4 auditor, more institutional ownership). More specifically, in a study addressing the use of industry membership combined with size to identify peer firms for use in relative performance evaluation, Albuquerque (2009) provides extensive discussion and empirical evidence supporting the view that size subsumes several characteristics affecting accruals, including diversification, operating leverage, and growth options. In addition, Kothari et al. (2005) document a correlation between size and discretionary accruals estimated using an industry-cross-sectional accruals model.

In analyses of the detection-power-related objective, we consider how and why previous research uses industry-based estimation samples. Our reading of the literature on detecting discretionary accruals suggests researchers have implicitly or explicitly maintained a three-part presumption: (1) estimation samples for detecting discretionary (managed) accruals should be homogeneous with respect to the normal (unmanaged) accruals generating process because (2) similarity in the normal accruals process is believed to be associated with the power to detect abnormal accruals, and (3) firms in the same industry have congruent normal accruals generating processes.\(^6\)

Our research mainly aims to shed light on the first two components of this three-part presumption.\(^7\) With regard to the first component, our results suggest both size-based samples and industry-based samples exhibit reasonable, and reasonably similar, evidence of accruals homogeneity. For example, the explanatory power of the Jones model with intercept for industry-based estimation samples is about 51–52% while the explanatory power of this model for the lagged assets-based estimation samples is about 44%. We conclude that if sample attrition is not a concern, industry-based samples produce higher explanatory power in the estimation of normal accruals; if sample attrition is a concern, the researcher would trade off a reduction in explanatory power with a loss in sample size. However, inconsistent with the second component of the three-part presumption, we find lagged assets-based estimation samples produce higher detection rates of abnormal (or discretionary) accruals, as compared to industry-based estimation samples. That is, our evidence supports the first presumption and refutes the second presumption: while industry-based estimation samples have the highest

---

\(^2\) The event firm is the firm whose discretionary accruals are the subject of the researcher’s analysis.

\(^3\) The researcher could pool data across countries, within industry, to increase the number of firms in each industry. This pooling has the potential to introduce noise that lowers the power of the test. If the researcher aims to examine jurisdictional influences on discretionary accruals, pooling observations across jurisdictions is either not feasible or will bias the results against detecting such influences.

\(^4\) Consistent with the notion that estimation samples for both normal accruals estimation and abnormal or discretionary accruals detection should have a reasonable degree of homogeneity with respect to the accruals generating process, we sometimes refer to estimation samples as “peer firms.” For example, estimation samples based on industry membership are “industry peers” or “industry-based peers” and estimation samples based on size are “size peers” or “size-based peers.”

\(^5\) Our tests also address the fact that the larger population of U.S. traded firms relative to any non-U.S. market implies size-based neighbors in the U.S. will (likely) be more similar to the target firm than are neighbors in non-U.S. markets.

\(^6\) Researchers have questioned the empirical validity of the assumption that firms in the same industry (same SIC code) have congruent normal accruals (e.g., Bernard and Skinner, 1996; Brickley and Zimmerman, 2010; Depuch et al. (2012)) find little support for the assumption that firms in the same industry have a homogeneous accruals-generating process; they do not propose an alternative to industry for defining an estimation sample.

\(^7\) We provide supplementary analyses in Section 6.6 on the third component.
explanatory power for normal accruals, these same models are not the best at detecting abnormal accruals (rather, size-based samples generally dominate here).

We compare the power to detect discretionary accruals using industry-based estimation samples with the detection power of size-based samples using both simulations and archival tests. Our tests modify the approach used by Dechow et al. (1995) to examine the power of several accruals models to detect earnings management. Their aim was to compare accrual models, whereas we compare the effects of estimation sample selection for a given set of models. Dechow et al. estimated accrual models using time-series data,9 while our focus is on cross-sectional estimation using industry-based versus size-based criteria for estimation sample selection. We use Dechow et al.’s framework to guide our analysis and to assist the reader in interpreting our results.

Our first simulations use all firms, both U.S. and Canadian, with available data on Compustat North America over 1951–2009; we refer to these firms as the U.S. data or the U.S. sample. These simulation tests reveal that estimation samples based on lagged assets are at least as powerful as industry peers at detecting induced discretionary accruals, and they are often more powerful. Our first archival tests consider the power of size-based and industry-based peer groups to detect discretionary accruals for two samples of restatement firms and firms named in an SEC Accounting and Auditing Enforcement Release (AAER firms).9 These firms are known to have discretionary accruals, because a restatement or AAER is after-the-fact evidence of purposeful or inadvertent deviation from a normal (GAAP-based) accruals-generating process, but the amounts may be unknown to the researcher. These tests show that models estimated using lagged-asset-based estimation samples have higher rates of discretionary accruals detection than do estimations using industry-based estimation samples. Our second simulation tests focus on non-U.S. data, specifically, Compustat Global data for 1988–2009. We find size-based peers perform as well as, or better than, industry-based peers at detecting induced discretionary accruals and impose far less sample attrition. Our second archival tests, using German restatement data, support the conclusions based on our analyses of U.S. restatement/AAER data: using industry-based estimation samples results in significant sample loss, and size-based estimation samples yield better detection results than do industry-based estimation samples.

These results are robust to a battery of sensitivity tests and additional analyses, including variations in the numbers of firms included in industry-based and size-based estimation samples. We also provide evidence that the performance of estimation samples based on lagged assets is not induced by the inclusion of lagged assets in the accruals models we analyze.

Viewed as a whole, our results indicate size-based estimation samples perform at least as well as industry-based estimation samples in terms of detecting discretionary accruals in both U.S. data and non-U.S. data. The use of size-based estimation samples for U.S. data avoids sample losses (up to 22% in some cases) arising from using industry-based estimation samples. Our results have even greater practical value for estimating discretionary accruals models using non-U.S. data, where basing estimation samples on industry membership results in sample attrition as high as 90%. Entire countries excluded from industry-based estimations can be included in size-based estimations, with no loss of power to detect discretionary accruals.

The remainder of the paper is organized as follows. Section 2 provides descriptive evidence on our sample. Section 3 reports the results of a simulation analysis investigating the power of size-based and industry-based peer groups to detect induced discretionary accruals using U.S. data. Section 4 analyzes the ability of estimations using industry-based peers and size-based peers to detect discretionary accruals of restatement firms and AAER firms. Section 5 reports results of analyses of non-U.S. data, including both simulation tests and tests for detection of discretionary accruals in German restatement firms. Section 6 reports additional tests and section 7 concludes.

2. Sample and descriptive data

We provide descriptive information about total accruals, discretionary accruals and size measures for the sample of Compustat North America firm-years covering 1951–2009 meeting the data requirements of our simulation tests. The requirements for firm-years to be included in the U.S. simulation sample are described in Table 1. Other than satisfying plausibility checks and reporting the data items needed for the discretionary accruals models,10 firm-years are required to have the data on each criterion we use to create estimation samples. The criteria are industry membership, which requires data on SIC code, and size, which requires data on assets, lagged assets, sales, lagged sales, market capitalization and firm age. Because our analyses require a minimum number of firms to estimate each cross-sectional discretionary accruals model, we require each peer group to have at least 11 firm-year observations (10 non-event firms and one event firm).11 The final simulation sample contains 143,584 firm-years, representing 59 annual cross-sections and 1972 distinct SIC2-year peer

---

8 Using the firm as its own control offers the advantage of holding firm-specific factors constant, but results in substantial, and potentially unacceptable, sample loss because sample firms must have a sufficient time series to estimate the accruals models. For example, requiring at least 10 consecutive time-series observations eliminates about 57% of the 239,896 non-bank firm-year observations on Compustat over 1950–2009 with the necessary data to estimate the variables in a Jones-style accruals model. This corresponds to a sample loss of 59% distinct firms. In addition, the resulting sample is biased towards bigger and more profitable survivors.

9 We describe our data sources for restatement firms and AAER firms in Section 4.

10 For restatement and AAER observations, we estimate the accruals models using data from the unrestated Compustat database (“as first reported”).

11 This requirement, or a similar one, is common in the literature. For example, Kothari et al. (2005) require 10 firms in each 2-digit SIC code.
groups (4349 distinct SIC3-year peer groups and 5486 distinct SIC4-year peer groups). On average, the industry-year peer groups contain 73 firms (SIC2), 33 firms (SIC3) and 26 firms (SIC4).

Table 1: Selection of data observations for simulation using U.S. data 1950–2009.

<table>
<thead>
<tr>
<th>Selection criteria</th>
<th># Firm-years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unique firm-years on Compustat North America</td>
<td>408,245</td>
</tr>
<tr>
<td>Observations for firms with more than 1 year of data</td>
<td>407,726</td>
</tr>
<tr>
<td>With lagged total assets, total assets, sales ( &gt; -1 ), ROA ( &gt; -1 )</td>
<td>299,414</td>
</tr>
<tr>
<td>With data for total accruals calculation</td>
<td>247,841</td>
</tr>
<tr>
<td>Non-bank firm-years</td>
<td>239,896</td>
</tr>
<tr>
<td>With data for identifying ALL peer groups</td>
<td>203,842</td>
</tr>
<tr>
<td>With at least 11 firm-year observations per SIC4</td>
<td>143,584</td>
</tr>
</tbody>
</table>

This table reports the sample restrictions imposed by requirements to have the necessary data to calculate an accruals observation, to identify peer groups, and to estimate the accruals models. The most restrictive criterion requires 10 non-event firms in the same SIC4 code. The bottom rows of the table report the sizes of the industry peer groups, e.g., out of the 143,584 firm-year observations meeting our requirements, there are 1972 distinct SIC2-year peer groups.

Table 2: Descriptive statistics for U.S. samples.

<table>
<thead>
<tr>
<th># Obs.</th>
<th>Mean</th>
<th>Std. dev.</th>
<th>P5</th>
<th>P25</th>
<th>Median</th>
<th>P75</th>
<th>P95</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total accruals</td>
<td>143,584</td>
<td>-0.0310</td>
<td>0.3312</td>
<td>-0.2215</td>
<td>-0.0897</td>
<td>-0.0391</td>
<td>0.0126</td>
</tr>
<tr>
<td>Total accruals (GAO restatements)</td>
<td>1337</td>
<td>-0.0702</td>
<td>0.1895</td>
<td>-0.2491</td>
<td>-0.1070</td>
<td>-0.0572</td>
<td>-0.0139</td>
</tr>
<tr>
<td>Total accruals (AA restatements)</td>
<td>5032</td>
<td>-0.0455</td>
<td>0.3451</td>
<td>-0.2331</td>
<td>-0.1004</td>
<td>-0.0487</td>
<td>-0.0008</td>
</tr>
<tr>
<td>Total accruals (AAER)</td>
<td>372</td>
<td>0.0174</td>
<td>0.0804</td>
<td>-0.2966</td>
<td>-0.0722</td>
<td>-0.0034</td>
<td>0.1225</td>
</tr>
<tr>
<td>Abs. total accruals</td>
<td>143,584</td>
<td>0.1059</td>
<td>0.3154</td>
<td>0.0063</td>
<td>0.0309</td>
<td>0.0642</td>
<td>0.1188</td>
</tr>
<tr>
<td>Abs. total accruals (GAO restatements)</td>
<td>1337</td>
<td>0.1051</td>
<td>0.1726</td>
<td>0.0060</td>
<td>0.0334</td>
<td>0.0705</td>
<td>0.1221</td>
</tr>
<tr>
<td>Abs. total accruals (AA restatements)</td>
<td>5032</td>
<td>0.1151</td>
<td>0.3285</td>
<td>0.0068</td>
<td>0.0325</td>
<td>0.0682</td>
<td>0.1208</td>
</tr>
<tr>
<td>Abs. total accruals (AAER)</td>
<td>372</td>
<td>0.2310</td>
<td>0.7686</td>
<td>0.0065</td>
<td>0.0368</td>
<td>0.0879</td>
<td>0.2072</td>
</tr>
<tr>
<td>Total assets</td>
<td>143,584</td>
<td>1662</td>
<td>8672</td>
<td>5.28</td>
<td>26.83</td>
<td>104.77</td>
<td>531.32</td>
</tr>
<tr>
<td>Lagged total assets</td>
<td>143,584</td>
<td>1528</td>
<td>8091</td>
<td>4.42</td>
<td>22.68</td>
<td>90.45</td>
<td>466.25</td>
</tr>
<tr>
<td>Sales</td>
<td>143,584</td>
<td>1364</td>
<td>7698</td>
<td>4.06</td>
<td>25.27</td>
<td>103.60</td>
<td>499.23</td>
</tr>
<tr>
<td>Lagged sales</td>
<td>143,584</td>
<td>1264</td>
<td>7269</td>
<td>3.05</td>
<td>21.54</td>
<td>90.51</td>
<td>447.74</td>
</tr>
<tr>
<td>Market cap.</td>
<td>143,584</td>
<td>1647</td>
<td>11,050</td>
<td>2.87</td>
<td>18.80</td>
<td>88.03</td>
<td>482.85</td>
</tr>
<tr>
<td>Firm age (in years)</td>
<td>143,584</td>
<td>15.54</td>
<td>14.39</td>
<td>6.00</td>
<td>11.00</td>
<td>21.00</td>
<td>46.00</td>
</tr>
<tr>
<td>ROA</td>
<td>143,584</td>
<td>-0.01</td>
<td>0.20</td>
<td>-0.39</td>
<td>-0.02</td>
<td>0.04</td>
<td>0.07</td>
</tr>
</tbody>
</table>

