The Effect of Board Structure on Firm Disclosure and Behavior: A Case Study of Korea and a Comparison of Research Designs

Bernard Black
Northwestern University, Law School and Kellogg School of Management
bblack@northwestern.edu

Woochan Kim
Korea University Business School
wckim@korea.ac.kr

Julia Nasev
Ludwig-Maximilians-Universität Munich
nasev@lmu.de

Draft October 2020

Northwestern University School of Law
Law and Econ Research Paper Number 12-13

This paper can be downloaded without charge from SSRN at:
http://ssrn.com/abstract=2133283

The Appendix can be downloaded without charge from SSRN at:
http://ssrn.com/abstract=3693354

Acknowledgments: We thank seminar participants at McCombs School of Business (July 2012), Chicago Booth School of Business (Nov. 2012), 7th International Conference on Asia-Pacific Financial Markets (Dec. 2012), American Accounting Association annual meeting (August 2013), Financial Management Association annual meeting (2013), Shuping Chen, Martin Dierker, Dain Donelson, Zi Jia, Christian Leuz, Jim Naughton, Joshua Ronen, and Ira Yeung for comments.

The Effect of Board Structure on Firm Disclosure and Behavior: A Case Study of Korea and a Comparison of Research Designs

Abstract

We exploit a large legal shock to the board structure of Korean firms, using a strong research design – combined difference-in-differences and regression discontinuity – to study whether this board structure change affects firm financial reporting (disclosure, MD&A length, and abnormal accruals), investment and growth (sales growth and capital expenditures), and firm value (proxied by Tobin's q). We also compare results from the annual DiD/RD design to those from simpler panel and "causal" methods, and assess how results vary across methods. We find robust evidence across methods that the shock predicts improved scores on a Disclosure Subindex, confirm prior findings of an increase in Tobin's q, find some evidence for a drop in sales growth, but no convincing evidence of significant change for other outcomes. By comparing results across methods, we illustrate how using multiple causal designs can provide insight and evidence on robustness not available from a single design, as well as case study evidence that panel methods, simple DiD, and its close cousin, shock-based IV, can produce apparent false positives.

JEL classifications: C01; G3; G34; M41.

Keywords: corporate governance, board structure, financial reporting quality, investment, sales growth, earnings management, accrual quality, Korea, causal inference, panel data, difference-in-differences, instrumental variables, regression discontinuity.

1 Introduction

We study two research questions. The first involves the effects of a shock to the board structure of Korean firms on firm behavior. We study whether this board structure shock affects other aspects of governance (disclosure, board procedure, shareholder rights), financial reporting (absolute and signed abnormal accruals, MD&A length), firm financial outcomes (profitability, leverage, growth, and investment), and firm value (proxied by Tobin's *q*). We do so using a strong, "causal" research design – a design that combines annual difference-in-differences ("annual DiD") regressions with a regression discontinuity (RD) design, that draws strength from both design aspects. With this annual DiD/RD benchmark design, we find evidence that the shock predicts improved scores on a Disclosure Subindex and higher Tobin's q, some evidence for lower sales growth, but no significant change in the other outcomes.

We also provide case study evidence on the sensitivity of research results to choice of research design. Finance and accounting research has moved strongly toward relying on natural experiments, using various research designs, including (DiD) and RD separately (not often combined), and instrumental variables (IV) with the shock as an instrumental variable. But natural experiments are often hard to find or answer limited questions. Classic panel data approaches – with firm fixed effects (FE) or random effects (RE) – are often the best available designs. We assess how our results would change if we used either panel methods or simpler "causal" methods. The potential limitations of panel methods, as a guide to causal effects, are known (e.g., Leamer, 1983; LaLonde, 1986). But we know little about how important these limitations are in applied research. We also know little about how reliable simpler causal designs are as guides to causation, or about how results differ across methods, including different causal methods which rely on the same shock.

We exploit a large, exogenous shock to the board structure of large, public Korean firms (assets over 2 trillion won, about US$2 billion, "2T"), with no similar shock to smaller firms. Korean rules, adopted in 1999 following the East Asian financial crisis of 1997-1998, require large firms to have at least 50% outside directors, an audit committee (with at least 2/3 outside directors and an outside chair), and an outside director nominating committee (with at least 50% outside directors). Before the rules were adopted, essentially no Korean firm had *any* of these governance elements. Thus, this legal shock strongly affects two core governance institutions – outside directors and audit committees. Prior research has found that the shock strongly affects the market value of Korean firms, reflected in both share prices and Tobin's *q* (Black et al., 2006; Black and Kim, 2012).

For the first research goal, we find evidence that this shock leads to higher scores for a "Disclosure Subindex", but no evidence for a significant change in other aspects of firm governance (a Board Procedure Subindex or a Shareholder Rights Subindex) or other aspects of disclosure (signed and absolute abnormal accruals, and word length of the Management's Discussion and Analysis (MD&A) section of annual reports). Turning to financial outcomes, we confirm prior findings that the board structure shock predicts higher Tobin's *q* (Black and Kim, 2012), find some (less than definitive) evidence of a drop in sales growth, but no overall evidence of significant change in other outcomes, including profitability measured by return on assets (ROA), or investment (measured by capital expenditures and research and development spending (R&D)). This evidence, taken as a whole, is consistent with independent directors and audit committees having a real, although modest effect on selected aspects of firm behavior.

For the methods goal, true causal effects are unknown and, as with any case study, our results may not generalize to other settings. However, the Korean shock provides an unusually "clean" shock. It is plausibly exogenous (the 2T threshold is apparently arbitrary), it strongly alters board structure, and it occurs at a specific point in time.[1] We compare assumed "truth" from our benchmark annual DiD/RD design to results from simpler "causal" designs – including "simple DiD" (comparing after to before the shock), annual DiD, RD, and instrumental variables (IV) using the shock as an instrument.

With regard to whether our benchmark design can recover approximate truth, studies across different fields provide evidence that RD can often recover results close to those from a randomized experiment.[2] From a theoretical perspective, within a bandwidth around the discontinuity, one can treat the RD sample as if it came from a randomized experiment, with the running variable providing the assignment mechanism which determines which units are treated (Mattei and Mealli, 2016). Combining RD with DiD addresses a principal limitation of RD, that results are valid only near the discontinuity (Mealli and Rampichini, 2012), and permits use of a broader bandwidth, which increases sample size and power. Using a combined DiD/RD design also allows us to relax the usual RD assumption of similar *levels* between treated and control and rely instead on the weaker DiD assumption of parallel *changes* (parallel trends). And using annual DiD allows one to assess whether the treatment effect appears *when* it should relative to the shock, and whether there are non-parallel pre-treatment trends, that could explain an apparent treatment effect. Thus, our benchmark annual DiD/RD design provides a reasonable basis for comparison with other methods. Moreover, given the rarity of true

---

[1] We discuss below other limitations of this setting. A principal concern is limits on statistical power due to the modest number of Korean firms above the 2T size threshold.

[2] Buddelmeyer and Hielke (2004); Black et al. (2007); Cook and Wong, (2008); Cook et al. (2008); Green et al., (2009); Berk et al., (2010); Shadish et al. (2011); Gleason et al. (2012); Moss et al. (2014).

randomized experiments in finance and accounting, we know of no other way to address the reliability of results across different research design.

We compare results from the benchmark design to simpler causal designs and to panel data designs using either firm fixed effects (FE) or random effects (RE). Overall, as we move from our benchmark design to annual DiD alone (which we view as the next strongest design if RD is not available) to other causal designs, firm FE, and finally to firm RE, there is a general tendency for more outcomes to become statistically significant. We term these results "apparent" false positives. We do not know truth and thus cannot be sure that these results are false positives, but they would not be convincing if one also had results from stronger designs.

For our methods goal – assessing whether results are consistent across methods – there is no clear winner in terms of *not* generating apparent false positives. However, simple DiD is a clear loser – it generates more apparent false positives than other causal methods; indeed more than firm FE. In the common case where an RD design is not available, a moderately conservative suggestion might be to take seriously only results found with annual DiD and at least one other method. Firm FE also produces a number of apparent false positives, and firm RE produces even more. Thus, our case study is consistent with the view that RE alone has limited credibility as a guide to causation.

We find that RD generates several apparent false negatives, likely due to limited power. For the other methods, our case study suggests that apparent false positives are an important risk, but apparent false negatives are smaller concern.

The differences in results across different research designs suggest the importance of applying multiple designs, including multiple causal designs when available, and assessing the consistency of results across designs. Often, researchers report results from a single design. This can lead to reporting apparent false positives. Moreover, the differences in results across designs increase the risk that researchers may engage in specification search and report results from the design that produces statistically significant results.

Two examples can illustrate the differences across methods for some outcomes. Consider first Abs(AA). Substantial prior research treats abnormal accruals as reflecting, in part, management discretion to manage earnings and examines whether governance appears to affect abnormal accruals. Much of that research finds that better governance predicts lower abs(AA). But all prior studies use non-causal designs. Consistent with those studies, we find that higher KCGI scores predict lower abs(AA) with panel data designs. But we find no evidence of an effect on abs(AA) with causal designs. Consider next sales growth, for which the panel designs provide some evidence that stronger governance predicts slower growth, which in Korea at time is likely efficiency-enhancing (significant with firm RE, marginally significant with firm FE). This result survives with some simpler causal designs (simple and annual DiD), but is only marginally significant with our benchmark design. Considering the evidence from all causal methods together, we think it fair to say that we find some, but less than definitive evidence of a decline in sales growth.

We proceed as follows. Part 2 discusses the setting. Part 3 provides an overview of the research designs and outcomes we study. In Part 4, we present results with panel designs, in Part 5 results with simpler causal designs and in Part 6 results with combined causal designs, including our benchmark design. Part 7 discusses the results and concludes.

2 Korean Setting

2.1 Korean Governance Shock

We study a 1999 legal shock to the governance of large Korean firms. We describe that shock briefly here and discuss shock exogeneity below. We provide further details in an online appendix; see also Black and Kim (2012). In response to the 1997-1998 East Asian financial crisis, the Korean government pursued corporate governance reforms requiring that "large" firms (assets > 2 trillion won, below "2T") have (i) a minimum of 50% outside directors; (ii) a minimum of three outside directors, (iii) an audit committee with an outside director as chair and at least two-thirds outside members; and (iv) an outside director nominating committee (with at least 50% outside members) to select new outside directors. Smaller firms were required to have at least 25% outside directors. Prior to the reforms, almost no firms had 50% outside directors and none had an audit committee; indeed, Korean Company Law had to be amended to allow for board committees. This shock strongly affected two core governance institutions – the board of directors and the audit committee – that plausibly affect financial reporting, investment, firm growth, and firm value.

The rules came into force partly in 2000 and fully in 2001.[3] The impact of the reforms on share price, and thus Tobin's *q*, should appear in the second half of 1999 (2H 1999) in an efficient market, as investors anticipate the future effects of the reforms on board structure and thus on firm behavior. Black and Kim (2012) find evidence of a large share price response to the reforms. Large firm share prices increase by about 30% relative to mid-sized firms (0.5 trillion won < assets < 2T) during the reform adoption period. We expect the impact of the reforms on other outcomes to appear with a lag, thus in 2001 or later.

[3] Large firms had to have at least three outside directors and the two committees in 2000, and had to have 50% outside directors in 2001. A 2003 rule change required large firms to have a majority of outside directors beginning in 2005 (exactly 50% was no longer allowed).

The two principal sources of bias in regression estimates of causal effects are reverse causation and omitted variable bias. Reverse causation is not plausible. The legal reforms were imposed involuntarily on large Korea firms, which opposed them. The more serious concern is omitted variable bias, but as we discuss below, our benchmark design should limit most sources of omitted variable bias.

2.2 Evidence on Exogeneity of the Reform Shock

We summarize here evidence that the 1999 reform can be reasonably treated as "exogenous", in the qualitative sense of coming from outside an otherwise endogenous system of financial outcomes and corporate governance practices.[4] Black and Kim (2012) provide additional details. First, the reforms cause a major change in board structure at large firms. Figure 1 shows how Board Structure Subindex – the part of KCGI directly affected by the reforms – changes over 1998-2004 for large and mid-sized firms. We exclude banks and former state-owned enterprises (SOEs) from the sample; and exclude all financial firms when using capital expenditures as the outcome variable. The vertical line shows the 2T threshold; the horizontal line is at 11.67, which is the minimum Board Structure Subindex score for firms that comply with the large-firm rules. No firms have any of the three reform elements at year-end 1998; only one mid-sized and one large firm have adopted any of these elements at year-end 1999. Some large firms comply with the new rules in 2000, ahead of the spring 2001 deadline; all comply by 2001. A few over-comply and are above the horizontal line.

The reforms were not adopted due to industry pressure – the major Korean chaebol groups opposed the reforms. At the same time, some mid-sized firms voluntarily adopted board structure changes; the tendency for voluntary adoption increases over 2000-2004. The discontinuity at 2T is thus "fuzzy", and increasingly so in later years.

