

# Exploring the Accrual Landscape by Open Science\*

*Joachim Gassen*

*Humboldt-Universität zu Berlin*

*December 6, 2018*

## **Abstract**

In an influential study, Bushman, Lerman, and Zhang (2016) document that the cash flow accrual association diminishes over time in the U.S. They identify non-timing related accruals as the main determinant of this trend. I reproduce, explore and extend their findings, both within the U.S. and internationally, to provide three contributions to the literature. First, I document how multicollinearity affects the reliability of their findings. Second, by exploring U.S. data as well as an international panel sample, I show that changes in the underlying cash flow distribution explain a large part of the observed variation in the cash flow accrual association. The remaining time-trend is relatively robust across countries. It is confined to the nineties and associated with changes in the number of listed firms as well as with the relative share of intangible intensive industries. All these observations seem largely at odds with accounting standards driving the changes in the cash flow accrual association. Third, by providing an interactive presentation of my key results online (click here to ExPanD) and by disclosing all of my material together with an Appendix detailing the necessary steps to reproduce my work, I hope to promote the benefits of open science to accounting research.

---

\*This project started as co-authored work together with Martin Bierer. I thank Martin for his contributions and passion that got this project off the ground. I take sole responsibility for the current version and appreciate feedback from Ulf Brüggemann, Maximilian Muhn, Jochen Pierk, Harm Schütt as well as from the audiences of seminars at Tilburg, Bristol, Exeter, the AMIS conference in Bucharest, the 2018 EAA Doctoral Colloquium and HU Berlin. The author can be reached at [gassen@wiwi.hu-berlin.de](mailto:gassen@wiwi.hu-berlin.de).

# 1 Introduction

The scientific community widely agrees that reproduction (understood here as reproducing findings using the same sample and method) and replication (understood here as retesting findings using a new sample while applying the same method) of empirical results are a central building block of sound empirical evidence. Yet, published reproduction or replication studies in economics are rare (Duvendack, Palmer-Jones, and Reed (2017)). While several reasons for this observation have been discussed by prior literature, most commentators agree that the lack of established reproducible workflows makes reproductions and replications inherently costly to do (Gertler, Galiani, and Romero (2018)).

This study aims to promote reproductions and replications by demonstrating that reproducing the findings of an influential accounting study can provide additional insights into such a fundamental research question as the association between accruals and cash flows. It makes extensive use of exploratory data analysis and visualization tools. Maybe most importantly, it reduces the costs of future reproductions and extensions of archival studies in two ways: For one, it uses an open science workflow that allows researchers to immediately use and build on the underlying code and data. For another, it features the ExPanDaR toolbox, developed especially for this study. This open source R package allows interactive exploratory panel data analysis in a web-based environment. It thereby provides interested parties (like readers, editors and reviewers) the opportunity to assess the robustness of findings without providing them access to potentially proprietary or commercialized data.

The resulting reduction in reproduction costs should have the following two effects: (a) Using ExPanD, it becomes significantly easier for interested parties to assess the robustness of empirical archival work, reducing the risk of p-hacking and increasing the confidence in published results. (b) Interested researchers can more efficiently reproduce and build on prior evidence, potentially yielding quicker evolution of knowledge.

To demonstrate the usefulness of this workflow, I reproduce and extend a study addressing a key area of interest to empirical archival researchers in the area of financial accounting: the statistical properties of the cash flow accrual association. The central objective of accruals is to smooth temporary fluctuations in cash flows, thereby increasing the informativeness of earnings.

This accounting activity results in a negative association between accruals and concurrent cash flows from operations. Bushman, Lerman, and Zhang (2016) (BLZ from now on) analyze the time trend of this association for U.S. firms and provide compelling evidence that the negative association between accruals and cash flow from operations diminishes over time. They also document that this trend cannot easily be explained by economics-based cash flow shocks, changes in the sample composition, changes in the asymmetric timeliness of earnings or by industry trends. Their main explanation for the finding is the increase of non-timing related accrual recognition and the increasing frequency of loss years.

I revisit and extend the findings of BLZ by applying the following three steps:

- (i) *Reproduction* (same research design, same population, same sample): I reproduce the sample and key variables of BLZ. I am able to reproduce the diminishing trend in the cash flow accrual association but fail to reproduce their findings with respect to its determinants, likely because their time-series sample is affected by extremely high levels of multicollinearity.
- (ii) *Robustness testing* (modified research design, same population): As BLZ openly discuss that their determinants tests are affected by multicollinearity, I document how this multicollinearity affects the reliability of their main inferences. I show that, based on the reproduced sample, one cannot make any reliable inference about which factors explain the declining cash flow accrual association, thereby qualifying some of BLZ's key findings. By using state of the art tools for data visualization I provide additional exploratory insights into the driving forces behind the diminishing cash flow accrual association in the U.S., showing that the changing cash flow distribution and new intangible intensive firms entering the market in eighties and the nineties are influential for the changes in the cash flow accrual association. To enable the interested reader to explore the robustness of the BLZ findings him/herself, I provide interactive online access to the reproduced sample via the ExPanD tool.
- (iii) *Replication and extension* (same and modified research design, different population, different sample): In order to address the multicollinearity issue that affects the U.S. analysis of BLZ, I replicate their methodology for an international sample, thereby moving from 51 yearly observations to up to a maximum of 788 country-year observations. Based on this sample, I confirm my exploratory findings from U.S. data by documenting that a large part of the time trend in the cash flow accrual association can be explained by changes in the underlying cash

flow distribution. The remaining time trend is relatively stable across countries, confined to the nineties, and associated with changes in the number of listed firms and the relative share of intangible intensive industries. Again, I provide readers access to the data underlying my analysis via the ExPanD app.

My findings highlight the benefits of applying an open science work flow for empirical archival studies. I successfully reproduce a current and prominent financial accounting study to show that its main findings are reproducible but subject to careful interpretation because of underlying data issues. My exploration reveals that the distribution of operating cash flows is changing significantly over time and that this appears to be strongly associated with the shift in the cash flow accrual association observed by BLZ. The change in the cash flow distribution and the diminishing cash flow accrual association is also observable for the international sample and the associated time-trend is surprisingly robust across countries. It is associated with the global IPO wave in the nineties, leading to an inflow of firms from intangible intensive industries, in particular the health sector.

The results from BLZ are commonly interpreted to indicate that changes in accounting rules have triggered an increase in non-timing related accruals, ultimately driving the weakening cash flow accrual association. My evidence is largely at odds with this explanation. Instead, it supports the findings of Srivastava (2014) as it is consistent with changing firm demographics, meaning more young intangible intensive firms becoming publicly listed around the globe, being influential for the observed changes in the cash flow accrual association.

A concurrent working paper which also builds on the findings of BLZ is Nallareddy, Sethuraman, and Venkatachalam (2018). The authors investigate the role of earnings and cash flow from operations for the predictability of future cash flows. Interestingly, they find that the explanatory power of operating cash flows (to predict future operating cash flows) increases over time. They show that this observation generalizes to an international sample of firms. Using a similar methodology as BLZ, they observe that the explanatory power of operating cash flows is associated with “decreasing operating cycles, decreasing working capital and increasing intangible intensity”. While Nallareddy, Sethuraman, and Venkatachalam (2018) explore the effect of BLZ’s findings on the predictive power of operating cash flows and accruals, my focus is on understanding

the drivers of the diminishing association between accruals and cash flows better.<sup>1</sup>

The code used to generate this paper and its results is publicly available on GitHub. The code has been optimized for reproducibility. The open science workflow enables researchers to build on my findings and to continue exploring the landscape of accrual accounting. Readers who have access to national and international Compustat data via the WRDS interface should be able to reproduce all of my findings easily by following the Reproduction Guide provided in Appendix B.

In addition, the interactive panel data exploration toolbox ExPanD developed as part of this project has the potential to reduce the need for manual reproduction of prior work: It facilitates systematic data exploration and robustness checking in a web-based environment. The code for ExPanD has been developed with the objectives of accessibility and re-usability in mind and is available as a separate R package (“ExPanDaR”). It can be used with generic panel data sets online (click here) and easily adapted to host specific panel data for secure online user-side exploration. It enables researchers to assess the robustness of published work without having access to the underlying data. I hope that it will reduce p-hacking, improve the efficiency of the review process and speed up the generation of new exploratory insights.

## 2 Reproduction of Bushman, Lerman, and Zhang (2016)

I choose with BLZ a study for my reproduction and replication exercise that (i) provides an important finding, which is broad in its scope and has the potential to substantially influence future financial accounting research, (ii) is exploratory in nature and raises related research questions that I am able to address by further data exploration within-sample or out-of-sample and (iii) which facilitates reproduction (because of using publicly available data as well as thorough documentation of data collection, sample selection and research design).

BLZ document that the negative association between accruals and concurrent cash flows in the U.S. decreased substantially over time and virtually disappeared in recent years. This finding is very relevant for our field as the properties of accruals are key drivers of established measures for earnings quality and earnings management (e.g., Jones (1991), Dechow (1994), Dechow and Dichev (2002), Dechow, Ge, and Schrand (2010) for an overview). As BLZ indicate in their study, work

---

<sup>1</sup>The changes in the temporal distribution of cash flows documented in this study might also be influential for the findings of Nallareddy, Sethuraman, and Venkatachalam (2018). However, I am not exploring this link here.

that is using such measures might be picking up overall time trends when designing tests (partially) relying on the time dimension for identification such as difference-in-differences designs. Given that the common trend assumption is critical for these designs, a key research question of BLZ (and also of this study) is the identification of cross-sectional variation that helps to explain this striking time-series pattern.

Another reason why the findings of BLZ are a good candidate for exploration and extension is the magnitude of the additional tests and extensions that the authors themselves provide. To some extent, BLZ can itself be classified as exploratory. By using visual exploratory data analysis tools, I try to make the findings of BLZ more accessible to the reader, potentially allowing future work to chart the accrual landscape at an even higher level of detail.

I start reproducing the sample of BLZ by following the sample selection steps and the variable definitions of BLZ as closely as possible. First, I create a November 27, 2018 dump of the Compustat US North America fundamentals data base. Second, I generate the variables used in BLZ by following their definitions (The resulting variable definitions are presented in Appendix A). Third, I apply the sample screens and the by-year winsorizing procedure described in BLZ, p.49f. This means that I delete observations from the financial services industry, observations from years with significant mergers/acquisitions activity and observations with missing cash flow or accrual data. In addition, I limit the sample to the periods 1964 to 2014. I refer the resulting sample as the reproduced sample. Its descriptive statistics are displayed in Table 1.

[Table 1 about here.]

For reference, Figure 1 displays the descriptive statics of the original BLZ sample as presented in Panel A of BLZ's Table 1.<sup>2</sup> I note several differences between the sample of BLZ and my sample. First, my sample is larger than the sample of BLZ (261,154 versus 217,164 observations of BLZ). Second, firms in my sample seem to be significantly smaller than the firms of BLZ (mean market value in reproduced sample M US-\$ 1,845 versus 2,139 of BLZ).

[Figure 1 about here.]

I assume that my initial Compustat data screens differ from the screens that BLZ applied, as my sample appears to include additional firms that tend to be relatively small in terms of market

---

<sup>2</sup>I note that both, my reproduced sample as well as the sample of BLZ, report a maximum value for *MV* that is likely due to a data error. It can be attributed to the firm "Delhaize America Inc.", fiscal year 2001. As *MV* is presented here for reference only and to maintain consistency with BLZ, I do not delete this observation.

value. In order to make my sample more comparable to the one used in BLZ, I apply an additional screen and exclude firms with average total assets below M US-\$ 7.5 from my sample (prior to winsorizing the data). I label the resulting sample as the test sample. Table 2 reports its descriptive statistics.

[Table 2 about here.]

After applying this additional screen, the test sample appears to be reasonably close, both in terms of sample size and descriptive statistics, to the BLZ sample presented in Figure 1. The only noteworthy exemption is that the extreme values of TACC presented by BLZ (Minimum -1.717 and Maximum 1.249) are substantially larger compared to the extreme values presented in my test sample (Minimum -1.03 and Maximum 0.371).

Despite these remaining differences, I continue by reproducing the main finding of BLZ. Therefore, I estimate yearly regressions of the following model:

$$TACC_{t,i} = \beta_{0,t} + \beta_{1,t}CFO_{t,i} + \epsilon_{t,i} \quad (1)$$

[Figure 2 about here.]

The left-hand side of Figure 2 presents the  $\beta_1$  coefficient over time along with the original coefficients reported in Table 2 of BLZ. It shows that my results are very similar to the key result reported in Table 2 of BLZ. In both cases, one observes a steady increase in  $\beta_1$  over time indicating that the cash flow/accrual association weakens over time. Furthermore, my results in terms of the adjusted  $R^2$  (right-hand side of Figure 2) mirror the findings of BLZ and illustrate the pronounced decline in the correlation between accruals and cash flows over time. The successful reproduction of the key BLZ findings provides some assurance that the test sample is close enough to the original sample of BLZ.

[Figure 3 about here.]

