Taking as given the discretionary accruals models that exist in the literature, we examine how the selection of peer firms affects the researcher’s ability to detect earnings management. Researchers commonly estimate accruals models in cross section, and define the peer set (the firms used to estimate the cross section) as all firms in the same industry. We challenge this view by examining whether other factors – notably firm size – are at least as important as industry membership in selecting peers. Using U.S. data, we document that lagged asset peers perform as well as (or better than) industry peers at detecting earnings management both in simulations where we induce increasing levels of accruals (of between 2% and 100%) and in tests examining restatement data (where the existence of a restatement proxies for observed earnings management). Finally, we show that lagged asset peers continue to perform well for non-U.S. data, where it can be applied with significantly less sample loss than any industry peer definition.

March 2011

All authors are at the Fuqua School of Business, Duke University. Corresponding author: jfrancis@duke.edu, Fuqua School of Business, Duke University, Durham, NC 27705. We appreciate helpful comments from workshop participants at Copenhagen Business School, Duke University, Hong Kong Polytechnic University, and Lancaster University.
Peer Firm Selection for Discretionary Accruals Models

1. Introduction

Taking as given the discretionary accruals models that exist in the literature (i.e., the Jones model and various modifications of the Jones model, including performance adjustment), we examine how the selection of peer firms affects the power of these models to detect earnings management. We note that not all accruals model approaches require the selection of peer firms: in particular, one could use the firm as its own control and estimate an accruals model for each firm using the firm’s own time-series of data. This time-series approach has the virtue of holding firm-specific factors constant, but results in significant sample loss because it requires that sample firms have a sufficient time series to estimate the accruals models. For example, requiring at least 10 consecutive time-series observations results in a loss of 57% of the 251,367 firm-year observations on Compustat over 1950-2009 (with the necessary data to for the accruals models1). As a consequence of the sample restrictions of this time-series approach, most researchers estimate discretionary accruals models in cross-section, which requires the selection of a peer set of firms to constitute the cross-section. Researchers commonly define the peer set as all firms in the same industry (also requiring that the industry have a minimum of 8-10 firms to be included in the sample). 2 The maintained assumption of this industry cross-sectional approach is that industry is the most important determinant of the innate, or normal, accruals generating process, and that within each industry, firms are homogeneous with respect to the process that generates innate accruals.

We probe this view by examining whether other factors – notably firm size – are as least as important as industry in estimating innate accruals. Our analysis of this question is motivated by three observations. The first is that it is not obvious that industry definitions are most determinative of the forces driving accruals. Several researchers, for example, argue that industry definitions do not do a good job of capturing the heterogeneity that actually exists within industries (Bernard and Skinner, 1996; DeFond and Jiambalvo (1994) were the first to use the industry cross-sectional approach to estimate accruals models.

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.

2 DeFond and Jiambalvo (1994) were the first to use the industry cross-sectional approach to estimate accruals models.
Brickley and Zimmerman, 2010). Empirical support for the effects of within-industry heterogeneity for accruals is provided by Dopuch, Mashruwala, Seethamraju and Zach (2010) who find little support for the maintained assumption that firms in the same industry have a homogeneous accrual-generating process. Dopuch et al. do not, however, put forth an alternative to industry for defining the cross-section.

The second observation is that firm size is likely to be correlated with at least some factors believed to affect accruals, such as growth, complexity and monitoring. In particular, relative to smaller firms, larger firms are likely to be more mature (so more stable, with lower growth rates), have more segments (so more complex) and be more closely monitored (larger analyst following, more regulatory oversight, greater likelihood of Big-4 auditor, more institutional holdings). To the extent these, and other factors correlated with size, are important in defining the accruals generating process, we expect that grouping firms on size sharpens the estimated accruals parameters.

The third observation is that while the use of industry peers is not terribly restrictive for U.S. markets where there are many firms, the use of industry peers in non-U.S. markets results in significant sample loss. For example, the average number of firms per industry in the U.S. over 1988-2009 (which also have the necessary data for 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 (SIC2), 1.8 (SIC3) and 1.6 (SIC4). Imposing the estimation requirements for the accruals models (i.e., that 10 observations besides the event firm are available for an industry to be included) on non-U.S. data will, therefore, lead to very different restrictions and samples. For U.S. data, the use of industry peers results in sample loss of between 1% and 22% (depending on how industry is defined); for the 69 non-U.S. countries with at least one year with 11 firm-year observations, the sample loss is between 31% and 90% on average (depending on how industry is defined and the weighting of countries). To put this more starkly, requiring sufficient industry data for a given country causes between 29 and 40 countries (out of 69) to be eliminated entirely from an earnings management study that estimated discretionary accruals by industry (with the range depending on the specificity of the industry
In contrast, size-based peers impose no sample loss incremental to that imposed by the accruals models themselves, because size-peers can be defined as those firms that are closest to the target firm in terms of size. Since firms can be ranked on a size continuum, there will always be a set of firms who are closest to, or in the neighborhood of, the target firm. Whether those firms are close enough for purposes of detecting earnings management is the research question that we explore. Importantly, even if firm size performs only about as well as industry, we believe the sample expansion opportunities provided by using firm size strictly dominate the sample restrictiveness of using industry.

We investigate the selection of peer firms for estimating accruals using both simulations and empirical archival tests. Our simulations modify the simulation approach used by Dechow, Sloan and Sweeney (1995) who examined the differential power of discretionary accruals models to detect earnings management. Our study differs from theirs’ in two key respects. First, Dechow et al. are interested in comparing accruals models (e.g., Jones model versus modified Jones model), whereas we are interested in comparing the effects of peer selection for a given set of models. Second, whereas Dechow et al. estimate accruals models using time-series data (which as noted earlier avoids the peer issue entirely), our focus is on cross-sectionally estimated discretionary accruals generated by different peers. These differences in focus result in appropriate differences in the simulation analyses that we conduct, although we use their framework to guide our analysis and to assist the reader in interpreting our results.

Our simulations are conducted using all U.S. firms with available data on Compustat over 1950-2009. We focus initially on U.S. data because our tests there are not compounded by the sample restrictions that industry definitions impose; that is, we have a cleaner comparison of the power of different peer selections to detect earnings management. Our tests reveal that peers based on lagged assets

---

3 An alternative would be to pool data across countries, and increase the number of firms included in each industry. The problem with that approach is that to the extent countries are not homogenous on dimensions that affect accruals, this pooling creates noise. More importantly, if the point of the research is to examine jurisdictional influences on earnings management (such as institutional arrangements at the national level), pooling observations across jurisdictions will bias the results against finding such influences.

4 As described in section 2, our tests also address the fact that the significantly greater population of U.S. traded firms relative to any non-U.S. market implies that neighbors in the U.S. will (likely) be more similar to the target firm than are neighbors in non-U.S. markets.
are as powerful at detecting induced earnings management than are industry peers, or any other peer
definition considered. We extend the simulation tests to consider the detection power of peer groups in an
“observed” earnings management setting, proxied by restatements. Specifically, we examine the rates at
which each peer group detects earnings management for a sample of firms with restatements over the
period 1996-2006. These tests show that lagged asset peers have significantly higher rates of earnings
management detection than do industry-peers: on average lagged asset peers detect restatements 26% of
the time versus 5-6% for industry peers.

We extend our U.S. based tests to consider the power to detect earnings management using non-
U.S. data. For non U.S. firms, we are able to perform only the simulation aspect of our tests because
restatement data are not widely available outside of the U.S. Our simulation tests, which are based on
Compustat Global data for the period 1988-2009, show that lagged asset peers continue to perform well at
detecting induced earnings management. As a whole, these results indicate that a size-based definition of
peers not only performs well for U.S. data (where industry imposes comparatively small sample
restrictions), but also for non-U.S. data (where industry imposes much greater sample restrictions).

The remainder of the paper is organized as follows. Section 2 reports the results of our simulation
analysis where we investigate the power of various peer groups to detect induced earnings management.
Section 3 extends the simulation analysis to the restatement setting, where we examine the ability of the
peer definitions to detect earnings management in a sample of firms with known restatements of their
financial reports (i.e., our proxy for observed rather than induced earnings management). Section 4
extends the simulations one step further, by probing the peer groups’ abilities to detect earnings
management in non-U.S. samples. Section 5 reports additional tests and section 6 concludes.

2. Simulations with Seeded Earnings Management

This section describes the simulations we performed to assess the influence of various peer
definitions on researchers’ ability to detect earnings management using discretionary accruals. Sections
2.1-2.6 detail the specific aspects of our simulation, and section 2.7 describes the rationale for key design choices in the simulation. Section 2.8 reports the results.