This table presents descriptive information about accruals (signed and absolute) for firm-years in the U.S. simulation sample, the U.S. GAO and Audit Analytics (AA) restatement samples and the U.S. AAER sample. Total accruals are as defined in the text. Total assets and sales are from Compustat. Firm age is defined as the maximum difference, in years, between the date of the fiscal year end and the founding year as available on Jay Ritter's website (http://bear.warrington.ufl.edu/ritter/ipodata.htm), or the first year with data on Compustat, or the first year with data on CRSP. ROA is net income before extraordinary items scaled by total assets.

On average, the industry-year peer groups contain 73 firms (SIC2), 33 firms (SIC3) and 26 firms (SIC4).
3. Simulations with seeded discretionary accruals

This section describes our simulation analyses of how choosing estimation samples based on size versus industry membership affects the ability to detect seeded discretionary accruals. We first describe the models for discretionary accruals and the estimation samples we consider. Section 3.2 outlines the setup of our simulations, Section 3.3 describes our design choices and Section 3.4 presents the simulation results. Section 3.5 presents results of analyses shedding light on the reasons for those results.

3.1. Discretionary accruals estimation and estimation samples

In our main tests, we conduct tests on three cross-sectional discretionary accruals models used in the literature. The models are based on the following three equations:

\[
\frac{Total\ accruals_{i,t}}{Total\ Assets_{i,t-1}} = \frac{\alpha_0}{Total\ Assets_{i,t-1}} + \alpha_1 \frac{\Delta Sales_{i,t}}{Total\ Assets_{i,t-1}} + \alpha_2 \frac{Net\ PPE_{i,t}}{Total\ Assets_{i,t-1}} + DA_{Jones} 
\]

(1)

\[
\frac{Total\ accruals_{i,t}}{Total\ Assets_{i,t-1}} = \alpha_0 + \alpha_1 \frac{1}{Total\ Assets_{i,t-1}} + \alpha_2 \frac{\Delta Sales_{i,t}}{Total\ Assets_{i,t-1}} + \alpha_3 \frac{Net\ PPE_{i,t}}{Total\ Assets_{i,t-1}} + DA_{Jones(\text{intercept})} 
\]

(2)

\[
\frac{Total\ accruals_{i,t}}{Total\ Assets_{i,t-1}} = \alpha_0 + \alpha_1 \frac{1}{Total\ Assets_{i,t-1}} + \alpha_2 \frac{\Delta Sales_{i,t} - \Delta AR_{i,t}}{Total\ Assets_{i,t-1}} + \alpha_3 \frac{Net\ PPE_{i,t}}{Total\ Assets_{i,t-1}} + DA_{Mod.\ Jones} 
\]

(3)

where Total accruals_{i,t} = firm j’s total accruals in year t, measured as the change in current assets (adjusted for the change in cash) minus the change in current liabilities (adjusted for current liabilities used for financing) minus depreciation expense; Total Assets_{i,t-1} = firm j’s total assets in year t – 1; \Delta Sales_{i,t} = firm j’s change in sales between year t – 1 and year t; Net PPE_{i,t} = firm j’s net property, plant and equipment in year t; \Delta AR_{i,t} = firm j’s change in accounts receivable between year t – 1 and year t; ROA_{i,t} = firm j’s return on assets in year t.

Eqs. (1) and (2) are the Jones (1991) model without and with an intercept, introduced and recommended by Kothari et al. (2005). Eq. (3) is a modified Jones model which includes an intercept and an adjustment for the change in accounts receivable. The residuals from estimating Eqs. in the cross section provide three types of discretionary accruals.

Our benchmark estimation sample is a peer group formed by a random selection of firms from the entire sample cross section. We define industry-based estimation samples (industry peers) based on 2-digit, 3-digit and 4-digit SIC codes, because our reading of the literature suggests these are the most commonly-used approaches to determining estimation samples for discretionary accruals models. We define size-based estimation samples (size peers) based on total assets, lagged total assets, sales, lagged sales, market capitalization and firm age. Varying the cross section (“peer group”) over which the discretionary accruals models are estimated yields a peer-group-dependent measure and distribution of discretionary accruals.

3.2. Simulation design, seeding and estimation sample evaluation via rejection rates

Using the firm-years from the sample described in Table 1, we perform 100 iterations, each of which has four steps:

1. We randomly select 500 firm-years and define them as event-firm-years. The event-firm-years remain constant throughout the iteration.
2. For each event-firm-year, we select initial peer firms from which estimation samples will be chosen:
   a. Sample cross section.
   b. Industry-based peer groups (SIC2 industry, SIC3 industry, SIC4 industry) chosen by matching by year and industry, regardless of the number of firm-years in the industry.
   c. Size-based peer groups (total assets, lagged total assets, sales, lagged sales, market capitalization, firm age) chosen by matching the year and the 25 adjacent (to the event firm) lower-ranked firms and the 25 adjacent (to the event firm) higher-ranked firms; these are the event firm’s “closest neighbors.”
3. Step 2 yields varying numbers of observations in the sets of initial peer firms; they vary for the four industry-based peer groups and are fixed at 50 for the six size-based peer groups.
4. To equalize the number of observations in each peer group, we randomly select 10 firms from the initial peer groups identified in Step 2. This step is intended to equalize the power of tests, which is affected by sample size. We use a

---

12 We discuss results based on other models, including variants of performance-adjusted accruals based on Kothari et al. (2005) and Larcker and Richardson’s (2004) extension of the Jones model, in Section 6.3.
13 We do not consider double-sorted peer groups (firms from the same industry and same size decile), because our goal is to relax sample restrictions.
14 Results of sensitivity tests related to industry definitions are discussed in Section 6.
15 We also examined ROA-based peer groups. Results (not tabulated) indicate ROA-based estimation samples have discretionary accruals detection power that is marginally better than reported for the entire cross section.
constant number of peer firms (10), both to estimate the discretionary accruals regression and for the significance test on the difference in discretionary accruals between event and non-event observations.

4. We repeat Steps 2 and 3 for each of the 500 event-firm-years.

Our simulation tests are based on 100 iterations, yielding 50,000 event-firm-years, each matched with 10 peer-firm-years from each of the 10 peer group definitions. While this design does not impose restrictions on how often a firm-year can be selected, either as an event observation or a non-event observation, the two-layer sampling minimizes the likelihood an entire subsample would be replicated.

For each event-firm-year, we seed discretionary accruals into the data. First, we calculate the event-firm-year’s ratio of total accruals to lagged total assets (the dependent variable in the accruals model regressions). We add between 2% and 20% of lagged total assets, in two percentage point increments, to yield 10 “positively managed” accruals figures for each event-firm-year. As a reference, Dechow et al. (1995) seed discretionary accruals in 10 percentage point increments, from 0% to 100%. While we also calculate results for 20–100% seed levels16 we focus on smaller amounts of induced discretionary accruals in the tables (20% or less), because we believe they are more descriptive of observed levels of discretionary accruals. In particular, Table 2 shows the 5th and 95th percentiles of scaled signed accruals for the U.S. simulation sample are about –22% and +19%, respectively; the unsigned percentiles are 0.6% (5th percentile) and 30% (95th percentile). The 0% seed case reflects 0% discretionary accruals as proxied by the raw Compustat data.

Our main simulations investigate the ability to detect an amount (between 2% and 20%) of seeded discretionary accruals at the subsample level. For each peer firm definition, we obtain 50,000 subsamples, each one containing one event-firm matched with 10 peer firms, and each one containing the 10 levels of seeded accruals. These seeded amounts represent the discretionary accruals that are the object of the discretionary accruals detection tests. Our tests compare the estimated discretionary accruals of the event-firm-years with the average of the estimated discretionary accruals of the 10 non-event-firm-years, using the following regression, separately for each seed level:

\[
\text{Discretionary accruals}_{it} = \alpha_0 + \alpha_1 \text{Event Dummy}_{it} + \eta_{it}
\]

where \(\text{Event Dummy}_{it} = 1\) for the single event-firm observation in each subsample. We assess detection power by counting the number of positive \(\alpha_1\) coefficients significant at the 10% level.17 The detection rate is the fraction, out of 50,000, of significant coefficients.

3.3. Explanation of simulation choices

In this section, we describe the reasoning for four design choices used in the simulation. First, we believe the two-layer selection of peer firms, as described in Section 3.2, captures a typical setting faced by a researcher whose objective is to identify discretionary accruals. For example, there are typically a limited number of “event” firm-years, each of which is hypothesized to contain discretionary accruals of some (perhaps unspecified) amount and sign. In all likelihood, the sample and data restrictions in a typical discretionary accruals study are more constraining than suggested by the comprehensive dataset from which we initially select peer firms in Step 2. We use the Step-3 selection to capture the effects of the constrained sample size of a typical discretionary accruals study. The second selection layer ensures a constant number of peer firms and ensures our results are not driven by using the closest size-based neighbors in the estimation sample.18

Second, in selecting neighbors for the size-based peer groups, we form relative peer groups (i.e., peers are defined as the event firm’s closest-in-size neighbors) not absolute peer groups (i.e., peers are determined using an absolute size cutoff). For example, an absolute size peer group would contain firms from the same asset decile or market capitalization decile as the event firm, where the cutoff values for those deciles are determined using the entire cross-section of firms; the event-firm is then matched to firms from its same decile. We believe the absolute peer group approach has at least two disadvantages compared to the relative peer group approach. First, the absolute approach does not ensure symmetry in the selection of peer firms; for example, non-event firms from the same decile will be systematically smaller (larger) when the event firm is a large (small) firm in that decile. Second, the absolute peer group approach requires the full cross-section of firms to determine the initial partitions (e.g., asset deciles or market capitalization deciles). Forming initial deciles/partitions using a subsample of 500 firms yields different results than does forming deciles/partitions using a subsample of 5000 firms, particularly if the 500 firm subsample is not distributed uniformly across the deciles of the 5000-firm subsample, but, say, biased towards bigger and more profitable firms.

A third design choice concerns the interaction between seeded discretionary accruals and the models of discretionary accruals. We seed discretionary accruals by adding between 2 and 20 percentage points to the event-firm’s ratio of total accruals to lagged assets. We do not adjust other variables such as sales or total assets. Our seeding, therefore, is most

16 These results are not tabulated; inferences based on these results are similar to those reported in the paper.
17 We find similar results (not tabulated) using 5% and 1% significance levels.
18 As a sensitivity test, we select the 10 closest-in-size peers. Results (not tabulated) show this lagged asset peer group performs better than the one we select based on the random sampling in Step 3.
similar to Dechow et al.’s (1995) “expense manipulation” view of discretionary accruals (although our approach is cross-sectional and theirs is time-series).