As a next step, I reproduce Table 4 of BLZ (presented as Figure 3) that contains their test for potential determinants of the cash flow accrual association trend. For this Table, BLZ calculate a yearly time series of adjusted  $R^2$  values from by year Dechow and Dichev (2002) type regressions:

$$TACC_{t,i} = \gamma_{0,t} + \gamma_{1,t}CFO_{t-1,i} + \gamma_{2,t}CFO_{t,i} + \gamma_{3,t}CFO_{t+1,i} + \epsilon_{t,i} \quad (2)$$

They then regress this time-series of 51 observations on a set of covariates. First, they use the yearly standard deviation of cash flows (*CFO\_SD*) and the one-period autocorrelation of cash flow changes (*DCFO\_ACORR*) as indicators for economic and timing related cash-flow shocks. Second, they use the standard deviation of one-time and non-operating items (*SD\_OI\_PTI*) to assess the effect of non-timing related accruals. Third, they include the yearly frequency of loss firms (*PCTLOSS*) that they also interpret as a proxy for non-timing related accruals. Forth, they use the adjusted  $R^2$  of Dichev and Tang (2008) regressions (*DT\_ADJRSQ*):

$$SALES_{t,i} = \delta_{0,t} + \delta_{1,t}EXPENSES_{t-1,i} + \delta_{2,t}EXPENSES_{t,i} + \delta_{3,t}EXPENSES_{t+1,i} + \epsilon_{t,i} \quad (3)$$

as a measure for the matching between revenues and expenses. Forth and finally, they use the annual average of selling, general and administrative expenses, deflated by total expenses, (*SGAINT\_MEAN*) as an indicator of intangible intensity.

I follow the variable definitions and the regression setup of BLZ (variable definitions are included in Appendix A). The resulting regression results are reported in Table 3.

[Table 3 about here.]

Comparing the findings from Table 3 with the original BLZ findings presented in Figure 3 it is apparent that, while the coefficients for *TIME* are very similar in terms of magnitude and equally significant, some coefficients exhibit relatively large differences. Most strikingly, the coefficient for *SGAINT\_MEAN* is robustly negatively associated with the adjusted  $R^2$  of Dechow and Dichev (2002) regression in my analysis, whereas it is insignificant in the analysis of BLZ. Also, in the full specification presented in the left-most column, the standard deviation of cash flows is significantly positively related with the outcome variable based on my analysis while BLZ find it to be insignificant. Finally, in the full specification, *PCTLOSS* is insignificant in my specification but significantly negative in BLZ's analysis.

It is non-trivial to explain these differences, also as BLZ do not report descriptive statistics for their independent variables. Omitting yearly winsorizing of the independent firm-year variables that serve as inputs to the yearly statistics<sup>3</sup> brings the results closer to the original findings of BLZ

<sup>3</sup>While Table 1 of BLZ indicates that all variables besides market value are winsorized the footnote to Table 4 refers only to accrual and cash flow data as been winsorized

but not surprisingly these differences are driven by very few extreme observations. In addition, BLZ limit the firm-year sample to construct Table 4 to observations that have lead and lag values of cash flow available. This leads to their time-series only covering the periods 1964-2013 (50 observations). Omitting 2014 has only a very limited impact on the results presented in Table 3. The online representation of this paper (click here to ExPanD) presents an alternative sample that is based on an attempt to reproduce the methodology of BLZ as closely as possible.

### 3 Robustness

The reproduction indicates that the findings of Table 3 are very sensitive to design choices. A main reason for this are the high to extremely high variance inflation factors that are presented in squared brackets in Table 3. BLZ openly discuss that multicollinearity is an issue in their study:

"One caveat in identifying potential explanations for the attenuation in the correlation between accruals and cash flows is that the explanatory variables considered in table 4 are highly correlated. The magnitude of Pearson correlations between the six explanatory variables ranges from 0.5 to 0.9, while the time-series examined provide only 51 observations for each variable. Multicollinearity could increase the standard error and thus render the coefficient estimate on one or more variables to be statistically insignificant. The second caveat is that any variable that is mechanically related to the time trend, such as the authors' age over time, is likely to exhibit significant coefficients. In light of these two caveats, we adopt the following criteria when interpreting the results. First, each explanatory variable should have a strong economic theory or intuition. Second, the explanatory variable should have a significant coefficient with the expected sign in the stand-alone regression, and the coefficient must not flip its sign from the stand-alone regression to the full model presented in column 6. Under these criteria, we interpret the results in table 4 to mean that one-time and nonoperating items and loss firms are mostly responsible for the attenuation in the overall correlation between accruals and cash flows and explain the majority of the effect (about 63 % =  $(0.016-0.006)/0.016$ )."

(BLZ, p. 65-66)

To explore the multicollinearity issue in more detail, Table 4 reports univariate regression

results based on the variables used in Table 3 and Table 5 the Pearson and Spearman correlations for the time series sample.

[Table 4 about here.]

[Table 5 about here.]

Based on the reproduced time-series sample, the multicollinearity of the independent variables seems even more severe than based on the analysis of BLZ. The absolutely smallest Pearson (Spearman) correlation is 0.77 (0.79) and most of the correlations are larger than 0.9. The univariate results in Table 4 support the observation of BLZ. Only *SD\_OI\_PTI*, *PCTLOSS* and *SGAINT\_MEAN* pass the “smell test” suggested by BLZ in the sense that they are significant in the expected suggestion direction univariately and at least have the same sign in the multiple regression of Table 3. However, the magnitude of the *SD\_OI\_PTI* and *PCTLOSS* coefficients changes drastically across specifications. The *SGAINT\_MEAN* coefficient, which proxies for intangible intensity, stays relatively constant but differs from the BLZ analysis.<sup>4</sup>

Based on the above, the last column of Table 4 presents the results of a multiple regression that contains only the independent variables that pass the BLZ “smell test”. I make two observations. First, the variance inflation factors are all close to or clearly above 10 in this regression, indicating strong multicollinearity. Second, the coefficient for *SD\_OI\_PTI* flips its sign and becomes insignificant. Based on standard econometrics, multicollinearity implies that estimates for each individual coefficient are imprecise as variation in each variable is not sufficiently orthogonal to the other covariates. Thus, the only valid conclusion that I can draw so far from reproducing the analysis of BLZ is that *TIME*, *SD\_OI\_PTI*, *PCTLOSS* and *SGAINT\_MEAN* are jointly associated with the reduction of  $R^2$  of Dechow and Dichev (2002) regression in the U.S. over time.<sup>5</sup> Any conclusions beyond that, and in particular that at the end of the quote above, are not supported by my reproduction.

Therefore, and to contribute to our understanding about the potential reasons for the

---

<sup>4</sup>While I can only speculate as of why, using the ExPanD platform and the BLZ reproduced sample it becomes apparent that its time-series of *SGAINT\_MEAN* observations is drastically influenced by two outliers if it is calculated based on non-winsorized data.

<sup>5</sup>On a more conceptual note, the constructs *SD\_OI\_PTI*, *PCTLOSS* and *SGAINT\_MEAN* could also be challenged as they are driven by accruals and as such are (in part) mechanically related to the dependent variable. Also, the support for *TIME* is not existent ex ante but is to be established by the empirical analysis. In the notion of the argument quoted from BLZ above, this would require a higher reliability hurdle for the *TIME* coefficient. I am not following these arguments here as I concentrate on the general feasibility of the research design.

diminishing cash flow accrual association, I now revisit the changes in the cash flow accrual association documented by BLZ more fundamentally. I do this by first exploring the association with visual tools. After that, I use an international sample that allows me to identify covariates that are associated with this phenomenon.

## 4 Exploration

From the results of the previous section, I take special interest in the flipping coefficient for  $CFO\_SD$ . Based on the definition of  $R^2$ , one should expect a mechanical positive association between the variance of the dependent variable of the Dechow and Dichev (2002) regression ( $CFO$ ) and  $R^2$ , assuming that the variance of the residual stays constant. For the multiple regression, we observe this positive association but univariately, the association is strongly negative. This can be the case if the functional form of the cash flow accrual association varies along the cash flow distribution over time or when the error term of the cash flow accrual association is heteroscedastic. In both cases this indicates that the distribution of  $CFO$  is related to the cash flow accrual association.

To explore this further, I partition the firm-year test sample by cash flow deciles. For each decile, I re-run yearly regressions based on model (1), the univariate model regressing accruals on cash flows. In the second step, I use the generated  $\beta_1$  and adjusted  $R^2$ s and regress them on the variable  $TIME$ :

$$\beta_{1,t}[Adj.R_t^2] = b_0 + b_1 TIME_t + \epsilon_t \quad (4)$$

Figure 4 plots the  $b_1$  estimates of model (4) across the cash flow deciles. It shows that the time trend (of the diminishing cash flow accrual association) is strongly moderated by the magnitude of operating cash flows. Negative cash flow firms experience the diminishing cash flow accrual association but the trend is much less pronounced for positive cash flow firms.

[Figure 4 about here.]

I next explore whether this observation extends to the time trend with respect to the adjusted  $R^2$  of model (1). Figure 5 communicates the result.

[Figure 5 about here.]

This pattern is striking as it indicates that the drop in  $Adj.R^2$  described in BLZ seems to

be driven by extreme cash flow observations. To explore this fact a little bit further, Figure 6 plots the time trends separately for firms within the first, fifth and tenth decile of *CFO*. It becomes apparent from the plot that the explanatory power of cash flows for accruals tends to lie in the extremes. Given that the bins at the extreme deciles are wider, this is not entirely surprising. What seems more relevant is that the coefficient trend differs across the extremes. For the sub-sample with particularly low cash flows the negative association between accruals and cash flows clearly diminishes at the end of the sample period. In contrast, for firms with particularly high cash flows, the association seems to be more persistent (positive  $R^2$ s and negative coefficient throughout the sample period).

[Figure 6 about here.]

While one observes a negative association between cash flows and accruals for all sub-samples at the beginning of the sample period, the association becomes positive for the low cash flow sub-sample and remains negative for the high cash flow sub-sample. This indicates a change in the functional form of the association between cash flows and accruals.

To explore the functional form of the cash flow accrual association over time, I estimate locally estimated smooth regressions for yearly samples. The result can be visualized by a video ([click here to see video](#)) or by comparing the functional forms across two samples, one spanning the beginning of the sample period (1964-1973) and one spanning the end of the sample period (2005-2014).

[Figure 7 about here.]

Figure 7 and the video display a stark contrast in the estimated functional form. For the beginning of the sample period the relation is almost perfectly linear. However, towards the end of the sample period the relation is highly non-linear, exhibiting a kinked regression curve very much in line with the idea of asymmetric timeliness of accruals as suggested by Ball and Shivakumar (2006). Note that the strong difference in the functional forms across periods is not driven by an increasing sample size, as the plot is based on 20,000 observations for each period.

The change in the functional form seems to be mainly driven by two types of firm-year observations that were not observable in the early sample periods: (i) observations with extremely low operating cash flows and (ii) firm-year observations with moderate levels of operating cash flows but particularly low accruals. This results in a positive (negative) cash flow accrual association for

firm-year observations with negative (positive) cash flows towards the end of the sample period. Again, this finding is in line with Figure 6, which indicates that the cash flow accrual association is strongly moderated by the magnitude of operating cash flows.

What also becomes apparent is that the distribution of cash flows and accruals becomes more dispersed towards the end of the sample period. To visualize the change in the distribution of operating cash flows over time, Figure 8 plots density estimates over the sample years.

[Figure 8 about here.]

Towards the end of the sample period the distribution of cash flow from operations becomes significantly wider and more left skewed. Particularly, the increase of relatively large negative cash flows over time is striking and likely to be influential for the observed change in functional form. As becomes apparent from the video and as has been documented by prior literature (Srivastava (2014), Fama and French (2004)), many of these negative cash flow firms entered the U.S. market during the 80s and the 90s. Therefore, I explore next how the changes in the cash flow accrual association relate to sample composition over time.

[Figure 9 about here.]

Figure 9 presents a visual reporting the *TIME* coefficient from model (4) based on adjusted  $R^2$ s from Dechow and Dichev (2002) regressions based on balanced samples ending in the respective period. It documents that the time trend was strongest for the period 1973 to 1992 and is last observed to be significant for the time period 1984 to 2003.<sup>6</sup> This implies, first, that the trend is driven by 80s and the 90s in the U.S. and second, in line with the analysis presented by BLZ, that it is not (only) due to new firms entering the market.

Finally, visual inspection of Figure 7 also indicates that a sizable fraction of these negative cash flow firms are based in the health sector. As can be easily confirmed by studying a random sample of annual reports of these, most of them tend to be relatively small, and invest heavily in in-process research and development. As this expenses do not materialize as tangible capitalizable assets, they result in negative cash flows that are large relative to the firms' asset base. Since many of firms seek new financing regularly, it is not uncommon for them to report persistently large

---

<sup>6</sup>One potential concern about this analysis is that  $R^2$  is bounded by zero and that, as the association between cash flow and accruals weakens,  $R^2$ s converge to zero. I address this concern by also analyzing fisher-transformed values of  $R^2$  that are not bound by zero. The graph looks almost identical with different magnitudes of coefficients and I present the untransformed version here for expositional ease.

negative cash flows over multiple years.

[Figure 10 about here.]

Figure 10 explores this observation more formally and also tests whether new IPO firms contribute to the change in the cash flow accrual association. It reports the standard deviation of  $CFO$  and the Pearson correlation between  $CFO$  and  $TACC$  for different industry samples over time. In combination, these measures help to understand the drivers of the  $R^2$  metric of model (1):

$$R^2 = \rho(CFO, TACC)^2 = \frac{cov(CFO, TACC)^2}{\sigma^2(CFO)\sigma^2(TACC)} \quad (5)$$

The left panel of Figure 10 shows the data for the full test sample, divided by time in two sub-samples up to and post 1989. It becomes apparent that the upward trend in the Pearson correlation measure is present in all industries but particularly strong in the health, business equipment, telecommunication and chemical sectors. All these sectors are regularly classified as being very intangible intensive. The right panel is using a balanced panel containing observations from the ten year period 1985 to 1994. This ten year windows keeps the observational drain from balancing the observations limited but as can be gauged by the dot sizes, it still results in significantly fewer observations than the left panel. Not surprisingly, for the balanced sample, the time trend is much less pronounced. For the extractive industry it runs into the opposite direction. For consumer products it is very weak. It is strongest for the health sector followed by the chemical sector. This observation is consistent with new firms contributing to the diminishing cash flow accrual association.