2.1. Peer group definitions

Our interest in this paper is on whether and how the definition of peers for purposes of estimating and evaluating discretionary accruals influences the researcher’s ability to detect earnings management. While there are many ways that peer groups could be defined, we restrict our investigation to ways that are both supported by some conceptual basis and impose fewer, not more, restrictions on samples.

Because of their ubiquitous use in estimating cross sectional accruals models, we begin with industry, where we define peer groups separately for 2-digit, 3-digit and 4-digit SIC codes. Because of both the conceptual and the practical reasons for considering a peer selection variable based on a readily available measure of firm size, we focus on several size peer group measures: assets, lagged assets, sales, lagged sales, market capitalization and firm age. Finally, we consider as a base case a peer group formed by a random selection of firms from the entire cross section of firms with available data.

2.2 Data generation

The requirements for firm-years to be included in the simulation dataset are described in Table 1. Other than plausibility checks and the data items needed for the discretionary accruals models, firm-years are required to have the necessary information to sort firms by each of the peer definitions (i.e., firm-years must have SIC code, total assets, lagged total assets, sales, lagged sales market capitalization and firm age). In addition, because our analyses require a minimum number of firms to estimate each cross-sectional discretionary accruals model, we impose the requirement that each peer group have at least 11 firm-year observations (10 non-event firms and one event firm). When applied to the SIC4 peer group,

---

5 We repeated our tests using NAICS codes rather than SIC codes. Results (not reported) show that NAICS codes never improve on the results found for SIC codes, and are often worse than them (i.e., NAICS codes show lower detection rates than found for SIC codes).

6 In unreported tests, we also examined ROA peer groups. ROA peers do not outperform lagged asset peers. In general, ROA peers’ performance is similar to that reported for the entire cross-section.

7 We do not consider double-sorted peer groups (i.e., firms selected from the same industry code and same asset decile) because our goal is to relax the peer group assumptions, not put further restrictions on them.

8 This requirement, or one similar, is common in the literature. For example, Kothari et al. require 10 firms per 2-digit SIC code.
this requirement imposes the most sample loss. The final simulation sample contains 146,665 firm-years. This sample represents 59 cross-sections (one for each year) and 2,114 distinct SIC2-year peer groups (4,504 distinct SIC3-year peer groups and 5,658 distinct SIC4-year peer groups). On average, the industry-year peers groups contain 69 firms (SIC2), 33 firms (SIC3) and 26 firms (SIC4).

2.3 Random samples

From the simulation sample, we perform 100 iterations, where each iteration consists of the following steps:

1. We randomly select 500 unique 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 an initial set of peer firms:
   a. From the four non-metric peer-groups (entire cross-section, SIC2 industry, SIC3 industry, SIC4 industry) by matching the year and the entire industry, irrespective of the number of firm-years in the industry.
   b. From the six metric peer-groups (total assets, lagged total assets, sales, lagged sales, market cap, firm age) by matching the year and the 25 adjacent lower-ranked peers and the 25 adjacent higher-ranked peers, i.e., what we term the event firm’s “closest neighbors”.

3. Because the number of observations in the sets of peer firms created by Step 2 are not equal (they vary for the four non-metric peer groups and are fixed at 50 for the six metric peer groups), we impose a second sampling procedure to ensure that the number of firms in each peer group is constant. Holding the size of the peer group constant is important because the number of peer firms influences the power tests that we conduct. We therefore randomly select 10 firms from the matched peers identified from Step 2, for each peer group. This ensures that there is always a constant number of peer firms (10), both to estimate the discretionary accruals regression and for the analysis of the discretionary accruals.
4. We repeat Steps 2 and 3 for each of the 500 event-firm-years. Note that while event-firm-years are unique, our design does not rule out the possibility that the same peer-firm-year is used for multiple event-firm-years.

Our final data set (after 100 iterations) consists of 50,000 event-firm-years, each matched with 10 peer-firm-years.

2.4 Seeding earnings management

For each event-firm-year, we seed earning management into their data as follows. 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 then add between 2% and 20% of this ratio, in two percentage point increments, to yield 10 “positively managed” accruals figures for each event-firm-year. We focus on the lower levels of induced earnings management (i.e., 20% or less) because they are likely to be more descriptive of the levels of earnings management that managers actually engage in.

2.5 Discretionary accruals estimation

Because we are agnostic about which discretionary accruals model should be used, we conduct tests on six models that have been used in the literature. The six models are based on variations of the following four equations:

\[
\frac{\text{Total accruals}_{i,t}}{\text{Total Assets}_{i,t-1}} = \alpha_0 + \alpha_1 \frac{1}{\text{Total Assets}_{i,t-1}} + \alpha_2 \frac{\Delta \text{Sales}_{i,t}}{\text{Total Assets}_{i,t-1}} + \alpha_3 \frac{\text{Net PPE}_{i,t}}{\text{Total Assets}_{i,t-1}} + \epsilon_{i,t} \tag{1}
\]

\[
\frac{\text{Total accruals}_{i,t}}{\text{Total Assets}_{i,t-1}} = \alpha_0 + \alpha_1 \frac{1}{\text{Total Assets}_{i,t-1}} + \alpha_2 \frac{\Delta \text{Sales}_{i,t} - \Delta \text{AR}_{i,t}}{\text{Total Assets}_{i,t-1}} + \alpha_3 \frac{\text{Net PPE}_{i,t}}{\text{Total Assets}_{i,t-1}} + \epsilon_{i,t} \tag{2}
\]

\[
\frac{\text{Total accruals}_{i,t}}{\text{Total Assets}_{i,t-1}} = \alpha_0 + \alpha_1 \frac{1}{\text{Total Assets}_{i,t-1}} + \alpha_2 \frac{\Delta \text{Sales}_{i,t} - \Delta \text{AR}_{i,t}}{\text{Total Assets}_{i,t-1}} + \alpha_3 \frac{\text{Net PPE}_{i,t}}{\text{Total Assets}_{i,t-1}} + \alpha_4 \text{ROA}_{i,t} + \epsilon_{i,t} \tag{3}
\]

\[
\frac{\text{Total accruals}_{i,t}}{\text{Total Assets}_{i,t-1}} = \alpha_0 + \alpha_1 \frac{1}{\text{Total Assets}_{i,t-1}} + \alpha_2 \frac{\Delta \text{Sales}_{i,t}}{\text{Total Assets}_{i,t-1}} + \alpha_3 \frac{\text{Net PPE}_{i,t}}{\text{Total Assets}_{i,t-1}} + \alpha_4 \text{ROA}_{i,t} + \epsilon_{i,t} \tag{4}
\]

---

9 Dechow et al. seed earnings management in 10 percentage point increments, from 0% to 100%. We focus on the lower levels of induced earnings management (i.e., 20% or less) because they are likely to be more descriptive of the levels of earnings management that managers actually engage in.
where Total accruals\(_{j,t}\) = 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\(_{j,t-1}\) = firm j’s total assets in year t-1; \(\Delta Sales_{j,t}\) = firm j’s change in sales between year t-1 and t; Net PPE\(_{j,t}\) = firm j’s net property, plant and equipment in year t; \(\Delta AR_{j,t}\) = firm j’s change in accounts receivable between year t-1 and t; ROA\(_{j,t}\) = firm j’s return on assets in year t.

Equations (1) and (2) capture the Jones model without an intercept (the original model proposed by Jones, 1991) and with an intercept (a modification introduced by Kothari, Leone and Wasley, 2005). Equation (3) is the modified Jones model which includes an intercept and an adjustment for the change in accounts receivable. Equation (4) adds a performance adjustment to equation (2), in the form of the firm’s current period ROA as an explanator of accruals. The residuals obtained from estimating equations (1)-(4) are the discretionary accruals. As described by Kothari et al., these residuals can be used to construct “performance-adjusted discretionary accruals” (PADA), equal to the difference between two error terms whereby the second error term is selected from an ROA-matched firm in the same peer group. The final two measures of discretionary accruals that we consider are PADA-variants of the discretionary accruals obtained from equations (2) and (3).

2.6. Assessing detection power

Our simulations investigate the ability to detect a given amount (between 2% and 20%) of seeded earnings management at the subsample level. For each peer definition, we obtain 50,000 subsamples, where each subsample consists of one event-firm matched with 10 peer firms; each subsample also has each of the 10 levels of seeded earnings management. Our tests essentially compare the discretionary accruals of the event-firm-years with the average discretionary accruals of the 10 non-event firm-years by running the following dummy regression, separately for each seed level:

\[
Discretionary\ accruals_{j,t} = \alpha_0^{[0\%, 2\%, 4\%, ... , 20\%]} + \alpha_1^{[0\%, 2\%, 4\%, ... , 20\%]} Event\ Dummy_{j,t} + \eta_{j,t} \tag{5}
\]

where EventDummy = 1 for the single event-firm observation in each subsample.
We assess detection power by counting the number of positive $\alpha_i$ coefficients that are significant at the 10% level.\textsuperscript{10} The detection rate equals the fraction, out of 50,000, of significant coefficients.