It is unlikely that the 2T threshold is magic, with firms just above the threshold responding to treatment differently than firms just below. That magic is especially implausible here, because the government proposed a 1T threshold, but later raised the threshold to 2T in response to political pressure from large firms. An event study of the initial proposal provides evidence that investors reacted positively for firms with assets between 1T and 2T (Black and Kim, 2012). Yet, because the threshold was later raised, these firms are part of the control group for our study.

---

[4] This is the typical meaning given to the term "exogenous" in the causal inference literature. See, for example, Angrist and Pischke (2009); Imbens (2004). It differs from the technical definition of an "exogenous" variable in parametric regressions, as presented in standard econometrics textbooks.

Could the 1999 legal reform affect firms through a channel other than board structure? A possible alternative channel for our setting is that many large firms belonged to large business groups, known as *chaebol*. Could the apparent response to the shock really be a response to general government pressure on *chaebol* firms following the East Asian financial crisis? We cannot entirely exclude this channel, but Black and Kim (2012) run a horserace between a *chaebol* dummy and a large firm dummy; the effect of large firm dummy on share price response to the reforms survives, while *chaebol* dummy is insignificant.

A concern for any RD design is whether the threshold can be manipulated (McCrary, 2008). We find no evidence that firms reduce or limit their size to avoid the rules. There is no bunching of firms above or below either the 1T or 2T thresholds. The few firms who shrink from above to below the 2T threshold after the reforms either do not abandon the large-firm board structure elements (so avoiding these rules cannot be the reason for shrinking) or continue to shrink (thus, business reversals rather than avoidance explains shrinking).[5]

2.3 Korean Corporate Governance Index (KCGI)

For our non-causal analyses, we rely on a broad Korea Corporate Governance Index (*KCGI*), developed in Black and Kim (2012), and summarized in Table 1. We construct *KCGI* over 1998 to 2004, for the vast majority of public companies listed on the Korea Stock Exchange. *KCGI* (0 ~ 100) consists of five equally weighted subindices: Board Structure (5 elements), Disclosure (3 elements), Shareholder Rights (4 elements), Board Procedure (14 elements), and Ownership Parity (one element). Within each subindex, all elements are equally weighted, except that Board Structure Subindex is composed of Board Independence Subsubindex (2 elements, 0 ~ 10), and Board Committee Subsubindex (3 elements, 0 ~ 10). For details on construction of KCGI, see Black and Kim (2012).

These subindices and elements are Korea-specific. They cover aspects of governance which we judged to be important in Korea during this period, and for which we have data. Almost all Korean firms had a controlling shareholder or group, so takeover defenses were unimportant. As a result, KCGI is quite different from U.S.-centric indices, which focus heavily on takeover defenses (e.g., Gompers et al., 2003; Bebchuk et al., 2009), and its elements are quite different from those studied by Larcker et al. (2007).

Table 2, Panel A provides summary statistics for *KCGI* and subindices; Panel B provides correlation coefficients. All subindices except Ownership Parity are strongly correlated.

---

[5] Of the seven firms that are large at year-end 1999 but shrink below the 2T threshold during our sample period, four retain the reforms, one soon becomes large again, and one continues to shrink. Only one large firm shrinks to moderately below the threshold and abandons the large-firm reforms, and thus might have shrunk for this purpose.

2.4 Dataset and Background on Korean Firms and Accounting Rules

Our data comes from various sources. We take balance sheet, income, cash flow statement data, foreign ownership data, related-party transactions, and original listing year from the TS2000 database maintained by the Korea Listed Companies Association (KLCA); MD&A word count is based on annual reports; information on chaebol groups from the Korea Fair Trade Commission (KFTC); stock price and return data from DataGuide 3.0 maintained by FnGuide; information on ADRs from JP Morgan and Citibank websites; and industry classification from the Korea Statistics Office (KSO). Share ownership comes from the Korea Stock Exchange (KSE) supplemented by a hand-collected database.

Most Korean firms have a controlling family. Many firms, especially large firms belong to business groups, known as *chaebol*. The incentives of these firms' managers in making decisions on financial disclosure, investment, and growth could differ from those at firms without a controlling shareholder; similarly, incentives could differ between *chaebol* and non-*chaebol* firms. But it is not obvious why any differences between Korean and, say, U.S. firms should affect our core research questions, which involve the robustness of classic panel and simpler causal research designs, versus more careful designs.

We measure Korean governance over 1998-2004. We measure outcomes other than Tobin's *q*, with a one-year lag, over 1999-2005, because we expect the governance change to affect these outcomes with a lag. Our causal research designs assume that the 1999 governance reforms to large firms is the only external shock affecting large firm outcomes, relative to those for smaller firms. We therefore note the principal changes to Korean accounting rules during this period. Korea adopted an analog to Regulation FD (Fair Disclosure) in 2002, and an analog to the Sarbanes-Oxley Act in 2003. Most of these reforms became effective in 2004, and apply to both large and small firms, so they should not compromise our research design.

A wide variety of economic factors can affect firm-value and firm governance (e.g., Himmelberg, Hubbard, and Palia, 1999). We therefore use an extensive set of covariates to limit omitted variable bias. We use the following covariates, and discuss here their importance with Tobin's *q* as the outcome of interest. *Firm size:* natural logarithm (*ln*) of assets, asset size is the running variable for the RD design and because size is known to predict Tobin's *q*. *Firm age*: *ln*(years listed +1), because younger firms are likely to be faster-growing and more intangible asset-intensive; *Leverage*, which can influence Tobin's *q* by affecting income tax and reducing free cash flow problems, and is mechanically related to Tobin's *q* (when leverage is an outcome, we remove it from the covariates). *Growth prospects and profitability,* which predict Tobin's *q*. We control for sales growth (but remove this covariate when sales growth is the outcome) and for profitability using *EBIT*/sales (but remove this covariate when ROA is the outcome). *Capital* *intensity and asset tangibility:* Asset tangibility can both predict Tobin's *q* and affect what type of governance a firm needs. We control for PPE/sales, capex/PPE, advertising/sales, and R&D/sales (removed when R&D/sales is the outcome). *Liquidity*: annual share turnover (traded shares/total shares). *Ownership*: fractional ownership by the largest shareholder, foreign investors, and the state. *Product market competition*, which can directly affect value and substitute for governance in imposing discipline on managers: exports/sales and domestic market share in the firm's principal 4-digit industry. With RE, we also use several firm-level variables which can predict both governance and *q*: *US cross-listing dummies* and *MSCI index dummy* to proxy for liquidity and foreign investor interest. *Business group (chaebol) dummy,* because group firms may behave differently than stand-alone firms. We also use industry*year FE with both the RE and FE specifications. Table 3 defines the principal outcome and covariates we use in this study and indicates our winsorization choices. Table 4 provides a "covariate balance" table for the outcomes and covariates. Some covariates can be seen as outcomes of the board structure shock; if so, controlling for them could bias estimated treatment effects. We address this concern by reporting results both with limited covariates (only ln(assets) and ln(years listed)) and with full covariates.

A limitation of this study is the modest number of large firms. Our main RD results use mid-sized and large Korean public firms, with assets from 0.5-8 trillion won (a factor of 4 on either side of the 2T threshold). This size band reflects a compromise between desire for a narrow band, to make treated and control firms more similar, and need for a reasonable sample size. The tradeoff between bandwidth and sample size is common in RD designs, but is acute for us because we have a limited number of large firms. For example, with Disclosure Subindex as the outcome and full covariates, we have 53 large firms in the [2T, 8T] range, and 41 in the [2T, 4T] range, depending on year and specification. Sample size limits statistical power and could lead to failure to reject the null even if an effect is present.

Many causal designs rely on limited sources of variation in the predictor variable, limited samples, or both. This limitation is acute for RD – as one narrows the RD bandwidth, the basis for causal inference improves but the sample and thus statistical power shrinks at the same time. Concern with statistical power is also important for our study because of the limited number of large firms. We therefore explore the balance between sample size and cleaner causal inference by reporting results with both a broader and a narrower RD bandwidth.

3 Panel and Causal Research Designs; Outcome Variables

An important goal of this paper is to compare estimates across different research designs. We summarize here the designs that we examine.

3.1 Benchmark Annual DiD/RD Design and the Meaning of an "Exogenous Shock"

It is often said that a shock must be "exogenous" to permit causal inference –the shock must come from outside the endogenous system one seeks to study. The shock should assign treatment in an "as if random" manner, unrelated to their potential outcomes (y_i^1 if treated, y_i^0 if control). More realistically, one can hope that the shock is unrelated to potential outcomes after controlling for observed covariates $\boldsymbol{x}$. This cannot be directly tested, so researchers look for indirect evidence.

An RD design uses a shock which separates firms into treated and control based on a sharp threshold. The 2T threshold we study presumably reflects legislative judgment as to the benefits and costs of different thresholds. Even so, the threshold conveys no magic – firms close to but above the threshold are likely similar to firms close to but below the threshold, on everything but the "running" variable that defines the threshold (in our study, firm assets), and any related variables (here, other size-related variables). Thus, if one limits the sample to a bandwidth around the threshold, and controls for the running variable, one can recover a good approximation to a randomized experiment.

Because the running variable determines which firms are treated, the RD design rules out reverse causation. Less obviously, the RD design also blocks omitted variable bias near the threshold (Angrist and Pischke, 2015; Angrist and Rokkanen, 2015). Omitted variable bias arises if an omitted variable u is *partially* correlated (controlling for other covariates) with both the outcome and the treatment. But if the running variable fully determines treatment, u will be partially uncorrelated with the treatment dummy.

In practice, matters are not this simple. Near the threshold, it is common to use a linear functional form for the running variable (e.g., Cattaneo et al., 2019). A more flexible specification is also possible, but can absorb the treatment effect. Thus, researchers must exercise judgment as to both the bandwidth around the RD threshold and the functional form for the running variable. Still, the RD design should sharply limit omitted variable bias.

Our combined annual DiD/RD design reinforces the RD design, and further limits omitted variable bias; so does our use of extensive time-varying covariates. As is typical in a DiD design, we use firm FE to control for time-invariant firm characteristics, and use changes over time in control group outcomes to control for any changes in the treatment group, other than those resulting from the treatment. By graphing the annual DiD estimates, we can assess whether pre-treatment trends appear parallel, and whether the response to the shock appears *when* it should relative to the shock. For example, we would expect the Korean reform, adopted in 2H 1999, to affect share prices and thus Tobin's q immediately, but since the reforms were implemented in 2000 and 2001, we would expect any effect on abs(AA) to appear more slowly.

We cannot know, of course, how close our benchmark design comes to the "truth" that a natural experiment would uncover, if one were feasible. Still, if results from weaker designs disappear with the benchmark design, this suggests that the weaker designs are probably generating false positives.

3.2 Classic Panel Data Designs

We use two classic panel data designs: RE and FE. We use an unbalanced panel, and cluster standard errors on firm.[6] For outcomes other than Tobin's q, we lag KCGI one period, because it seems likely that governance changes will not immediately affect our outcome variables. We do not lag KCGI for Tobin's q because we expect (as Black and Kim, 2012, find) investors to react when the large firm rules are adopted, in 2H 1999.

These panel models are well-known. We review here aspects relevant for our study. The firm effects model is:

$$y_{i,t} = \beta_0 + \beta_1 \cdot KCGI_{i,t-1} + \boldsymbol{\beta_2} \cdot \boldsymbol{x}_{i,t} + g_t + f_i + \varepsilon_{i,t} \qquad (1)$$

Here $\boldsymbol{x_{i,t}}$ is a vector of covariates, which we assume to be "exogenous" (unaffected by the values of the outcome or other covariates),[7] and the g_t are industry × year dummies.[8]

The FE model can be seen as a "time-demeaned" specification. Let $\boldsymbol{x}_{i,t}^{dm} = (\boldsymbol{x_{i,t}} - \boldsymbol{\bar{x}_i})$, and similar for other variables. The FE model is:

$$y_{i,t}{}^{dm} = \beta_1 \cdot KCGI_{i,t-1}^{dm} + \boldsymbol{\beta_2} \cdot \boldsymbol{x}_{i,t}^{dm} + g_t^{dm} + \varepsilon_{i,t}^{dm} \qquad (2)$$

---

6 We prefer one-way to two-way clustering. Black et al. (2020) report that, in their setting, two-way clustered standard errors can be severely downward biased with a short time dimension, but standard errors clustered on firm come very close to reproducing results from randomization inference.

7 We term this "(technical) exogeneity," to avoid confusion between exogeneity in this sense, and the qualitative meaning of an "exogenous shock." In our experience, in most shock-based research, no one takes the technical exogeneity assumption very seriously; a typical view is that the shock must be qualitatively exogenous and the $\mathbf{x}_{i,t}$ should not be (strongly) affected by the shock.