Summing up, the exploratory analysis of the cash flow accrual association reveals the following insights: (1), the distribution of cash flows from observations has changed significantly in the U.S. over the period of 1964 to 2014. It has become more dispersed and more left skewed, leading to a sizable fraction of firms with large negative cash flows, relative to their asset base. (2), this new type of firms has given rise to a new functional form of the cash flow accrual association. It is basically flat and uninformative for negative cash flow firms but remains negative and significant for positive cash flow firms. (3) While this is also consistent with an increased level of conditional conservatism as indicated by Ball and Shivakumar (2006), it seems that sample composition, in particular new firms entering in the U.S. market in the 80s and the 90s is one, if not the main

reason for this observation (consistent with Srivastava (2014)). (4) The trend is also observable for incumbent firms. (5) In terms of industries, it seems to be driven by intangible intensive industries, in particular the health sector. In the next section, I will test whether these observations also hold out of sample by using an international sample.

## 5 International Tests

As became apparent from the reproduction of BLZ, it is inherently hard, if not impossible, to infer anything about the determinants that are associated with the diminishing cash flow accrual relation from analyzing U.S. yearly data. In order to address this concern and in order to verify whether my exploratory findings hold out of sample, I now test for determinants using international data. Going international allows to use a set of countries as separate observations, moving from a time-series analysis to a panel analysis. To make sure that my exploratory findings based on the U.S. do not bias my inferences, I omit the U.S. from the analysis.

The data for the international sample is obtained from Compustat Global. I apply the original BLZ screens, but not the additional size screen, in order to generate the international sample. In addition, I do not drop firm-year observations with significant acquisition activity, given that the Compustat item (`# aqs`) is not available internationally. As original cash flow data is sparse in the early years of international data, I replace original cash flow and accruals data with balance sheet based measures whenever original cash flow data are not observable. To be included in the international country-year sample, I require (i) each country-year to have at least 30 firm-year observations and after applying this filter (ii) each country to have at least 20 country-years. For country-years with at least (less) than 100 firm-years I winsorize the extreme 1% (5%) observations on either side for all continuous variables. The resulting sample period spans from 1989 to 2016.

[Table 6 about here.]

Table 6 shows the distribution of the test sample across countries. In a first step, I examine the extent to which the diminishing cash flow accrual association represents an international phenomenon. For this reason, I estimate the model (4) based on data from annual Dechow and Dichev (2002) regressions for each country and plot the respective time coefficient in Figure 11. For the sake of comparison, this figure also contains a *TIME* estimate for the U.S., based on data

spanning the same sample period 1989 to 2016. A more negative coefficient in Figure 11 indicates a more pronounced decline of the association between cash flows and accruals over time. It can be seen that the diminishing cash flow and accrual association is indeed an international phenomenon. In the majority of countries, the time-trend is even more pronounced than in the US. However, it has to be noted that, as can be assessed by Figure 2, the trend in the U.S. is most prominent in the period prior to 1989.

[Figure 11 about here.]

In order to estimate to which extent intangible intensity, non-timing related accruals and market composition explain the diminishing cash flow accrual association, I use two additional variables besides the metrics *SGAINT\_MEAN*, *SD\_OI\_PTI*, *PCTLOSS* from the BLZ analysis. First, I assess the frequency of IPOs and delistings by measuring the relative number of listed firms in a given country year, relative to the maximum of listed firms for each country (*REL\_MSIZ*E). Second, I measure the share of firms from intangible intensive industries in each country year (*SHARE\_INT\_IND*). Based on my exploratory findings from the last section, I classify the sectors health, business equipment, telecommunication and chemical as intangible intensive sectors. In addition to these variables, I also include the first three moments of the cash flow distribution (*CFO\_MEAN*, *CFO\_SD* and *CFO\_SKEW*) to control for the effect of the changing cash flow distribution over time.

Table 7 reports the descriptive statistics for this set of variables and Table 8 presents the Pearson and Spearman correlations.

[Table 7 about here.]

[Table 8 about here.]

The correlations displayed in Table 8 are still large but much smaller than for the U.S. time-series sample. The descriptive statistics are of reasonable economic magnitude. Missing observations for certain variables are caused by variables that require lead and lag data, as for example the *Adj.R<sup>2</sup>* measure based on country-yearly Dechow and Dichev (2002) regressions (Model 2). To maximize power, I allow the sample to float across different model specifications. The sample is available online for user exploration ([click here to Expand](#)).

[Table 9 about here.]

Table 9 reports the findings from various multiple regression models. The overall uncon-

trolled time trend for the international panel sample is very similar to the U.S. sample, which seems at odds with unique features of the U.S. accounting regime driving the reduction of the cash flow accrual association. Next, the variance inflation factors across all models are still sizable but much less severe compared to the U.S. time-series analysis. This implies that it is now feasible to discuss separate coefficients independently. Turning to the moments of the cash flow distribution, it can be observed that they are robustly related to adjusted  $R^2$  across all specifications. The only exception is column (5) where *CFO\_MEAN* flips its sign, possibly due to *PCTLOSS* subsuming its predictive power. This is consistent with the changing cash flow distribution being a key driver of the changes in cash flow accrual association. As cash flows become smaller (more negative) and more left skewed, the association becomes weaker. After controlling for these factors, the standard deviation of cash flows is consistently positively related to  $R^2$ , which is in line with its mechanical relation with  $R^2$ .

The variables introduced by BLZ load consistently as far as *SD\_OI\_PTI* and *PCTLOSS* are concerned. *SGAINT\_MEAN* is not loading, contrary to the results for the U.S. time-series analysis. Given the variation in international income statement classifications, it is a priori doubtful whether *SGAINT\_MEAN* is able to identify country years with more intangible intensive activity. While the findings for *SD\_OI\_PTI* and *PCTLOSS* are interesting and in line with the finding of BLZ that on-time negative accruals are contributing to explaining the diminishing cash flow accruals association, I once again urge the reader to take care when interpreting these coefficients as they are mechanically related to reported accruals and thus, to the dependent variables of the models.

Finally, the measures picking up changes in market size and industry composition are robustly associated with the adjusted  $R^2$ s of country-yearly Dechow and Dichev (2002) regressions. This confirms the exploratory evidence for the U.S. sample. In terms of relative explanatory power, they contribute about as much as the cash flow moments to the model. It seems however, that they are explaining most of the time trend as they reduce the coefficient for *TIME* from -0.016 to -0.0057. This might imply that they are explaining a significant part of the time trend observed in the data. I will focus on this aspect when discussing my last series of tests.

[Table 10 about here.]

This last round of tests is presented by Table 10. For these analyses, I include yearly fixed effects into the models. This fixed effect structure will absorb any common time trend in the

data. To the extent that the dependent variables in Table 9 indeed explain the common trend, I expect their coefficients to lose significance once firm fixed effects are included. While the other coefficients remain largely stable, this is exactly what happens with *REL\_MSIZES* and, to a somewhat lesser extent, *SHARE\_INT\_IND*. The coefficient picking up the relative size of the market (in terms of firms listed) is dwarfed in magnitude and loses its significance once fixed effects are included, suggesting that global IPO trends are a prime candidate for explaining the time trend. The coefficient for the market share of intangible intensive industries is reduced to about half its size and its significance now varies across specifications. Taken together, that makes both variables strong candidates for explaining the observed global time trend in the cash flow accrual association. The other variables largely remain unaffected from including yearly fixed effects. My interpretation is that, while the cash flow distribution is by nature a key determinant of the cash flow accrual association, it varies across countries and thus might be explaining local trends but not the global trend. A similar statement applies to the determinants suggested by BLZ with the caveat that they might be picking up mechanical relations.

## 6 Conclusion

This project documents that reproduction and replication can generate new insights. While the downwards trend in the cash flow accrual association documented by Bushman, Lerman, and Zhang (2016) can be reproduced, their conclusions about its determinants are subject to careful interpretation. My reproduced sample is showing a level of multicollinearity that makes it essentially impossible to identify individual determinants. Thus, the results of BLZ about the potential reasons for the diminishing trend of the cash flow accrual association are open for debate. Building on an exploratory analysis of the U.S. data and an associative test of my findings on international data, I come to the following conclusions.

- (1) The drastic changes in cash flow distribution over time in the U.S. but also internationally are a key associative determinant of the changes in the cash flow accrual association. The functional form of this association varies across the cash flow distribution with negative cash flow firms having no robust link of cash flows with accruals and also generally more dispersed accruals.

- (2) The diminishing cash flow accrual association and also the changes in cash flow distribution seem to be linked to new firms entering the market and to a general trend towards a more intangible intensive business model. This conjecture is supported by exploratory U.S. evidence showing that new firms from intangible intensive industries, in particular from the health sector, appear to have extremely left skewed cash flows and low cash flow accrual associations. Incumbent firms show weaker patterns but also experience a decline in the cash flow accrual association. Overall, the decline is confined to the eighties and nineties in the U.S.
- (3) Based on international data, I document a global time trend in the cash flow accrual association that seems to be best explained by global IPO waves and, to a somewhat lesser extent, by changes in the global industry composition. The distributional properties of cash flows are also very influential internationally, but do not *per se* seem to drive the common global trend.

Taken together, these three observations are consistent with Srivastava (2014) and inconsistent with accounting rules being the key driver behind the diminishing cash flow accrual association documented by BLZ. This does not imply that accounting rules do not matter at all for this phenomenon. I observe large negative accruals for normal cash flow firms becoming more common in later sample years in the U.S. and these accruals might well be the consequence of stricter impairment and/or revaluation rules. However, these rules were largely adopted in the early 2000s and thus cannot drive the fundamental changes in the cash flow accrual association that materialized earlier.

Obviously, my findings are descriptive in nature and remain open for debate and alternative explanations. One potential concern is that my measure for IPO activity, which is based on the number of firms that are included with Compustat in a given country-year, is affected by changes in database coverage bias. Also, the industry level proxy for intangible intensive business models is crude. These short-comings might serve as a motivation for fellow researchers to make use of the open science tools that I introduce into the accounting literature with this project. Everybody is invited to directly build on my work by either re-running and extending this project's code that is available on Github (see reproduction guide in Appendix B) and/or by using the online panel data exploration tool developed as a part of this project. It allows researchers to explore data and findings across different sample specifications and with different dependent or independent variables (click here to ExPanD).

## Appendix A: Variable Definitions

**Table A1: Firm-Year Level Variable Definitions**

| Variable         | Compustat Definition   | Description   |
|------------------|--|---|
| $AVG\_AT_{i,t}$  | $0.5AT_{t-1} + 0.5AT_t$  | Average total assets in M US-\$   |
| $E_{i,t}$        | $IB_t / AVG\_AT_t$   | Earnings deflated by average total assets   |
| $TACC_{i,t}$     | If ( $year > 1987$ )<br>$(IBC_t - OANCF_t) / AVG\_AT_t$ else<br>$(ACT_t - ACT_{t-1} - (CHE_t - CHE_{t-1}) - ((LCT_t - LAG_{t-1}) - (DLC_t - DLC_{t-1})) - DP_t) / AVG\_AT_t$ | Total accruals, derived from earnings and cash flow from operations as reported in the cash flow statement after 1987 and from the balance sheet before, deflated by average total assets |
| $CFO_{i,t}$      | If ( $year > 1987$ )<br>$OANCF_t / AVG\_AT_t$ else<br>$E_t - TACC_t$   | Cash flow from operations, as reported in the cash flow statement after 1987 and derived from the balance sheet before, deflated by average total assets                                  |
| $MV_{i,t}$       | $PRCCF_t * CSHO_t$   | Market value of common stock in M US-\$   |
| $BM_{i,t}$       | $CEQ_t / MV_t$   | Book-to-market ratio based on common stock  |
| $SGAINT_{i,t}$   | $XSGA_t / (SALE_t - IB_t)$   | SG&A Intensity: Share of total cost that can be attributed to selling, general and administrative expenses  |
| $EXPENSES_{i,t}$ | $SALE_t / AVG\_AT_t$   | Net sales, deflated by average total assets   |
| $SALES_{i,t}$    | $(SALE_t - IB_t) / AVG\_AT_t$  | Expenses, deflates by average total assets  |

Note: The variable definitions follow BLZ, p. 75 f. All data and data mnemonics are from Compustat. Winsorization is done by year at the five percent level when the respective country year has less than 100 observations and at the one percent level otherwise. For non-U.S. data,  $CFO$  and  $TACC$  are calculated following the balance sheet approach whenever cash flow statement data is not available.

**Table A2: Country-Year Level Variable Definitions**

| Variable                 | Description   |
|--------------------------|---|
| <i>CFO_MEAN</i>          | Mean of <i>CFO</i> for the respective country year  |
| <i>CFO_SD</i>            | Standard deviation of <i>CFO</i> for the respective country year  |
| <i>CFO_SKEW</i>          | Skewness of <i>CFO</i> for the respective country year  |
| <i>DCFO_ACORR</i>        | One year auto-correlation of change in <i>CFO</i> (values winsorized by year) for the respective country year   |
| <i>SD_OI_PTI</i>         | Standard deviation of the difference between operating income after depreciation (#oiadp) and pretax income (#pi), both terms deflated by <i>AVG_AT</i> and winsorized by year, for the respective country year |
| <i>PCTLOSS</i>           | Share of country year observations with negative earnings before extraordinary items  |
| <i>DT_ADJR2</i>          | Adjusted $R^2$ for country year Dechow and Tang regressions, regressing <i>SALES</i> on current, lagged, and lead <i>EXPENSES</i>   |
| <i>SGAINT_MEAN</i>       | Country year mean <i>SGAINT</i>   |
| <i>REL_MSIZE</i>         | Country year number of firms included in sample relative to maximum of country firms included in any year over the sample period  |
| <i>SHARE_INT_IND</i>     | Share of firms that are in the Health, Business Services, Chemical Products or Telecommunications Sector (as defined by Fama and French 12 industry groups) for each country year                               |
| <i>Adj.R<sup>2</sup></i> | Adjusted $R^2$ of an Dechow/Dichev model, estimated for the respective country year, with all input variables winsorized by year  |
| <i>TIME</i>              | Time trend indicator, defined as $year - 1989$ for the international panel and as $year - 1964$ for the U.S. time-series sample   |

Note: The variable definitions follow BLZ, p. 75 f. All data and data mnemonics are from Compustat. Winsorization of input variables is done by year at the five percent level when the respective country year has less than 100 observations and at the one percent level otherwise.