2.7. Rationale for simulation choices

Any simulation requires that the researcher make a number of design choices. In this section, we describe our reasoning for four design choices reflected in the previously described simulation.

Our first design choice concerns two-layer selection of peer firms, as described by Steps 1 and 2 of the iteration (section 2.3). We choose this design because we think it better reflects the setting faced by a researcher studying earnings management. Specifically, such a researcher typically has a limited sample of “event” firm-years which she hypothesizes to have engaged in earnings management of some (often unspecified) form. In all likelihood, the sample and data restrictions faced by that researcher are more constraining than suggested by the comprehensive dataset from which we initially select peer firms (in Step 1 of the simulation). We introduce the second selection layer (Step 2 of the simulation) to more closely mimic the constrained sample size of the typical earnings management study. Specifically, the second selection layer not only provides a constant number of peer firms, it ensures that our results are not driven by having “perfect” neighboring firms available -- which may not be the case in actual earnings management settings.\textsuperscript{11}

A second design choice concerns the selection of neighbors for the metric peer groups. We form relative peer groups (i.e., peers are defined as the event firm’s closest neighbors) not absolute peer groups (i.e., peers are determined using some absolute cutoff). For example, an absolute size peer group would select peer 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 a set of peer firms from its same decile, once the event-firm information is known. We believe the absolute peer group approach has at least two disadvantages compared to the relative peer

\textsuperscript{10} Our results are not influenced by the choice of 10%; we find similar results (not reported) using 5% and 1%.
\textsuperscript{11} As a sensitivity test, we eliminate the second layer and simply select the ten closest in size peers in Step 1. As expected, results (not reported) show that this lagged asset peer group performs better than the one we select based on the random sampling in Step 2.
group approach that we take. 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 is dependent on having a full cross section of firms to determine the initial partitions (e.g., asset deciles or market capitalization deciles). Stated differently, forming initial deciles/partitions on a subsample of 500 firms yields different results than forming deciles/partitions using a subsample of 5,000 firms, particularly if the 500 firm subsample is not distributed equally across the 5,000-based deciles, but, say, biased towards bigger and more profitable firms.

A third design choice concerns the confluence of how we seed earnings management (section 2.4) and how we model earnings management (section 2.5). Our current modeling choice seeds earnings management by adding between 2% and 20% of the ratio of total accruals to lagged assets to the event-firm’s existing ratio. We do not adjust any other variables in the models, such as sales or total assets. Our seeding, therefore, essentially assumes that the balance sheet and the income statement of the event-firm are unchanged. This approach is most similar to Dechow et al.’s “expense manipulation” view of earnings management, albeit our approach is conducted in cross-section whereas theirs’ is performed in time-series. Modeling other financial statement effects requires other assumptions; these assumptions, summarized by Dechow et al., essentially affect the values of the independent variables in the accruals models, differentially so across models. Because we are not interested in comparing accruals models (as was the focus of Dechow et al.), we focus on the simplest (from a modeling perspective) view of earnings management which does not adjust the values of the independent variables in the accruals models.

---

12 Dechow et al. also model earnings management as “revenue manipulation” and “margin manipulation”. In the revenue manipulation scenario, they essentially add the seed level not only to total accruals, but to sales and accounts receivable. Hence, for the Jones models (given by equations 1, 2 or 3), the change in sales would increase; however, the modified Jones model (given by equation 4) is not affected because the increase in sales is offset by the same increase in accounts receivable. In the margin manipulation scenario they introduce a magnifier, defined as the firm’s net profit margin, to translate an X dollar change in total accruals to a (bigger) change in sales and accounts receivable. Again, this affects the Jones model which contains the change in sales, but does not affect the modified Jones model because the bigger change in sales is offset by the same bigger change in accounts receivable.
Moreover, we can think of no reason why our results concerning peer firm selection would differ if we chose a different earnings management perspective.

A fourth design choice concerns the level of the analysis. As described in section 2.6, we perform our analyses of detection rates at the subsample level (where we have 50,000 subsamples each consisting of 1 event firm and 10 non-event firms) rather than at the more aggregate sample level (where we have 100 samples each consisting 500 event firms matched with 5,000 non-event firms). We select the subsample level for two reasons. First, as a practical matter, the discretionary accruals estimation must be performed on the subsample level, so the finer data are available to us. Second, it is hard to observe differences at the sample level because at this level firm-specific idiosyncrasies are averaged out, implying that for all seeded earnings management levels (even the smallest ones, of 2% or 4%), detection rates will be close to 100% for all models. Stated differently, we believe that 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 often used in earnings management research. By focusing our analysis at the subsample level, we believe we more closely approximate the issues faced in this research.

2.8. Results of simulation

Table 2 shows the results of the simulation; here we report for each model (panels of table) and peer group definition (columns of table) the fraction of times, as a percentage of the 50,000 subsamples, that the $\alpha_1$ coefficient from equation (5) is significant at the 10% level for increasing levels of seeded earnings management (the seed levels are the rows of each panel).

We begin by discussing results for the 0% seed level. Recall that this level corresponds to the benchmark case of no earnings management (i.e., no data manipulation by us) and can be used to gauge the validity of the approach. Specifically, if there is no earnings management and if the data are truly random, we should see 5% of cases revealing significant positive earnings management (5% of cases will also show significant negative earnings management). The 0% seed level is important not only because it serves as a measure of Type I errors, but because it speaks to whether our initial model is unbiased. For
all model specifications and peer definitions, Table 2 shows that the detection rates for the 0% seed level are very close to 5%; statistical tests, not reported, reveal that in no case is the detection rate for the 0% level reliably different from the predicted 5% that would occur by chance at a 10% level of significance.

Turning to the non-zero seed levels, our interest is in which peer group (column) shows the highest detection rates across various accruals model specifications. We first illustrate the differential detection rates of the peer groups graphically for one of the accruals models. Figure 1 shows the detection rates achieved for each peer group and seed level for the Jones model with intercept, given by equation (2). 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 formed based on firm age, sales and market capitalization also consistently perform poorly across all seed levels. We next turn to which peer groups perform the best. At the lowest seed levels (i.e., those less than 10%), the industry peers and lagged asset peers have the highest detection rates, and these rates are quite close to each other. At seed levels between 10% and 20%, the lagged asset definition of peers begins to strictly dominate all others.

Because the graphs such as the one in Figure 1 differ across the accruals models, and because it is not easy to visually perceive the differences in detection rates at the lowest seed levels, we also consider an aggregate measure which assesses the average performance of each peer group across the 10 seed levels examined for each model. We measure the performance of a given peer-model-seed combination as the detection rate achieved by the noted peer group, for each model and seed level, divided by the maximum detection rate achieved by any peer group for that seed-model combination. We then average this peer-model-seed performance measure across the seed levels for each model and peer group, to yield an “effectiveness score” for each peer group and model combination. The closer is the effectiveness score to one (the value that would result if a single peer group had the highest detection rates for every seed level), the better at detection is that peer group. Note that this measure also allows us to incorporate how far a given detection rate is from the maximum detection rate achieved because it includes their distance from the maximum.
The effectiveness scores for each peer group are reported in the last row of each panel of Table 2. Consistent with Figure 1, we find that the cross-section peer performs the worst, with effectiveness scores ranging between 78.1% and 80.4%. Peer groups based on firm age, sales and market capitalization perform slightly better, with effectiveness scores of between 84.0% and 87.3%. In general, the industry peer groups do reasonably well, with effectiveness scores of 93.9% to 100%. With one exception, the lagged assets peer group has the highest effectiveness scores (98.7% to 99.5%); the exception is the Jones model without intercept, where the industry peer groups outperform the lagged asset peer group. As an additional test, we extend the analyses in Table 2 for seed levels greater than 20% (i.e., 10 percentage point increments between 20% and 100%; also see footnote 10). Results of this analysis for the Jones model with intercept are graphed in Figure 2. The graph shows clear evidence that for seed levels higher than 20%, lagged asset peers strictly dominates all other peer groups. Effectiveness scores that include the higher seed levels (not reported) confirm the dominance of lagged asset peers as having the highest effectiveness scores of any peer group.