Modeling other financial statement effects requires additional assumptions, as summarized and applied by Dechow et al. (1995). A specific form of earnings management affects specific income statement and/or balance sheet accounts that might or might not be included as independent variables in the normal accruals model, and a guiding principle of such models is that the independent variables should not be affected by earnings management. Put another way, the ability to detect a given type of earnings management will be affected by the choice of the normal accruals model itself, because certain types of earnings management might affect the variables in one model, but not those in another model. In an extreme case in which both accruals (the dependent variable) and an independent variable change by equal amounts, earnings management would remain undetected. We acknowledge that the choice of a specific accruals model is important for the detection of a given type of earnings management. Because we do not compare accruals models (the focus of Dechow et al., 1995), we focus on the simplest (from a modeling perspective) view of discretionary accruals which does not adjust the values of the independent variables in the accruals models. The independent variables of all models can thus be assumed to be unaffected by (our seeded) earnings management, and hence we do not induce varying detection power across models.

A fourth design choice concerns the level of the analysis. As previously described, we analyze detection rates at the subsample level (50,000 subsamples each consisting of 1 event firm and 10 non-event firms) rather than at the more aggregated sample level (100 samples each consisting of 500 event firms and 5000 non-event firms). We select the subsample level for two reasons. First, as a practical matter, the discretionary accruals estimation must be performed at the subsample level, so the finer data are readily available. Second, it is hard to observe differences at the sample level because firm-specific idiosyncrasies are averaged out, implying that even at small seeded discretionary accruals levels detection rates will approach 100% quickly for all models. Stated differently, we believe performing our analysis at the sample level (where the number of event firms and non-event firms is large) would mislead readers about the generalizability of our findings to the smaller samples typically used in discretionary accruals research. By analyzing subsamples, we believe we more closely approximate the issues faced in discretionary accruals research.

3.4. Results of simulations

Table 3 reports for each discretionary accruals model (panels of table) and peer group definition (columns of table) the fraction of times, as a percentage of the 50,000 subsamples, the $\alpha_{1}$ coefficient from Eq. (4) is significantly positive at the 10% level for increasing levels of seeded discretionary accruals (the seed levels are the rows of each panel). We refer to the rejection rate, at the 10% level, as the discretionary accruals detection rate. Results for the 0% seed level correspond to the benchmark case of no seeded discretionary accruals and can be used to gauge the validity of the approach. Specifically, if there are no discretionary accruals and if the data are truly random, 5% of cases should exhibit significant positive discretionary accruals at the 10% level (5% of cases will show significant negative discretionary accruals). A 95% confidence interval around this theoretical rejection rate ranges from 0.7% to 9.3% for samples of 100 observations. Table 3 shows the detection rates for the 0% seed level are close to 5% for all peer definitions and models, although there is some evidence the 0% detection rates are higher for the lagged asset-based estimation sample. These rejection rates are well within the limits of the 95% confidence interval and are not statistically different from 5% or each other.

Turning to the non-zero seed levels, our interest is in which peer group (column) yields the highest detection rates across accruals models. These results show that for all seed levels, the peer group formed using the cross-section performs the worst; this result is expected since this peer group does not attempt to match non-event firms with event-firms on any dimension except the event year. Peer groups based on firm age, sales and market capitalization also perform poorly across all seed levels. At seed levels less than 10%, the industry-based peers and lagged asset-based peers have the highest detection rates, and these rates are similar. At seed levels between 10% and 20%, the detection rate of the lagged asset-based peers begins to exceed the others.

We create two measures of the aggregate performance of each peer group across the 10 seed levels for each model. Our first measure of the performance of a given peer group/model/seed level combination is the detection rate achieved by a specified peer group, for each model and seed level, divided by the maximum detection rate achieved by any peer group for that seed level/model combination. We average this peer group/model/seed level performance measure across the seed levels for each model and peer group, to yield an “effectiveness score” for each peer group/model combination. The closer the effectiveness score to 100% (the value if the peer group had the highest detection rate for every seed level), the better at...
The detection rate for each seed level of induced discretionary accruals (2–20% of lagged assets) and peer definitions, for three accruals models. The detection rate is the fraction of times, as a percentage of the 50,000 subsamples, the slope coefficient in Eq. (4) is significantly positive at the 10% level. The 0% seed level is a specification check, insofar as a well-specified model should show a 5% detection rate of positive discretionary accruals (when none is induced) at a 10% significance level. Effectiveness scores are reported for each peer group and model; the effectiveness score is the average, across seed levels, of the absolute value of the distance between the peer group’s detection rate and the maximum (across all peer groups) detection rate when none is induced at a 10% significance level. Effectiveness scores are reported for each peer group and model; the effectiveness score is the average, across seed levels, of the absolute value of the distance between the peer group’s detection rate and the maximum (across all peer groups) detection rate when none is induced at a 10% significance level. Effectiveness scores are reported for each peer group and model; the effectiveness score is the average, across seed levels, of the absolute value of the distance between the peer group’s detection rate and the maximum (across all peer groups) detection rate when none is induced at a 10% significance level. Effectiveness scores are reported for each peer group and model; the effectiveness score is the average, across seed levels, of the absolute value of the distance between the peer group’s detection rate and the maximum (across all peer groups) detection rate when none is induced at a 10% significance level. Effectiveness scores are reported for each peer group and model; the effectiveness score is the average, across seed levels, of the absolute value of the distance between the peer group’s detection rate and the maximum (across all peer groups) detection rate when none is induced at a 10% significance level. Effectiveness scores are reported for each peer group and model; the effectiveness score is the average, across seed levels, of the absolute value of the distance between the peer group’s detection rate and the maximum (across all peer groups) detection rate when none is induced at a 10% significance level. Effectiveness scores are reported for each peer group and model; the effectiveness score is the average, across seed levels, of the absolute value of the distance between the peer group’s detection rate and the maximum (across all peer groups) detection rate when none is induced at a 10% significance level. Effectiveness scores are reported for each peer group and model; the effectiveness score is the average, across seed levels, of the absolute value of the distance between the peer group’s detection rate and the maximum (across all peer groups) detection rate when none is induced at a 10% significance level. Effectiveness scores are reported for each peer group and model; the effectiveness score is the average, across seed levels, of the absolute value of the distance between the peer group’s detection rate and the maximum (across all peer groups) detection rate when none is induced at a 10% significance level. Effectiveness scores are reported for each peer group and model; the effectiveness score is the average, across seed levels, of the absolute value of the distance between the peer group’s detection rate and the maximum (across all peer groups) detection rate when none is induced at a 10% significance level.

The effectiveness scores and effectiveness ranks for each peer group are reported in the last two rows of each panel of Table 3. We find the cross-section peer group performs the worst, with effectiveness scores between 78.2% and 79.1% and an
effectiveness rank of 10 throughout. Peer groups based on firm age, sales, lagged sales and market capitalization perform better, with effectiveness scores between 84.1% and 85.7% and ranks between 6.20 and 8.30. The industry peer groups have effectiveness scores between 94.0% and 98.3% and ranks of 1.90–4.60; effectiveness ranks monotonically decrease over the 2-digit, 3-digit and 4-digit industry peer definitions. The lagged assets-based peer group has the highest effectiveness scores (98.9–99.6%) and lowest effectiveness ranks (1.60–1.70), and the assets-based peer group has effectiveness scores between 92.6% and 94.8% and effectiveness ranks between 3.6 and 4.5. We extend the analyses in Table 3 to seed levels between 20% and 100%, in 10 percentage point increments. Results of this analysis for the Jones model with intercept (results untabulated) show that when seeded discretionary accruals exceed 20%, the lagged asset-based peer group has the highest detection power in terms of both effectiveness scores and effectiveness ranks, among the estimation samples we consider.

We perform several sensitivity checks on these results. First, we include all non-event firms in the event-firm’s industry in the estimation sample. Therefore, there are more peer firms in the industry-based peer groups than in the size-based peer groups (where we continue to select 10 observations). Adding more firms to the industry peer groups adds degrees of freedom for estimating the accruals model, which should improve detection rates. Results of this analysis (not tabulated) are mixed: increasing the number of firms in the industry-based estimation sample reduces the detection rate for SIC2, does not affect the detection rate for SIC3, and increases the detection rate for SIC4. The relation between detection rates for these expanded samples of industry-based peers and detection rates for lagged assets-based peers is unaffected: the detection rate for lagged asset-based peers is indistinguishable from the detection rates for industry-based peers at seed levels less than 10%, and is higher for seed levels above 10%.

In a second sensitivity test, we align the size of the size-based peer groups with the size of the SIC4 peer group, for each event firm. This approach maximizes the estimation sample for the industry peers and ensures size-based peer groups are the same size as the industry-based peer groups. However, if multiple event-firm-years are selected from the same SIC4-year, the industry peer firms selected will be the same for the event-firm (with the exception of the event-firm itself), but the size-based estimation samples might shift considerably. Results (not tabulated) are similar to those reported.

In Table 3, we present results from estimating the modified Jones model using the change in sales and the change in accounts receivable in the estimation of the regression coefficients. Dechow et al. (1995) chose an alternative specification, and imported the coefficient estimates from a “regular” Jones model (without the adjustment for the change in accounts receivable) into the modified Jones model. We probe the sensitivity of our results to this variation in estimating the modified Jones model discretionary accruals, both allowing for an intercept and suppressing the intercept in the (regular) Jones model. In comparison to the modified Jones model results presented in Table 3, we find our results are largely unaffected by this design variation (results not tabulated). The effectiveness scores (effectiveness ranks) for lagged total asset peers are 99.3% (1.50) in the specification with intercept, and 99.6% (1.40) in the specification without an intercept, as compared to an effectiveness score (effectiveness rank) of 99.6% (1.60) in Table 3. For the industry peer samples, we find effectiveness score and ranks are very slightly lower. For example, the effectiveness score of SIC4 peers (SIC3 peers) [SIC2 peers] drops to 96.9% (96.7%) [93.5%] in the intercept specification, and slightly lower when the intercept is suppressed. We conclude that our qualitative results are not affected by this design choice, and the results reported in Table 3 are if anything conservative. As a result, we use the traditionally estimated modified Jones model residual in further tests.

To summarize, our simulation tests indicate that lagged asset-based peers perform at least as well as industry-based peers at detecting induced discretionary accruals. At low levels of seeded discretionary accruals differences in detection rates between industry-based peers and lagged asset-based peers are small; the differences become larger and favor lagged asset peers for higher seed levels.

3.5. Analysis of the simulation results

To shed light on the Table 3 finding that lagged assets-based estimation samples have discretionary accruals detection rates at least as good as the detection rates of industry-based estimation samples, we perform three analyses. These analyses are predicated on the view that industry membership and size are alternative indicators of similarity, albeit with different properties. For example, industry membership is dichotomous; two firms are either in the same industry or they are not. Size, however, is continuous, suggesting the possibility of gradations in similarity as the size-neighborhood of a given firm is expanded or contracted. In the absence of a theory as to which of these two indicators of similarity would function better in the context we consider, we aim to provide econometric insights into the Table 3 results.

Our first two analyses consider (1) what portion of an amount seeded into total accruals is detected in the discretionary accruals estimate, i.e., the regression residual; and (2) how seeding affects the standard deviation of discretionary accruals across the 11 firms in a simulation subsample. The regression residual forms the numerator, and the standard deviation forms the denominator, of the test statistics underlying the results in Table 3 of the difference between event-firm and non-event firms’ discretionary accruals. Our third analysis probes the link between estimation samples and the explanatory power of normal accruals models (with 0% seeded accruals) as compared to discretionary accruals models (with specified amounts of seeded accruals). Table 4 shows results of these analyses for the Jones model with intercept. Results (not tabulated) for the other models and seeding levels we consider are qualitatively similar. The table shows results of estimating the normal accruals models with no seeding (denoted ‘0%’) in columns 1, 4 and 7; columns 2, 5 and 8 report results for models with 20 percent (‘20%’) seeded accruals.
This table provides descriptive statistics from the Jones model (with intercept) regressions in the simulations of Table 3, averaged across the 50,000 samples. The first three columns contain the average event-firm discretionary accruals estimate in the benchmark case with no seeding (‘0%’) and in the case with the maximum seed (‘20%). The third column shows the increase in the discretionary accruals estimate, as a fraction of the 20% initial seed into total accruals. Columns 4 through 6 contain data on the average within-subsample standard error of the discretionary accruals. Columns 7, 8 and 9 show the explanatory power of the regression in the benchmark case, the maximum-seed case and the decrease in explanatory power after seeding (in percent).