## Appendix B: Reproduction Guide

To reproduce this paper and its findings, you need admin access to computational resources (a PC running Windows or Linux as well as a Mac will do) and Compustat fundamental annual data, both national and international. For the following steps I assume that you have an active subscription to WRDS.

1. If you have access to a Docker installation (<https://www.docker.com>), continue with step 10.
2. Download the installer of the statistical package R for your system: <https://cran.r-project.org/>
3. Install R to your system
4. Download RStudio: <https://www.rstudio.com/products/rstudio/download/>
5. After R has finished installing, install RStudio
6. If you have a Github account, you can link RStudio to it (follow the instructions under <http://happygitwithr.com/>) and then fork the Github repository <https://github.com/joachim-gassen/ExpAcc> as a new project. Continue with step 14.
7. If you do not have a Github account, download the ZIP-File containing the repository for this paper from <https://github.com/joachim-gassen/ExpAcc> and extract it on your computer (click the top right green button named “clone or download”).
8. Start RStudio. Use the file window (normally placed as a tab window in the lower right part of the screen) to navigate where you have stored the repository. Set your working directory to that directory by using “More/Set As Working Directory”. You can also change the working directory in RStudio by clicking on “session” and then “Set working directory”.
9. Continue with step 14.
10. Check the Docker settings to make sure that you have at least 2 CPUs, 8 GB of memory and 2 GB of swap space available for Docker
11. Issue the following command to pull the container from Docker Hub and run it in your environment:  

```
docker run -d -p 8787:8787 -e PASSWORD=yourpass --name expacc joegassen/expacc
```
12. Open your browser and point it to: <http://localhost:8787>
13. Logon with username “rstudio” and the password you selected in step 11.
14. Open the file “produce\_findings.R” in the main directory. Run all (Ctrl + A, Ctrl + Enter).

This will install necessary R packages, including the ExPanDaR package, download necessary data from WRDS (takes a while), and generate all samples, tables and figures that are presented in the paper. As a last step, it will open a local version of the ExPanD app that is also available online (<https://jgassen.shinyapps.io/ExpAcc/>). To download the data from WRDS, it will prompt you for your WRDS username and your password. The password will not be stored.

15. After you downloaded the necessary data, you might want to set the variables “refresh” and “pull\_wrds\_data” in “produce\_findings.R” to “FALSE” to avoid re-pulling the data. If you re-pull WRDS data, the old data files will be renamed.
16. If you want to reproduce the paper itself (requires an installed Latex distribution), open the file “ExpAcc\_paper.Rmd” in the paper directory and knit it (Ctrl + Shift + K, Icon in the tray above the text editor). Please note: The code in “ExpAcc\_paper.Rmd” is not downloading data from WRDS and other sources.
17. Where do you go from here? A good first step would be to explore and audit the code of the repository. It is contained in the code directory. The naming of the files should be somewhat self-explanatory. As an example for what to do: Experiment with the sample screens and the outlier treatment of the respective samples.
18. If you stumble across something interesting, please get in touch. The most efficient way to do that would be to open an issue on Github or to issue a pull request with your suggested changes. But I also cater to traditional email: [gassen@wiwi.hu-berlin.de](mailto:gassen@wiwi.hu-berlin.de)

## References

- Ball, Ray, and Lakshmanan Shivakumar. 2006. “The Role of Accruals in Asymmetrically Timely Gain and Loss Recognition.” *Journal of Accounting Research* 44 (2): 207–42. doi:10.1111/j.1475-679X.2006.00198.x.
- Bushman, Robert M., Alina Lerman, and Frank X. Zhang. 2016. “The Changing Landscape of Accrual Accounting.” *Journal of Accounting Research* 54 (1): 41–78. doi:10.1111/1475-679X.12100.
- Dechow, Patricia M. 1994. “Accounting Earnings and Cash Flows as Measures of Firm Performance: The Role of Accounting Accruals.” *Journal of Accounting and Economics* 18 (1): 3–42. doi:10.1016/0165-4101(94)90016-7.
- Dechow, Patricia M., and Ilia D. Dichev. 2002. “The Quality of Accruals and Earnings: The Role of Accrual Estimation Errors.” *The Accounting Review* 77 (s-1): 35–59. doi:10.2308/accr.2002.77.s-1.35.
- Dechow, Patricia M., Weili Ge, and Catherine Schrand. 2010. “Understanding Earnings Quality: A Review of the Proxies Their Determinants and Their Consequences.” *Journal of Accounting and Economics* 50 (2-3): 344–401. doi:10.1016/j.jacceco.2010.09.001.
- Dichev, Ilia D., and Vicki W. Tang. 2008. “Matching and the Changing Properties of Accounting Earnings over the Last 40 Years.” *The Accounting Review* 83 (6): 1425–60. doi:10.2308/accr.2008.83.6.1425.
- Duvendack, Maren, Richard Palmer-Jones, and W. Robert Reed. 2017. “What Is Meant by Replication and Why Does It Encounter Resistance in Economics?” *American Economic Review* 107 (5): 46–51. doi:10.1257/aer.p20171031.
- Fama, Eugene F., and Kenneth R. French. 2004. “New Lists: Fundamentals and Survival Rates.” *Journal of Financial Economics* 73 (2): 229–69. doi:10.1016/j.jfineco.2003.04.001.
- Gertler, Paul, Sebastian Galiani, and Mauricio Romero. 2018. “How to Make Replication the Norm.”

*Nature* 554 (February): 417–19. doi:10.1038/d41586-018-02108-9.

Jones, Jennifer J. 1991. “Earnings Management During Import Relief Investigations.” *Journal of Accounting Research* 29 (2): 193–228. doi:10.2307/2491047.

Nallareddy, Suresh, Mani Sethuraman, and Mohan Venkatachalam. 2018. “Earnings or Cash Flows: Which Is a Better Predictor of Future Cash Flows?” Duke University. doi:10.2139/ssrn.3054644.

Srivastava, Anup. 2014. “Why Have Measures of Earnings Quality Changed over Time.” *Journal of Accounting and Economics* 57 (2-3): 196–217. doi:10.1016/j.jacceco.2014.04.001.

**TABLE 1**  
*Descriptive Statistics*

| <b>Panel A: Descriptive statistics</b> |          |        |           |         |        |        |        |           |
|--|----------|--------|-----------|---------|--------|--------|--------|-----------|
| Variable                               | <i>N</i> | Mean   | Std. Dev. | Min.    | Q1     | Median | Q3     | Max.      |
| $E_{i,t}$                              | 217,164  | -0.004 | 0.186     | -1.425  | -0.013 | 0.039  | 0.077  | 0.392     |
| $TACC_{i,t}$                           | 217,164  | -0.054 | 0.129     | -1.717  | -0.093 | -0.045 | -0.001 | 1.249     |
| $CFO_{i,t-1}$                          | 186,109  | 0.055  | 0.145     | -0.904  | 0.013  | 0.074  | 0.127  | 0.420     |
| $CFO_{i,t}$                            | 217,164  | 0.050  | 0.153     | -0.904  | 0.008  | 0.072  | 0.127  | 0.420     |
| $CFO_{i,t+1}$                          | 184,881  | 0.056  | 0.144     | -0.904  | 0.015  | 0.075  | 0.128  | 0.420     |
| $MV_{i,t}$                             | 189,864  | 2,139  | 12,763    | 0       | 28     | 119    | 619    | 1,819,782 |
| $BM_{i,t}$                             | 188,999  | 0.724  | 1.549     | -47.674 | 0.321  | 0.597  | 1.032  | 10.825    |
| $SG\&A\ Intensity$                     | 180,566  | 0.268  | 0.216     | 0.001   | 0.119  | 0.210  | 0.346  | 1.905     |

Figure 1: Table 1, Panel A of BLZ

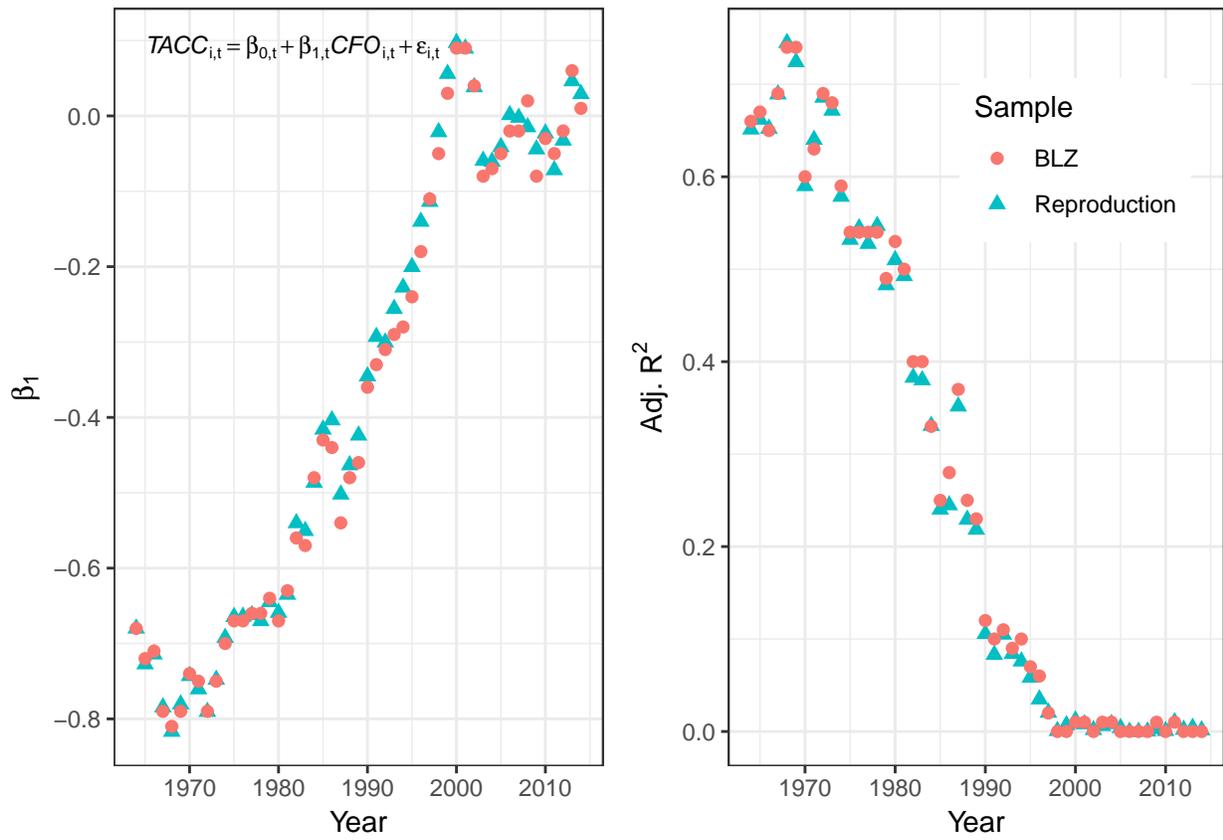


Figure 2: Regression results for yearly estimates of model (1) compared to results of BLZ, Table 2

**TABLE 4**  
*The Impact of a Temporal Change in Economic-Based Cash Flow Shocks  
and Non-Timing-Related Accruals*

| Regression                        | 1                  | 2                 | 3                 | 4                 | 5                 | 6                 |
|-----------------------------------|--------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| Intercept                         | 0.763<br>(36.17)   | 0.554<br>(3.63)   | 0.777<br>(40.57)  | 0.993<br>(1.01)   | 0.932<br>(7.22)   | 5.322<br>(4.41)   |
| Time                              | -0.016<br>(-22.16) | -0.014<br>(-7.78) | -0.006<br>(-3.54) | -0.017<br>(-9.39) | -0.012<br>(-3.54) | -0.006<br>(-2.02) |
| <i>Std(CFO)</i>                   |                    | -0.080<br>(-0.12) |                   |                   |                   | -0.227<br>(-0.27) |
| <i>Auto_ΔCFO</i>                  |                    | -0.412<br>(-1.68) |                   |                   |                   | -0.278<br>(-1.54) |
| <i>Std(OI-PTI)</i>                |                    |                   | -0.011<br>(-0.01) |                   |                   | -2.264<br>(-2.05) |
| <i>PctLoss</i>                    |                    |                   | -1.098<br>(-3.77) |                   |                   | -0.657<br>(-2.18) |
| Adj. $R^2$ ( <i>Dichev-Tang</i> ) |                    |                   |                   | -0.226<br>(-0.23) |                   | -4.335<br>(-3.85) |
| <i>SG&amp;A Intensity</i>         |                    |                   |                   |                   | -1.109<br>(-1.32) | -1.655<br>(-1.48) |
| Adj. $R^2$                        | 0.909              | 0.912             | 0.947             | 0.907             | 0.910             | 0.958             |