We perform a number of sensitivity checks on our simulation. First, we relax our selection procedures and allow all non-event firms in the event-firm’s industry to be included as a peer firm. This means that there will be more peer firms in the industry peer groups than in the metric peer groups (where we continue to select 10 observations). A priori it is not obvious that more peers should bias towards or away from higher detection rates. On the one hand, more firms imply more degrees of freedom for estimating the accruals model which should aid the detection rates for industry peers. On the other hand, if the larger sample is more heterogeneous than the (smaller) constant peer sample, then the larger peer sample will introduce noise into the estimation, which should reduce detection rates. The results of this analysis (not reported) show that increasing the number of firms in the industry peer sample reduces the detection rate for SIC2, leaves unchanged the detection rate for SIC3, and increases the detection rate for

\footnote{Consistent with the better comparability of higher-digit industry definitions relative to lower-digit definitions, effectiveness scores generally increase over the 2-digit, 3-sigit and 4-digit industry peer definitions. There is single instance where this ordering is not observed: for the Jones model that includes ROA as an explanatory variable, the effectiveness score for 3-digit SIC code (98.2%) is greater than that for 4-digit (98.0%).}
SIC4. In all cases, though, the industry peers show the same relation with respect to the detection rates for lagged assets: the detection rate for lagged asset peers is indistinguishable from the detection rates for industry peers for seed levels less than 10%, and strictly dominates the industry peers for seed levels above 10%.

In a second sensitivity test (not reported), we take the SIC4 peer group’s size and perfectly match the size of the metric peer group, for each event firm. This approach maximizes the number of peers available for the industry peer at the same time that it ensures equivalent-size metric peer groups. It has the disadvantage that, if multiple event-firm-years are selected from the same SIC4-year, the industry peer firms selected will essentially be the same for the event-firm (with the exception of the event-firm itself), but the metric peers might shift considerably. Results are similar to those reported.

In summary, our simulation tests indicate that while industry peers do not perform poorly, they are dominated by lagged asset peers. This finding is observed for all but one of the six models of discretionary accruals that we examine, and is robust to a variety of sensitivity tests.

3. Tests on Restatement Firms

In this section we examine how well the different peer groups perform at detecting unusual levels of absolute discretionary accruals for a sample of firm-years where we observe restatements. Thus, in contrast to our simulation tests where we induce earnings management (in the form of increasing seed levels), for this sample, we take the existence of a restatement as observable evidence of managed earnings. We then assess what percentage of the time the restatement firm’s level of absolute discretionary accruals is significantly greater than its peer firms. Under the view that restatement firms have, in fact, managed earnings such that their absolute discretionary accruals are larger than their peers, we associate larger detection rates as indicating that the peer group is better at detecting observed earnings management.

In order to assess statistical significance, our analyses of discretionary accruals are performed at the iteration level; in total, we perform 500 iterations. Each iteration consists of selecting 200 event-firm-
years from the population of Compustat firms with restatements over the period 1996-2006; an event-firm-year is a firm-year with a restatement announcement in the 11 months following the fiscal-year end. For each event-firm-year, we randomly select 10 non-restating firms from the same year and the same peer-group, where 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 value of these residuals at the iteration level, generating 200 event-firm absolute residuals and 2,000 peer firm absolute residuals per iteration.

For each iteration, we then compute a Z-score comparing the mean absolute residual for the 200 event-firms with the mean absolute residual of the 2,000 non-event firms. The Z-score is a statistical measure of whether the event-firm residuals are reliably different from the non-event firms’ residuals. After 500 iterations, we have a series of 500 Z-scores. Recall that because we focus on observed restatements (where we assume accruals have been managed), we expect Z-scores to be significant; that is, we expect the Z-score to detect the earnings management. Our analyses of peer groups focus on how the choice of peer groups affects this detection rate: better (worse) peer groups will have a greater (smaller) frequency of significant Z-scores.

The maintained assumptions implied by this analysis are twofold. First, we assume that restatement firms have managed their accruals. Second, we assume that this accruals management affects the fiscal year prior to the restatement announcement. Because these assumptions will not hold for all observations, we do not expect to observe 100% detection rates in our simulated analyses.

Table 3 reports the detection rates, defined as the fraction of iterations (out of 500 total) where the Z-score is significant at the 10% level or better. The rows in Table 3 correspond to the accruals model and the columns correspond to the peer groups. For each row, we are looking for the peer group with the highest detection rate. With one exception (the Jones model without intercept, where market capitalization is the best peer), lagged asset peers has the highest detection rates, ranging from 17.4% to 37.2%, compared to a range of 1.6% to 29.4% for all other peer groups. Importantly, these tests show that lagged asset peers strictly dominate all definitions of industry peers: detection rates for industry peers are
never more than half the detection rate of lagged asset peers, and are generally much smaller. Moreover, industry peers even underperform the entire cross-section.

In summary, our findings from the restatement analysis are consistent with our simulation results. In both cases, we find lagged asset peers dominates industry (and other) peers in detecting earnings management.

4. Peer Group Selection for Analyses Using Non-U.S. Data

This section considers the relative power of industry peers versus other peer definitions at detecting earnings management in investigations that focus on non-U.S. data. In our view, there are at least two issues to consider when extending our analyses to non-U.S. data. First is the relative restrictiveness of industry peer definitions when applied to markets where there are considerably fewer firms than exist in the U.S. markets (section 4.1). Second is for the subset of countries where sufficient data exists to use industry peers, what is the detection power of industry peers versus other peer definitions (section 4.2)? We summarize the key findings and inferences in section 4.3.

4.1 Restrictiveness of industry peer definitions on non-U.S. data

As discussed in the introduction, the restrictiveness of an industry definition of peers is much more problematic when applied to non-U.S. data than to U.S. data. To see this, we impose the same data requirements on non-U.S. data that our section 2 simulation imposed on U.S. data: we require that each firm-year observation have the data necessary for the accruals models and we require that there be at least 11 firms per industry group (one event firm plus 10 non-event firms). Importantly, we perform our analyses by country, as opposed to combining data across countries (that is, we do not combine same-industry observations from two or more countries). We start by requiring a firm to have at least 11 firm-year observations (the minimum needed to estimate an accruals model) on Compustat Global for the
period 1988-2009. Table 4 shows that these requirements result in 218,575 firm-years, representing 69 countries.\footnote{In total, Compustat Global contains firm-year observations for 99 distinct countries. Thirty countries do not meet the minimal restrictions we impose, leaving us with the 69 countries listed in Table 4.}

Table 4 also reports the distribution of firm-year observations, by country, after imposing the requirement that each country have the necessary data to determine peer firms using the three industry definitions based on 2-digit, 3-digit, and 4-digit SIC codes; these data are reported in the columns labeled SIC2, SIC3 and SIC4, respectively. As a benchmark for these non-U.S. data, the first row of Table 4 shows the application of these same requirements to Compustat North America data, for the same time period 1988-2009. Applied to U.S. data, there are a total of 124,787 firm-year observations with the data to estimate the accruals models; applying the additional industry requirements associated with 2-digit, 3-digit and 4-digit SIC codes reduces this sample by 1%, 12% and 22%, respectively.

Our focus in Table 4 is on how the number of firm-year observations, and their distribution across industries within each country, influences the sample loss imposed by industry definitions of peer groups. In terms of sample size, Japan has the largest fraction of the total, with 47,488 observations, while Bangladesh has the least, 11 observations; the median country (New Zealand) has 1,104 firm-year observations. For the least restrictive SIC2 definition, only two countries have sample loss rates that are in the single digits – Japan with 5% and China with 9%; this compares to 1% for the U.S. Even the country with the second largest sample size, Great Britain with 20,867 firm-year observations, experiences sizable sample loss from an SIC2 requirement (17%).

The bottom three rows of Table 4 report statistics on sample loss, aggregated across the 69 countries. The first measure is the number of countries (out of 69) that would be eliminated entirely because of insufficient data to estimate the accruals models using industry peer definitions. For the least restrictive SIC2 definition, we lose 29 countries, with this count 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 (i.e., it weights each country equally). Using this measure, we find
that 76% of the sample observations are lost using an SIC2 definition, with this percentage 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 the country’s sample size as a proportion of the total; the weighted average version therefore will produce smaller measures of sample loss because it will count the sample loss experienced by Japan (Bangladesh) more (less) in the overall measure. Weighted sample losses are 32% (SIC2), 60% (SIC3) and 70% (SIC4).

In summary, the evidence in Table 4 suggests that using industry peers (that is, requiring the necessary data to estimate accruals models at the industry level) imposes significant sample restrictions on non-U.S. data. Recall that using a metric peer group (such as peers based on a measure of firm size) generates no sample loss incremental to that imposed by the accruals model itself. Metric peer definitions, therefore, have the potential to allow for much larger samples.15 It is unclear, however, whether the detection power of metric peers is as good as industry peers. We probe this issue in the next section.