Table 4
Average statistics on the simulated samples from Jones model.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Event (DA) (0%)</td>
<td>Event (DA) (20%)</td>
<td>Increase (in % of 20%)</td>
<td>SE (Event) (0%)</td>
<td>SE (Event) (20%)</td>
<td>Increase (in %)</td>
<td>$R^2$ (0%)</td>
<td>$R^2$ (20%)</td>
<td>Decrease (in %)</td>
</tr>
<tr>
<td>Entire cross-section</td>
<td>−0.0003</td>
<td>0.1272</td>
<td>63.76%</td>
<td>0.0906</td>
<td>0.0946</td>
<td>4.45%</td>
<td>0.5170</td>
<td>0.4747</td>
<td>−8.18%</td>
</tr>
<tr>
<td>SIC2</td>
<td>−0.0005</td>
<td>0.1272</td>
<td>63.81%</td>
<td>0.0835</td>
<td>0.0881</td>
<td>5.59%</td>
<td>0.5211</td>
<td>0.4669</td>
<td>−10.40%</td>
</tr>
<tr>
<td>SIC3</td>
<td>−0.0006</td>
<td>0.1266</td>
<td>63.60%</td>
<td>0.0828</td>
<td>0.0876</td>
<td>5.86%</td>
<td>0.5134</td>
<td>0.4616</td>
<td>−10.10%</td>
</tr>
<tr>
<td>SIC4</td>
<td>−0.0007</td>
<td>0.1266</td>
<td>63.62%</td>
<td>0.0823</td>
<td>0.0871</td>
<td>5.89%</td>
<td>0.5130</td>
<td>0.4618</td>
<td>−9.98%</td>
</tr>
<tr>
<td>Total asset neighbors</td>
<td>−0.0005</td>
<td>0.1281</td>
<td>64.28%</td>
<td>0.0821</td>
<td>0.0868</td>
<td>5.73%</td>
<td>0.5327</td>
<td>0.4786</td>
<td>−10.14%</td>
</tr>
<tr>
<td>Lagged total assets neighbors</td>
<td>−0.0007</td>
<td>0.1394</td>
<td>70.07%</td>
<td>0.0961</td>
<td>0.1002</td>
<td>4.26%</td>
<td>0.4449</td>
<td>0.3911</td>
<td>−12.10%</td>
</tr>
<tr>
<td>Sales neighbors</td>
<td>−0.0008</td>
<td>0.1268</td>
<td>63.83%</td>
<td>0.0931</td>
<td>0.0977</td>
<td>4.95%</td>
<td>0.4808</td>
<td>0.4394</td>
<td>−8.60%</td>
</tr>
<tr>
<td>Lagged sales neighbors</td>
<td>−0.0002</td>
<td>0.1274</td>
<td>63.80%</td>
<td>0.0942</td>
<td>0.0988</td>
<td>4.97%</td>
<td>0.4721</td>
<td>0.4312</td>
<td>−8.66%</td>
</tr>
<tr>
<td>Market capitalization neighbors</td>
<td>−0.0002</td>
<td>0.1272</td>
<td>63.72%</td>
<td>0.0882</td>
<td>0.0927</td>
<td>5.09%</td>
<td>0.5103</td>
<td>0.4645</td>
<td>−8.97%</td>
</tr>
<tr>
<td>Firm age neighbors</td>
<td>−0.0002</td>
<td>0.1271</td>
<td>63.65%</td>
<td>0.0910</td>
<td>0.0954</td>
<td>4.82%</td>
<td>0.4956</td>
<td>0.4528</td>
<td>−8.64%</td>
</tr>
</tbody>
</table>
The average event firm residual from estimating a normal (no seeded discretionary accruals) model, reported in column 1, shows all estimation samples function well in terms of yielding minimal evidence of abnormal accruals; that is, the measure of abnormal accruals is miniscule for all estimation samples, with average residuals ranging from $-0.0002$ to $-0.00008$. The largest negative residuals are associated with the SIC4, lagged assets and sales-based estimation samples, suggesting if anything, there is a bias against detecting seeded positive discretionary accruals for these estimation samples. Column 2 displays the residuals resulting from seeding 20% discretionary accruals into the event firm’s total accruals. Column 3 shows the portion of the seeded amount transferred into the discretionary accruals estimate as a percentage of the 20% seed. To the extent the seeded accrual influences the regression coefficients (i.e., the seeding changes the results of estimating the model of normal accruals), the residuals will increase by less than 20%; if the regression coefficients are entirely unaffected by the presence of the seeded accrual, the entire seeded amount will translate into the residual. For all but the lagged-asset-based estimation samples, the models detect between 63.6% and 64.3% of the seeded accruals; for the lagged-asset-based estimation samples, the residuals increase by 70% of the seed. This result suggests that more of the seeded accrual appears in the regression residual for estimation samples based on lagged assets, thus contributing to the detection power of the size-based estimation samples.

We next consider how the seeded accrual affects the standard deviation of discretionary accruals estimates within the sample of 11 firms. Columns 4 and 5 report the average standard deviation of residuals in the no-seed case and the 20%-seed case, respectively, followed by column 6 containing the percentage increase due to the seeding. These standard deviations, which would be the denominators in tests of differences in the discretionary accruals, will increase in the presence of seeded abnormal accruals, because the seeding adds an amount to the dependent variable that should not be explained by the explanatory variables. Results show the increases in standard deviation range from 4.3% (lagged assets) to 5.9% (SIC4), with increases generally smaller for the size-based estimation samples than for the industry-based estimation samples. Combined with the previously discussed finding, this result suggests the detection power of lagged-assets-based estimation samples derives from both a numerator effect (a larger mapping of the seeded accruals amount into the regression residual, the discretionary accruals estimate) and a denominator effect (a smaller standard deviation of residuals).

The preceding analyses show that differences in discretionary accruals detection rates for different types of estimation samples are linked to the extent to which the seeding of accruals affects the regression model itself, and therefore the regression residuals. To provide more evidence on this point, we focus on the average (over 50,000 samples) explanatory power for the zero-seed benchmark accruals model (column 7) and the 20%-seed accruals model (column 8). Column 9 reports the decrease in explanatory power when we seed 20% accruals, as an inverse measure of the extent to which the regression is influenced by the seeding. The lagged-assets-based estimation sample shows the largest decrease in explanatory power when abnormal accruals are seeded, about $-12.1\%$ (from column 9) and the lowest cross-sectional explanatory power for normal accruals, about 44.5% (from column 7); explanatory power for the other normal accruals models ranges from about 53.3% (total assets) to 47.2% (lagged sales) and the decline in explanatory power ranges from 10.4% (SIC2) to 8.2% (entire cross section).

We interpret these results as follows. First, viewing size and industry membership as two alternative measures of similarity, we would expect both size-based estimation samples and industry-based estimation samples to have substantial explanatory power for normal accruals models. We confirm this expectation, and also find industry-based models tend to have higher explanatory power for normal accruals. Second, seeding discretionary accruals increases only a single dependent variable observation (the event firm); detection power stems from the extent to which this single-observation increase in the dependent variable appears in the error term as opposed to influencing the regression coefficients. If the regression coefficients do not change at all, the entire seeded amount appears in the error term and would therefore be detected as discretionary accruals. The larger the decline in explanatory power of a model with seeded accruals, as compared to a model of normal accruals with zero seeding, the greater the portion of the seeded amount appearing in the error term (the unexplained portion of the shifted dependent variable). We find both industry-based and size-based estimation samples exhibit over 8 percentage point decreases in explanatory power in the presence of seeded accruals, and the decrease is largest for the lagged-assets-based estimation sample—that is, the seeding shifts the lagged-assets-based regression coefficients least.

Econometrically, therefore, the detection power of estimation samples based on lagged total assets derives from their relatively more robust, relatively more stable, regression coefficients in the presence of seeded discretionary accruals. Put another way, using estimation samples based on size results in a greater portion of a seeded amount appearing in the error term, the unstandardized measure of discretionary accruals, as opposed to in a shift in the regression coefficients. Although prior research has, perhaps implicitly, assumed that explanatory power for normal accruals is tantamount to detection power for abnormal accruals, our results suggest a tradeoff between two alternative indicators of similarity. Specifically, if the focus is on the explanatory power of a model of normal accruals, using industry-based estimation samples increases explanatory power and likely results in smaller sample sizes. However, if the focus is on detection power, as is the case in much earnings management research, using lagged assets-based estimation samples does not sacrifice detection power and substantially increases sample sizes.

4. Tests on restatement firms and AAER firms

In this section, we examine the performance of industry-based and size-based estimation samples in detecting unusual levels of absolute discretionary accruals in firm-years with after-the-fact acknowledged abnormal accruals. We analyze
Table 5
Detecting earnings management II – U.S. samples of negative-event firms.


<table>
<thead>
<tr>
<th>Entire cross section</th>
<th>SIC2</th>
<th>SIC3</th>
<th>SIC4</th>
<th>Total assets neighbors</th>
<th>Lagged total assets neighbors</th>
<th>Sales neighbors</th>
<th>Lagged sales neighbors</th>
<th>Market capitalization neighbors</th>
<th>Firm age neighbors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jones model</td>
<td>14.5%</td>
<td>9.5%</td>
<td>6.5%</td>
<td>8.5%</td>
<td>19.0%</td>
<td>39.5%</td>
<td>22.0%</td>
<td>26.0%</td>
<td>14.0%</td>
</tr>
<tr>
<td>Jones model (with intercept)</td>
<td>7.5%</td>
<td>6.5%</td>
<td>4.5%</td>
<td>7.5%</td>
<td>15.0%</td>
<td>26.0%</td>
<td>14.5%</td>
<td>15.5%</td>
<td>9.0%</td>
</tr>
<tr>
<td>Modified Jones model</td>
<td>8.0%</td>
<td>7.0%</td>
<td>3.5%</td>
<td>6.5%</td>
<td>10.5%</td>
<td>25.0%</td>
<td>16.0%</td>
<td>15.5%</td>
<td>9.0%</td>
</tr>
<tr>
<td>Average</td>
<td>10.0%</td>
<td>7.7%</td>
<td>4.8%</td>
<td>7.5%</td>
<td>14.8%</td>
<td>30.2%</td>
<td>17.5%</td>
<td>19.0%</td>
<td>10.7%</td>
</tr>
</tbody>
</table>

Panel B: Analysis of absolute discretionary accruals of restatements from Audit Analytics, 1994–2009

<table>
<thead>
<tr>
<th>Entire cross section</th>
<th>SIC2</th>
<th>SIC3</th>
<th>SIC4</th>
<th>Total assets neighbors</th>
<th>Lagged total assets neighbors</th>
<th>Sales neighbors</th>
<th>Lagged sales neighbors</th>
<th>Market capitalization neighbors</th>
<th>Firm age neighbors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jones model</td>
<td>5.5%</td>
<td>1.5%</td>
<td>1.5%</td>
<td>1.0%</td>
<td>8.0%</td>
<td>20.5%</td>
<td>10.0%</td>
<td>10.5%</td>
<td>6.0%</td>
</tr>
<tr>
<td>Jones model (with intercept)</td>
<td>2.5%</td>
<td>1.0%</td>
<td>1.5%</td>
<td>0.5%</td>
<td>3.5%</td>
<td>20.0%</td>
<td>4.5%</td>
<td>7.0%</td>
<td>3.5%</td>
</tr>
<tr>
<td>Modified Jones model</td>
<td>5.0%</td>
<td>1.5%</td>
<td>2.0%</td>
<td>1.5%</td>
<td>6.0%</td>
<td>19.0%</td>
<td>3.5%</td>
<td>8.0%</td>
<td>5.0%</td>
</tr>
<tr>
<td>Average</td>
<td>4.3%</td>
<td>1.3%</td>
<td>1.7%</td>
<td>1.0%</td>
<td>5.8%</td>
<td>19.8%</td>
<td>6.0%</td>
<td>8.5%</td>
<td>4.8%</td>
</tr>
</tbody>
</table>