Figure 3: Table 4 of BLZ

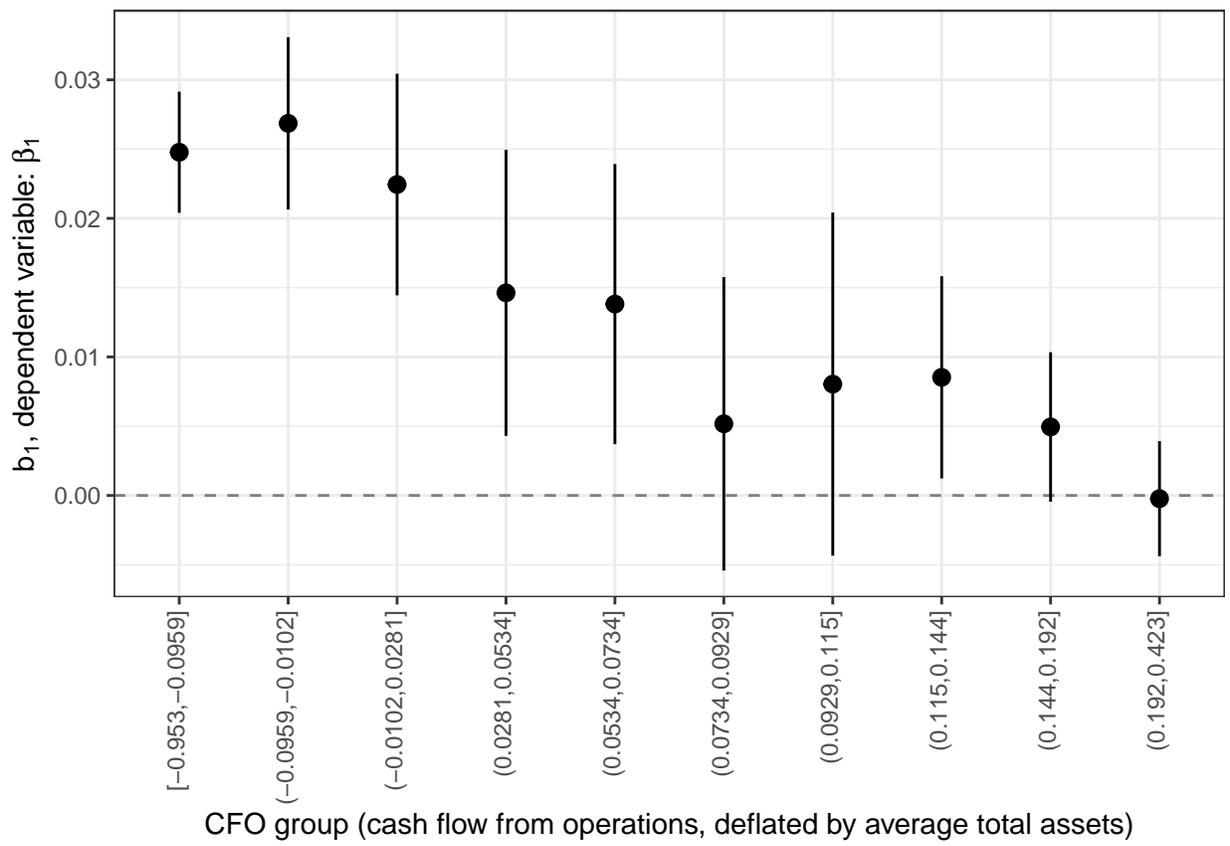


Figure 4: *TIME* coefficient of model (4) by *CFO* deciles

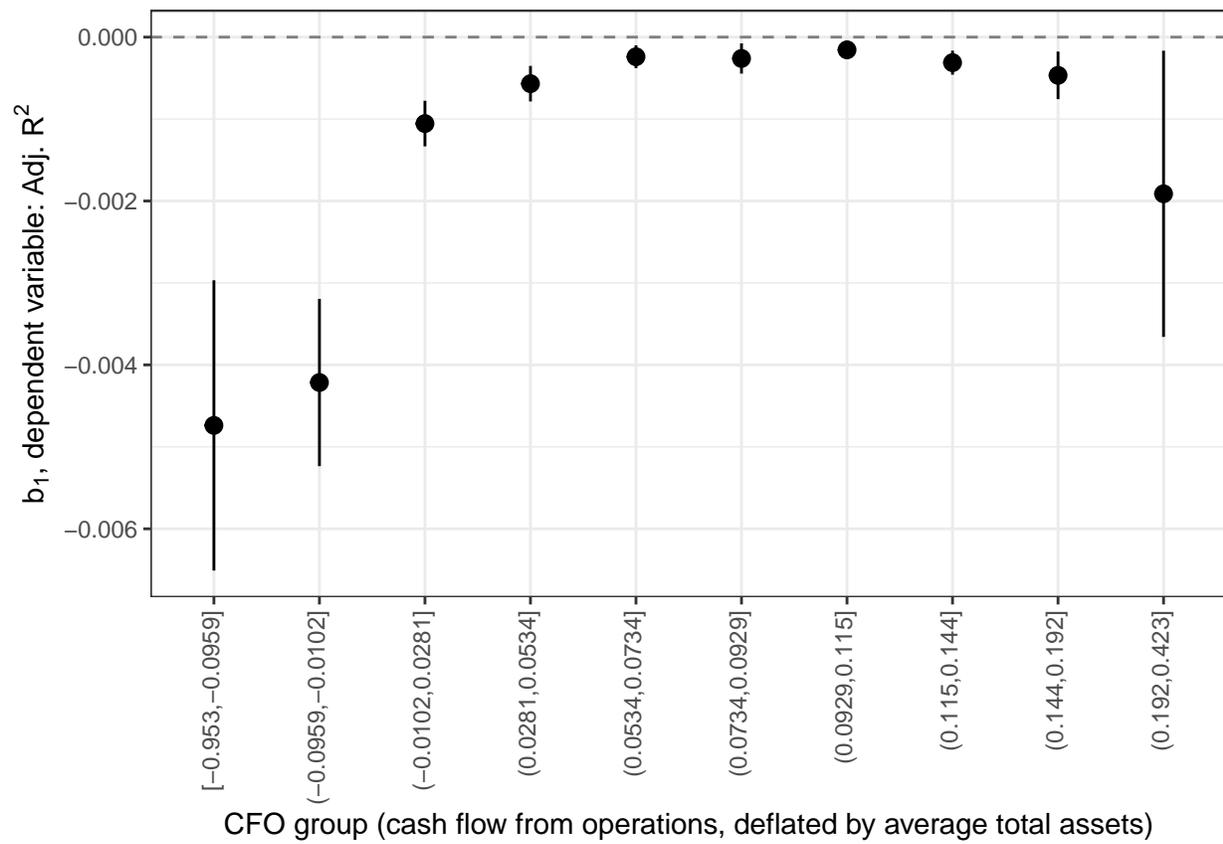


Figure 5: *TIME* coefficient of model (4) with *Adj. R<sup>2</sup>* as dependent variable by cash flow deciles

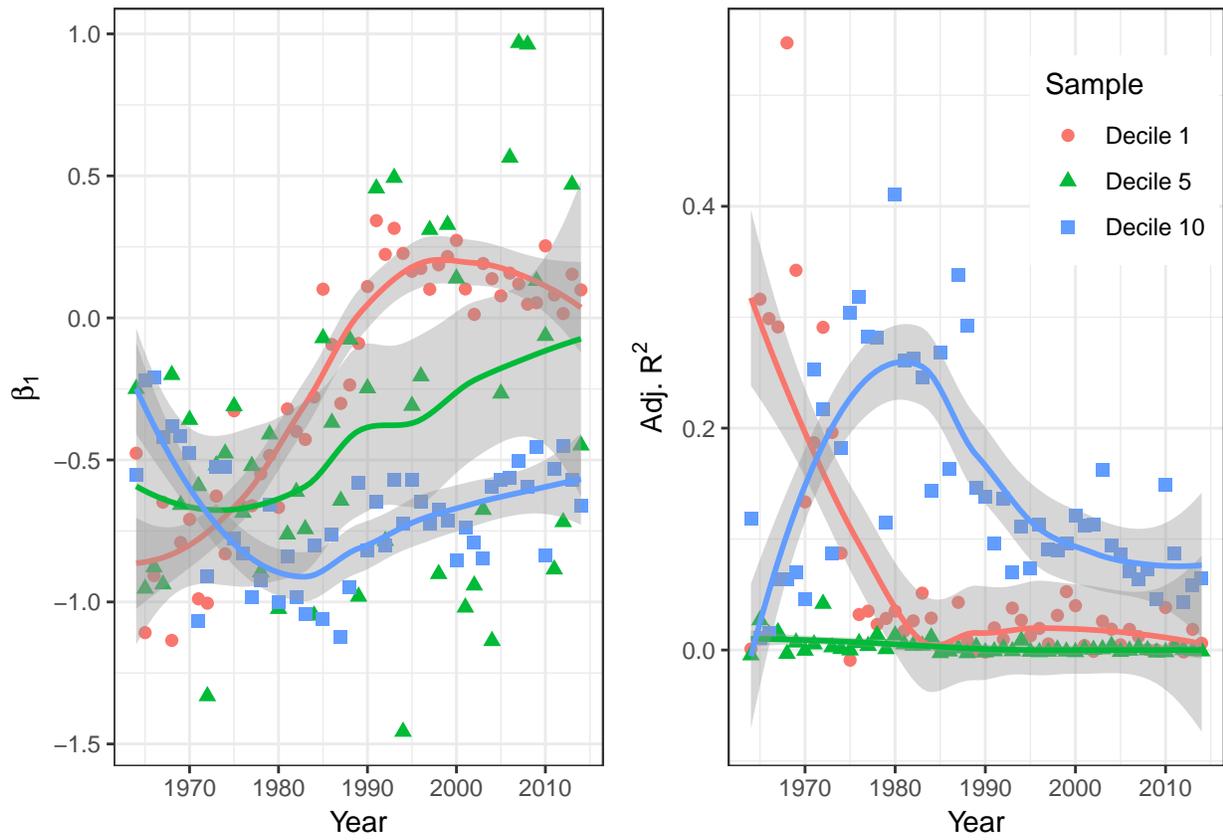


Figure 6: *CFO* coefficients and *Adj. R<sup>2</sup>* of model (1) by cash flow deciles

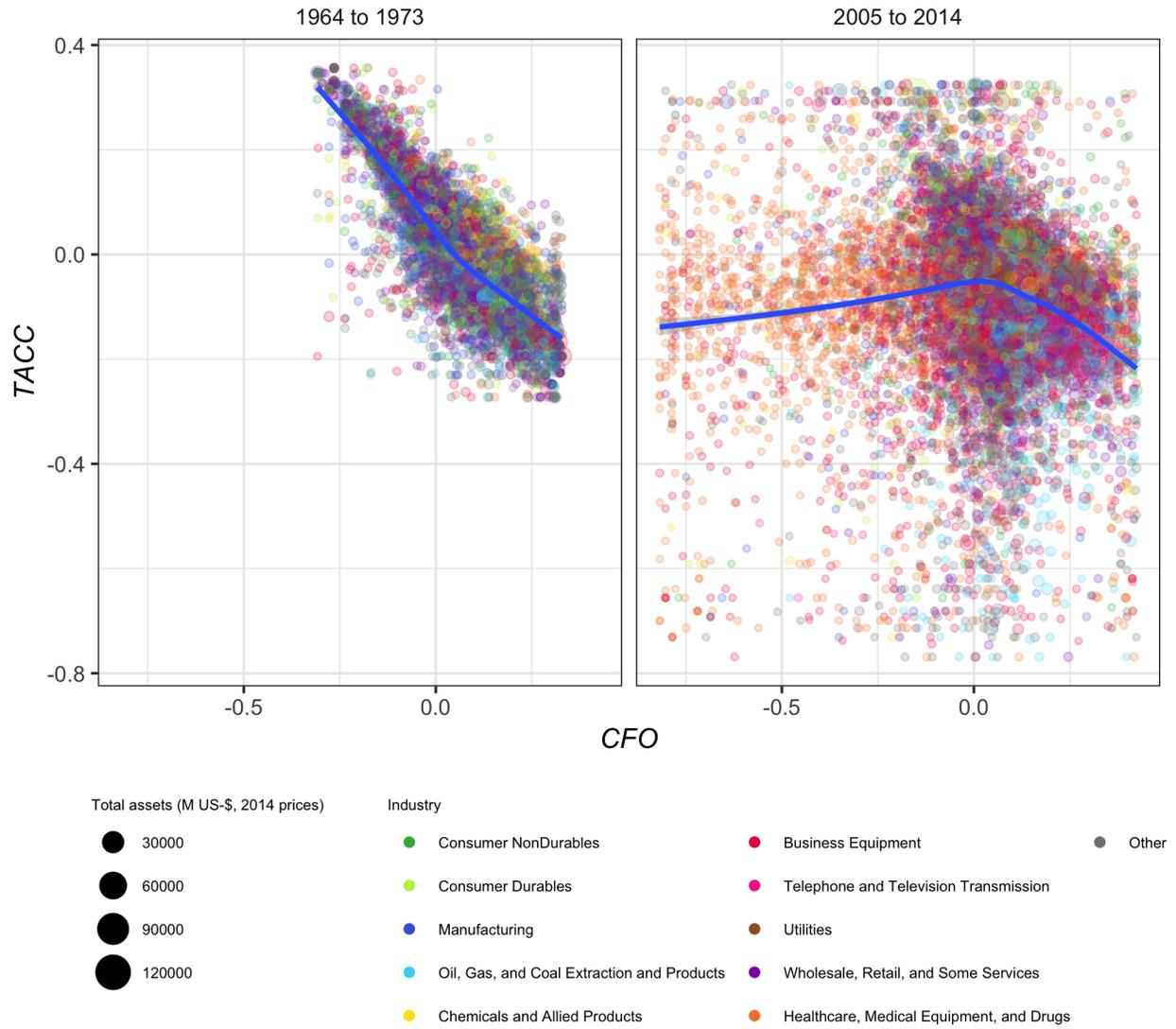


Figure 7: *CFO* and *TACC* scatter plots (sample of 20,000 observations for each period)