4.2. Detecting induced earnings management in non-U.S. data

In this section we examine how well various peer groups detect earnings management in non-U.S. data. Our analysis here uses the same simulation tests as reported in Table 2, except now we apply those tests to each of the 69 countries with available data. The most restrictive requirement imposed is the requirement that for each randomly selected event firm there be 10 non-event firm observations in the 4-digit SIC code for that country. The observations that meet this requirement are shown in the SIC4 column in Table 4. As shown there, we lose 40 of the 69 countries. In addition, nine of the remaining 29 countries have relatively few observations (less than 100 firm-years) for us to perform the bootstrapping required by our simulation. We therefore restrict our analysis to the 20 countries with at least 100 firm-year observations identifiable under SIC4.

For this “restricted sample” of 20 countries, we analyze how each of the peer groups performs at detecting induced earnings management. For purposes of reporting the results, we tabulate results for a

---

15 For countries with few firm-year observations (i.e., only 10 firm-event years for a given year), the peer group is the entire cross-section.
single model (the Jones model with intercept) and for a subset of peer groups (entire cross-section, three industry peers, and lagged asset), even though we perform the tests for the full set of models and peers. Further, although we analyze all seed levels, we report results only for the effectiveness score, which averages detection rates (measured relative to the best detection rate for that seed level) across all seed levels. The results, reported in Table 5, show that lagged asset peers has the highest effectiveness score for 12 of the 20 countries in the restricted sample. Lagged asset peers also has the highest average effectiveness score (96.5%) calculated across countries; this compares to about 95% for industry peers and 89% for the entire cross section.

Tables 5’s comparisons of peer group detection power hold the sample sizes of each peer group constant at 11 observations (one event firm plus 10 non-event firms). As noted earlier, applying this requirement to industry definitions (especially SIC4) results in substantial sample loss. Given that we find that lagged asset peers is preferable to industry peer definitions for the constant-size peer groups, we next investigate the performance of lagged asset peers in research designs where we do not impose the industry requirements. Note that because lagged asset peers impose no incremental sample loss, we can theoretically perform this extension for the 69 countries with at least 11 observations in a given year (listed in Table 4). Because we impose the requirement that each country have at least 100 firm-year observations (for us to perform the bootstrapping required by our simulation), we perform our analysis on the 58 countries listed in Table 4 that have at least 100 firm-year observations (the “maximized sample”). Table 6 shows the detection rates for these 58 countries, for a 10% seed level (results for other seed levels are similar and are not tabulated). On average, lagged peers detect elevated levels of discretionary accruals 24.2% of the time. This detection rate is comprised of a 24.5% detection rate for countries included in the industry tests and a 24.0% rate for those not included in those tests; statistical tests reveal that these detection rates are indistinguishable from each other. We can also compare the detection rates for those countries that are in both the maximized and restricted samples; any difference in detection rates

---

16 In unreported tests, we also ensure that the models are unbiased for the non-U.S. data (i.e., we observe 5% detection rates for the 0% seed level).
here relate purely to the additional data included in the maximized sample. For the 20 countries in both samples, the detection rates are when industry data are required (25.1%, restricted sample) versus when they are not (24.5%, maximized sample) are statistically indistinguishable.

4.3. Summary of analyses involving non-U.S. data

Overall, we draw the following observations from the evidence in Tables 4-6. First, non-U.S. countries are restricted in their ability to estimate accruals models using industry peer definitions because of the data requirements imposed by industry definitions. Therefore, studies that use non-U.S. data and industry definitions will have some subset of limitations. They might, for example, experience significant sample losses (and therefore, likely have low power to detect earnings management), or they might be restricted to the few countries whose markets are large enough to support these requirements (Japan, China or Great Britain). Alternatively, the researcher could maintain the industry peer requirements but aggregate his/her data across countries in order to achieve the necessary number of firm-year observations per industry. The problem with this approach is that, by treating all observations for a given industry (but across countries) as similar, it ignores the influence of institutional specific factors on earnings management (i.e., accounting standards, legal codes, culture, etc.). This approach is particularly problematic if the researcher is interested in whether and how jurisdictional factors affect earnings management because combining observations across countries will obscure the very phenomenon that the researcher wishes to investigate. Finally, researchers could drop the industry peer definition entirely and allow all firms in a given country to serve as peers. The problem with this “entire cross section” peers definition is that we find it performs the worst (i.e., is least powerful) at detecting earnings management.

The second observation is that, among all peer definitions considered, lagged asset peers exhibits the highest ability to detect earnings managements, both in U.S. data and in non-U.S. data. This finding is especially valuable for non-U.S. samples because it means that researchers could estimate accruals models, using lagged asset peers, with no sample loss incremental to that imposed by the requirement that the firm have data on the variables included in the accruals models. No incremental loss means that entire countries which could not be analyzed using industry peers can be analyzed using lagged asset peers.
Increasing the set of countries with the necessary data to estimate accruals models should increase the power of research designs examining whether and how jurisdictional specific factors influence managers’ ability and incentives to engage in earnings management, because the larger samples should include both more and more diverse countries.

5. Additional Tests

In addition to the sensitivity tests already reported, we performed several other checks. First, we investigate whether the superiority of lagged asset peers is driven by the use of lagged assets as the scalar in the accruals models; that is, are our peer results mechanically induced by the accruals model specification? To probe this possibility, we repeat our tests after modifying the accruals models to use lagged sales as the scalar. If the choice of scalar drives our results concerning peer group, we should find that lagged sales peers has the highest detection power, and that this detection power rivals the previously documented detection power of lagged asset peers. Results (not reported) show that lagged sales peers do not have the highest detection power in the lagged sales scalar setting. Moreover, the effectiveness scores for lagged sales peers are 80-85%, well below the range reported for lagged asset peers (91-99.5% as reported in Table 4). These findings suggest that the scalar used in the accruals model is not driving the detection rates of the peer groups.

Second, having documented that a metric peer group performs best at detecting earnings management, a natural next question to ask is what is the optimal number of non-event firms, from that metric peer group, to select as non-event firms? Recall that our prior tests require 10 non-event firms (11 including the event firm) as the minimum needed to estimate the accruals models. However, because the number of metric peers is defined by the number of neighbors, one could easily expand the neighborhood’s size beyond 10. Increasing the neighborhood’s size increases the number of observations used to estimate the accruals models (and so increases the degrees of freedom) and also increases the likely heterogeneity across the peer observations (and so decreases the precision of the accruals estimates). It is an empirical question which of these two forces dominates. To provide evidence on the
peer size issue, we focus on lagged asset peers, and we repeat our simulation analyses increasing the number of lagged asset peers from 10 firms (base case) to 20, 50, 100, 250, 500, 1000 and 2000 firms.

Results of this test are shown in Table 7 where we tabulate detection rates for the Jones model with intercept (other models yield similar inferences and are not reported). The first row (5% seed level) provides a measure of the unbiasedness of the model for each peer group size. Consistent with intuition, we find that for very large peer groups (n>250), detection rates deviate from the expected 5%; in particular, results show low Type I errors for large peer groups. This finding indicates that the accruals models are not well-specified for large peer groups. Turning to the induced earnings management levels, we see that the highest detection rates are achieved with peer sizes of n=10 and n=20.17 The finding that the optimal number of non-event firms to include in the peer group (for purposes of detecting earnings management) is about 10-20 is consistent with research in finance which shows that the benefits of risk diversification are largely reached when 10-20 stocks are included in a portfolio (Breeley, Myers and Allen, 2006; Elton and Gruber, 1977; 1984).18 Our finding is also consistent with Gong et al.’s finding that the average number of peer firms (reported by firms for compensation purposes) is 15. We conjecture, but do not know, that Gong et al.’s finding is driven by the same underlying phenomenon – diversification of idiosyncratic effects.

6. Summary and Conclusion

We examine the ability of various definitions of peer firms to detect earnings management, with an emphasis on how well firm size peer definitions do relative to industry peer definitions. Our main finding is that defining an event firm’s peer set based on lagged assets performs best, insofar as it has the highest detection rates both in simulations where we induce earnings management and in research settings where we observe earnings management (as proxied by the existence of a restatement).

17 In unreported tests, we considered peer sizes of between 10 and 20, in increments of one. In general, no consistent optimization pattern emerged within that interval.

18 Elton and Gruber’s analysis uses weekly stock return data for firms traded on the NYSE and AMSE during 1971-1974. We replicated their tests on more current returns data, and find similar results (not reported).
A firm-size definition of peers has much practical value because it imposes no incremental sample loss. For U.S. data, this means a sample size savings of anywhere from 1-3% (SIC2 definitions) to 22-30% (SIC4 definitions). The bigger effect comes for non-U.S. data where the sample size savings range from 30% to 90% (depending on industry definitions and weighting schemes). We document that the superior detection power of lagged asset peers extends to non-U.S. data as well. Further, we show for samples not constrained by the availability of industry peers, lagged asset peer detection rates are very similar to the detection rates observed for samples where we can perform a controlled comparison of peer group detection rates.