8 We use industry × year FE rather than simpler year FE to control for possible time trends within particular industries, but face a practical constraint on the fineness of the industry categories: For our causal specifications, especially within-bandwidth specifications, we need to maintain a minimum number of firms in each industry (or else the industry × year FE will approach firm × year FE, which would absorb an actual treatment effect). Based on assessing the number of large firms in particular industries, we report results using three broad industry groups: manufacturing, financial, and other. We obtain similar results with six industry groups (three within manufacturing, plus construction, financial, and other), see Appendix for results, and with only year FE (results not reported). Industries are based on 2- and 3-digit Korean Industrial Classification codes.

RE leads to a "quasi-demeaned" feasible GLS estimate. Let σ_ε and σ_f be the standard deviations of $\varepsilon_{i,t}$ and f_i, T be the number of periods, and define:

$$\lambda = 1 - \frac{\sigma_\varepsilon}{\sqrt{\sigma_\varepsilon^2 + T \cdot \sigma_f^2}}$$

Also define quasi-demeaned variables $\boldsymbol{x}_{i,t}^{qdm} = (\boldsymbol{x}_{i,t} - \lambda \cdot \bar{\boldsymbol{x}}_i)$ and similar for other variables. The RE model is:

$$y_{i,t}{}^{qdm} = \beta_1 \cdot KCGI_{i,t-1}^{qdm} + \boldsymbol{\beta}_2 \cdot \boldsymbol{x}_{i,t}^{qdm} + g_t^{qdm} + f_i^{qdm} + \varepsilon_{i,t}^{qdm} \qquad (3)$$

The RE models makes a (technical) "strict exogeneity" assumption that the firm effects are uncorrelated with the covariates in all time periods. The FE estimator makes a weaker (technical) exogeneity assumption (confusingly also called strict exogeneity) that the $\boldsymbol{x}$'s (including *KGGI*) are exogenous in all time periods (thus ruling out feedback between $y_{i,t}$ and future $\boldsymbol{x}$'s), and relies only on within-firm variation across time, which reduces power.

(Technical) exogeneity of the $\boldsymbol{x}$'s and *KCGI* requires, among other things that neither *KCGI* nor current values of the outcome variable influence future $\boldsymbol{x}$'s, and that current values of the outcome do not influence future *KCGI*. This is unlikely to be strictly true, but one might hope that it is a reasonable approximation. First, for Korean firms, time varying characteristics only weakly predict *KCGI* (Black, Jang and Kim, 2006). Bhargava and Sargan (1983) suggest that assuming (technical) strict exogeneity of the $\boldsymbol{x}$'s is more reasonable if the data has a "short" time dimension, and the principal variable of interest (here, *KCGI*) is time-persistent.

3.3 Simple Causal Designs

The Korean shock permits several distinct causal research designs. We summarize these designs following the notation in Imbens and Rubin (2015); see Atanasov and Black (2016) for more details on these designs as used in finance and accounting research. For each firm *i,* we assume a binary "treatment" ($w_i = 1$ if treated; 0 if not). The firm-level causal effect τ_i of treatment on y_i is defined as the value of y_i if firm *i* is treated, minus the value of y_i if firm *i* is not treated:

$$\tau_i = y_i(w_i = 1) - y_i(w_i = 0), \text{ or, more compactly: } \tau_i = y_i^1 - y_i^0$$

The "fundamental problem of causal inference" (Holland, 1986) is that for each firm, we observe only one of the two "potential outcomes," y_i^1 and y_i^0. The causal challenge is to credibly impute the missing potential outcomes for treated firms from control firms (and potentially vice versa).

The central challenge to imputation is "selection bias": the treated and control firms differ in unobserved ways, which biases the treatment effect estimates.

3.3.1 Simple DiD

The simplest causal design is "simple" DiD, in which one compares the after-minus-before change in the outcome for treated firms to the after-minus-before change for control firms. If one has panel data with several pre-treatment and several post-treatment periods, one compares the post- and pre-treatment means. Suppose first that we had only two time periods, one before and one after the reform, with large firms "treated" and smaller firms as "controls." One can then estimate the average treatment effect for the treated (ATT) as (after-minus-before change for treated firms) minus (after-minus-before change for control group).

$$ATT_{\text{DiD}} = E_{treated}[\bar{y}_{i,a}^{1} - \bar{y}_{i,b}^{1}] - E_{controls}[\bar{y}_{j,a}^{0} - \bar{y}_{j,b}^{0}]$$

This estimate can be implemented using an FE regression. Without covariates, we estimate:

$$\{simple\ DiD\}: y_{i,t} = \alpha + f_i + g_t + \delta_{DiD} \cdot w_{i,t} + \varepsilon_{i,t} \qquad (4)$$

Here the sample period runs from $-n_{pre}$ to $+n_{post}$ with $t = 1$ as the first post-treatment period; $w_{i,t}$ (= 1 for treated firms if $t > 1$, 0 otherwise); and $\hat{\delta}_{DiD}$ is the empirical estimate of ATT. The Korean reforms were adopted in 1999 but took effect partly in 2000 and fully in 2001. For Tobin's q as outcome, we follow Black and Kim (2012), use semiannual data (defining 1H 1999 to end in May 1999 because the reform period starts in June 1999) and define $w_{i,t} = 1$ for large firms for 2H 1999 and later. For other outcomes, we use annual data and define $w_{i,t} = 1$ for 2001 and later years for other outcomes, but find similar results if we define $w_{i,t} = 1$ starting in 2000.

One can add "exogenous" covariates that are not affected by the treatment to eqn. (4). To obtain an unbiased estimate of the treatment effect, one must make an untestable "parallel trends" assumption; that the outcomes for treated firms (had they not been treated) and control firms would have followed parallel paths over time. This assumption cannot be tested in the treatment period, because we observe only treated outcomes for the treated firms. However, annual DiD graphs provide an important plausibility check – one can assess whether outcomes for the treatment and control groups were parallel during the pre-treatment period.

The DiD and firm FE models are similar. The principal differences are: (i) the independent variable of interest is the treatment dummy $w_{i,t}$, instead of a possibly continuous variable such as *KCGI* in eqn. (2); and (ii) we have reason to believe that $w_{i,t}$ is plausibly exogenous, because it comes from an external shock, rather than being a firm choice.

3.3.2 Annual DiD

Simple DiD assumes a one-time jump from before to after the shock. A more convincing alternative is an annual DiD specification (sometimes called "leads and lags"), which lets one assess when, during the treatment period, the apparent shock effect emerges, and also lets one assess whether pre-treatment trends appear parallel. This specification is, for outcomes for which we have data from 1998-2005:

$$\{annual\ DiD\}:\ y_{i,t} = \alpha + f_i + g_t + \sum_{k=1998}^{2005} \delta_{DiD}^{k} \cdot w_i^{k} + \varepsilon_{i,t} \qquad (5)$$

Each w_i^k turns on for treated firms in period k and then off again. One interaction term must be omitted. For financial outcomes, for which we expect a potential treatment effect starting in 2001, we omit 2000; for share price related outcomes, for which the potential treatment effect should occur in 1999 (the reform adoption year), we omit 1998. The annual DiD coefficients can be graphed, with confidence intervals. If pre-treatment trends are parallel, the coefficients on w_i^k during the pre-treatment period should be small and insignificant, with no apparent trend. During the treatment period, the "lag" coefficients will map out the treatment effect over time.

3.3.3 Shock-Based Instrumental Variables

A third available causal design is shock-based IV, with large firm dummy as an instrument for Board Structure Subindex. One can either use IV during the treatment period alone (with large firm dummy as the instrument) or during the entire period, with large firm dummy × post as the instrument. We prefer the latter approach. The model is two-stage least squares (2SLS), during the post-reform period. The instrument z substitutes for the instrumented variable; and we *assume* that the power of the *instrument* to predict the outcome reflects the true power of the *instrumented variable*. The *2SLS* estimate of the coefficient on *gov*, without covariates, is

$$\hat{\beta}_{2SLS} = \frac{Cov(z, y)}{Cov(z, x)}$$

This coefficient $\hat{\beta}_{2SLS}$ can also be expressed in terms of the "intent-to-treat" DiD coefficient δ_{DiD}; this is known as a Wald estimate:

$$\hat{\beta}_{2SLS} = \frac{\hat{\delta}_{DiD}}{\hat{\beta}_{1S}} = \frac{effect\ of\ instrument\ (shock)\ on\ y}{effect\ of\ instrument\ (shock)\ on\ x} \qquad (6)$$

Here $\hat{\beta}_{1S}$ is the first-stage coefficient from regressing x on z.

A classic statement of the requirements for a valid instrument z (e.g., Wooldridge, 2010, ch. 5) is that z is correlated with x (strongly enough to avoid weak instrument issues); and $Cov(z, \varepsilon) = 0$, where ε is the unobserved true error in the original regression. The first condition can be tested in the sample. The second condition is not testable. An alternate statement of the requirements for a valid instrument would be, following Angrist and Pischke (2009, § 4.1):

(i) *instrument strength*: z is correlated with x (strongly enough to avoid weak instrument problems); and conditioned on other covariates;

(ii) *instrument as good as randomly assigned*: z cannot be influenced by the outcome variable y (thus ruling out reverse causation and simultaneity). This is sometimes loosely phrased as z being "exogenous" to the variables in the original regression. But z must also be as good as randomly assigned – it must be independent of the potential outcomes, either fully ($z \perp\!\!\!\perp y_1, y_0$) or conditioned on covariates ($z \perp\!\!\!\perp y_1, y_0 | \boldsymbol{x}$); and

(iii) *only through condition* (often called an "exclusion restriction"): z predicts the outcome y *only through* the instrumented variable, not directly nor through unobserved variables $\boldsymbol{u}$. This condition cannot be directly tested; one can only defend it through logic and a design that seeks to exclude other channels.

Shock-IV makes the same parallel trends assumption as DiD. This is part of assumption (ii) (as if random assignment), and is reflected in eqn. (6), in which the DiD coefficient forms the numerator for the shock-IV estimate. It estimates a causal effect only for "compliers" whose behavior is affected by the instrument (Angrist et al., 1996).

Given the uncertainty that often exists as to whether the only through condition is satisfied, we would be uncomfortable using shock-IV alone, but view it as both a valuable complement to simple DiD, and as important to study for our methods goal, given the frequent use of IV in finance and accounting research. Both DiD and IV require making the parallel trends assumption, and assessing the plausibility of this assumption in the pre-treatment period.

3.3.4 Regression Discontinuity

The RD model is available when an arbitrary threshold determines which firms are treated and which are control. Assume that a legal rule causes a predictor variable to change if a forcing variable (here, firm assets) exceeds a threshold level (here, 2T). Firms near the threshold should be similar on all variables, both observed and unobserved, except the forcing variable and related variables (here, other size variables). The below threshold firms form a control group for the above-threshold firms. See Imbens and Lemieux (2009) and Lee and Lemieux (2010) for reviews of RD designs. Let $x^{forcing}$ be the "forcing variable" and $\boldsymbol{x}^{other}$ be the other pre-treatment covariates. Near the discontinuity, assignment to treatment should be independent of everything except $x^{forcing}$ (and related variables, which we ignore for simplicity):

$$w_i \perp\!\!\!\perp \left(y_{0i}, y_{1i}, \boldsymbol{x}_i^{other}, \boldsymbol{u}_i\right)$$

One must still control for the direct effect of the forcing variable in predicting the outcome. A regression-based estimate, with a linear control for $x^{forcing}$ and no other covariates, is:

$$y_i = \alpha + \delta_{RD} \cdot w_i + \beta \cdot x_{i,b}^{forcing} + \varepsilon_i$$

Here $\hat{\delta}_{RD}$ is the estimated treatment effect. Judgment is needed on how flexibly to control for the forcing variable. For our setting, with ln(assets) as the forcing variable and a 2T threshold, a specification that allows different slopes above and below the threshold is:

$$\{RD\}\text{:}\ y_{it} = \alpha + g_t + \delta_{RD} \cdot w_i + \beta \cdot \ln\left(\frac{assets}{2T}\right) + \gamma \cdot w_i \cdot \ln\left(\frac{assets}{2T}\right) + \varepsilon_{it} \quad (7)$$

As for any shock-based design, the shock must be exogenous. For RD, this means that firms cannot manipulate which side of the threshold they fall on. We discuss above evidence on lack of manipulation. The shock should also be strong. Strength can be assessed graphically – the proportion of complier firms (for us, firms which comply with the large firm rules) should be visibly higher just above the threshold than just below it (see Figure 1).[9]

If some above-threshold firms do not comply with the treatment, some below-threshold firms voluntarily comply, or both, one has a "fuzzy" discontinuity. An above-threshold dummy can then be used as an instrument for actual treatment. One measures a local average treatment effect for firms who would comply if above the threshold, but not if below it. The usual IV conditions apply.