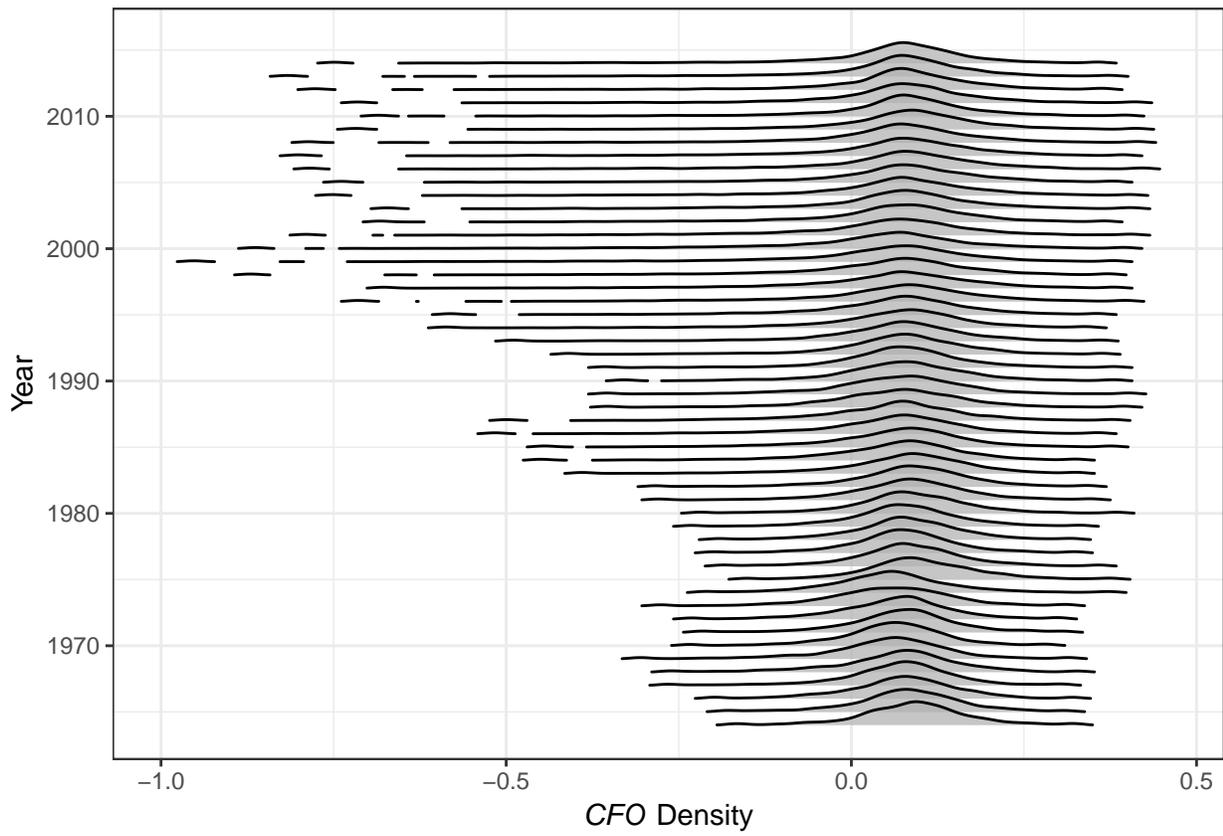


Figure 8: Density of *CFO* over time

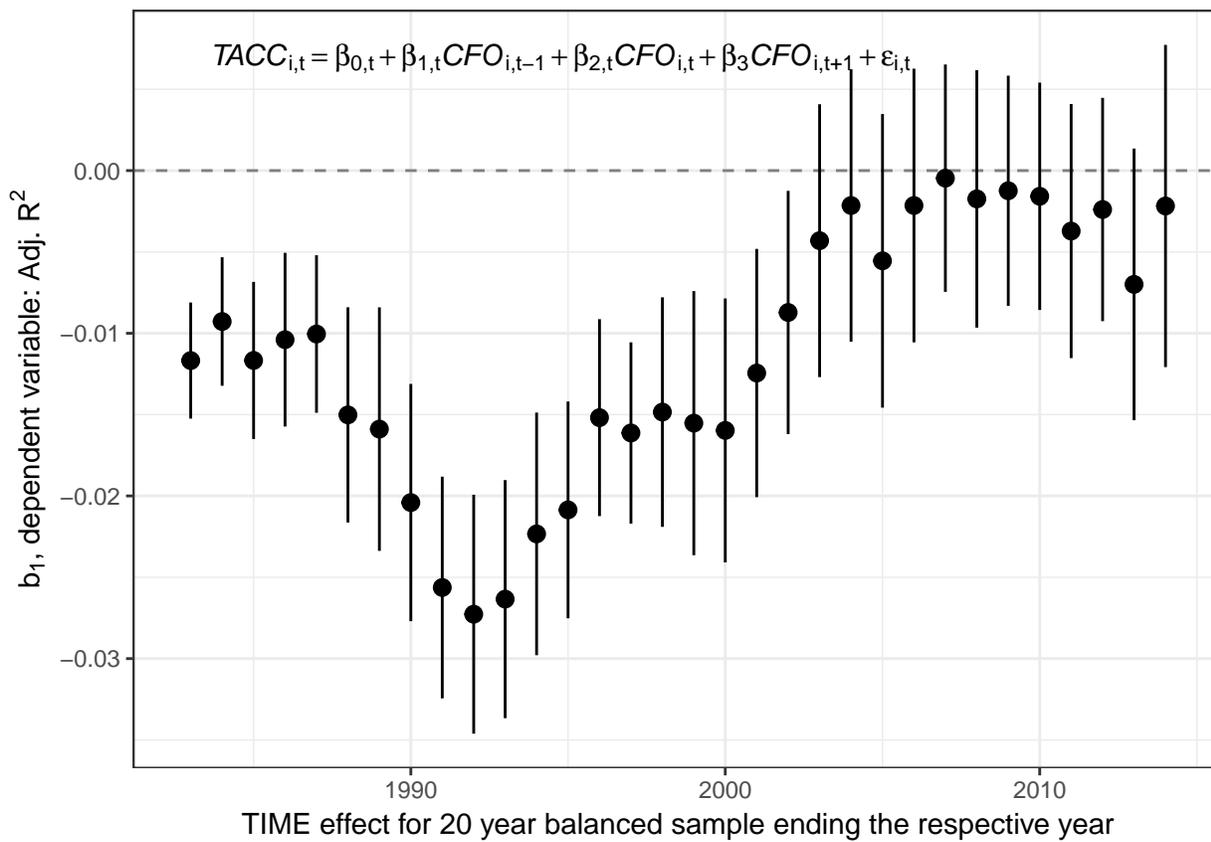


Figure 9: Development of adjusted  $R^2$  over time for balanced samples

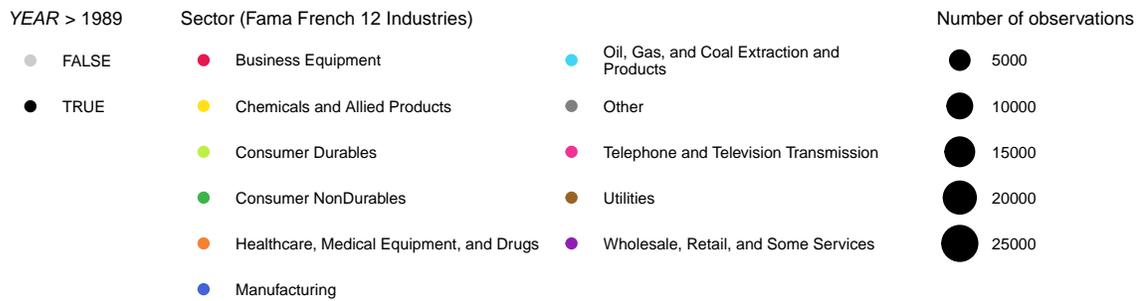
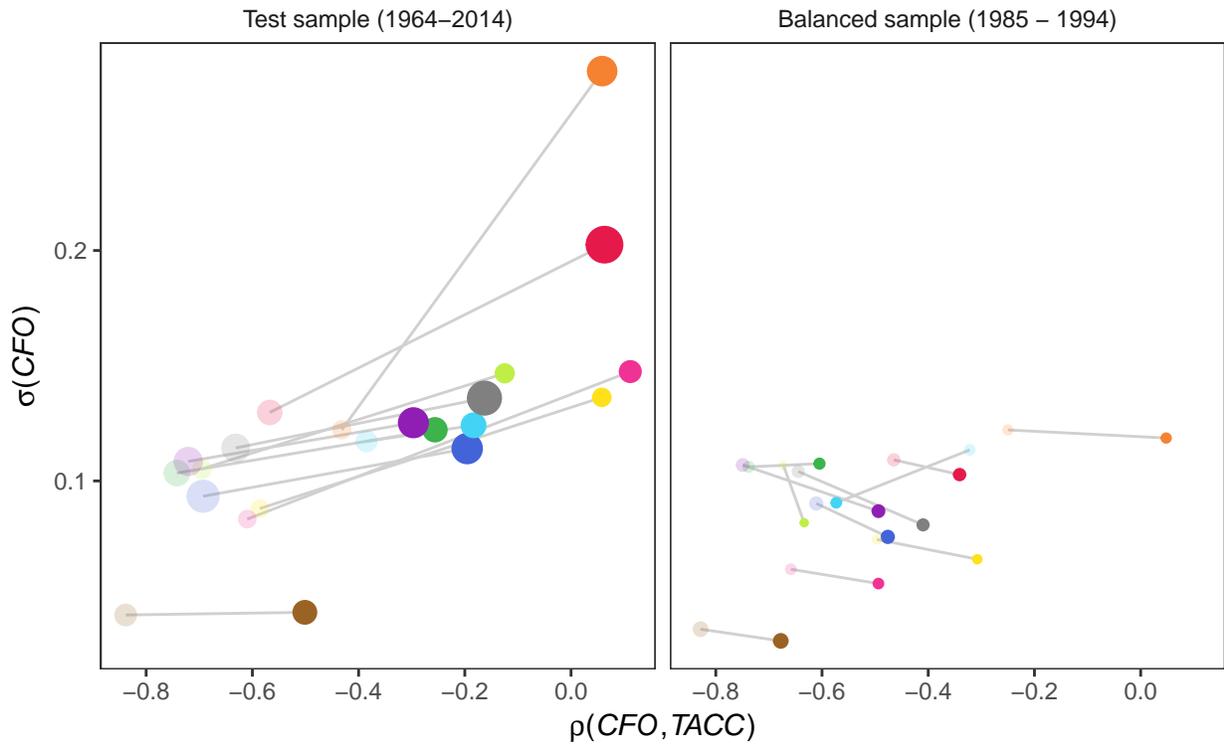


Figure 10: Industry and sample composition effects combined

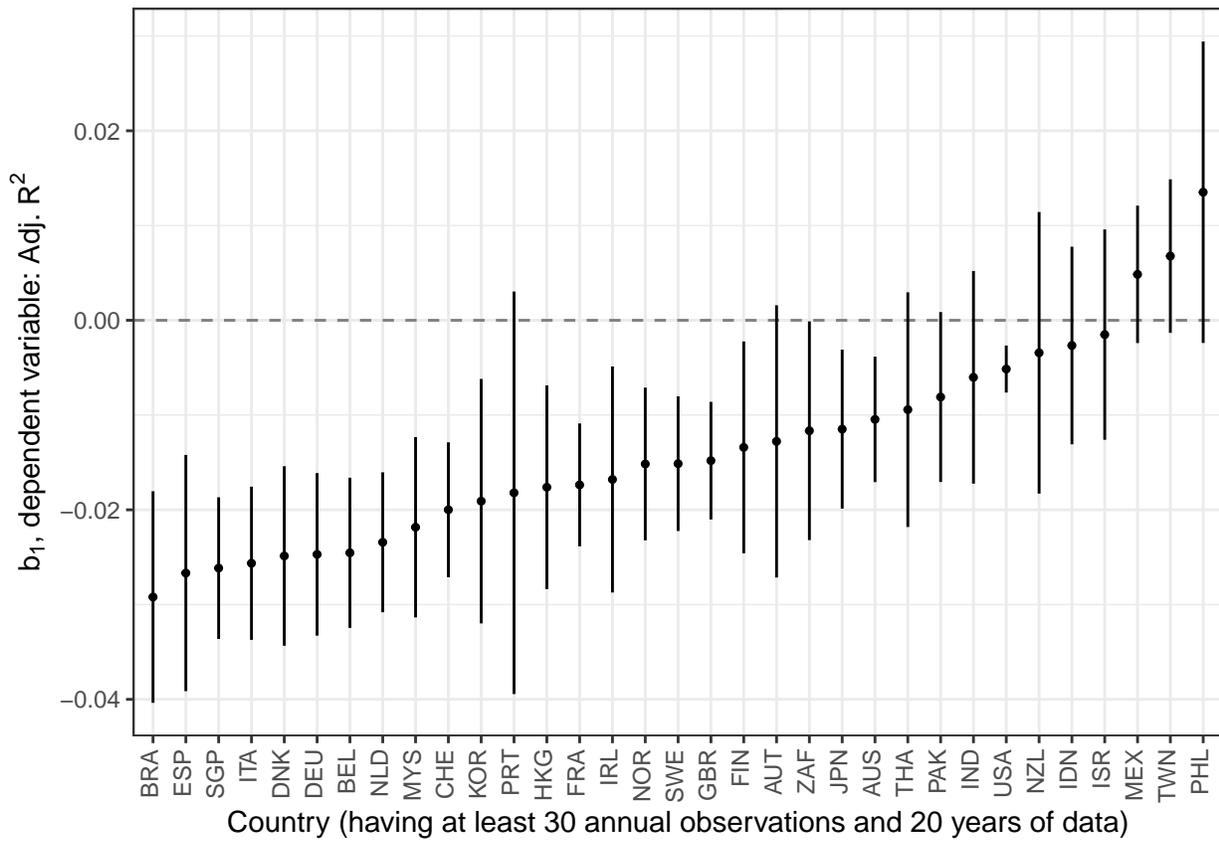


Figure 11: *TIME* coefficient by country for the adjusted  $R^2$ s of Dechow and Dichev regressions