Panel C: Analysis of absolute discretionary accruals of AAER firms, 1979–2002

<table>
<thead>
<tr>
<th>Entire cross section</th>
<th>SIC2</th>
<th>SIC3</th>
<th>SIC4</th>
<th>Total assets neighbors</th>
<th>Lagged total assets neighbors</th>
<th>Sales neighbors</th>
<th>Lagged sales neighbors</th>
<th>Market capitalization neighbors</th>
<th>Firm age neighbors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jones model</td>
<td>39%</td>
<td>25%</td>
<td>22%</td>
<td>23%</td>
<td>25%</td>
<td>52%</td>
<td>45%</td>
<td>38%</td>
<td>27%</td>
</tr>
<tr>
<td>Jones model (with intercept)</td>
<td>44%</td>
<td>22%</td>
<td>18%</td>
<td>19%</td>
<td>25%</td>
<td>48%</td>
<td>42%</td>
<td>35%</td>
<td>26%</td>
</tr>
<tr>
<td>Modified Jones model</td>
<td>49%</td>
<td>27%</td>
<td>23%</td>
<td>31%</td>
<td>22%</td>
<td>58%</td>
<td>60%</td>
<td>38%</td>
<td>30%</td>
</tr>
<tr>
<td>Average</td>
<td>44.0%</td>
<td>24.7%</td>
<td>21.0%</td>
<td>24.3%</td>
<td>21.3%</td>
<td>52.7%</td>
<td>49.0%</td>
<td>37.0%</td>
<td>27.7%</td>
</tr>
</tbody>
</table>

Panel A reports the fraction of the time GAO (Government Accountability Office) restatement firms’ absolute discretionary accruals differ significantly (at the 10% level) from the absolute discretionary accruals of non-restating peer firms, for three models of discretionary accruals. Panels B and C repeat the analysis for Audit Analytics restatement firms and for AAER firms, respectively. We assume restatement firms and AAER firms reported substantial discretionary accruals, so better (worse) peer groups should have a greater (smaller) frequency of detecting discretionary accruals.

three samples: 1337 restatement firm-years from two Government Accountability Office (GAO) reports covering fiscal years 1996–2006; 5032 restatement firm-years from the Audit Analytics database (AA firms) covering 1994–2009; and 372 firm-years named in SEC Accounting and Auditing Enforcement Releases during 1979–2002 (AAER firms).22 The event years in our analyses are all fiscal years during 1996–2006 (GAO restatements); all fiscal years ending within the restatement period (Audit Analytics restatements), or all fiscal years named in the Accounting and Auditing Enforcement Release (AAER firms).

In contrast to simulation tests with specified levels of induced (seeded) discretionary accruals, we take the existence of a restatement/AAER as evidence of material but unspecified amounts of discretionary accruals in reported earnings (we refer to both intentional and unintentional misstatements as discretionary accruals).23 In these analyses, we assess how often the

22 GAO restatement data are used by Hennes et al. (2008) and are available at Andrew Leone’s website (http://sbaleone.bus.miami.edu). We thank him for making these data available. We obtain the detailed dataset of AAER announcements during 1979–2002 from Dechow et al. (2011). We thank them for these data.

23 Hennes et al. (2008) conclude that about 24% of restatements are due to irregularities and 76% to errors, i.e., unintentional misapplications of authoritative guidance. For our purposes either would result in poor earnings quality that the models for detecting discretionary accruals should detect.
restatement/AAER firm’s estimated absolute discretionary accruals exceed the estimated absolute discretionary accruals of its peer firms. Under the view that restatement/AAER firms have, in fact, managed earnings such that their absolute discretionary accruals are larger than those of their peers, larger detection rates indicate the peer group is better at detecting discretionary accruals.

Our analyses of discretionary accruals are performed at the iteration level; we perform 200 iterations. To analyze the GAO restatement sample, we first select 100 event-firm-years from the population of Compustat firms with restatements; an event-firm-year is a firm-year with a restatement announcement in the 11 months following the fiscal-year end.\(^\text{24}\) For each event-firm-year, we randomly select 10 non-restating firms from the same year and from each peer group, where the 10 peer groups are as previously defined. We estimate the discretionary accruals models for each sample consisting of one event-firm and 10 peer firms, generating residuals for each of the 11 firms. We pool the absolute values of these residuals at the iteration level, generating 100 event-firm absolute residuals and 1000 peer firm absolute residuals per iteration. We focus on absolute discretionary accruals to avoid directional predictions of the intentional earnings manipulation or the unintentional error leading to the restatement.

For each iteration, we compare the mean absolute residual for the 100 event-firms with the mean absolute residual of the 1000 non-event firms, calculating a \(t\)-statistic for the difference. After 200 iterations, we have 200 \(t\)-statistics for the differences in mean absolute residuals.\(^\text{25}\) Because we focus on observed restatements/AAERs containing presumed material discretionary accruals, we expect the \(t\)-statistics to be significantly positive. The detection rate is the frequency of significant \(t\)-statistics, as a percentage of the 200 iterations. Our analyses focus on how the choice of peer groups affects this detection rate: holding the event firms constant, better (worse) peer groups will have larger (smaller) detection rates. Because our assumptions that restatement firms have managed or misstated their accruals and the accruals management affects the rate: holding the event firms constant, better (worse) peer groups will have larger (smaller) detection rates. Because our assumptions that restatement firms have managed or misstated their accruals and the accruals management affects the fiscal year prior to the restatement announcement may not hold for all observations, we do not expect to observe 100% detection rates.

Table 5 reports the detection rates, defined as the fraction of iterations where the \(t\)-statistic is significant at the 10% level or better; rows correspond to the accruals models and columns correspond to the peer groups. In panel A, the results for the GAO restatement sample show the lagged asset-based peer group has the highest average detection rate of 30.2%, ranging between 25.0% and 39.5%, depending on the accruals model. The other estimation samples yield average detection rates between 4.8% (SIC3) and 19.0% (lagged sales). In panel B, the results for the Audit Analytics restatement sample support similar inferences; the average detection rate for the lagged asset-based peer group is 19.8% (the range is 19.0–20.5%), while the average detection rates for other estimation samples range from 1.0% (SIC4) to 8.5% (lagged sales). Panel C, contains the results of the AAER analysis. Compared to the samples of restatement firms, overall detection rates increase substantially, consistent with an increased severity of accounting violations of AAER firms as compared to (possibly voluntary) restatements. The average (across accruals models) detection rate for the lagged assets estimation sample is 52.7%, ranging from 48% to 58%, and the other estimation samples show average detection rates between 21% (SIC3) to 49% (sales).

These findings from analyses of restatement firms and AAER firms are consistent with our simulation results: both show lagged assets work at least as well as, and sometimes better than, estimation samples based on industry membership in detecting discretionary accruals. We corroborate these findings by repeating our analyses using signed discretionary accruals (results not tabulated). In these analyses, we assess the frequency of significant differences in discretionary accruals at the 10% level, regardless of whether the differences are positive or negative. Overall detection rates across all estimation samples for both the two restatement samples and the AAER sample tend to increase compared to the tabulated results, although the detection rates for the entire cross-section and lagged asset-based estimation sample decrease slightly. Industry-based estimation sample detection rates average 22.5% across three industry definitions, models and samples of restatement/AAER firms, and the lagged-asset-based estimation sample shows the highest average detection rate, 31.8%.

5. Estimation sample selection for analyses of discretionary accruals in non-U.S. data

This section considers the relative power of industry-based estimation samples versus size-based estimation samples in detecting discretionary accruals in non-U.S. data. Section 5.1 provides evidence on the restrictiveness of industry membership as the sample selection criterion in non-U.S. markets, characterized by considerably fewer firms than the U.S. markets. Section 5.2 provides simulation-based evidence on the discretionary accruals detection power of industry-based peers versus size-based peers using non-U.S. data, and Section 5.3 summarizes key findings and inferences. Section 5.4 reports the results of our analysis of German restatement firms.

5.1. Restrictiveness of industry-based estimation samples in non-U.S. data

As discussed in the introduction, using industry-based peers to estimate discretionary accruals models imposes more substantial sample attrition for non-U.S. data than for U.S. data. To see this, we impose the same data requirements on

\(^\text{24}\) Analyses of the Audit Analytics restatement sample and the AAER sample follow the same design: because of differences in population size, we adjust the size of the event sample to 200 (Audit Analytics sample) and 50 (AAER sample). To decrease the probability of duplicating an event sample, we use 100 iterations of the simulation for the AAER sample (as opposed to 200 for the GAO and Audit Analytics samples).

\(^\text{25}\) This test is statistically equivalent to the approach used in the simulations tests, a regression on an event-firm dummy variable.
non-U.S. data that our simulation analysis imposed on U.S. data: we require each firm-year observation to have the data necessary for the accruals models and at least 11 firms per industry group (one event firm plus 10 non-event firms). We perform our analyses by country (we do not combine same-industry observations from two or more countries or same-country observations across industries). We start by requiring a country-year to have at least 11 firm-year observations on Compustat Global for the period 1988–2009. We use Compustat Global because it has more consistent and complete industry identifiers than other non-U.S. databases such as Datastream. These requirements result in 217,153 firm-years, representing 69 countries.

We calculate three measures of sample attrition for the 69 Compustat Global countries (results not tabulated). The first measure is the number of countries (of 69) eliminated entirely because of insufficient data to estimate the accruals models using industry-based peer groups. The SIC2 definition eliminates 29 of 69 countries, increasing to 37 and 40 for SIC3 and SIC4, respectively. The second measure of sample loss averages the country-specific losses of firm-year observations across the 69 countries, weighting each country equally. Using this measure, 76% of the sample observations are lost using an SIC2 definition, increasing to 89% and 93% for SIC3 and SIC4 definitions. The third measure is a weighted average version of the second measure, where the weights are each country’s sample size as a proportion of the total; the weighted average version

---

**Table 6**

Earnings management detection — effectiveness scores by country.