When the Korean reforms were adopted in 1999, the discontinuity at the 2T threshold was very close to being "sharp": Essentially no firms, large or not, complied with any of the large firm rules prior to 1999; yet all large firms had to do so by 2001. See Figure 1. Over time, some smaller firms voluntarily adopted some reforms. This creates a "fuzzy RD" setting: the probability of compliance jumps at the threshold, but not from 0 to 1. One can then use the discontinuity as an instrument for treatment (Angrist and Pischke, 2009, § 6.2). The regression model is again 2SLS, within the RD bandwidth. Covariates aside, we can express the fuzzy RD estimate as a Wald ratio:

$$\hat{\beta}_{fuzzy\ RD} = \frac{\hat{\delta}_{sharp\ RD}}{\hat{\beta}_{1S}} = \frac{effect\ of\ crossing\ threshold\ on\ y}{effect\ of\ crossing\ threshold\ on\ prob.(compliance)}$$

We present results using a relatively broad bandwidth of [0.5T, 8T] (a factor of four on either side of the 2T threshold) and a narrower [1T, 4T] bandwidth. The bandwidth choice reflects a

[9] The shock must also satisfy "only through" conditions: it must be isolated in time from other shocks that might affect the outcome, and – if we want to assign a causal effect to a forced variable, the shock must predict the outcome only through the forced variable.

compromise between treated and control firms becoming less similar as we expand the bandwidth, and the need to preserve sample size and this power.

3.4 Combined Causal Designs

The Korean setting allows us to also use combined causal designs. In particular, we can run simple DiD, annual DiD, or IV *within an RD bandwidth*. Our benchmark design is such a combined design (see section 3.1). A cost of the combined design is loss of sample size, and thus loss of statistical power, by limiting the bandwidth and thus the number of treated firms.

Combining RD with DiD has an important advantage. DiD makes a parallel trends assumption – that *changes* in outcome levels are as good as randomly assigned, but for treatment. RD alone makes a stronger assumption that outcome *levels* are as good as randomly assigned, but for treatment, conditioned on the running variable, but makes this assumption more plausible by limiting the bandwidth. DiD with an RD bandwidth uses the bandwidth limitation to make treated and control more similar, and allows one to rely on the weaker, more plausible assumption of parallel changes rather than similar levels.

3.5 Outcome Measures

We study whether the board structure shock affects other aspects of firm governance (proxied by firm scores on the Disclosure, Board Procedure, and Shareholder Rights subindices); Financial reporting (MD&A length; absolute and signed abnormal accruals), firm value (proxied by Tobin's q) and firm financial outcomes (ROA, book and market leverage, sales growth, capital expenditures, and R&D spending).[10]

3.5.1 Subindices of KCGI

We study three subindices of KCGI, for which a change in board structure could plausibly lead to board action that would change these indices as well: for Disclosure, Board Procedure, and Shareholder Rights. Disclosure Subindex consists of three elements: Firm conducted investor relations activity in last year; firm website includes resumes of board members; and firm provides English language financial disclosure. In panel regressions where Disclosure Subindex is the outcome, we remove it from KCGI. Board Procedure Subindex consists of 14 elements (for a full list see Table 1), examples include: board chairman is an outside director or (from 2003) firm has outside director as lead director; firm holds at least 4 board meeting

[10] MD&A length is, to be sure, a crude proxy for the quality of MD&A disclosure. It was not feasible to use textual analysis methods, because these have been developed for English language disclosure, and cannot be readily adapted to Korean language disclosure.

annually; system for evaluating directors exists. Shareholder Rights Subindex consists of the following four elements: Firm uses cumulative voting to elect directors; firm permits voting by mail; firm discloses director candidates to shareholders in advance of shareholder meeting; and board approval is required for related party transactions.

3.5.2 MD&A Word Count

We obtain annual reports for the firms in our sample from the Korean DART (Data Analysis, Retrieval, and Transfer) database, and count the number of words in the MD&A (management's discussion and analysis) section of these reports, to assess whether governance predicts more extensive MD&A disclosure. The dependent variable is *ln*(MD&A words).[11]

3.5.3 Tobin's *q*

Improved governance could lead to firms generating higher cash flows from the same assets, or could lead investors to assign higher value to the same cash flows. Either result could lead to an increase in Tobin's *q*, controlling for other factors that can affect Tobin's *q*, including firm size and growth. Black et al. (2015) report evidence that the Korean board structure shock leads investors to assign higher value to firms that were at higher risk of engaging in related party transactions that were adverse to firm value. We measure Tobin's *q* as year-end market value of assets/book value of assets, with market value of assets estimated as [book value of debt + book value of preferred stock + market value of common stock], and study ln(*q*) to reduce the influence of high-*q* outliers.[12]

3.5.4 Profitability and Leverage

We assess whether the board structure shock affected firm profitability, using ROA as our measure of profitability. We assess whether the shock affected firm decisions on leverage using as measures both book leverage (total debt/book value of assets) and, given evidence that the shock affected share values, market leverage (total debt/market value of assets).

[11] We study MD&A word count rather than a more complex measure such as sentiment, because the disclosures are in Korean, and we are not aware of reliable measures of sentiment for Korean language disclosure. We considered a number of other measures of financial reporting quality but do not study them because they are measured within each firm over a multiyear period, so are not compatible with a DiD research design. These include conditional conservatism, timeliness, value relevance, avoidance of small losses, and earnings persistence.

[12] While Tobin's *q* is a commonly used outcome in governance research, concerns have been raised about its use (e.g., Dybvig and Warachka (2015), Bartlett and Partnoy (2018)). One alternative is to study the numerator of Tobin's *q* (market value of assets). However, in our regressions we (mechanically) obtain identical results with ln(*q*) as the outcome, controlling for ln(assets) (the denominator of Tobin's *q*) or using ln(market value of assets) as the outcome, without this control.

3.5.5 Sales Growth and Investment

In the period leading up to the East Asian financial crisis, many Korean *chaebol* groups emphasized growth over profitability (e.g., Campbell and Keys, 2002). Many *chaebol* firms had Tobin's q values well below 1, implying that a dollar of invested capital was producing less than a dollar of market value on average. The board structure reforms could potentially lead to a greater focus on firm value, which might lead to lower growth and investment. We test this possibility using sales growth, capital expenditures, and R&D spending as outcomes. We measure sales growth as $\left[\left(revenue_{i,t} - revenue_{i,t-1}\right)/revenue_{i,t}\right]$, capex as $\left[\left(capex_{i,t}/assets_{i,t-1}\right)\cdot 100\right]$ and R&D as $\left(R\&D_{i,t}/sales_{i,t}\right)$.

3.5.6 Signed and Absolute Abnormal Accruals

Accruals are accounting adjustments that turn cash flow into earnings. Variation in accruals across industries should reflect industry specific conditions. In contrast, variation across firms within an industry is likely to reflect a mix of the firm's specific circumstances and managerial discretion. A large accounting literature uses firm-level accruals as a proxy for earnings management and treats abnormal accruals (measured relative to other firms in the same industry) as reflecting management discretion – indeed these are often called "discretionary accruals." This literature generally treats lower abs(AA) as evidence of lower earnings management and thus improved disclosure.

A core problem in using accruals as a proxy for earnings management is estimating normal, "unmanaged" accruals. This is a question on which theory provides little guidance. The central approach is to develop a model of "normal" accruals, based on the accruals reported by other firms in the same industry. A number of models of normal accruals have been developed. We use the model in eqns. (8)-(9), which is closely related to what can be called a Jones-Dechow-Larcker model, based on Jones (1991); extended by Dechow et al. (1995) and Larcker et al. (2007), but confirm in unreported results that we obtain similar results with the Jones or Dechow models. We first regress accruals within each industry-year on covariates that can predict accruals (eqn. (8)). We use these regression coefficients to predict each firm's "normal" accruals (eqn. (9)). The remaining accruals are termed "abnormal" and reflect a combination of firm-specific circumstances and earnings management (eqn. (10)).[13] In these equations, ppe is property, plant and equipment; btm is book-to-market ratio, and cfo is cash flow from operations.

[13] In eqn. (8), the terms in square brackets come from the Jones-model, as extended by Dechow et al. (1995); the controls for book-to-market ratio and operating cash flow/assets come from Larcker et al. (2007). We investigated

$$\frac{accruals_{i,t}}{assets_{i,t-1}} = \left[\alpha_i + \beta_{i1}\left(\frac{1}{assets_{i,t-1}}\right) + \beta_{i2}\left(\frac{\Delta sales_{i,t}}{assets_{i,t-1}}\right) + \beta_{i3}\left(\frac{ppe_{i,t}}{assets_{i,t-1}}\right)\right] + \beta_{i4} \cdot btm_{i,t} + \beta_{i5}\left(\frac{cfo_{i,t}}{assets_{i,t-1}}\right) + \beta_{i6}\left(\frac{cfo_{i,t}}{assets_{i,t-1}}\right)^2 + \varepsilon_{i,t} \quad (8)$$

We use the estimated parameters from eqn. (8) to predict normal accruals for each firm:

$$normal\ accruals(na)_{i,t} = \left[\hat{\alpha}_i + \hat{\beta}_{i1}\left(\frac{1}{assets_{i,t-1}}\right) + \hat{\beta}_{i2}\left(\frac{\Delta sales_{i,t} - \Delta receivables_{i,t}}{assets_{i,t-1}}\right) + \hat{\beta}_{i3}\left(\frac{ppe_{i,t}}{assets_{i,t-1}}\right)\right] + \hat{\beta}_{i4} \cdot btm_{i,t} + \hat{\beta}_{i5}\left(\frac{cfo_{i,t}}{assets_{i,t-1}}\right) + \hat{\beta}_{i6}\left(\frac{cfo_{i,t}}{assets_{i,t-1}}\right)^2 \quad (9)$$

For each model, signed abnormal accruals are:

$$abnormal\ accruals(aa)_{i,t} = \frac{accruals_{i,t}}{assets_{i,t-1}} - normal\ accruals(na)_{i,t} \quad (10)$$

Absolute Abnormal Accruals are the absolute value of the signed abnormal accruals.

4 Overview and Classic Panel Data Results

4.1 Overview of Results Across Methods

In the remainder of this article, we develop results first with classic panel methods (this section), then with simpler causal methods (section 5), and then combined causal designs, including our benchmark annual DiD/RD design (section 6). Table 5 provides an overview of which results, for which outcomes, are statistically significant (5% level, 2-sided test).

For our substantive goal – to assess the effects of the Korean board structure shock: we find consistent evidence across most designs, including the benchmark design, that this shock leads to higher scores for Disclosure Subindex. We also find reasonably strong evidence that the board structure shock predicts higher Tobin's q, and some evidence that it leads to lower sales growth, but no evidence of significant change in our other outcomes.

For the methods goal, we obtain scattered results for different outcomes, with no clear patterns. For causal designs, the scattered nature of those results is the takeaway. Causal designs can generate apparent false positives that would not be convincing if one had the benchmark design available as a stronger guide to truth. Firm FE can also generate apparent false positives, beyond those one might find with a causal design – MD&A word count and

---

whether adjusted R^2 in predicting normal accruals would improve if we include squares or interactions of the terms in this model; this led us to add (cash flow/lagged assets)^2 to the model. Results are similar with the Jones-Dechow and Jones-Dechow-Larcker models. We estimate eqn. (8) separately for each firm, excluding that firm from its industry-year group. We use two-digit Korean industry classification codes for industries other than manufacturing and four-digit codes for manufacturing. We require a minimum of 8 firms per industry-year, but obtain similar results with a 10-firm minimum.

abs(AA) provide examples. And firm RE generates some apparent false positives that are not confirmed with firm FE or causal methods. While panel data designs could, in theory, also generate apparent false negatives, we do not find that pattern in Table 5.

4.2 Firm FE and RE Results

In this section, we present results with classic panel methods for our outcome variables. A high-level overview focusing on firm FE results (Table 6): KCGI predicts significantly higher scores on the Disclosure and Board Procedure subindices, higher MD&A word count, higher firm value (proxied by Tobin's q), lower sales growth, and lower abs(AA).[14] Thus, it would be easy for a researcher, using only classic panel methods, to report that either KCGI as a whole, or Board Structure Subindex, predicts a number of outcomes which suggest that governance improves financial reporting (measured by Disclosure Subindex, MD&A word count, and abs(AA)), and also affects board procedure, firm value, and growth. Yet, most of those results fall away with causal methods, especially with our benchmark design and with the annual DiD specification.

In Table 6, Panel A. We present firm RE and FE results with ln(q) as an outcome, including results for covariates. We present similar regressions for the other outcomes in Panel B, but in this and later tables, suppress results for covariates to save space. All regressions include industry × year FE. Regressions (1) and (2) provide firm RE and FE results for KCGI with limited covariates – only *ln*(assets), and *ln*(years listed). Regressions (3) and (4) are similar but include full covariates. We provide both sets of results for two reasons. First, doing so provides readers with a sense for the sensitivity of results to the included covariates, and thus to potential omitted variable bias due to omitted covariates. Second, if some covariates are outcomes of the treatment, including them could lead to biased estimates of the treatment effect.