**Table 1: Descriptive Statistics Reproduced Sample**

|                | N       | Mean      | Std. dev.  | Min.    | 25 %   | Median | 75 %    | Max.          |
|----------------|---------|-----------|------------|---------|--------|--------|---------|---------------|
| $E_{i,t}$      | 261,154 | -0.076    | 0.521      | -9.120  | -0.037 | 0.035  | 0.077   | 0.483         |
| $TACC_{i,t}$   | 261,154 | -0.083    | 0.304      | -6.330  | -0.103 | -0.046 | 0.003   | 0.528         |
| $CFO_{i,t-1}$  | 225,241 | 0.019     | 0.261      | -3.403  | -0.005 | 0.068  | 0.125   | 0.473         |
| $CFO_{i,t}$    | 261,154 | 0.009     | 0.285      | -3.403  | -0.012 | 0.066  | 0.125   | 0.473         |
| $CFO_{i,t+1}$  | 228,087 | 0.019     | 0.267      | -3.403  | -0.004 | 0.069  | 0.126   | 0.473         |
| $MV_{i,t}$     | 219,979 | 1,844.522 | 11,875.633 | 0.000   | 14.362 | 70.545 | 446.054 | 1,819,781.926 |
| $BM_{i,t}$     | 218,992 | 0.619     | 1.789      | -47.853 | 0.272  | 0.556  | 0.997   | 9.982         |
| $SGAINT_{i,t}$ | 219,609 | 0.279     | 0.204      | 0.002   | 0.130  | 0.226  | 0.373   | 1.134         |

Note: This reproduced sample contains annual Compustat North America data for the period 1964 to 2014. The sample selection process follows the procedure described in BLZ, p.49f: I delete observations from the financial services industry, observations from years with significant mergers/acquisitions activity and observations with missing cash flow or accrual data. Also in line with BLZ, I by-year winsorize all data besides  $MV_{i,t}$  to the top and bottom percentile. Variables are as defined in Appendix A.

**Table 2: Descriptive Statistics Test Sample**

|                | N       | Mean      | Std. dev.  | Min.    | 25 %   | Median  | 75 %    | Max.          |
|----------------|---------|-----------|------------|---------|--------|---------|---------|---------------|
| $E_{i,t}$      | 220,104 | -0.001    | 0.184      | -1.487  | -0.009 | 0.040   | 0.078   | 0.406         |
| $TACC_{i,t}$   | 220,104 | -0.054    | 0.125      | -1.032  | -0.093 | -0.044  | 0.000   | 0.371         |
| $CFO_{i,t-1}$  | 188,888 | 0.058     | 0.143      | -0.953  | 0.016  | 0.075   | 0.129   | 0.423         |
| $CFO_{i,t}$    | 220,104 | 0.052     | 0.151      | -0.953  | 0.011  | 0.073   | 0.128   | 0.423         |
| $CFO_{i,t+1}$  | 191,570 | 0.058     | 0.142      | -0.953  | 0.017  | 0.076   | 0.129   | 0.423         |
| $MV_{i,t}$     | 191,747 | 2,114.360 | 12,697.553 | 0.000   | 24.488 | 107.468 | 593.720 | 1,819,781.926 |
| $BM_{i,t}$     | 190,831 | 0.712     | 1.540      | -47.070 | 0.317  | 0.594   | 1.028   | 10.019        |
| $SGAINT_{i,t}$ | 184,068 | 0.255     | 0.186      | 0.006   | 0.119  | 0.209   | 0.339   | 1.025         |

Note: The test sample is based on the reproduced sample reported in Table 1. In addition to the sampling process described there, observations with  $AVG\_AT$  below M US-\$ 7.5 are excluded. Variables are as defined in Appendix A.

**Table 3: Replication of Table 4 of BLZ**

|                               | <i>Dependent variable:</i> |                                |                                 |                                |                                |                                 |
|-------------------------------|----------------------------|--------------------------------|---------------------------------|--------------------------------|--------------------------------|---------------------------------|
|                               | <i>Adj.R<sup>2</sup></i>   | <i>Adj.R<sup>2</sup></i>       | <i>Adj.R<sup>2</sup></i>        | <i>Adj.R<sup>2</sup></i>       | <i>Adj.R<sup>2</sup></i>       | <i>Adj.R<sup>2</sup></i>        |
|                               | (1)                        | (2)                            | (3)                             | (4)                            | (5)                            | (6)                             |
| <i>TIME</i>                   | -0.016***<br>(0.001)       | -0.012***<br>(0.002)<br>[5.09] | -0.007***<br>(0.001)<br>[6.73]  | -0.016***<br>(0.002)<br>[5.86] | -0.005***<br>(0.002)<br>[9.57] | -0.005***<br>(0.002)<br>[14.46] |
| <i>CFO_SD</i>                 |                            | -0.927<br>(0.650)<br>[6.26]    |                                 |                                |                                | 1.919**<br>(0.755)<br>[23.73]   |
| <i>DCFO_ACORR</i>             |                            | -0.414*<br>(0.243)<br>[3.11]   |                                 |                                |                                | -0.147<br>(0.154)<br>[3.54]     |
| <i>SD_OI_PTI</i>              |                            |                                | -0.123<br>(0.960)<br>[16.38]    |                                |                                | -2.140**<br>(0.981)<br>[26.08]  |
| <i>PCTLOSS</i>                |                            |                                | -1.103***<br>(0.261)<br>[19.84] |                                |                                | -0.043<br>(0.291)<br>[37.61]    |
| <i>DT_ADJR2</i>               |                            |                                |                                 | 0.274<br>(0.865)<br>[5.86]     |                                | -1.161<br>(0.742)<br>[14.09]    |
| <i>SGAINT_MEAN</i>            |                            |                                |                                 |                                | -3.723***<br>(0.518)<br>[9.57] | -4.293***<br>(0.995)<br>[55.67] |
| Constant                      | 0.757***<br>(0.022)        | 0.611***<br>(0.154)            | 0.791***<br>(0.018)             | 0.478<br>(0.879)               | 1.379***<br>(0.088)            | 2.471***<br>(0.732)             |
| Estimator                     | ols                        | ols                            | ols                             | ols                            | ols                            | ols                             |
| Fixed effects                 | None                       | None                           | None                            | None                           | None                           | None                            |
| Std. errors clustered         | No                         | No                             | No                              | No                             | No                             | No                              |
| Observations                  | 51                         | 51                             | 51                              | 51                             | 51                             | 51                              |
| <i>R<sup>2</sup></i>          | 0.902                      | 0.918                          | 0.955                           | 0.902                          | 0.953                          | 0.973                           |
| Adjusted <i>R<sup>2</sup></i> | 0.900                      | 0.912                          | 0.952                           | 0.898                          | 0.951                          | 0.969                           |

Note: This table reproduces Table 4 from BLZ. The dependent variable is the adjusted  $R^2$  from annual Dechow and Dichev regressions (Model 2). All independent variables are as defined in Appendix A. Variance inflation factors are reported in square brackets below the standard errors, which are reported in parentheses. \*/\*\*/\*\* indicate two-sided significance levels of 10/5/1 %, respectively.

**Table 4: Univariate Associations of BLZ Variables**

|                                | <i>Dependent variable:</i> |                           |                           |                           |                           |                           |                           |                                 |
|--------------------------------|----------------------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------------|
|                                | <i>Adj.R</i> <sup>2</sup>  | <i>Adj.R</i> <sup>2</sup> | <i>Adj.R</i> <sup>2</sup> | <i>Adj.R</i> <sup>2</sup> | <i>Adj.R</i> <sup>2</sup> | <i>Adj.R</i> <sup>2</sup> | <i>Adj.R</i> <sup>2</sup> | <i>Adj.R</i> <sup>2</sup>       |
|                                | (1)                        | (2)                       | (3)                       | (4)                       | (5)                       | (6)                       | (7)                       | (8)                             |
| <i>TIME</i>                    | -0.016***<br>(0.001)       |                           |                           |                           |                           |                           |                           | -0.004***<br>(0.002)<br>[9.79]  |
| <i>CFO_SD</i>                  |                            | -5.553***<br>(0.395)      |                           |                           |                           |                           |                           |                                 |
| <i>DCFO_ACORR</i>              |                            |                           | -2.639***<br>(0.280)      |                           |                           |                           |                           |                                 |
| <i>SD_OI_PTI</i>               |                            |                           |                           | -7.224***<br>(0.374)      |                           |                           |                           | 0.033<br>(0.886)<br>[16.43]     |
| <i>PCTLOSS</i>                 |                            |                           |                           |                           | -1.833***<br>(0.071)      |                           |                           | -0.706**<br>(0.273)<br>[25.59]  |
| <i>DT_ADJR2</i>                |                            |                           |                           |                           |                           | 6.884***<br>(0.555)       |                           |                                 |
| <i>SGAINT_MEAN</i>             |                            |                           |                           |                           |                           |                           | -5.182***<br>(0.180)      | -2.049***<br>(0.668)<br>[19.42] |
| Constant                       | 0.757***<br>(0.022)        | 1.099***<br>(0.055)       | -0.746***<br>(0.118)      | 0.821***<br>(0.027)       | 0.785***<br>(0.019)       | -6.306***<br>(0.537)      | 1.610***<br>(0.045)       | 1.120***<br>(0.108)             |
| Estimator                      | ols                        | ols                       | ols                       | ols                       | ols                       | ols                       | ols                       | ols                             |
| Fixed effects                  | None                       | None                      | None                      | None                      | None                      | None                      | None                      | None                            |
| Std. errors clustered          | No                         | No                        | No                        | No                        | No                        | No                        | No                        | No                              |
| Observations                   | 51                         | 51                        | 51                        | 51                        | 51                        | 51                        | 51                        | 51                              |
| <i>R</i> <sup>2</sup>          | 0.902                      | 0.801                     | 0.645                     | 0.884                     | 0.931                     | 0.758                     | 0.944                     | 0.963                           |
| Adjusted <i>R</i> <sup>2</sup> | 0.900                      | 0.797                     | 0.638                     | 0.881                     | 0.930                     | 0.753                     | 0.943                     | 0.960                           |

Note: This table reports univariate associations for the variables used in Table 4 of BLZ. The dependent variable is the adjusted  $R^2$  from annual Dechow and Dichev regressions (Model 2). All independent variables are as defined in Appendix A. Variance inflation factors are reported in square brackets below the standard errors, which are reported in parentheses. \*/\*\*/\*\* indicate two-sided significance levels of 10/5/1 %, respectively.

**Table 5: Correlations for U.S. Time-Series Sample**

|                             | A            | B            | C            | D            | E            | F            | G            | H            |
|-----------------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| A: <i>Adj.R<sup>2</sup></i> |              | <b>-0.95</b> | <b>-0.90</b> | <b>-0.80</b> | <b>-0.94</b> | <b>-0.96</b> | <b>0.87</b>  | <b>-0.97</b> |
| B: <i>TIME</i>              | <b>-0.94</b> |              | <b>0.89</b>  | <b>0.77</b>  | <b>0.90</b>  | <b>0.92</b>  | <b>-0.91</b> | <b>0.95</b>  |
| C: <i>CFO_SD</i>            | <b>-0.89</b> | <b>0.89</b>  |              | <b>0.82</b>  | <b>0.91</b>  | <b>0.89</b>  | <b>-0.92</b> | <b>0.96</b>  |
| D: <i>DCFO_ACORR</i>        | <b>-0.81</b> | <b>0.79</b>  | <b>0.82</b>  |              | <b>0.81</b>  | <b>0.81</b>  | <b>-0.81</b> | <b>0.82</b>  |
| E: <i>SD_OI_PTI</i>         | <b>-0.96</b> | <b>0.94</b>  | <b>0.93</b>  | <b>0.81</b>  |              | <b>0.97</b>  | <b>-0.92</b> | <b>0.94</b>  |
| F: <i>PCTLOSS</i>           | <b>-0.96</b> | <b>0.92</b>  | <b>0.90</b>  | <b>0.82</b>  | <b>0.97</b>  |              | <b>-0.88</b> | <b>0.96</b>  |
| G: <i>DT_ADJR2</i>          | <b>0.96</b>  | <b>-0.96</b> | <b>-0.91</b> | <b>-0.80</b> | <b>-0.98</b> | <b>-0.94</b> |              | <b>-0.90</b> |
| H: <i>SGAINT_MEAN</i>       | <b>-0.93</b> | <b>0.93</b>  | <b>0.94</b>  | <b>0.81</b>  | <b>0.95</b>  | <b>0.94</b>  | <b>-0.93</b> |              |

This table reports Pearson correlations above and Spearman correlations below the diagonal. Number of observations: 51. Correlations with significance levels below 5% appear in bold print. All variables are as defined in Appendix A.