<table>
<thead>
<tr>
<th>Country</th>
<th>Entire cross-section</th>
<th>SIC2</th>
<th>SIC3</th>
<th>SIC4</th>
<th>Lagged total assets neighbors</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUS</td>
<td>85.8%</td>
<td>93.9%</td>
<td>92.3%</td>
<td>93.8%</td>
<td>99.4%</td>
</tr>
<tr>
<td>BMU</td>
<td>95.4%</td>
<td>97.5%</td>
<td>95.5%</td>
<td>97.4%</td>
<td>97.3%</td>
</tr>
<tr>
<td>BRA</td>
<td>93.7%</td>
<td>99.3%</td>
<td>98.8%</td>
<td>99.3%</td>
<td>94.4%</td>
</tr>
<tr>
<td>CHL</td>
<td>98.0%</td>
<td>97.3%</td>
<td>98.6%</td>
<td>97.8%</td>
<td>97.6%</td>
</tr>
<tr>
<td>CHN</td>
<td>81.0%</td>
<td>85.0%</td>
<td>87.3%</td>
<td>87.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>CYM</td>
<td>93.8%</td>
<td>90.2%</td>
<td>94.4%</td>
<td>87.6%</td>
<td>99.6%</td>
</tr>
<tr>
<td>DEU</td>
<td>86.1%</td>
<td>93.4%</td>
<td>97.8%</td>
<td>100.0%</td>
<td>96.6%</td>
</tr>
<tr>
<td>FRA</td>
<td>88.6%</td>
<td>87.2%</td>
<td>85.6%</td>
<td>85.9%</td>
<td>99.7%</td>
</tr>
<tr>
<td>GBR</td>
<td>82.5%</td>
<td>96.6%</td>
<td>95.0%</td>
<td>96.1%</td>
<td>98.0%</td>
</tr>
<tr>
<td>IND</td>
<td>86.2%</td>
<td>91.4%</td>
<td>92.6%</td>
<td>96.0%</td>
<td>99.8%</td>
</tr>
<tr>
<td>JPN</td>
<td>89.8%</td>
<td>94.8%</td>
<td>96.3%</td>
<td>96.0%</td>
<td>99.9%</td>
</tr>
<tr>
<td>KOR</td>
<td>84.3%</td>
<td>94.2%</td>
<td>96.7%</td>
<td>100.0%</td>
<td>91.9%</td>
</tr>
<tr>
<td>MYS</td>
<td>89.2%</td>
<td>96.0%</td>
<td>97.8%</td>
<td>96.8%</td>
<td>96.3%</td>
</tr>
<tr>
<td>NOR</td>
<td>81.4%</td>
<td>93.5%</td>
<td>99.9%</td>
<td>98.6%</td>
<td>82.5%</td>
</tr>
<tr>
<td>PAK</td>
<td>99.4%</td>
<td>95.1%</td>
<td>96.6%</td>
<td>98.0%</td>
<td>96.6%</td>
</tr>
<tr>
<td>RUS</td>
<td>99.3%</td>
<td>96.5%</td>
<td>97.9%</td>
<td>96.7%</td>
<td>95.8%</td>
</tr>
<tr>
<td>SGP</td>
<td>82.5%</td>
<td>97.3%</td>
<td>98.8%</td>
<td>99.0%</td>
<td>91.1%</td>
</tr>
<tr>
<td>SWE</td>
<td>93.6%</td>
<td>92.1%</td>
<td>88.4%</td>
<td>100.0%</td>
<td>94.5%</td>
</tr>
<tr>
<td>THA</td>
<td>93.1%</td>
<td>99.6%</td>
<td>99.6%</td>
<td>99.4%</td>
<td>92.4%</td>
</tr>
<tr>
<td>TWN</td>
<td>86.0%</td>
<td>88.9%</td>
<td>88.8%</td>
<td>91.8%</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

Average: 89.5% 94.0% 94.9% 95.9% 96.2%

Count if highest: 2 2 4 4 8

Count if lowest: 13 2 2 1 2

This table reports the effectiveness scores for selected peer group definitions, for the 20 Compustat Global countries with at least 100 firm-year observations under the SIC4 definition. We tabulate results for the Jones model with intercept; other accruals models produce similar inferences and are not shown. The effectiveness score is the average, across seed levels, of the absolute value of the distance between the peer group’s detection rate and the maximum (across all peer groups) detection rate for that seed level. An effectiveness score of 100% indicates the peer group is always the best at detecting discretionary accruals. The last row in the table reports the average effectiveness score for each peer group, calculated across the 20 countries. The Count if highest (lowest) reports the number of countries (of 20) for which the specified peer group has the highest (lowest) effectiveness score. The SIC3 peer group has the highest score in 5 countries and the lagged total assets peer group has the highest score in 11 countries.

---

26 Compustat Global contains firm-year observations for 99 countries, of which 30 countries do not meet the minimal restrictions we impose.
produces smaller measures of sample loss because it counts the sample loss for Japan (Bangladesh and Romania) more (less) in the overall measure. Weighted average sample losses are 32% (SIC2), 59% (SIC3) and 70% (SIC4).

The results of this analysis show that requiring the necessary data to estimate accruals models at the industry level imposes significant sample attrition on non-U.S. data, and even eliminates a substantial number of countries. In contrast, using a size-based estimation sample generates no sample attrition beyond the loss imposed by the accruals model itself. Basing estimation samples on similarity in lagged assets instead of industry membership, therefore, offers the possibility of much larger samples, including more countries. That said, in a country with relatively few firms, the average size-spread within a size-based peer group could be large, which could reduce the power of size-based peers to detect discretionary accruals. Therefore, it is an empirical question whether the detection power of size-based peers is as good as the detection power of industry-based peers in countries other than the U.S.

5.2. Detecting induced discretionary accruals in non-U.S. data

We compare industry-based and size-based peer groups in terms of their ability to detect discretionary accruals in non-U.S. data. Our analysis applies the simulation tests reported in Table 3 to each of the 69 countries with available data, and restricts the simulations to 50 event firms in each of 100 iterations. The most restrictive requirement is that for each randomly selected event firm there be 10 non-event firm observations in a 4-digit SIC code for that country. Only 29 of 69 non-U.S. countries meet this requirement; of these, nine countries have too few observations (less than 100 firm-years) to perform the bootstrapping required by our simulation. We therefore analyze the “restricted sample” of 20 countries with at least 100 firm-year observations identifiable under SIC4.

For this “restricted sample,” we analyze the performance of industry-based and size-based peer groups at detecting induced discretionary accruals. We tabulate results for the Jones model with intercept and for the entire cross-section, three industry-based peers, and lagged asset-based peers. Results (not tabulated) for the other models and peer groups yield similar inferences. We analyze all seed levels and report results for the aggregate effectiveness score which averages detection rates (measured relative to the best detection rate for that seed level) across all seed levels.27 We also report a count of how many times (of the 20 restricted sample countries) a specified peer group has the highest effectiveness score. The results, reported in Table 6, show the lagged asset-based peer group has the highest average effectiveness score (96.2%) calculated across the 20 countries and a higher effectiveness score for more individual countries than any industry-based peer group (for example, the lagged asset peer group has the highest effectiveness score in 8 countries versus 4 countries for both the SIC3 and SIC4 peer groups).

The tests reported in Table 6 hold the size of each peer group constant at 11 observations (one event firm plus 10 non-event firms). Applying this requirement to industry-based peer groups results in substantial sample loss, especially for SIC4. Given that lagged assets-based peers perform better than industry-based peers for the constant-size peer groups, we investigate the performance of lagged asset-based peers when we do not impose the industry requirements. Basing estimation samples on similarity in lagged assets imposes no incremental sample loss, so in theory we can perform this analysis for the 69 Compustat Global countries with at least 11 observations in a given year. Because we require each country to have at least 100 firm-year observations (to perform the bootstrapping required by our simulation), we analyze the 58 non-U.S. countries with at least 100 firm-year observations. The detection rates for these 58 countries (not tabulated) show the lagged asset-based peers detect the seeded discretionary accruals 25.9% of the time at a 10% seed level in the Jones model with intercept. For the 20 countries in both the restricted sample and the 58-country sample, there will be differences in detection rates from including additional data when there are no industry restrictions. For these 20 countries, the average detection rate at a 10% seed level in the Jones model with intercept when industry data are required is 24.7%; this detection rate is not reliably different at conventional levels from the 25.9% detection rate for the 58-country sample. Results for other seed levels and other models yield similar inferences.

5.3. Summary of simulation analyses involving non-U.S. data

We draw two inferences from simulation analyses of non-U.S. data. First, estimating discretionary accruals models using non-U.S. data and industry-based peers results in significant sample attrition. As a result, researchers may face the problem of low detection power for discretionary accruals and may find the analysis restricted to the few countries with markets large enough to support the data requirements (Japan, China, Great Britain). Alternatively, the researcher could aggregate data across countries, within industry, to increase the number of firm-year observations per industry. This approach treats all observations for a given industry, across countries, as similar, so it ignores the influence of jurisdiction-specific factors (e.g., accounting standards, legal systems, degree of market development). This approach is problematic if the researcher is interested in whether and how jurisdictional factors affect discretionary accruals because combining observations across countries will obscure the phenomenon being investigated. Finally, researchers could use a country-cross-section as the estimation sample. Our results indicate this “entire cross-section” peer group is less powerful at detecting discretionary accruals than either industry-based peers or size-based peers.

27 Untabulated analyses confirm the models are unbiased for the non-U.S. data (i.e., we observe 5% detection rates for the 0% seed level).
The second inference is that size-based estimation samples perform at least as well as, and often better than, industry-based estimation samples for detecting discretionary accruals in non-U.S. data. This finding means researchers can estimate accruals models using lagged asset-based peers with no sample loss incremental to the loss from requiring the firm to have data to calculate the variables included in the accruals models. Therefore, countries which cannot be analyzed using industry-based peers can be analyzed using lagged asset-based peers. Increasing the number of countries for which researchers can estimate discretionary accruals models should increase the power of research designs examining whether and how jurisdiction-specific factors influence managers’ ability and incentives to engage in discretionary accruals, because the larger samples should include both more and more diverse countries.

5.4. International restatement analysis

We analyze restatements for German listed firms during 2004–2009. The Deutsche Prüfstelle für Rechnungslegung (German Financial Reporting Enforcement Panel, FREP), established in 2005, examines financial reports of publicly listed entities, selected randomly or upon request from the Bundesanstalt für Finanzdienstleistungsaufsicht (BaFin, Federal Financial Supervisory Authority) and has, effectively, the power to impose a restatement. From public sources, we hand-collected data on 169 announcements about financial misrepresentation in 2004–2009, made public between February 2006 and June 2012. From these 169 announcements, we identified 83 unique restatements by German firms with the necessary Compustat Global data to estimate the accruals models.

The second inference is that size-based estimation samples perform at least as well as, and often better than, industry-based estimation samples for detecting discretionary accruals in non-U.S. data. This finding means researchers can estimate accruals models using lagged asset-based peers with no sample loss incremental to the loss from requiring the firm to have data to calculate the variables included in the accruals models. Therefore, countries which cannot be analyzed using industry-based peers can be analyzed using lagged asset-based peers. Increasing the number of countries for which researchers can estimate discretionary accruals models should increase the power of research designs examining whether and how jurisdiction-specific factors influence managers’ ability and incentives to engage in discretionary accruals, because the larger samples should include both more and more diverse countries.

5.4. International restatement analysis

We analyze restatements for German listed firms during 2004–2009. The Deutsche Prüfstelle für Rechnungslegung (German Financial Reporting Enforcement Panel, FREP), established in 2005, examines financial reports of publicly listed entities, selected randomly or upon request from the Bundesanstalt für Finanzdienstleistungsaufsicht (BaFin, Federal Financial Supervisory Authority) and has, effectively, the power to impose a restatement. From public sources, we hand-collected data on 169 announcements about financial misrepresentation in 2004–2009, made public between February 2006 and June 2012. From these 169 announcements, we identified 83 unique restatements by German firms with the necessary Compustat Global data to estimate the accruals models.

Our analysis of the German restatement data parallels the analysis of U.S. restatement data, reported in Table 5. The “maximized sample,” with no additional restrictions, contains 83 event observations and 3085 German listed peer non-
restatement firms; the "restricted sample," which requires at least 10 peer firms in the same industry, has 15 event observations in the SIC4 sample and 607 peer firms. Because of the much smaller samples used in this analysis, as compared to the analysis of U.S. data, we limit the number of event firms per simulation to 80% of the total event firm count (i.e., 12 for the restricted sample and 66 for the maximized sample). We perform 100 runs of the simulation.

Results of the analysis are in Table 7, including descriptive data (Panel A) and detection rate results for the restricted (Panel B) and maximized (Panel C) samples. Generally, the restatement samples have larger mean and median absolute accruals, with less dispersion, than the respective peer samples. Compared to the maximized sample, the restricted sample of German restatement firms has larger mean and median values of absolute total accruals. The detection rate results in Panel B indicate no meaningful detection power on average for the industry-based samples using the Jones model or its modifications. The entire cross-section and the lagged total assets peer samples provide average detection rates of 2.3% and 4.0%, respectively. We conclude there is little detection power in tests using this restricted sample. Panel C presents results for the larger 'maximized' sample of 83 restatement firms and 3085 peers, for which industry-based peer definitions are, by construction, not feasible. The average detection rates are 4.3% for the estimation sample randomly drawn from all non-restatement firms and 39.3% for the estimation sample based on lagged total assets.