Overall, we believe our finding concerning the detection power of lagged asset peers is important both in the U.S. context and in the non-U.S. context. For both settings we show that researchers would increase their ability to detect earnings management by using lagged asset peers rather than industry peers. While lagged asset peers increases the sample size in both contexts, that increase is more dramatic for non-U.S. settings and more valuable, because in those settings, entire countries might be included that would not be if an industry peer design was used.
Table 1 reports the sample restrictions imposed by requirements to have the necessary data for an accruals observation, for identifying all peer groups, and for estimating the accruals models. The most restrictive criterion requires 10 non-event firms in the same SIC4 code. The bottom rows of the table describe the sizes of the industry peer groups, e.g., out of the 146,665 firm-year observations that meet our requirements, there are 2,114 distinct SIC2-year peer groups.

<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>Firms with more than one year of data</td>
<td>407,726</td>
</tr>
<tr>
<td>With lagged total assets, total assets, sales &gt;= 1</td>
<td>303,194</td>
</tr>
<tr>
<td>With data for total accruals calculation</td>
<td>251,367</td>
</tr>
<tr>
<td>With data for identifying ALL peer groups</td>
<td>213,807</td>
</tr>
<tr>
<td>With at least 11 firm-year observations per SIC4</td>
<td>146,665</td>
</tr>
</tbody>
</table>

Table 1
Selection of Data Observations for Simulation Using U.S. Data 1950-2009

<table>
<thead>
<tr>
<th>Peer group</th>
<th># Firm-years</th>
<th># Peergroup-years</th>
<th>Mean</th>
<th>Min</th>
<th>Median</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entire cross-section</td>
<td>146,665</td>
<td>59</td>
<td>2,486</td>
<td>15</td>
<td>2,467</td>
<td>5,833</td>
</tr>
<tr>
<td>SIC2</td>
<td>146,665</td>
<td>2,114</td>
<td>69</td>
<td>11</td>
<td>34</td>
<td>958</td>
</tr>
<tr>
<td>SIC3</td>
<td>146,665</td>
<td>4,504</td>
<td>33</td>
<td>11</td>
<td>18</td>
<td>824</td>
</tr>
<tr>
<td>SIC4</td>
<td>146,665</td>
<td>5,658</td>
<td>26</td>
<td>11</td>
<td>17</td>
<td>436</td>
</tr>
</tbody>
</table>
Table 2 reports the detection rates for each seed level of induced earnings management (2%-20%) and peer definitions, for each of the six accruals models. Note that the 0% seed level is a specification check, insofar as a well-specified model should show detection of positive earnings management (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 for that seed level. Thus, an effectiveness score of 100% indicates that, for that model, the peer group was always the best at detecting earnings management.

Table 2 is continued on the next page.
Table 2 reports the detection rates for each seed level of induced earnings management (2%-20%) and peer definitions, for each of the six accruals models. Note that the 0% seed level is a specification check, insofar as a well-specified model should show detection of positive earnings management (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 for that seed level. Thus, an effectiveness score of 100% indicates that, for that model, the peer group was always the best at detecting earnings management.
Table 3 reports the fraction of the time that restatement firms’ absolute discretionary accruals differ significantly from non-restating peer firms, for the various definitions of peers we consider. Because we assume that restatement firms engaged in earnings management, better (worse) peer groups should have a greater (smaller) frequency of detecting earnings management.

<table>
<thead>
<tr>
<th></th>
<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>MktCap Neighbors</th>
<th>Firm Age Neighbors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jones Model</td>
<td>20.0%</td>
<td>11.2%</td>
<td>11.4%</td>
<td>11.2%</td>
<td>23.2%</td>
<td>26.6%</td>
<td>27.6%</td>
<td>29.4%</td>
<td>24.0%</td>
</tr>
<tr>
<td>Jones Model (with intercept)</td>
<td>10.8%</td>
<td>7.8%</td>
<td>5.4%</td>
<td>6.4%</td>
<td>16.8%</td>
<td>37.2%</td>
<td>15.6%</td>
<td>16.0%</td>
<td>13.4%</td>
</tr>
<tr>
<td>Modified Jones Model</td>
<td>11.4%</td>
<td>7.0%</td>
<td>6.8%</td>
<td>6.0%</td>
<td>15.0%</td>
<td>36.2%</td>
<td>16.6%</td>
<td>16.2%</td>
<td>14.4%</td>
</tr>
<tr>
<td>Jones Model + ROA</td>
<td>5.0%</td>
<td>3.2%</td>
<td>2.4%</td>
<td>1.6%</td>
<td>9.0%</td>
<td>20.2%</td>
<td>5.4%</td>
<td>6.0%</td>
<td>7.8%</td>
</tr>
<tr>
<td>PADA (based on Jones)</td>
<td>4.4%</td>
<td>2.8%</td>
<td>3.4%</td>
<td>2.2%</td>
<td>6.4%</td>
<td>20.0%</td>
<td>5.6%</td>
<td>4.2%</td>
<td>6.6%</td>
</tr>
<tr>
<td>PADA (based on Modified Jones)</td>
<td>4.6%</td>
<td>2.6%</td>
<td>2.8%</td>
<td>2.4%</td>
<td>5.6%</td>
<td>17.4%</td>
<td>7.2%</td>
<td>4.2%</td>
<td>5.2%</td>
</tr>
<tr>
<td>Average</td>
<td>9.4%</td>
<td>5.8%</td>
<td>5.4%</td>
<td>5.1%</td>
<td>12.8%</td>
<td>26.3%</td>
<td>13.0%</td>
<td>12.7%</td>
<td>11.9%</td>
</tr>
<tr>
<td>Country</td>
<td>Entire Cross-Section</td>
<td>SKC2</td>
<td>SKC3</td>
<td>SKC4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>----------------------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USA</td>
<td>210,555</td>
<td>2,607</td>
<td>204,923</td>
<td>3%</td>
<td>5,197</td>
<td>174,383</td>
<td>17%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>USA (1988 - 2009)</td>
<td>124,707</td>
<td>1,258</td>
<td>123,243</td>
<td>1%</td>
<td>2,804</td>
<td>110,110</td>
<td>12%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARE</td>
<td>212</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARG</td>
<td>674</td>
<td>6</td>
<td>70</td>
<td>90%</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AUS</td>
<td>9,405</td>
<td>229</td>
<td>6,257</td>
<td>33%</td>
<td>134</td>
<td>3,482</td>
<td>63%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AUT</td>
<td>1,185</td>
<td>4</td>
<td>49</td>
<td>96%</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BEL</td>
<td>1,437</td>
<td>12</td>
<td>141</td>
<td>90%</td>
<td>4</td>
<td>48</td>
<td>97%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BGD</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0%</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMU</td>
<td>4,996</td>
<td>125</td>
<td>2,545</td>
<td>49%</td>
<td>51</td>
<td>777</td>
<td>84%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BRA</td>
<td>3,154</td>
<td>71</td>
<td>1,368</td>
<td>57%</td>
<td>30</td>
<td>594</td>
<td>81%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHE</td>
<td>2,931</td>
<td>82</td>
<td>1,317</td>
<td>55%</td>
<td>12</td>
<td>137</td>
<td>95%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHL</td>
<td>1,484</td>
<td>24</td>
<td>414</td>
<td>72%</td>
<td>12</td>
<td>188</td>
<td>87%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHN</td>
<td>18,915</td>
<td>399</td>
<td>17,258</td>
<td>9%</td>
<td>498</td>
<td>13,147</td>
<td>31%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>COL</td>
<td>174</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYM</td>
<td>2,131</td>
<td>36</td>
<td>825</td>
<td>61%</td>
<td>18</td>
<td>382</td>
<td>82%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CZE</td>
<td>150</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEU</td>
<td>8,962</td>
<td>240</td>
<td>6,346</td>
<td>29%</td>
<td>132</td>
<td>2,815</td>
<td>69%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNK</td>
<td>1,798</td>
<td>19</td>
<td>244</td>
<td>86%</td>
<td>3</td>
<td>34</td>
<td>98%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ESP</td>
<td>1,852</td>
<td>15</td>
<td>168</td>
<td>91%</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EST</td>
<td>9</td>
<td>4</td>
<td>0</td>
<td>0%</td>
<td>1</td>
<td>0</td>
<td>0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FIN</td>
<td>2,131</td>
<td>36</td>
<td>825</td>
<td>61%</td>
<td>18</td>
<td>382</td>
<td>82%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FRA</td>
<td>8,891</td>
<td>223</td>
<td>5,448</td>
<td>39%</td>
<td>66</td>
<td>1,806</td>
<td>80%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GBR</td>
<td>20,867</td>
<td>562</td>
<td>17,219</td>
<td>17%</td>
<td>353</td>
<td>7,959</td>
<td>62%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRC</td>
<td>1,197</td>
<td>7</td>
<td>85</td>
<td>93%</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HRK</td>
<td>2,669</td>
<td>38</td>
<td>652</td>
<td>76%</td>
<td>17</td>
<td>250</td>
<td>91%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HUN</td>
<td>228</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IDN</td>
<td>2,649</td>
<td>53</td>
<td>480</td>
<td>69%</td>
<td>2</td>
<td>23</td>
<td>99%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IND</td>
<td>10,457</td>
<td>212</td>
<td>9,112</td>
<td>13%</td>
<td>220</td>
<td>5,952</td>
<td>43%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IRL</td>
<td>828</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISR</td>
<td>1,060</td>
<td>22</td>
<td>382</td>
<td>64%</td>
<td>12</td>
<td>225</td>
<td>79%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ITA</td>
<td>2,828</td>
<td>50</td>
<td>896</td>
<td>68%</td>
<td>10</td>
<td>157</td>
<td>94%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JAM</td>
<td>60</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JPN</td>
<td>47,488</td>
<td>791</td>
<td>45,164</td>
<td>5%</td>
<td>1,491</td>
<td>34,833</td>
<td>27%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KEN</td>
<td>61</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KOR</td>
<td>5,386</td>
<td>157</td>
<td>3,847</td>
<td>29%</td>
<td>82</td>
<td>1,555</td>
<td>71%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KWT</td>
<td>156</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LKA</td>
<td>391</td>
<td>6</td>
<td>117</td>
<td>70%</td>
<td>5</td>
<td>78</td>
<td>80%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LTV</td>
<td>180</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LUX</td>
<td>280</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LVA</td>
<td>116</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEX</td>
<td>156</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MYS</td>
<td>9,786</td>
<td>307</td>
<td>7,269</td>
<td>26%</td>
<td>184</td>
<td>3,291</td>
<td>66%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NGA</td>
<td>131</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NLD</td>
<td>2,133</td>
<td>50</td>
<td>150</td>
<td>58%</td>
<td>25</td>
<td>453</td>
<td>78%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPL</td>
<td>11,04</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OMN</td>
<td>273</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAK</td>
<td>1,359</td>
<td>34</td>
<td>628</td>
<td>54%</td>
<td>12</td>
<td>211</td>
<td>84%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PER</td>
<td>486</td>
<td>1</td>
<td>1</td>
<td>98%</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHL</td>
<td>1,270</td>
<td>12</td>
<td>182</td>
<td>86%</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>POL</td>
<td>1,506</td>
<td>26</td>
<td>383</td>
<td>75%</td>
<td>5</td>
<td>89</td>
<td>94%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>QAT</td>
<td>68</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ROU</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RUS</td>
<td>822</td>
<td>21</td>
<td>325</td>
<td>60%</td>
<td>16</td>
<td>236</td>
<td>71%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAU</td>
<td>164</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SGP</td>
<td>5,95</td>
<td>147</td>
<td>3,305</td>
<td>44%</td>
<td>86</td>
<td>1,348</td>
<td>77%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SVN</td>
<td>175</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SWE</td>
<td>3,80</td>
<td>70</td>
<td>1,660</td>
<td>56%</td>
<td>24</td>
<td>713</td>
<td>81%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>THA</td>
<td>4,292</td>
<td>134</td>
<td>2,208</td>
<td>49%</td>
<td>35</td>
<td>439</td>
<td>90%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TUR</td>
<td>820</td>
<td>3</td>
<td>35</td>
<td>90%</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TWN</td>
<td>9,526</td>
<td>145</td>
<td>8,458</td>
<td>11%</td>
<td>175</td>
<td>6,451</td>
<td>32%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VEN</td>
<td>69</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VNM</td>
<td>67</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZAF</td>
<td>2,77</td>
<td>53</td>
<td>894</td>
<td>68%</td>
<td>15</td>
<td>277</td>
<td>90%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZWE</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td>0</td>
<td>0</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