**Table 6: International Panel Sample**

| ISO code | Country                   | # years | # observations |
|----------|---------------------------|---------|----------------|
| GBR      | United Kingdom            | 28      | 32,938         |
| AUS      | Australia                 | 27      | 26,068         |
| HKG      | Hong Kong                 | 27      | 16,027         |
| MYS      | Malaysia                  | 27      | 15,891         |
| DEU      | Germany                   | 27      | 13,757         |
| FRA      | France                    | 27      | 13,366         |
| SGP      | Singapore                 | 27      | 9,706          |
| SWE      | Sweden                    | 27      | 8,584          |
| CHE      | Switzerland               | 27      | 4,420          |
| NOR      | Norway                    | 27      | 3,789          |
| NLD      | Netherlands               | 27      | 3,340          |
| DNK      | Denmark                   | 27      | 2,832          |
| FIN      | Finland                   | 27      | 2,587          |
| BEL      | Belgium                   | 27      | 2,184          |
| IRL      | Ireland                   | 27      | 1,394          |
| THA      | Thailand                  | 25      | 7,885          |
| BRA      | Brazil                    | 25      | 5,548          |
| ESP      | Spain                     | 25      | 2,775          |
| MEX      | Mexico                    | 25      | 2,056          |
| AUT      | Austria                   | 25      | 1,605          |
| ITA      | Italy                     | 24      | 4,581          |
| IDN      | Indonesia                 | 23      | 5,618          |
| ZAF      | South Africa              | 23      | 4,667          |
| JPN      | Japan                     | 22      | 54,830         |
| KOR      | Korea (Republic of)       | 22      | 12,855         |
| IND      | India                     | 21      | 50,418         |
| ISR      | Israel                    | 21      | 4,550          |
| PAK      | Pakistan                  | 21      | 4,259          |
| TWN      | Taiwan, Province of China | 20      | 21,909         |
| PHL      | Philippines               | 20      | 2,912          |
| NZL      | New Zealand               | 20      | 2,198          |
| PRT      | Portugal                  | 20      | 932            |
| Total    |                           | 788     | 346,481        |

Note: Country year sample for international analyses. All non-U.S. countries with at least 20 years of data are included. A minimum of 30 firm-year observations is required to estimate the sample statistics for each country year to be included in this sample.

**Table 7: Descriptive Statistics for International Panel Sample**

|                          | N   | Mean   | Std. dev. | Min.   | 25 %   | Median | 75 %  | Max.  |
|--------------------------|-----|--------|-----------|--------|--------|--------|-------|-------|
| <i>Adj.R<sup>2</sup></i> | 723 | 0.339  | 0.219     | -0.050 | 0.156  | 0.312  | 0.491 | 0.997 |
| <i>CFO_MEAN</i>          | 788 | 0.051  | 0.046     | -0.228 | 0.036  | 0.059  | 0.077 | 0.198 |
| <i>CFO_SD</i>            | 788 | 0.138  | 0.079     | 0.043  | 0.086  | 0.116  | 0.159 | 0.577 |
| <i>CFO_SKEW</i>          | 788 | -0.812 | 1.264     | -7.752 | -1.601 | -0.538 | 0.058 | 3.650 |
| <i>SD_OI_PTI</i>         | 788 | 0.080  | 0.065     | 0.000  | 0.040  | 0.061  | 0.096 | 0.583 |
| <i>PCTLOSS</i>           | 788 | 0.267  | 0.131     | 0.000  | 0.175  | 0.258  | 0.341 | 0.721 |
| <i>SGAIN_T_MEAN</i>      | 782 | 0.163  | 0.097     | 0.000  | 0.091  | 0.170  | 0.217 | 0.556 |
| <i>REL_MSIZ</i>          | 788 | 0.700  | 0.282     | 0.017  | 0.478  | 0.815  | 0.931 | 1.000 |
| <i>SHARE_INT_IND</i>     | 788 | 0.244  | 0.110     | 0.020  | 0.168  | 0.222  | 0.294 | 0.625 |

Note: This international panel sample contains country-year observations based on Compustat Global data for the period 1989 to 2016. Observations from the financial services industry and observations without data to calculate cash flow and accrual data are excluded. All firm-year data is by-year winsorized as explained in the paper before calculating the country-year measures. To be included in this panel, a country-year needs to contain at least 30 firm-year observations. Only countries with at least 20 years of data are included. All variables are as defined in Appendix A.

**Table 8: Correlations for International Panel Sample**

|                             | A            | B            | C            | D            | E            | F            | G            | H            | I            | J            |
|-----------------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| A: <i>Adj.R<sup>2</sup></i> |              | <b>-0.45</b> | <b>0.44</b>  | <b>-0.40</b> | <b>0.53</b>  | <b>-0.51</b> | <b>-0.58</b> | <b>-0.49</b> | <b>-0.39</b> | <b>-0.34</b> |
| B: <i>TIME</i>              | <b>-0.40</b> |              | <b>-0.35</b> | <b>0.32</b>  | <b>-0.35</b> | <b>0.30</b>  | <b>0.39</b>  | <b>0.61</b>  | <b>0.75</b>  | <b>0.43</b>  |
| C: <i>CFO_MEAN</i>          | <b>0.48</b>  | <b>-0.37</b> |              | <b>-0.77</b> | <b>0.59</b>  | <b>-0.58</b> | <b>-0.73</b> | <b>-0.56</b> | <b>-0.23</b> | <b>-0.36</b> |
| D: <i>CFO_SD</i>            | <b>-0.47</b> | <b>0.36</b>  | <b>-0.56</b> |              | <b>-0.68</b> | <b>0.68</b>  | <b>0.57</b>  | <b>0.51</b>  | <b>0.26</b>  | <b>0.37</b>  |
| E: <i>CFO_SKEW</i>          | <b>0.56</b>  | <b>-0.37</b> | <b>0.62</b>  | <b>-0.64</b> |              | <b>-0.52</b> | <b>-0.43</b> | <b>-0.45</b> | <b>-0.33</b> | <b>-0.37</b> |
| F: <i>SD_OI_PTI</i>         | <b>-0.68</b> | <b>0.39</b>  | <b>-0.54</b> | <b>0.71</b>  | <b>-0.53</b> |              | <b>0.55</b>  | <b>0.38</b>  | <b>0.25</b>  | <b>0.08</b>  |
| G: <i>PCTLOSS</i>           | <b>-0.60</b> | <b>0.40</b>  | <b>-0.71</b> | <b>0.51</b>  | <b>-0.46</b> | <b>0.62</b>  |              | <b>0.51</b>  | <b>0.29</b>  | <b>0.32</b>  |
| H: <i>SGAINT_MEAN</i>       | <b>-0.48</b> | <b>0.64</b>  | <b>-0.46</b> | <b>0.41</b>  | <b>-0.47</b> | <b>0.46</b>  | <b>0.45</b>  |              | <b>0.41</b>  | <b>0.43</b>  |
| I: <i>REL_MSIZES</i>        | <b>-0.30</b> | <b>0.71</b>  | <b>-0.23</b> | <b>0.25</b>  | <b>-0.33</b> | <b>0.27</b>  | <b>0.25</b>  | <b>0.41</b>  |              | <b>0.25</b>  |
| J: <i>SHARE_INT_IND</i>     | <b>-0.36</b> | <b>0.51</b>  | <b>-0.43</b> | <b>0.47</b>  | <b>-0.48</b> | <b>0.27</b>  | <b>0.39</b>  | <b>0.46</b>  | <b>0.34</b>  |              |

This table reports Pearson correlations above and Spearman correlations below the diagonal. The number of observations ranges from 718 to 788. Correlations with significance levels below 5% appear in bold print. All variables are as defined in Appendix A.

**Table 9: Test Results for International Panel Sample**

|                               | <i>Dependent variable:</i> |                                |                                |                                |                                |                                |
|-------------------------------|----------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|
|                               | <i>Adj.R<sup>2</sup></i>   | <i>Adj.R<sup>2</sup></i>       | <i>Adj.R<sup>2</sup></i>       | <i>Adj.R<sup>2</sup></i>       | <i>Adj.R<sup>2</sup></i>       | <i>Adj.R<sup>2</sup></i>       |
|                               | (1)                        | (2)                            | (3)                            | (4)                            | (5)                            | (6)                            |
| <i>TIME</i>                   | -0.016***<br>(0.002)       | -0.014***<br>(0.002)<br>[1.97] | -0.011***<br>(0.002)<br>[2.89] | -0.006***<br>(0.002)<br>[2.43] | -0.003*<br>(0.002)<br>[4.12]   | -0.002<br>(0.001)<br>[4.67]    |
| <i>CFO_MEAN</i>               |                            | 0.775**<br>(0.366)<br>[3.21]   |                                |                                | 0.610**<br>(0.309)<br>[5.16]   | -0.708**<br>(0.284)<br>[7.19]  |
| <i>CFO_SD</i>                 |                            | 0.837***<br>(0.246)<br>[4.25]  |                                |                                | 1.226***<br>(0.208)<br>[7.75]  | 1.404***<br>(0.215)<br>[14.43] |
| <i>CFO_SKEW</i>               |                            | 0.053***<br>(0.009)<br>[1.72]  |                                |                                | 0.054***<br>(0.008)<br>[2.03]  | 0.051***<br>(0.009)<br>[3.47]  |
| <i>SD_OI_PTI</i>              |                            |                                | -0.592**<br>(0.232)<br>[1.33]  |                                |                                | -0.962***<br>(0.211)<br>[6.44] |
| <i>PCTLOSS</i>                |                            |                                | -0.620***<br>(0.083)<br>[1.09] |                                |                                | -0.618***<br>(0.082)<br>[3.11] |
| <i>SGAINT_MEAN</i>            |                            |                                | 0.119<br>(0.211)<br>[3.05]     |                                |                                | -0.050<br>(0.089)<br>[4.65]    |
| <i>REL_MSIZE</i>              |                            |                                |                                | -0.154**<br>(0.070)<br>[2.65]  | -0.200***<br>(0.069)<br>[5.86] | -0.150***<br>(0.048)<br>[5.14] |
| <i>SHARE_INT_IND</i>          |                            |                                |                                | -0.871***<br>(0.164)<br>[2.19] | -0.925***<br>(0.133)<br>[2.14] | -0.808***<br>(0.133)<br>[4.16] |
| Estimator                     | ols                        | ols                            | ols                            | ols                            | ols                            | ols                            |
| Fixed effects                 | country                    | country                        | country                        | country                        | country                        | country                        |
| Std. errors clustered         | country                    | country                        | country                        | country                        | country                        | country                        |
| Observations                  | 723                        | 723                            | 718                            | 723                            | 723                            | 718                            |
| <i>R<sup>2</sup></i>          | 0.488                      | 0.530                          | 0.589                          | 0.540                          | 0.587                          | 0.680                          |
| Adjusted <i>R<sup>2</sup></i> | 0.464                      | 0.507                          | 0.568                          | 0.517                          | 0.565                          | 0.661                          |

Note: This table reports OLS regression results estimating the Adj.  $R^2$  measure on country-year Dechow and Dichev regressions (Model 2) on various sets of covariates. All variables are as defined in Appendix A. Variance inflation factors are reported in square brackets below the standard errors, which are clustered by country. \*/\*\*/\*\* indicate two-sided significance levels of 10/5/1 %, respectively.

**Table 10: Fixed Effect Results for International Panel Sample**

|                                | <i>Dependent variable:</i>       |                                  |                                  |                                  |                                  |
|--------------------------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|
|                                | <i>Adj.R</i> <sup>2</sup><br>(1) | <i>Adj.R</i> <sup>2</sup><br>(2) | <i>Adj.R</i> <sup>2</sup><br>(3) | <i>Adj.R</i> <sup>2</sup><br>(4) | <i>Adj.R</i> <sup>2</sup><br>(5) |
| <i>CFO_MEAN</i>                | 0.726*<br>(0.419)<br>[1.84]      |                                  |                                  | 0.662*<br>(0.400)<br>[2.22]      | -0.764***<br>(0.266)<br>[1.53]   |
| <i>CFO_SD</i>                  | 1.237***<br>(0.245)<br>[1.88]    |                                  |                                  | 1.246***<br>(0.243)<br>[2.23]    | 1.398***<br>(0.223)<br>[6.62]    |
| <i>CFO_SKEW</i>                | 0.053***<br>(0.010)<br>[1.21]    |                                  |                                  | 0.052***<br>(0.010)<br>[1.24]    | 0.050***<br>(0.010)<br>[2.65]    |
| <i>SD_OI_PTI</i>               |                                  | -0.425*<br>(0.244)<br>[1.27]     |                                  |                                  | -0.856***<br>(0.199)<br>[3.77]   |
| <i>PCTLOSS</i>                 |                                  | -0.520***<br>(0.115)<br>[1.17]   |                                  |                                  | -0.629***<br>(0.095)<br>[1.86]   |
| <i>SGAINT_MEAN</i>             |                                  | 0.091<br>(0.224)<br>[1.46]       |                                  |                                  | -0.167<br>(0.107)<br>[2.11]      |
| <i>REL_MSIZ</i>                |                                  |                                  | 0.053<br>(0.068)<br>[1.05]       | 0.007<br>(0.067)<br>[1.23]       | -0.050<br>(0.053)<br>[1.60]      |
| <i>SHARE_INT_IND</i>           |                                  |                                  | -0.246<br>(0.280)<br>[1.05]      | -0.288<br>(0.240)<br>[1.13]      | -0.371**<br>(0.183)<br>[1.63]    |
| Estimator                      | ols                              | ols                              | ols                              | ols                              | ols                              |
| Fixed effects                  | country, year                    |
| Std. errors clustered          | country, year                    |
| Observations                   | 723                              | 718                              | 723                              | 723                              | 718                              |
| <i>R</i> <sup>2</sup>          | 0.651                            | 0.648                            | 0.608                            | 0.654                            | 0.715                            |
| Adjusted <i>R</i> <sup>2</sup> | 0.620                            | 0.616                            | 0.574                            | 0.622                            | 0.687                            |

Note: This table reports OLS regression results estimating the Adj. *R*<sup>2</sup> measure on country-year Dechow and Dichev regressions (Model 2) on various sets of covariates. All variables are as defined in Appendix A. Variance inflation factors are reported in square brackets below the standard errors, which are clustered by country. \*/\*\*/\*\* indicate two-sided significance levels of 10/5/1 %, respectively.