The analyses of German restatement firms support our previous findings based on analysis of non-U.S data. First, using industry membership as the estimation sample selection criterion imposes substantial sample losses. Second, basing estimation samples on similarity in size as measured by lagged assets imposes no sample restrictions beyond those imposed by the accruals model itself and increases the power of widely-used accruals models to detect abnormal accruals.

6. Additional tests

This section reports the results of several investigations of the robustness of our main finding that peer groups based on similarity in size, as measured by lagged assets, perform at least as well as industry-based peer groups in detecting discretionary accruals. We investigate rejection rates in the 0% seed case, the effects of peer group size, alternative accruals models and industry definitions, and scaling. A final subsection provides supplementary analysis of within-industry accruals homogeneity.

6.1. Rejection rates in the 0% seed case

Our main test for detecting discretionary accruals is based on rejecting, at the 10% level, the hypothesis that the slope coefficient in Eq. (4) is zero. In the 0% seed case, we expect the rejection rate for positive (negative) accruals to be 5% (5%). Our first sensitivity analysis probes the rejection rates of 5.8–6.1% reported in Table 3 for lagged asset-based peers and the 0% seed level. We wish to determine whether the higher detection rates for positive seed levels observed for lagged asset-based peers are driven by the greater than 5% rejection rates found for the 0% seed level. The concern is over-rejection in the 0% seed case propagates, possibly nonlinearly, into higher detection rates for larger seed levels.

We first examine whether the over-5% rejection rates for the 0% seed level are observed across all lagged asset deciles, and find over-rejection rates are concentrated in smaller firms. If higher rejection rates at 0% seed levels for small firms drive our main findings about the detection power of lagged asset-based peer groups, we should observe the highest detection rates for small firms and seeded positive amounts of discretionary accruals. To test this conjecture, we calculate detection rates by lagged asset decile for all the seed levels and models we consider; we find no evidence that the highest detection rates are associated with small firms (results not tabulated). In fact, detection rates are higher for larger firms, whose results are better-specified in the 0% seed case (i.e., closer to the theoretical benchmark of 5%). We conclude from these results that the possibility of bias when estimation samples are based on lagged assets does not drive our finding that size-based peers perform as well as, or better than, industry-based peers in detecting discretionary accruals. Instead, the results point to a higher variation in the total accruals of small firms which is not explained by the combinations of discretionary accruals models and estimation sample selection criteria we consider.

6.2. Varying the estimation sample size

In this section, we first investigate the association between the total number of firms in an industry and detection rates for discretionary accruals. We hypothesize that using the entire industry as the estimation sample in a given year (i.e., maximizing the number of industry peer firms) leads to lower test power, which in turn supports our design choice to standardize the number of peers in our main tests (at 10 peers per event firm-year). Our tests repeat the simulations described in Section 3 using a stratified-by-industry design. Specifically, we sample 250 event-firms from each of the 54 SIC2 industries with replacement. Each event firm is matched with either all or, consistent with the design in the main tests, 10 randomly chosen peer firms from the same SIC2-year. The average number of peers per year (not tabulated) ranges from 369 (for SIC2 industry 73, Business Services) to 12 (SIC2 industry 17, Construction). Eight of the 54 SIC2 industries average more than 100 peers per year and 36 SIC2 industries contain fewer than 50.

For the benchmark (0% seed) rejection rates for each SIC2 industry and three accruals models, we find the average benchmark rejection rates across the 54 SIC2 groups are close to the expected 5%, with considerable cross-industry variation (results not tabulated) Average rejection rates range from 4.4% to 4.5% for estimation samples containing all industry peers
and from 5.2% to 5.3% when 10 industry peers are used. We also calculate the pairwise correlation coefficients ($\rho$) between the average number of peers and the rejection rate as well as the corresponding p-value of a two-sided test that $\rho = 0$ (results not tabulated).\footnote{28 This correlation is equal to the correlation between the average number of peer firms and the deviation of the rejection rate from its benchmark value of 5%. Using the overall number of firm-years in the industry instead of the average number of peer firms yields similar results (not tabulated).} When all peers in an industry-year are used, the correlations range from $-0.46$ to $-0.52$ (significant at the 0.0005 level or better), implying a greater incidence of rejection in industries with fewer observations. Moreover, the five industries with the largest (smallest) average number of peers show average rejection rates of 2.9% (6.1%) across the three models. When the test design is standardized to using 10 random peer firms, none of the correlations is significantly different from zero at conventional levels. The same five top (bottom) industries by size show rejection rates of 4.9% (5.2%) on average. In the context of our simulation tests, these results indicate that controlling the number of peer firms has a meaningful effect on detection rates under the null hypothesis of zero discretionary accruals when there are zero seeded accruals. These findings also suggest the number of peer firms may matter more than the choice of the accruals model in calibrating the design towards the theoretically correct 5% benchmark rejection rate in the 0% seed level case.

Given this finding, and our general finding that a size-based peer group often performs best at detecting discretionary accruals, a natural question is: what is the optimal number of estimation firms? To maximize the number of SIC4-years, we require 10 non-event firms (plus the event firm) to estimate the accruals models. In the second set of tests in this section, we probe the effects of expanding the size-based estimation sample beyond 10. To provide evidence on this question, we repeat our main simulation analyses increasing the number of lagged asset-based peers from 10 firms (base case) to 20, 30, 50, 100, 250, 500 and 1000 firms. We calculate detection rates for seeding levels between 0%, the benchmark case and 20% and the three models considered in Table 3 (results not tabulated).

Results for the 0% seed level, which provide a measure of the unbiasedness for each peer group size, show the rejection rate declines as the estimation sample increases from 10 firms, falling below 4% for samples exceeding 250. This is consistent with the across-industry evidence of a negative correlation between peer group size and rejection rates. For seeded positive discretionary accruals, the highest detection rates are generally achieved with estimation sample sizes of 10 and 20.\footnote{29 We also considered peer sizes between 10 and 20 in increments of one and found no consistent pattern.} The results provide additional support that a benchmark rejection rate exceeding 5% does not necessarily translate into higher rejection rates at higher seed levels. Specifically, the highest benchmark rejection rate is observed for $n = 10$ (5.8% for the 0% seed level), yet the highest detection rates for positive seed levels are not observed for $n = 10$ past the 8% seed level, and using 20 peers results in the highest rejection rate at the 20% seed level.

### 6.3. Alternative accruals models

We repeat our main analyses using several additional accruals models. First, we use a Healy-type (1985) model which uses the average accruals for the peer firms as the benchmark for normal accruals (i.e., there are no explanatory variables for normal accruals beyond peer group membership). Second, we use four different measures of “performance-adjusted discretionary accruals” (PADA). Based on Kothari et al. (2005), we define the first two PADA-variants as the difference between two regression residuals, where the second residual is selected from an ROA-matched firm in the same peer group. The regression residuals are based on the Jones model (with intercept) or the modified Jones model (Eqs. (2) or (3)). The third performance-adjusted measures are the residuals from Eq. (5), which adds a performance adjustment (current period ROA) to Eq. (2).

\[
\frac{\text{Total accruals}_{i,t}}{\text{Total Assets}_{i,t-1}} = \alpha_0 + \alpha_1 \frac{\Delta \text{Sales}_{i,t}}{\text{Total Assets}_{i,t-1}} + \alpha_2 \frac{\Delta \text{ROA}_{i,t}}{\text{Total Assets}_{i,t-1}} + \alpha_3 \frac{\text{Net PPE}_{i,t}}{\text{Total Assets}_{i,t-1}} + \alpha_4 \text{BM}_{i,t} + \alpha_5 \frac{\Delta \text{AR}_{i,t}}{\text{Total Assets}_{i,t-1}} + \text{DA}^{\text{ones}-\text{ROA}}_{i,t}
\]

(5)

Our last performance-adjusted discretionary accruals measure is based on Larcker and Richardson’s (2004) expansion of the modified Jones model to include proxies for expected operating growth (the book-to-market ratio, BM) and current operating performance (operating cash flows, CFO).

\[
\frac{\text{Total accruals}_{i,t}}{\text{Total Assets}_{i,t-1}} = \alpha_0 + \alpha_1 \frac{\Delta \text{Sales}_{i,t} - \Delta \text{AR}_{i,t}}{\text{Total Assets}_{i,t-1}} + \alpha_2 \frac{\Delta \text{ROA}_{i,t}}{\text{Total Assets}_{i,t-1}} + \alpha_3 \frac{\text{Net PPE}_{i,t}}{\text{Total Assets}_{i,t-1}} + \alpha_4 \text{BM}_{i,t} + \alpha_5 \frac{\text{CFO}_{i,t}}{\text{Total Assets}_{i,t-1}} + \text{DA}^{\text{IR}}_{i,t}
\]

(6)

We repeat our main tests using these five models, the simulations on U.S. data described in Section 3, the restatement/AER analyses of U.S. firms and the restatement analysis of German firms. Results of these tests are not tabulated. In the simulation tests, we find the results from the Healy model are similar to the tabulated results for the Jones model with intercept. Comparing detection rates across models, there is evidence of decreasing detection rates across all peer group definitions when performance is explicitly controlled for. The detection rates are particularly small for the two first-difference measures of performance-matched regression residuals. More importantly, we find that, holding the accruals model constant, lagged asset peers continue to perform at least as well as industry peers, and SIC4 continues to be the best-performing industry definition.
The results on the U.S. restatement/AAER firms and German restatement firms confirm the finding of lower overall detection power of the performance-adjusted measure; detection rates in the U.S. samples are not much different from 5%, and are mostly lower than 5%. Lagged asset-based estimation samples average a detection rate of about 14.5% compared to about 2.9% across all models and industry-based estimation samples. The pattern of lower detection rates for the performance adjusted measures also applies to the German restatement sample, where detection rates average around 2% for the lagged total asset peers. Within the models we consider, there is substantial variation in detection rates. The Larcker-Richardson model and the PADA model based on modified Jones yield a detection rate of 0%, while the Jones model augmented with ROA shows a 21% detection rate.

Overall, we conclude that our inferences about estimation samples are not sensitive to the choice of the discretionary accruals model. In a comparison across estimation samples, lagged total assets-based estimation samples continue to perform at least as well as estimation samples based on industry membership, independent of the chosen model.

6.4. Alternative industry definitions

We used SIC codes to define industry classification to connect to previous research in this area which primarily uses SIC codes. We examine the sensitivity of our results to three other industry definitions: NAICS, Global Industry Classification Standard (GICS), and historical SIC codes from the top ranked segment (in terms of sales revenues) as reported on the Compustat Segments database. Because we retain the requirement of at least 11 firms in the most narrow industry definitions, the samples necessarily change. We find discretionary accruals detection rates for these alternative industry definitions are lower than those reported for the SIC definitions in Table 3 (results not tabulated). Lagged assets-based estimation samples perform better than estimation samples based on these alternative industry definitions in terms of both effectiveness scores and effectiveness ranks. In the restatement tests, the discretionary accruals detection rates of the three alternative industry definitions are also considerably weaker than the SIC-based detection rates, and detection rates using lagged asset-based estimation samples continue to be higher than detection rates using industry-based estimation samples.

6.5. Scaling

To examine whether using lagged assets as the scaling factor (the denominator) in the accruals models influences our main finding, we repeat our tests using lagged sales as the scaling factor. If the choice of denominator drives our results, we should find lagged sales-based peers have approximately the same discretionary accruals detection power as lagged asset-based peers. Results (not tabulated) show lagged sales-based peers do not have the highest discretionary accruals detection power when we use lagged sales as the scaling factor. The effectiveness scores for lagged sales-based peers are about 85–93%, and are always below the scores for lagged asset-based peers. These results suggest the scaling factor used in the accruals model is not driving our detection rate results.