In Panel A, KCGI strongly predicts *ln*(Tobin's q); coefficients are somewhat larger with limited covariates and with RE. With firm FE and full covariates, *KCGI* takes a coefficient of 0.0030 (t = 3.37). Note that for Tobin's q, unlike our other outcomes, we do not lag *KCGI* because investors are forward looking, so the effect of a governance change should affect share prices immediately.[15]

---

[14] In panel regressions where we use KCGI as a whole to predict scores on a subindex, we remove the subindex from KCGI.

[15] One might be concerned about controlling for ln(assets) with ln(q) as the outcome, given the mechanical relationship between *ln*(q) and *ln*(assets): *ln*(q) = ln(market value of assets) – *ln*(assets). We believe that controlling for ln(assets) is nonetheless appropriate. The theory behind using Tobin's q as a measure of value is that well-governed firms can recover more value from the same assets. From that perspective, there is no necessary association between Tobin's q and firm size. As a robustness check, we ran our Tobin's q regressions

In Panel B, we also examine linear (instead of logged) q. With that specification, *KCGI* remains a strong predictor with firm RE, but weakens with firm FE. However, the firm FE coefficient is significant in a number of other specifications, including with year FE (instead of industry × year FE), using industry × year FE with finer 4-digit industries, or if we exclude outliers instead of winsorizing them.[16] Also, in causal specifications, results for ln(q) and linear q are similar and generally significant (see Tables 7-9).

Lagged *KCGI* strongly predicts Disclosure Subindex across all specifications. Lagged KCGI also predicts Board Procedure Subindex across specifications, but much more strongly with RE than with FE. Among other financial reporting outcomes, lagged KCGI predicts higher MD&A word count and lower abs(AA) but no change in signed AA. The finding that governance predicts lower abs(AA) is consistent with most, but not all, prior research on the impact of governance on accruals. For example, Klein (2002) finds that greater board and audit committee independence predict lower abs(AA) for S&P 500 firms over 1992 and 1993; Vafaes and Theodorou (1998) and Weir et al. (2001) find similar results for UK firms. Larcker et al. (2007) find more mixed results. However, below, this result will not survive with causal methods.

Among other financial outcomes, lagged KCGI predicts lower sales growth across specifications. Other outcomes, for book leverage, market leverage, capital expenditures and R&D are mostly insignificant with firm FE and full covariates.

5 Simpler Causal Methods

We turn next to simpler causal methods, and assess which outcomes have support, with which methods.

5.1 Simple DiD

Our first approach is simple DiD, using eqn. (4). The treatment group is large firms (assets > 2T); the control group is smaller firms (assets < 2T). The treatment period starts in 1999 for Tobin's q and in 2001 for other outcomes. We measure size at year-end 1999 for Tobin's q and year-end 2000 for other outcomes. We present results in columns (1) and (4) of Table 7. Col. (1) presents results with limited covariates (ln(assets) and ln(years listed)); Col. (4) presents

---

with and without controlling for *ln*(assets). Results for the predictor variable (KCGI in panel regressions, large firm dummy × post dumm in DiD regressions) barely changed.

[16] With year FE, coeff. = 0.0025 (t = 3.44); with 4-digit industry × year FE, coeff. = 0.0023 (t = 2.58); with excluding outliers [absolute values of studentized residuals from yearly regressions of linear q on KCGI greater than 1.96], coeff. = 0.0027 (t = 4.56).

results with full covariates. We discuss below the remaining columns in Table 7, which provide results from the combined DiD/RD approach. The DiD coefficients in Table 7 cannot be compared to those from Table 6 because the principal independent variable has changed from KCGI in Table 6 to post × large firm dummy in Table 7.

The results for Tobin's q (logged or not), Disclosure Subindex, and Board Procedure Subindex remain strong, with both limited and full covariates. With full covariates, post dummy predicts a 0.14 increase in ln(Tobin's q), a 4.5 point increase in Disclosure Subindex, and a 1.1 point increase in Board Procedure Subindex. The increase in Disclosure Subindex is large compared to the pre-shock mean for large firms of 6.1 (broad bandwidth). We also find, similar to Table 6, a significant increase in MD&A word count, and a significant drop in sales growth for large firms relative to small firms. However, a number of these results will not be robust across causal methods.

Several other results that we found with firm FE disappear with the simple DiD specification: for Shareholder Rights Subindex, the sign flips to negative (opposite from predicted; marginally significant); for abs(AA) the coefficient is small, insignificant, and positive (opposite from predicted). Conversely, market leverage strengthens; the coefficient is negative and significant. This is driven by an increase in share value, since the coefficient on book leverage is positive (and, with limited covariates, significant).

5.2 Evidence on Covariate Balance

A concern for any shock-based research design is how close the sample comes to "as good as random" assignment to treatment. For an RD design, there will necessarily be imbalance on the "forcing variable" and any related variables (for us, ln(assets) and other size-related variables), but one can hope for balance on other covariates. If the sample is balanced on pre-shock covariates and outcomes, this makes the parallel trends assumption more plausible and thus makes causal inference more credible. Table 4 provides two measures of covariate balance: a t-test for difference in means, and the sample-size-independent normalized difference (ND) measure suggested by Imbens and Rubin (2015): Let the covariates be x_k, their means be $\bar{x}_{kt}$ for treated and $\bar{x}_{kc}$ for controls, their standard deviations be σ_{kt} and σ_{kc}. The ND for x_k is:

$$ND_k = \frac{\bar{x}_{kt} - \bar{x}_{kc}}{\sqrt{\frac{\sigma_{kt}^2 + \sigma_{kc}^2}{2}}}$$

The left-hand part of Table 4 provides data for the full sample; the right-hand side provides data for firms within our broader, [0.5T, 8T] RD bandwidth.

As expected, Table 4 shows large differences in size variables such as *ln*(assets) and ln(market value of assets), and size-related variables, such as the MD&A disclosure length. These differences shrink when we restrict the sample to firms within the RD bandwidth. Apart from *ln*(assets) and *ln*(market value of assets), the only significant differences in outcomes are for Board Procedure and Shareholder Rights subindices, and the only significant differences in time-varying covariates are ownership-related variables: sole ownership (smaller for large firms), foreign ownership (larger for large firms), and chaebol membership (higher for larger firms). Still, even within the RD bandwidth, some normalized differences are often substantial. The pre-treatment differences in outcomes may be acceptable for inference from a combined DiD/RD design, where some degree of difference in pre-treatment *levels* between treated and control firms can be acceptable (the DiD design allows for differences in levels but assumes, in effect, that *changes* in potential outcomes (but for treatment) are as good as randomly assigned). But the differences in levels are problematic for a direct comparison of treated and control firms using RD alone, or post-shock IV alone.

5.3 Pre-Treatment Trends and Annual DiD

We assess whether the parallel trends assumption appears to hold during the pre-treatment period in two ways. First, in Appendix Figure 4, we plotted annual means for treated and control firms for each outcome, generally over 1994-2005. We did so for three different samples: the full sample (left-hand graphs), the broad [0.5T, 8T] RD bandwidth (middle graphs), and the narrow [1T, 4T] bandwidth (right graphs). Second, in Figure 2, we plotted annual DiD coefficients following eqn. (5) over the sample period for our main outcomes (see Table 10 for sample periods), for the same three samples. We plotted semiannual results for Tobin's q, because we want to capture more accurately the effect on share prices of the board structure reform, which takes place in the second half of 1999 (2H 1999). We pegged year 2000 values (May 1999 for Tobin's *q*) to zero. These graphs provide evidence on pre-treatment parallel trends and on whether there appears to be a treatment effect, using the annual DiD approach. Inference on pre-treatment from the annual means reported in the Appendix is similar.[17]

Parallel pre-treatment trends imply that the coefficients for 1998-2000 (1988 – 1H 1999 for Tobin's *q*) should be close to zero and show no time trend. One can also use the annual DiD graphs to assess whether any large-minus-small-firm difference appears when it should, relative to the board structure shock: in 2H 1999 for Tobin's q; no earlier than 2000 or 2001

[17] For space reasons, for selected outcomes, we report the annual DiD results only in the Appendix: linear q (results similar to ln(q)); market leverage (we report graphs for book leverage in the text); and signed AA (we report results for abs(AA) in the text).

for other outcomes. We discuss the full-sample results here, and the within bandwidth results in section 6.

For both ln(Tobin's q) and linear q, we see a rise in 2H 1999, the first post-shock period, which persists after that. This is evidence for a causal effect of the reform. Tobin's q also rises *when* it should, if Korean firms trade in a reasonably efficient market and the rise is caused by the board structure reforms. Black and Kim (2012) provide further evidence that share prices of large firm rise during the specific periods in 2H 1999 when the large firm reforms were being considered by Korea lawmakers.

For Disclosure Subindex, pre-treatment trends are reasonably parallel over 1998-2000, and then jump in 2001 and continue to rise steadily through the end of the sample period. This supports a causal effect of the shock on Disclosure Subindex. Disclosure Subindex rises substantially, and does so *when* it should, if the rise were caused by the legal reforms. The growing effect over time is consistent with the board structure reforms affecting disclosure gradually.

For Shareholder Rights Subindex, in contrast, the treated and control means are reasonably parallel and perhaps slightly declining throughout the sample period, with no evidence for a treatment effect.

Relative Board Procedure Subindex rises steadily from 1998-2002. This rise explains the positive coefficient using simple DiD. However, the annual graphs make it apparent that the higher average level during the treatment period could reflect continuation of a pre-treatment trend, rather than a causal effect of the board structure shock.

For MD&A word count, there is a rise in 2000, which is possible but early for a true treatment effect, and then a further rise in 2004 and 2005, which seems late, relative to the reforms, to be a true treatment effect. This late increase in MD&A disclosures by large firms in 2004 drive the significant coefficient in the simple DiD specification in Table 7. Yet it seems unlikely that this delayed effect is caused by the board structure shock. The ability to assess whether an apparent treatment effect occurs when it plausibly should, relative to the shock, is a great advantage of annual DiD regressions over the simple DiD results in Table 6.

For ROA, relative (large-minus-small-firm) ROA falls sharply from 1998-1999, which could be an effect of the East Asian crisis, is fairly flat over 1999-2001, and then rises starting in 2002. If we ignore the drop in 1999, this timing is consistent with a gradual effect of the board structure reforms on profitability. In Table 5, where we summarize results across methods, we code this as "some" evidence for a treatment effect.

For book leverage, there is a large relative jump in large firm leverage in 1999, which is too early to be a treatment effect, parallel trends after that, and thus no evidence of a treatment effect.

For sales growth, pre-treatment trends are reasonably parallel. Large firm growth falls relative to small firms in 2002, which is a plausible time for a treatment effect.

Capex/assets also shows non-parallel pre-treatment trends. Capex/assets for large firms, falls relative to small firms over 1998-2002, then rises in 2003-2004. This timing is not consistent with the board structure shock leading to lower post-shock investment.

For R&D/sales, the drop for large firms in 1999 is too early to be a treatment effect; and trends during the period from 1999 on are largely parallel, consistent with no treatment effect.

For abs(AA), there are annual fluctuations but no apparent relative trend, in either the pre-treatment or the treatment period.

Consider the combined evidence from the simple DiD and annual DiD regressions, there is evidence for a causal effect of the board structure shock on Disclosure Subindex, Tobin's *q*, and sales growth, but no consistent evidence of a treatment effect for the other outcomes. The differences between the two support one major methods finding: simple DiD appears prone to apparent false positives. For robust inference, it should be presented together with annual DiD, which can identify some of those apparent false positives.

5.4 Shock-IV

We turn next to shock-IV, applied to our panel dataset, over the full data period from 1998-2005. The IV is the interaction of post-reform dummy (=1 for 1999 on with Tobin's q as the outcome; for 2001 on for other outcomes) and large firm dummy (measured year-by-year). Table 8 reports results. The format of Table 8 is similar to Table 7; with results using both limited and full covariates, and results for the full sample, and broad and narrow-bandwidth subsamples. There is a choice to be made as to what the instrumented variable should be: Is the legal shock, which directly affects board structure, better seen as affecting only Board Structure Subindex or as affecting all of KCGI? We choose the second approach, which seems preferable given the evidence reported above that the board structure shock also affects Disclosure Subindex, which is part of KCGI.[18]

We report both first-stage and second-stage results. We report the first stage separately for Tobin's *q*, for which the IV turns on in 1999 because we do not expect a lag between the effect

[18] When we use a subindex of KCGI as the outcome variable, we remove that subindex from KCGI, similar to our approach in the panel regressions above.

of the board structure shock on share prices; and for ROA (representative of other outcomes), for which the IV turns on in 2001 because we expect a lag. The first stage is strong in all cases. The first stage is larger for outcomes other than Tobin's *q* because the time when the IV turns on matches more closely when *KCGI* changes as a result of the board structure shock.