# Countries dropped 29 37 40
Average Loss (equal weight for each country) 76% 89% 93%
Weighted average loss 32% 60% 70%
Table 4 reports the sample restrictions imposed on non-U.S. data by requirements to have the necessary accruals data to estimate the accruals models cross-sectionally for peer groups that have at least 11 firms (1 event and 10 non-event) per industry definition. The population consists of all firm-years reported on Compustat Global over 1988-2009. The column labeled “entire cross-section” shows the number of firm-years with data on accruals observations \( n=218,575 \). Of this number, 148,140 firm-year observations \( 88,374; 65,178 \) also have the necessary SIC2 data \( SIC3; SIC4 \). As a benchmark, the first two rows of the table show how the same restrictions affect U.S. data over the full period for which data are available \( 1950-2009 \), top row and for the same period for which we have non-U.S. data \( 1988-2009 \), second row.
Table 5
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>84.1%</td>
<td>94.0%</td>
<td>90.5%</td>
<td>88.1%</td>
<td>100.0%</td>
</tr>
<tr>
<td>BMU</td>
<td>96.6%</td>
<td>94.9%</td>
<td>97.2%</td>
<td>97.0%</td>
<td>98.5%</td>
</tr>
<tr>
<td>BRA</td>
<td>92.9%</td>
<td>98.9%</td>
<td>99.0%</td>
<td>98.3%</td>
<td>95.2%</td>
</tr>
<tr>
<td>CHL</td>
<td>99.1%</td>
<td>97.1%</td>
<td>98.5%</td>
<td>97.9%</td>
<td>99.4%</td>
</tr>
<tr>
<td>CHN</td>
<td>83.9%</td>
<td>87.2%</td>
<td>85.8%</td>
<td>85.2%</td>
<td>100.0%</td>
</tr>
<tr>
<td>CYM</td>
<td>97.3%</td>
<td>89.0%</td>
<td>87.7%</td>
<td>87.8%</td>
<td>99.0%</td>
</tr>
<tr>
<td>DEU</td>
<td>86.4%</td>
<td>99.8%</td>
<td>95.5%</td>
<td>97.8%</td>
<td>96.9%</td>
</tr>
<tr>
<td>FRA</td>
<td>84.7%</td>
<td>89.0%</td>
<td>90.2%</td>
<td>85.6%</td>
<td>98.4%</td>
</tr>
<tr>
<td>GBR</td>
<td>85.6%</td>
<td>97.9%</td>
<td>99.2%</td>
<td>98.6%</td>
<td>97.6%</td>
</tr>
<tr>
<td>IND</td>
<td>83.6%</td>
<td>89.6%</td>
<td>93.4%</td>
<td>93.1%</td>
<td>98.5%</td>
</tr>
<tr>
<td>JPN</td>
<td>87.1%</td>
<td>96.2%</td>
<td>96.3%</td>
<td>97.4%</td>
<td>100.0%</td>
</tr>
<tr>
<td>KOR</td>
<td>87.7%</td>
<td>96.9%</td>
<td>100.0%</td>
<td>96.5%</td>
<td>95.5%</td>
</tr>
<tr>
<td>MYS</td>
<td>88.8%</td>
<td>97.0%</td>
<td>96.9%</td>
<td>97.1%</td>
<td>98.7%</td>
</tr>
<tr>
<td>NOR</td>
<td>87.1%</td>
<td>96.6%</td>
<td>98.9%</td>
<td>99.3%</td>
<td>85.5%</td>
</tr>
<tr>
<td>PAK</td>
<td>93.5%</td>
<td>97.4%</td>
<td>96.8%</td>
<td>97.4%</td>
<td>99.1%</td>
</tr>
<tr>
<td>RUS</td>
<td>97.9%</td>
<td>94.7%</td>
<td>99.7%</td>
<td>97.1%</td>
<td>97.4%</td>
</tr>
<tr>
<td>SGP</td>
<td>82.1%</td>
<td>98.5%</td>
<td>99.0%</td>
<td>99.9%</td>
<td>85.9%</td>
</tr>
<tr>
<td>SWE</td>
<td>89.9%</td>
<td>91.0%</td>
<td>91.3%</td>
<td>97.1%</td>
<td>93.3%</td>
</tr>
<tr>
<td>THA</td>
<td>90.6%</td>
<td>99.6%</td>
<td>99.8%</td>
<td>99.7%</td>
<td>91.1%</td>
</tr>
<tr>
<td>TWN</td>
<td>84.9%</td>
<td>90.6%</td>
<td>90.9%</td>
<td>92.3%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Average</td>
<td>89.2%</td>
<td>94.8%</td>
<td>95.3%</td>
<td>95.2%</td>
<td>96.5%</td>
</tr>
</tbody>
</table>

Table 5 reports the effectiveness scores for selected peer definitions, for the 20 countries listed in Table 4 that have at least 100 firm-year observations under the SIC4 definition. We tabulate results for the Jones model with intercept; other accruals models produce similar results 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. Thus, an effectiveness score of 100% indicates that the peer group was always the best at detecting earnings management. The last row in the table reports the average effectiveness score for each peer group, calculated across the 20 countries.
Table 6 examines the performance of the lagged asset peer group when non-U.S. samples are constrained by industry definitions (the “restricted sample”) and when samples are not so constrained (the “maximized sample”). The 20 countries in the restricted sample are identical to those shown in Table 5. The 58 countries in the maximized sample include those 20, as well as the 38 other countries in Table 4 with at least 100 firm-year observations in cross-section. The 100 firm-year observation requirement is imposed (by us) to facilitate our simulation. Table 6 shows detection rates, for a 10% seed level using the Jones model with intercept; other seed levels and models produce similar results and are not reported.