These findings also speak to the question of whether the superior performance of lagged assets-based peers is driven by using lagged assets both as a parametric control in the accruals model (i.e., it is used to scale the variables) and as a non-parametric control (i.e., it is used to define the estimation sample). If any numerical variable would produce superior detection rates when used both parametrically and non-parametrically, then we would expect to find superior detection rates for lagged sales or any other variable used both as a parametric control and as the criterion for defining the estimation sample. We do not observe this result for lagged sales or for other variables we examine (results not tabulated). This finding suggests the detection rate results for lagged assets-based estimation samples are not due solely to econometric specification.

6.6. Supplementary analysis of within-industry accruals homogeneity

This section reports results of supplementary analyses of properties of accruals that we interpret as capturing homogeneity in the accruals process. The cross-sectional explanatory power results reported in Table 4 indicate industry-based peers are superior to size-based peers, if the research goal is an estimate of normal accruals.30 In this section, we provide additional evidence on the congruence of the total accruals and normal accruals generating processes within industry, relative to several size-based peer groups. Our analyses control for sample size to arrive at conclusions we believe would generalize to typical earnings management research settings. In a specific research setting designed to maximize homogeneity in accruals, however, a researcher would trade off increasing sample size by using size-based peers against increasing estimation efficiency by using a likely-smaller sample of industry peers.

The tests in this section focus on homogeneity in the firm-specific time-series behavior of total accruals, with the goal of assessing the effects of constraining accruals processes to be the same in the cross section of the estimation sample. For these tests, we require both a minimum of 10 consecutive fiscal years of accruals data and a standardized, directly comparable number of peer groups each year. Therefore, we calculate the number of distinct industry groups in each year.

---

30 This goal is not necessarily congruent with the goal, in this paper, of maximizing detection power for discretionary accruals.
Coefficients of the Jones model. The table reports the number of times, out of 1000 samples, this coefficient equality at the 0.01 level (out of 1000 samples). Fewer rejections indicate higher homogeneity within a group, as proxied by the regression.

### Panel A: Cluster analyses on total accruals

<table>
<thead>
<tr>
<th>Groups determined by</th>
<th># Clusters</th>
<th>SIC (as defined by grouping)</th>
<th>Average number of</th>
<th>( F ) test rejects coefficient equality at the 0.01 level ( F ) test rejects coefficient equality at the 0.01 level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Groups</td>
<td>Total assets</td>
<td>Lagged total assets</td>
</tr>
<tr>
<td>SIC2</td>
<td>10</td>
<td>58</td>
<td>40</td>
<td>52</td>
</tr>
<tr>
<td>SIC3</td>
<td>10</td>
<td>210</td>
<td>92</td>
<td>125</td>
</tr>
<tr>
<td>SIC4</td>
<td>10</td>
<td>319</td>
<td>116</td>
<td>148</td>
</tr>
<tr>
<td>SIC2</td>
<td>25</td>
<td>58</td>
<td>28</td>
<td>40</td>
</tr>
<tr>
<td>SIC3</td>
<td>25</td>
<td>210</td>
<td>51</td>
<td>65</td>
</tr>
<tr>
<td>SIC4</td>
<td>25</td>
<td>319</td>
<td>59</td>
<td>71</td>
</tr>
</tbody>
</table>

This table provides analyses of accruals homogeneity. For the test based on cluster analysis in Panel A, a pre-determined number of clusters (either 10 or 25) is formed based on the pairwise firm-specific time-series correlations of total accruals. The table contains the maximum number of distinct groups (‘Groups’) each year, depending on the industry definition. If a peer group is random with respect to the time-series correlations of accruals, the maximum of 58 (210, 319) peer groups would be in the average cluster. If a peer group exactly mimics the clusters, 58 (210, 319) groups would be in the average cluster. Columns 3 through 8 contain the average number of industry and size groups represented in the average cluster. Fewer peer groups represented in the clusters suggests more homogeneity, as proxied by the time-series behavior (correlations) of total accruals. Panel B contains the results of \( F \)-tests of equality of regression coefficients from firm-specific Jones model regressions using 1000 random samples of 10 firms each; we jointly test if the coefficients in Eq. (1) are equal across the firms in a given sample. The table reports the number of times, out of 1000 samples; this \( F \)-test rejects coefficient equality at the 0.01 level (out of 1000 samples). Fewer rejections indicate higher homogeneity within a group, as proxied by the regression coefficients of the Jones model.

and form the same number of peer groups for the size-based criteria (total assets, lagged total assets, sales, lagged sales, market capitalization). The number of groups per year depends on the industry definition, SIC2 or SIC3 or SIC4. The first test, which focuses on total accruals, is based on cluster analyses yielding a pre-determined number of clusters, either 10 or 25, from the time-series correlation of (scaled) total accruals. Each year, we estimate the pairwise correlation of total accruals over the preceding 10 years for each firm pair. We transform the resulting correlation matrix into a distance matrix to use in a clustering algorithm, which returns 10 (or 25) clusters, where within-cluster (between-clusters) variation of firm-specific correlations is minimized (maximized). For each annual cluster, we count the number of distinct peer groups (for example, industry groups, total assets groups) represented. This test is based on the following reasoning: if the firms in a peer group have the most homogeneous accruals processes, the entire peer group would be in a single cluster, and the number of distinct peer groups in each cluster will be small. The lower limit on the number of distinct peer groups in a cluster is the number of peer groups divided by the number of clusters formed, either 10 or 25. If, however, peer groups are random with respect to homogeneity in the accruals, as captured by the time-series correlation, we would expect each peer group to be represented in each cluster.

Table 8, Panel A, reports results from the annual cluster analyses and the associated peer group counts per cluster. The average year has 58 (210) (319) peer groups based on SIC2 (SIC3) (SIC4). When we impose 10 clusters per year, a peer group selection criterion resulting in highly homogeneous peer groups would yield an average of 5.8 (21.0) (31.9) peer groups per cluster, and random peer groups would be represented 58 (210) (319) times per cluster. We find 40 (92) (116) different SIC2 (SIC3) (SIC4) industries are represented in the average cluster on total accruals. While these amounts greatly exceed the minimums, they compare favorably to the results for size-based peer group definitions, where, for example, the average lagged-total-assets cluster contains 52 (124) (148) peer groups. We repeat the analysis with 25 clusters and obtain qualitatively similar results, reported in the second half of Panel A.

---

31 After imposing the additional data constraints, the firm age criterion does not yield a sufficient number of distinct groups in each year. We therefore omit firm age as a peer group criterion from the analyses in this section.

32 Industry groups do not contain equal numbers of firms in a given year; this variation in group sizes cannot be meaningfully mimicked in the construction of the size-based peer groups. We do not believe this variation in group size introduces a systematic bias into our cluster-analysis-based test, in particular not a bias favoring industry. We draw similar conclusions from our second test, based on coefficient equality, where we standardize the number of peer firms at 10.
Our second test, which focuses on normal accruals, applies an F-test for coefficient equality to the firm-specific estimated coefficients in the Jones model. The hypothesis is that two firms with the same normal accruals generating process will have statistically indistinguishable coefficients on the explanatory variables. For these tests, we require 10 firms per annual industry cross section. In the most extreme case, requiring 10 firms per SIC4-year, the sample contains 109,473 firm-year observations. We first randomly draw 1000 industry-and-year pairs, and then 10 firms for each industry-year. The combination of our time-series and cross-sectional data requirements ensures a full panel of data for each sample; i.e., each draw consists of 10 firms with 10 annual observations to estimate firm-specific regressions. We estimate a firm-specific Jones model for each firm in a given sample over the 10 observations; using a joint F-test, we test whether the slope coefficients \( \alpha_1, \alpha_2 \) and \( \alpha_3 \) in Eq. (1) are equal for all 10 firms.

Panel B of Table 8 reports the proportion of samples (out of 1000) for which the test rejects coefficient equality at the 0.01 level or better. The more homogeneous the accruals generating processes of the firms in a group, as captured by statistically indistinguishable Jones model coefficients, the more likely the test will fail to reject equality of the coefficients. That is, smaller numbers in the table suggest a more homogeneous normal accruals generating process within groups. Across the three industry definitions, we find that F-test rejects coefficient equality for 576–641 of the 1000 industry groups, or approximately 58–64%. This rejection rate is both high in absolute terms and substantially lower than the rejection rates for the size-based peer group definitions, averaging approximately 82.5% or higher.

To summarize, we perform two tests of homogeneity in the total accruals generating process and the normal accruals generating process to supplement our analysis of how estimation sample selection criteria affect the power to detect abnormal accruals. The first test defines homogeneity in terms of the time-series correlation of total accruals and the second test defines homogeneity in terms of equality of firm-specific regression coefficients from a Jones model. Both tests confirm greater homogeneity of industry-based estimation samples, with respect to the accruals attributes we consider, relative to size-based estimation samples. Industry-based estimation samples should therefore provide the most efficient cross-sectional estimates of normal accruals, subject to concerns about (possibly considerable) sample attrition. In addition, industry definitions of normal accruals seem far from perfect, confirming previous research, e.g., Bernard and Skinner, 1996; Brickley and Zimmerman, 2010; Dopuch et al., 2012, suggesting considerable accruals heterogeneity among firms in the same industry.

7. Summary and conclusions

We examine the ability to detect discretionary accruals using several variants and extensions of the Jones (1991) model of discretionary accruals and estimation samples based on two alternative indicators of similarity: industry membership (industry peers) and size (size peers). Our examination is motivated by the practical problem of sample attrition when estimation samples are based on industry membership, particularly the SIC4 industry definition, and particularly for non-U.S. data.

Our main finding is that estimation samples based on similarity in size as measured by lagged total assets perform at least as well as industry membership-based estimation samples, and often better, in detecting both seeded discretionary accruals and observed discretionary accruals (as proxied by the existence of a restatement or an Accounting and Auditing Enforcement Release). The superior discretionary accruals detection power of lagged asset-based estimation samples applies to both U.S. data and non-U.S. data. For non-U.S. samples not constrained by the availability of industry peers, lagged asset-based estimation sample detection rates are similar to the detection rates observed for samples where we can perform a controlled comparison of detection rates for industry-based peers and lagged asset-based peers.

We provide evidence of a tradeoff between increasing explanatory power of normal accruals models and increasing detection power for abnormal or discretionary accruals. While both size-based and industry-based estimation samples produce reasonable explanatory power in estimating normal accruals, the industry-based samples achieve higher explanatory power. This result is corroborated by additional analyses showing industry-based samples are characterized by normal accruals with greater congruity than are size-based samples, although neither sample selection criterion yields normal accruals that are wholly or even substantially congruent. In contrast, size-based estimation samples, in particular, samples based on similarity in lagged total assets, often yield higher detection rates for abnormal or discretionary accruals. Viewed as a whole, these results suggest the greater detection power of the size-based estimation samples is due to the greater stability of the accruals model regression estimated using size-based samples, as compared to using industry-based samples, and this greater stability comes at the cost of lower explanatory power.

Defining estimation samples (peer firms) based on similarity in lagged assets instead of industry membership has substantial practical value in estimating discretionary accruals models because application of the size-based criterion imposes no incremental sample loss, beyond the sample losses resulting from estimating the variables in the models. For U.S. data, this means avoiding sample attrition of anywhere from 1–3% (SIC2 definitions) to 22–30% (SIC4 definitions). The benefits, in terms of increased sample sizes, are much greater for non-U.S. data, where using lagged assets instead of industry membership to identify estimation samples avoids sample attrition ranging from 32% to 93% (depending on industry definitions and weighting schemes).

---

33 We do not impose \( \alpha_1 = \alpha_2 = \alpha_3 \) for each firm.
We believe our finding concerning the discretionary accruals detection power of lagged asset-based estimation samples is important both in the U.S. context and in the non-U.S. context. For both settings we show that using lagged asset-based samples rather than industry-based samples to estimate discretionary accruals models increases sample sizes, often substantially, and generally results in equal or better detection of discretionary accruals. The overall effect is more dramatic for non-U.S. settings and more valuable, because in those settings, entire countries dropped in an industry-based estimation sample design can be retained in a size-based estimation sample design.

References