We discuss here the full sample results in columns (1) and (4); we discuss below the within-bandwidth results. Simple DiD and shock-IV are similar specifications. If the instrumented variable is binary, then the DiD coefficients can be understood as an "intent-to-treat" effect, while the IV coefficients can be understood as providing a "local average treatment effect" for "complier" firms, whose behavior was changed by the instrument (Atanasov and Black, 2016, 2020). The correspondence is less close for a continuous instrumented variable, but we should still expect DiD and IV to have similar statistical strength. That expectation broadly holds in this case, although not for all outcomes. In the second stage, the instrumented variable strongly predicts *ln*(Tobin's *q*), Disclosure Subindex, Board Procedure Subindex, and lower sales growth, and less strongly predicts higher MD&A word count and higher ROA. Other outcomes are insignificant. Note, however, that a number of coefficients on instrumented KCGI are much larger in magnitude than the corresponding panel data coefficients from Table 6. This is a warning for potential violation of the only-through condition (Jiang, 2017). Such a violation could arise because the IV correlates with firm size, and *ln*(assets) is an imperfect control for the true nonlinear association between firm size and Tobin's *q*.

5.5 Post-Shock RD

We next turn to post-shock RD, as another available causal design in our setting. We present RD results, using a sample that is pooled over the post-treatment period (1999-2004 for Tobin's q and 2001-2004 for other outcomes), in Table 9. Regressions (1)-(3) use the broad [0.5T, 8T] bandwidth. Regression (1) does not control for size within the bandwidth; regression (2) includes *ln*(assets) as a size control; regression (3) allows for different slopes on ln(asses) below and above the 2T threshold. Regressions (4)-(6) are similar but use the narrower [1T, 4T] bandwidth.

Consider first the broader bandwidth, without a size control. We find significant increases in Tobin's q (linear or logged), all three subindices, MD&A word count, and book leverage. However, when we add a control for ln(assets), the only coefficient that remains significant is for board procedure. Yet we saw above that the post-reform increase in Board Procedure Index for large firms could merely reflect continuation of a non-parallel pre-treatment trend. Conversely, the increases in Tobin's *q* and Disclosure Subindex, which appears causal in the leads-and-lags graphs in Figure 2, take smaller, insignificant coefficients. One possible reason is that the linear control for *ln*(assets) competes with large firm dummy as a predictor of these outcomes, and can potentially absorb a true treatment effect. This concern is more acute with a small sample, and we have a limited number of large firms; and also more acute for the narrow bandwidth, because the within-bandwidth correlation between the running variable (ln(assets)) and the above-threshold dummy will be higher.

The DiD framework, because it includes firm FE and before and after time periods, lets us assess more effectively whether the legal shock, versus a general effect of firm size, predicts these outcomes. We therefore place limited weight on the mostly insignificant RD results with this control, and greater weight on other approaches, especially DiD within the RD bandwidth. In effect, we have a setting where, due to the limited number of large firms, RD alone has insufficient power to detect causal effects, which based on the evidence from other methods, likely exist.

5.6 Overview: Causal versus Panel Designs; Multiple Causal Designs

With the partial exception of RD, across our simpler causal methods, we find evidence for an effect of the legal shock on Disclosure Subindex, Tobin's q, and sales growth, but either weaker or no evidence for other outcomes.

One core theme from comparing panel methods to simple causal methods is the fragility of the panel results. Results that appear strong with panel data can vanish if we apply causal methods to the same data. The panel data results for abs(AA) and MD&A word count provide examples. To be sure, different methods make somewhat different assumptions, so it is possible that the panel data results could reflect a true effect. But given both sets of results, one would not find the panel result to be convincing. The converse is also possible – that a true causal effect exists, which would be found with causal methods but not with panel methods. We see a hint of this possibility for capex/assets, which is insignificant with panel methods, but takes a significant negative coefficient with simple DiD. However, this apparent result is not present with the annual DiD design, which we view as the most robust. On the whole, results that are insignificant with panel methods remain so with causal methods. We thus provide case study evidence that panel data methods can provide apparent false positives, but are less likely to provide apparent false negatives.

A further theme of this article involves the differences in results across causal methods, and thus the value of using multiple shock-based designs to exploit the same shock. In particular, the annual DiD design best exploits the time dimension of the data. The leads-and-lags graphs in Figure 2 let us assess: (i) the existence or absence of parallel trends during the pre-treatment period; (ii) whether post-shock differences appear *when* (relative to the shock) they should, if they are genuinely caused by the shock; and (iii) whether any apparent causal effect persist during the post-shock period.

6 Combined Causal Methods

In this part, we combine RD with other causal methods. We limit the sample to a bandwidth around the 2T threshold, and then apply simple DiD, IV, and finally our benchmark design, annual DiD regressions, within this bandwidth.

Limiting the sample to a bandwidth around the 2T threshold improves covariate balance and thus enhances credibility, at the cost of reduced sample size. Within the RD bandwidth, treated and control firms will differ on the RD running variable (in this study, assets) and related variables (here, such as market capitalization). But one can expect reasonable balance on other covariates. We indeed find improved balance within the RD bandwidth sample, as compared to the full sample (see Table 4).

The choice of RD bandwidth involves trading off improved balance from a narrower bandwidth versus loss of sample size as one restricts the bandwidth. The latter concern is acute for our setting, because of the limited number of large firms. We present results with two bandwidths: a broader [0.5T, 8T] bandwidth and a narrower [1T, 4T] bandwidth. Due to the concern with sample size, we prefer the results with the broader bandwidth and with limited covariates (ln(assets) and ln(years listed)). We note that use of RD combined with other causal designs lends additional credibility to results with a broader bandwidth than would be the case for RD used alone.

6.1 Combined Simple DiD/RD

In Table 7, columns (1) and (4), discussed above, present simple DiD results with the full sample and limited or full covariates, respectively. Columns (2) and (5) present broad bandwidth results with limited and full covariates, respectively, and columns (3) and (6) present narrow bandwidth results. Consider first Tobin's q. With limited covariates, the coefficient on the treatment dummy is significant and economically meaningful with both bandwidths, albeit smaller than for the full sample. The coefficient is also significant with the broad bandwidth and full covariates, but loses significance in the most restrictive specification, with narrow bandwidth and full covariates. Results for logged and non-logged q are similar. Overall, we still find reasonably strong evidence for a post-shock jump in Tobin's q.

Consider next Disclosure Subindex. Here too, we retain reasonable strong evidence for a treatment effect. The coefficient on the treatment dummy is smaller than with the full sample, but remains significant with either limited covariates or the broad bandwidth, and marginally significant in the most restrictive specification.

Sales growth takes a marginally significant and negative coefficient with the broader bandwidth, in the predicted direction, with both limited and full covariates, but loses significance with the narrower bandwidth. Board Procedure Subindex retains significance with the broad bandwidth and limited covariates, but becomes insignificant using the narrow bandwidth. Moreover, as noted above, the broad bandwidth results could reflect a continuation of non-parallel pre-treatment trends. All other outcomes become insignificant with either bandwidth.

6.2 IV/RD

In Table 8, we present shock-IV results, in a format similar to Table 7. We discuss the full sample results above, and focus here on the within bandwidth results.

With the IV specification, the results for Tobin's *q* weaken and become only marginally significant with the broad bandwidth and limited covariates, but insignificant with the narrow bandwidth or with full covariates. In contrast, Disclosure Subindex remains strong; the coefficient on treatment dummy is significant in all specifications. Board Procedure Subindex retains significance except in the most restrictive specification (narrow bandwidth, full covariates). For sales growth, we find significance with the broad bandwidth and full covariates, but not with limited covariates. There is no consistent evidence of a treatment effect for the other outcomes, indeed MD&A word count takes a significant negative coefficient (opposite from predicted) with the narrow bandwidth and limited covariates, albeit barely so.

6.3 Benchmark Design: Annual DiD /RD

The last combined research design we examine is our benchmark annual DiD /RD design, for which we provide graphs in Figure 2 with limited covariates. Figure 2 presents results for both the broad and narrow bandwidths; we discuss primarily the broad bandwidth results.[19] Disclosure Subindex remains significant with both bandwidths. We continue to have reasonable evidence for parallel pre-treatment trends, albeit for a limited pre-treatment period. The large-firm rules came into force partly in 2000 and partly in 2001, so 2000 can be seen as a partial-treatment year, with 2001 as the first full treatment year.

For Tobin's *q*, the full-sample results, shown in the left-hand figure, weaken somewhat with broad bandwidth, and weaken further with the narrow bandwidth. There is still evidence of a jump in Tobin's *q* in 2H 1999 (strong for logged than for non-logged *q*) but then a gradual falloff

[19] Figures with full covariates are included in the Appendix. Given the limited number of large firms and the further loss of power from measuring effects year-by-year, without averaging over the post-shock period, our judgment is that the power from the narrower bandwidth, or from using full covariates (which should not be important give that we are already within the RD bandwidth) is more likely to suppress a true effect due to limited power than to convey truth.

after that. This somewhat reduces our confidence in the existence of a true causal effect of the board structure reforms.

For sales growth, the drop in 2002, which is a plausible time for a treatment effect, remains strong with both bandwidths, but with some evidence for a rebound with the narrow bandwidth.

For the other outcomes, there is no substantial evidence for a causal effect. For Board Procedure Subindex, the full-sample and broad bandwidth results could reflect continuation of pre-treatment trends, and the evidence for any increase in this subindex weakens substantially with the narrow bandwidth.

6.4 Overall Assessment of Results from Combined Research Designs

On the whole, we retain reasonable confidence that there is a true causal effect of the board structure shock on Disclosure Subindex, and some confidence for Tobin's q and sales growth. For the other outcomes, there is no overall evidence of a causal effect of the shock.

7 Discussion and Conclusion

We exploit here a major exogenous shock to the board structure of large Korean firms. This shock is economically important – share prices of large firms rise by around 30% versus mid-sized firms during the period when the legal reforms take place (Black and Kim, 2012). Using a strong benchmark research design (annual DiD regressions within an RD bandwidth), we report reasonably strong evidence that the board structure shock leads to improved scores on a Disclosure Subindex, and milder evidence that it leads to higher Tobin's q and lower sales growth. We find no convincing evidence for an effect of the board structure shock on an array of other outcomes.

For our second, methods goal, we provide evidence on the reliability of panel methods and simpler causal methods, as a guide to causation. An association between governance and an outcome, such as Tobin's q or disclosure, with firm FE and extensive covariates, is evidence consistent with causation, but not proof. But how strong is that evidence? A statistically significant result, obtained with causal methods, such as DiD and shock-IV, is also evidence of causation, and is often considered to be stronger evidence than is available from classic panel designs. But again, how strong?

We provide case-study evidence on those questions here, by comparing results from panel methods and simpler causal methods to those from our benchmark design. With classic panel methods, a number of additional outcomes appear significant, but evidence for a treatment effect fades with the simpler causal designs, and is not present with the benchmark design. We thus provide case study evidence for two propositions: First, classic panel methods can be prone to apparent false positives, but are less prone to apparent false negatives. This suggests that for governance research, even careful panel designs will often provide only mild evidence on causation.

Second, causal designs can themselves be unreliable. Within the simpler causal designs, we find significant or marginally significant results for some outcomes with either simple DiD or IV designs, that lose significance with the annual DiD/leads-and-lags design and with combined designs. This suggests the value of exploiting multiple designs, including combined designs where available, as well as the limits on inference from any single causal research design.

To be sure, that panel data results weaken with stronger causal designs does not mean a causal result is not there in fact. Stronger designs can be statistically weaker; especially RD because the limited bandwidth reduces sample size. We do not argue against use of panel designs when they are the best available, as is often the case. We believe instead that results with classic panel methods should be interpreted cautiously – and that researchers should apply several research designs, including causal designs, when feasible.

The advice for researchers from the comparison of results across designs includes: (i) interpret panel data results, including FE results, with caution; (ii) when a shock-based design is available, seek to apply multiple causal designs using the same shock, including combined methods when available (such as DiD/RD); and (iii) examine period-by-period results, not only after-versus-before differences (as in simple DiD) or above versus below threshold differences (as in RD). While it is common for editors and referees to ask authors to run various robustness checks, it is less common for researchers to report, or for editors and referees to request, results using a variety of research designs. The extent to which our results generalize beyond Korea and this particular shock, or beyond corporate governance research, is, of course, unknown. A promising avenue for future accounting and finance research is to exploit other shocks, especially shocks that permit a combined DiD/RD design, compare results across causal and non-causal research designs, and report on what differences emerge.

References

Angrist, J. D., Pischke, J.-S., 2009. *Mostly Harmless Econometrics*, Princeton University Press.

Angrist, J. D., Pischke, J.-S., 2015. *Mastering 'Metrics: The Path from Cause to Effect,* Princeton University Press.

Angrist, J. D., Imbens, G. W., Rubin, D. B., 1996. Identification of Causal Effects Using Instrumental Variables, *Journal of the American Statistical Association* 91, 444-455.

Angrist, J. D., Rokkanen, M., 2015. Wanna Get Away? Regression Discontinuity Estimation of Exam School Effects Away from the Cutoff, *Journal of the American Statistical Association* 110, 133-1344.

Atanasov, V., Black, B., 2016. Shock-Based Causal Inference in Corporate Finance Research, *Critical Finance Review* 5, 207-304.