### Table 6

<table>
<thead>
<tr>
<th>Country</th>
<th>Restricted Sample</th>
<th>Maximized Sample</th>
<th>Country</th>
<th>Restricted Sample</th>
<th>Maximized Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARE</td>
<td>18.7%</td>
<td>14.8%</td>
<td>KWT</td>
<td>14.8%</td>
<td>22.9%</td>
</tr>
<tr>
<td>ARG</td>
<td>30.0%</td>
<td>27.4%</td>
<td>LKA</td>
<td>20.0%</td>
<td>20.0%</td>
</tr>
<tr>
<td>AUS</td>
<td>16.7%</td>
<td>19.8%</td>
<td>LTU</td>
<td>20.0%</td>
<td>17.1%</td>
</tr>
<tr>
<td>AUT</td>
<td>27.1%</td>
<td>25.3%</td>
<td>LUX</td>
<td>23.5%</td>
<td>25.5%</td>
</tr>
<tr>
<td>BEL</td>
<td>29.0%</td>
<td>17.7%</td>
<td>LVA</td>
<td>17.7%</td>
<td>29.5%</td>
</tr>
<tr>
<td>BMU</td>
<td>18.3%</td>
<td>19.8%</td>
<td>MAR</td>
<td>23.3%</td>
<td>22.9%</td>
</tr>
<tr>
<td>BRA</td>
<td>39.6%</td>
<td>24.7%</td>
<td>MEX</td>
<td>25.6%</td>
<td>27.7%</td>
</tr>
<tr>
<td>CHE</td>
<td>37.5%</td>
<td>24.7%</td>
<td>MYS</td>
<td>23.3%</td>
<td>27.7%</td>
</tr>
<tr>
<td>CHL</td>
<td>44.4%</td>
<td>30.4%</td>
<td>NGA</td>
<td>11.0%</td>
<td>24.0%</td>
</tr>
<tr>
<td>CHN</td>
<td>19.5%</td>
<td>20.3%</td>
<td>NLD</td>
<td>27.7%</td>
<td>27.7%</td>
</tr>
<tr>
<td>COL</td>
<td>46.6%</td>
<td>20.0%</td>
<td>NOR</td>
<td>24.0%</td>
<td>24.0%</td>
</tr>
<tr>
<td>CYM</td>
<td>15.8%</td>
<td>17.7%</td>
<td>NZL</td>
<td>25.6%</td>
<td>25.6%</td>
</tr>
<tr>
<td>CZE</td>
<td>20.5%</td>
<td>21.1%</td>
<td>OMN</td>
<td>20.7%</td>
<td>20.7%</td>
</tr>
<tr>
<td>DEU</td>
<td>16.3%</td>
<td>21.1%</td>
<td>PAK</td>
<td>27.5%</td>
<td>17.7%</td>
</tr>
<tr>
<td>DNK</td>
<td>28.0%</td>
<td>21.1%</td>
<td>PER</td>
<td>17.7%</td>
<td>17.7%</td>
</tr>
<tr>
<td>EGY</td>
<td>15.8%</td>
<td>28.0%</td>
<td>PHL</td>
<td>21.8%</td>
<td>21.8%</td>
</tr>
<tr>
<td>ESP</td>
<td>25.7%</td>
<td>28.0%</td>
<td>POL</td>
<td>17.3%</td>
<td>17.3%</td>
</tr>
<tr>
<td>FIN</td>
<td>25.7%</td>
<td>28.4%</td>
<td>PRT</td>
<td>22.4%</td>
<td>22.4%</td>
</tr>
<tr>
<td>FRA</td>
<td>28.0%</td>
<td>29.0%</td>
<td>RUS</td>
<td>19.1%</td>
<td>21.1%</td>
</tr>
<tr>
<td>GBR</td>
<td>26.2%</td>
<td>29.0%</td>
<td>SAU</td>
<td>18.6%</td>
<td>18.6%</td>
</tr>
<tr>
<td>GRC</td>
<td>21.4%</td>
<td>21.4%</td>
<td>SGP</td>
<td>22.2%</td>
<td>21.2%</td>
</tr>
<tr>
<td>HKG</td>
<td>21.1%</td>
<td>21.1%</td>
<td>SVN</td>
<td>42.3%</td>
<td>42.3%</td>
</tr>
<tr>
<td>HUN</td>
<td>25.3%</td>
<td>21.1%</td>
<td>SWE</td>
<td>16.0%</td>
<td>26.6%</td>
</tr>
<tr>
<td>IDN</td>
<td>19.8%</td>
<td>25.3%</td>
<td>THA</td>
<td>40.8%</td>
<td>22.1%</td>
</tr>
<tr>
<td>IND</td>
<td>20.0%</td>
<td>18.9%</td>
<td>TUR</td>
<td>16.8%</td>
<td>16.8%</td>
</tr>
<tr>
<td>IRL</td>
<td>27.6%</td>
<td>18.9%</td>
<td>TWN</td>
<td>24.0%</td>
<td>26.1%</td>
</tr>
<tr>
<td>ISR</td>
<td>23.8%</td>
<td>23.8%</td>
<td>ZAF</td>
<td>24.2%</td>
<td>24.2%</td>
</tr>
<tr>
<td>ITA</td>
<td>29.8%</td>
<td>29.8%</td>
<td>Average</td>
<td>25.4%</td>
<td>26.0%</td>
</tr>
<tr>
<td>JOR</td>
<td>17.1%</td>
<td>17.1%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JPN</td>
<td>50.9%</td>
<td>52.7%</td>
<td>Average if in restricted sample</td>
<td>24.5%</td>
<td>23.2%</td>
</tr>
<tr>
<td>KOR</td>
<td>24.6%</td>
<td>27.7%</td>
<td>Average if not in restricted sample</td>
<td>24.0%</td>
<td>24.0%</td>
</tr>
</tbody>
</table>

Table 6 Lagged Asset Peer Performance for Restricted and Maximized Samples
Table 7 examines how the size of the lagged asset peer group affects earnings management detection rates. Results are shown for the Jones model with intercept; other models produce similar results and are not tabled. Note that the 0% seed level is a specification check, insofar as a well-specified model should show detection of positive earnings management (when none is induced) at a 10% significance level.

<table>
<thead>
<tr>
<th>Magnitude of Accruals Management (in % of lagged total assets)</th>
<th>Number of (Non-Event) Peers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
</tr>
<tr>
<td>0% Jones Model (with intercept)</td>
<td>6.1</td>
</tr>
<tr>
<td>2%</td>
<td>8.6</td>
</tr>
<tr>
<td>4%</td>
<td>12.1</td>
</tr>
<tr>
<td>6%</td>
<td>16.9</td>
</tr>
<tr>
<td>8%</td>
<td>22.5</td>
</tr>
<tr>
<td>10%</td>
<td>28.4</td>
</tr>
<tr>
<td>12%</td>
<td>34.0</td>
</tr>
<tr>
<td>14%</td>
<td>39.7</td>
</tr>
<tr>
<td>16%</td>
<td>44.9</td>
</tr>
<tr>
<td>18%</td>
<td>49.6</td>
</tr>
<tr>
<td>20%</td>
<td>54.2</td>
</tr>
</tbody>
</table>

Table 7 Optimizing the Number of Non-Event Firms
Figure 1 graphs the detection rates for each peer definition and seed level (between 0% and 20%) for the Jones model with intercept; other models yield similar inferences and are not shown.
Figure 2 graphs the detection rates for each peer definition and seed level (between 0% and 100%) for the Jones model with intercept; other models yield similar inferences and are not shown.
References