Atanasov, V., Black, B., 2020. The Trouble with Instruments: The Need for Pre-Treatment Balance in Shock- IV Designs, *Management Science*, forthcoming.

Bartlett, R. P., Partnoy, F., 2018. The Misuse of Tobin's Q. Working Paper, UC Berkeley Public Law Research Paper, https://ssrn.com/abstract=3118020 or http://dx.doi.org/10.2139/ssrn.3118020.

Bebchuk, L., Cohen, A., Ferrell, A., 2009. What Matters In Corporate Governance, *Review of Financial Studies* 22, 783-827.

Berk, R., Barnes, G., Ahlman, L., Kurtz, E., 2010. When Second Best is Good Enough: A Comparison Between a True Experiment and a Regression Discontinuity Quasi-Experiment, *Journal of Experimental Criminology* 6, 191-208.

Bhargava, A., Sargan, J. D., 1983. Estimating dynamic random effects models from panel data covering short time periods, *Econometrica* 51, 1635–1660.

Black, B., Desai, H., Litvak, K., Yoo W., Yu J. J., 2020. Specification Choice in Randomized and Natural Experiments: Lessons from the Regulation SHO Experiment, Working paper, http://ssrn.com/abstract=3657196,

Black, B., Jang, H., Kim, W., 2006. Predicting firms' governance choices: Evidence from Korea, *Journal of Corporate Finance* 12, 660-691.

Black, B., Kim, W., 2012. The Effect of Board Structure on Firm Value: A Multiple Identification Strategies Approach Using Korean Data, *Journal of Financial Economics* 104, 203-226.

Black, B., Kim, W., Jang, H., Park, K.-S., 2015. Why Does Corporate Governance Affect Firm Value: Evidence on a Self-Dealing Channel from a Natural Experiment in Korea, *Journal of Banking and Finance* 51, 131-150.

Black, D., Galdo J., Smith, J., 2007. Evaluating the Regression Discontinuity Design Using Experimental Data, Working Paper, http://economics.uwo.ca/newsletter/misc/2009/smith_mar25.pdf.

Buddelmeyer, H., Soufias, E., 2004. An Evaluation of the Performance of Regression Discontinuity Design on PROGRESA, World Bank Policy Research Paper 3386, http://ssrn.com/abstract=434600.

Campbell, T., Keys P., 2002. Corporate Governance in South Korea: the chaebol experience, *Journal of Corporate Finance* 8, 373-391.

Cattaneo, M. D., Idrobo, N., Titiunik, R., 2019. *A Practical Introduction to Regression Discontinuity Designs: Foundations*, Cambridge Univ. Press.

Cook, T. D., Shadish, W. R., Wong, V. C., 2008. Three Conditions under Which Experiments and Observational Studies Produce Comparable Causal Estimates: New Findings from Within-Study Comparisons, *Journal of Policy Analysis and Management* 27, 724-750.

Cook, T. D., Wong, W. R., 2008. Empirical Tests of the Validity of the Regression Discontinuity Design, *Annales d'Economie et de Statistique* 91/92, 127-150.

Dechow, P., Sloan, R., Sweeney, A., 1995. Detecting earnings management, *The Accounting Review* 70, 193–225.

Dybvig, P. H., Warachka, M., 2015. Tobin's q Does Not Measure Firm Performance: Theory, Empirics, and Alternatives. Working Paper, https://ssrn.com/abstract=1562444 or http://dx.doi.org/10.2139/ssrn.1562444.

Gleason, O. P. M., Resch, A. M., Berk, J. A., 2012. Replicating Experimental Impact Estimates Using a Regression Discontinuity Approach, U.S. Department of Education Pub. NCEE 2012-4025.

Gompers, P., Ishii, J., Metrick A., 2003. Corporate Governance and Equity Prices, *Quarterly Journal of Economics* 118, 107-155.

Green, D. P., Leong, T. Y., Kern, H. L., Gerber, A. S., Larimer, C. W., 2009. Testing the Accuracy of Regression Discontinuity Analysis Using Experimental Benchmarks, *Political Analysis* 17, 400-417.

Holland, P., 1986. Statistics and Causal Inference, *Journal of the American Statistical Association* 81, 945-960.

Imbens, G. W., 2004. Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review, *Review of Economics and Statistics* 86, 4-29.

Imbens, G. W., Lemieux, T., 2009. Regression Discontinuity Designs: A Guide to Practice, *Journal of Econometrics* 142, 615-635.

Imbens, G. W., Rubin, D. B., 2015. *An Introduction to Causal Inference in Statistics, Social and Biomedical Sciences*, Cambridge University Press.

Jiang, W., 2017. Have instrumental variables brought us closer to the truth, *Review of Corporate Finance Studies* 6, 127-140.

Jones, J., 1991. Earnings management during import relief investigations, *Journal of Accounting Research* 29, 193–228.

Klein, A., 2002. Audit committee, board of director characteristics, and earnings management, *Journal of Accounting and Economics* 33, 375-400.

LaLonde, R. J., 1986. Evaluating the Econometric Evaluations of Training Programs with Experimental data, *American Economic Review* 76, 604-620.

Larcker, D., Richardson, S., Tuna, I., 2007. Corporate governance, accounting outcomes, and organizational performance, *The Accounting Review* 82, 963–1008.

Leamer, E. E., 1983. Let's Take the Con Out of Econometrics, *American Economic Review* 73, 31-43.

Lee, D., Lemieux, T., 2010. Regression Discontinuity Designs in Economics, *Journal of Economic Literature* 48, 281-355.

McCrary, J., 2008. Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test, *Journal of Econometrics* 1342, 698-714.

Mattei, A., Mealli, F. 2016. Regression Discontinuity Designs as Local Randomized Experiments, *Observational Studies* 2, 156-173.

Mealli, F., Rampichini, C., 2012. Evaluating the Effects of University Grants by Using Regression Discontinuity Designs, *Journal of the Royal Statistical Society A* 175, 775-798.

Moss, B. G., Yeaton, W. H., Lloyd, J. E., 2014. Evaluating the Effectiveness of Developmental Mathematics by Embedding a Randomized Experiment within a Regression Discontinuity Design, *Educational Evaluation and Policy Analysis* 36(2), 170-185.

Shadish, W. R., Galindo, R., Wong, V. C., Steiner, P. M., Cook, T. D., 2011. A Randomized Experiment Comparing Random and Cutoff-Based Assignment, Psy*chological Methods* 16, 179-191.

Weir, C., Laing, D., McKnight, P. J., 2001. An Empirical Analysis of the Impact of Corporate Governance Mechanisms on the Performance of UK Firms, Working Paper, Robert Gordon University.

Vafaes, N., Theodorou, E., 1998. The relationship between board structure and firm performance in the UK, *The British Accounting Review* 30(4), 383-407.

Wooldridge, J., 2010. *Econometric Analysis of Cross Section and Panel Data*, Second Edition, MIT Press.

Figure 1. Board Structure Subindex and Asset Size

The scatter plots show the relationship between ln(assets in billion Korean won) and Board Structure Subindex (0~20) from 1998–2004 for firms with assets $\in$ [0.5T, 8T]. The 1999 reforms require large firms (assets > 2 trillion won; ln(assets) = 7.60) to have a minimum index value ≥ 11.7 (5 points for 50% outside directors; 6.7 points for audit and outside director nomination committees). Audit committee is required in 2000; 50% outside directors and outside director nominating committee in 2001. Sample excludes banks and SOEs. Vertical line indicates 2 trillion Won; horizontal line indicates minimum index value for large firms. Firm size is measured separately for each year. Sample size in 2000 is 45 large firms and 79 mid-sized firms (assets between 0.5-2 trillion won). Note that for firms which first exceed 2T won in assets in year *t*, compliance is required the following year.

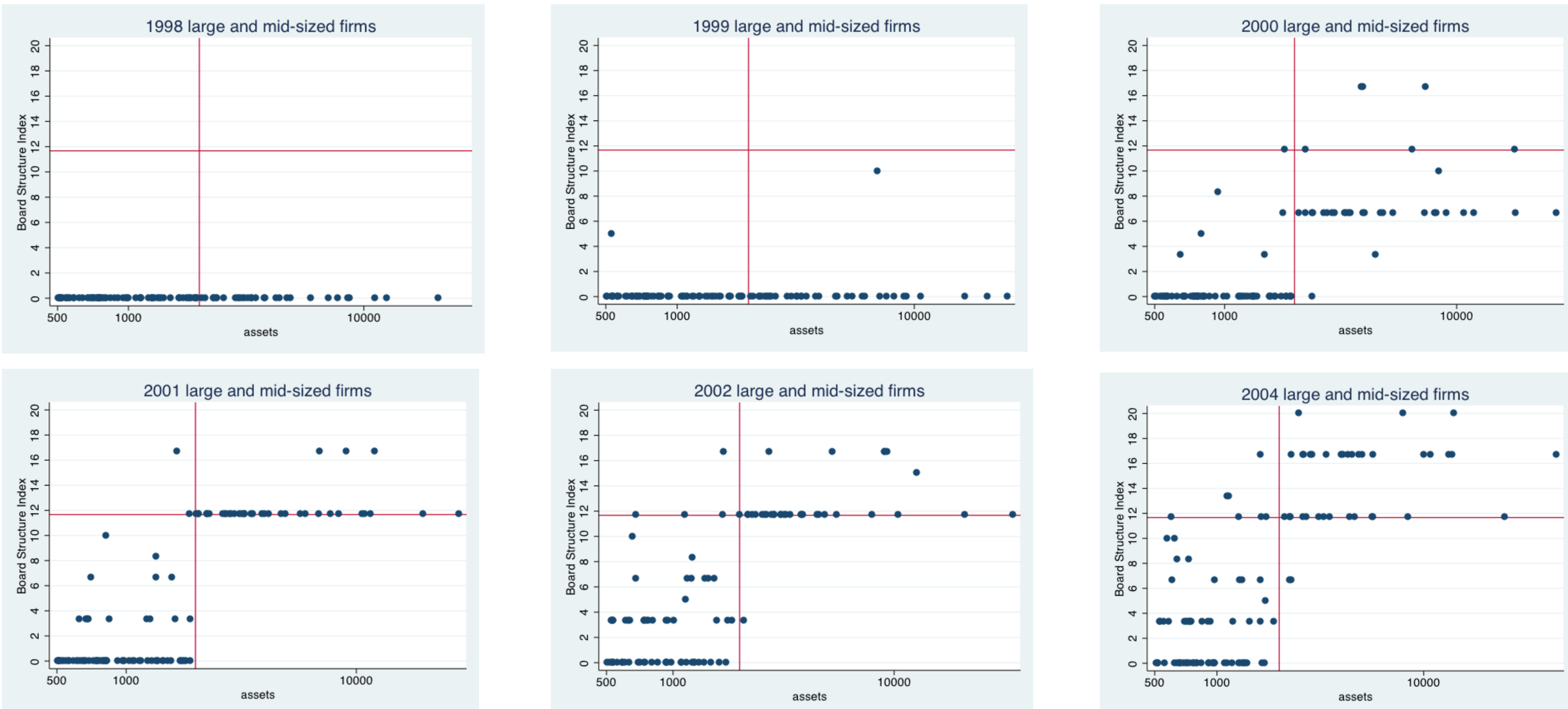
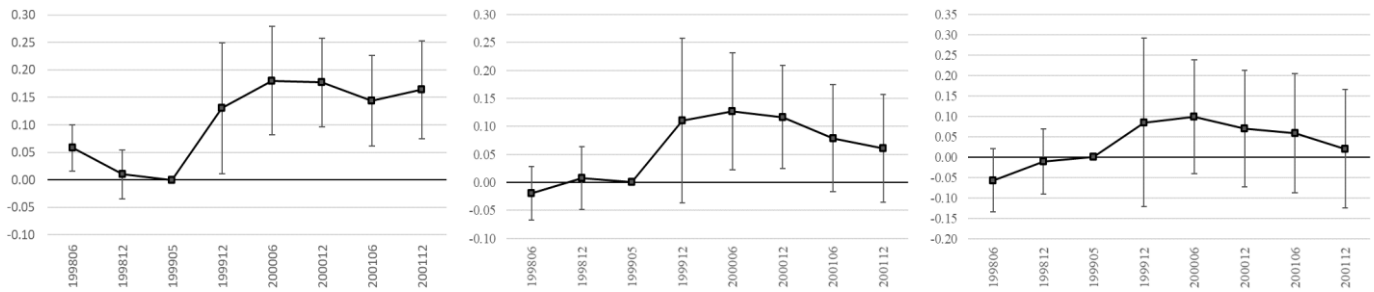


Figure 2. Annual DiD Results, Overall and within RD Bandwidth

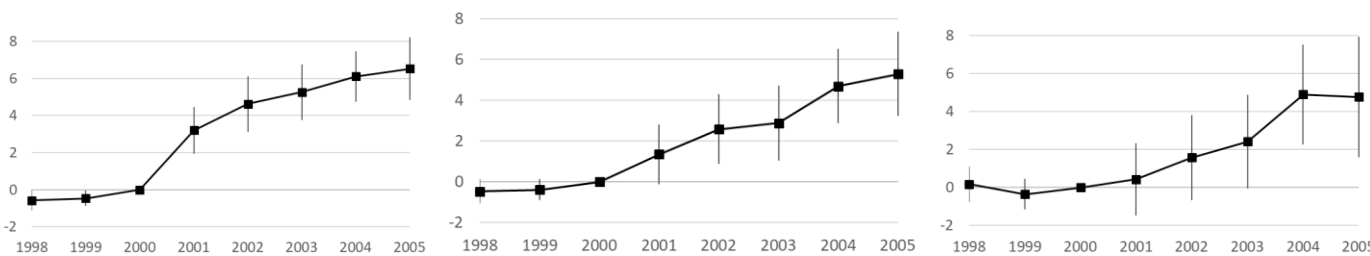
Figures present coefficients on large firm dummy from annual DiD regressions, with firm and industry × year FE, following eqn. (5), with limited covariates (ln(assets) and ln(years listed)). Point estimates are relative to May 1999 for Tobin's *q*, and 2000 for other outcomes. Vertical lines show 95% confidence interval, using standard errors clustered on firms. We use full sample for left-hand figures, broad bandwidth ([0.5T, 8T]) for middle figures, and narrow bandwidth ([1T, 4T]) for right-hand figures. Treatment and control groups are defined at May 1999 for Tobin's *q* and at year-end 2000 for other outcomes.

Full sample **Broad bandwidth [0.5T, 8T]** **Narrow bandwidth [1T, 4T]**

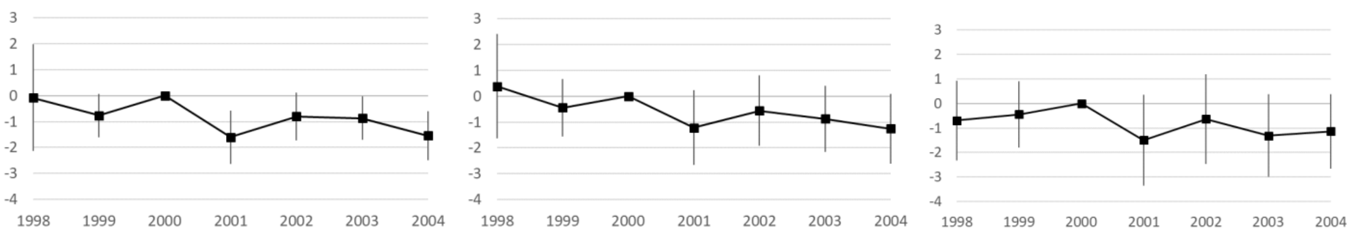
***ln*(Tobin's *q*)**



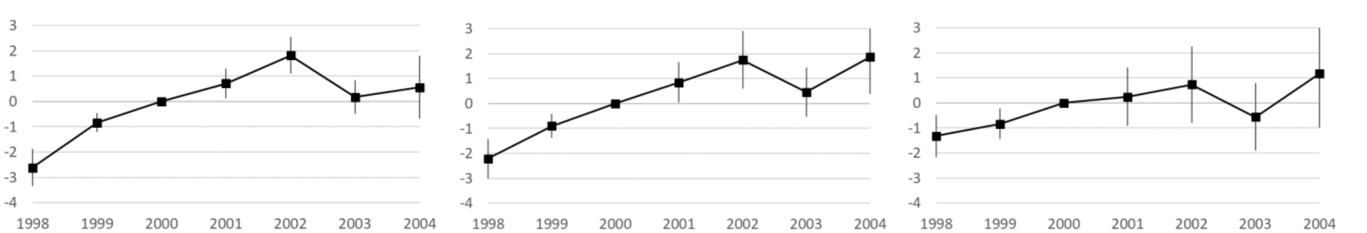
Disclosure Subindex



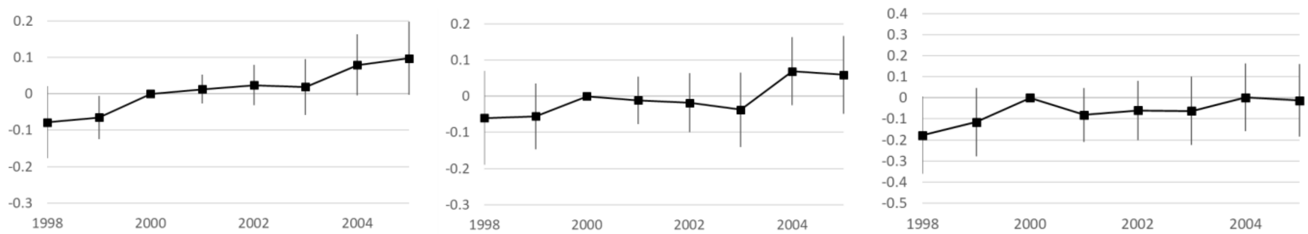
Shareholder Rights Subindex



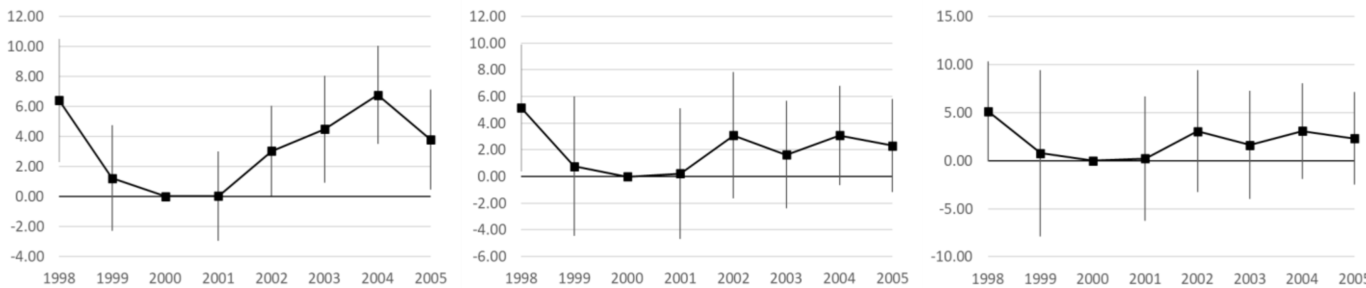
Board Procedure Subindex



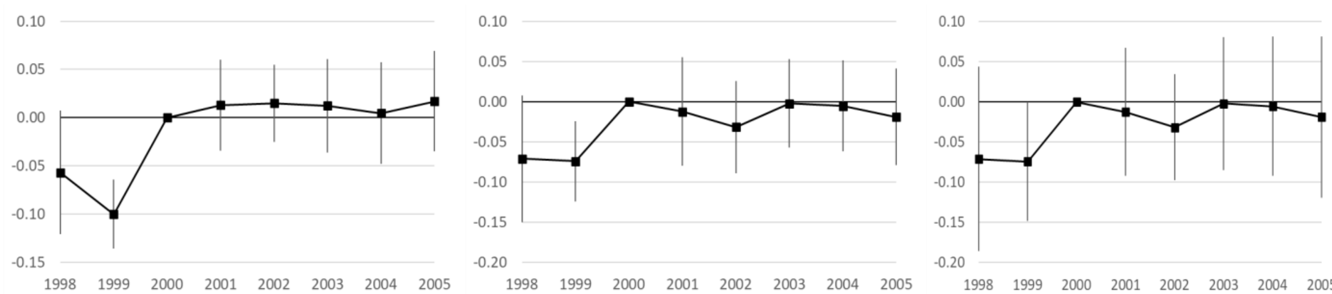
ln(MD&A word count)



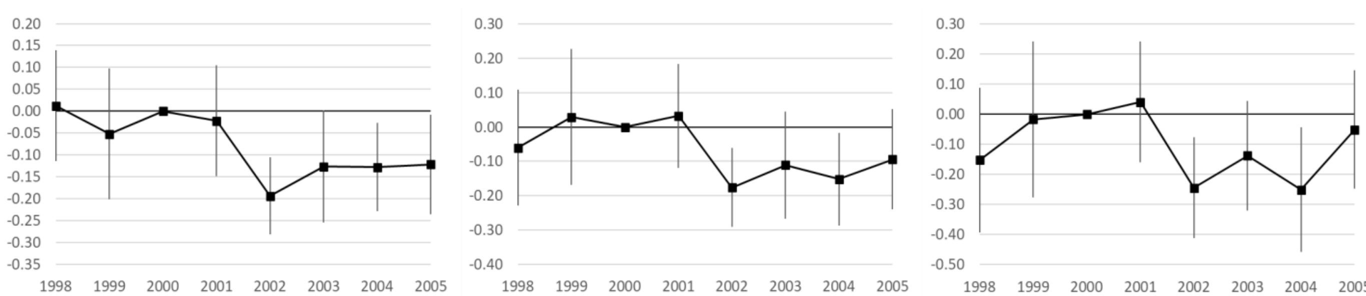
ROA



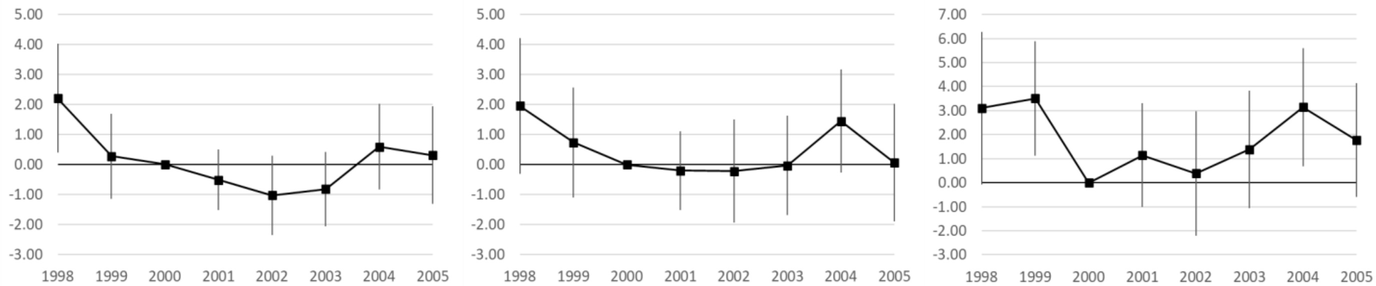
Book Leverage



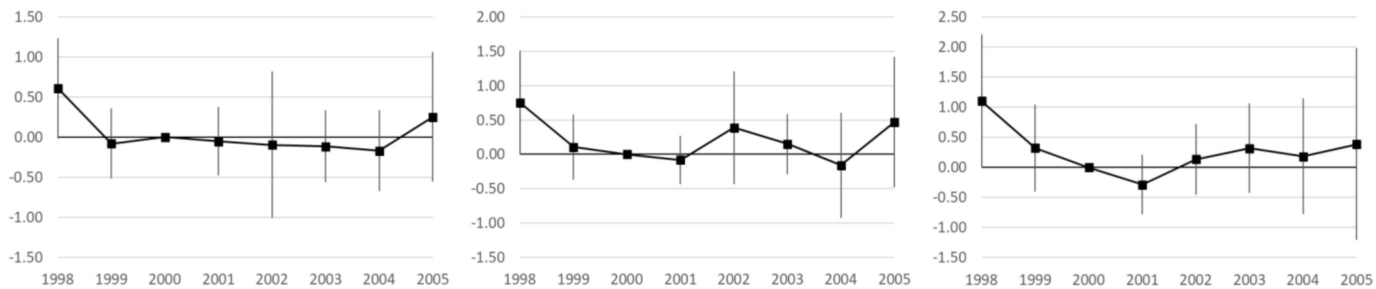
One-Year Sales Growth



(capex/assets)×100



R&D/sales



Abs(AA)×100

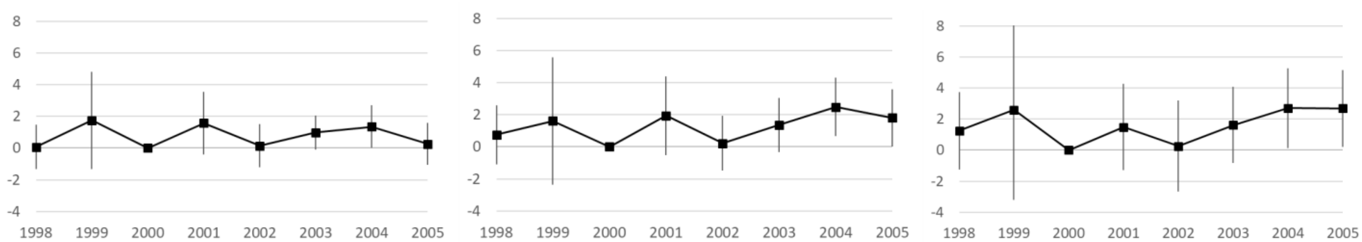


Table 1. Construction of KCGI

This table shows the governance elements included in KCGI. For details on data sources element and index construction, see Black and Kim (2012).

Shareholder Rights Subindex (A)
Firm permits cumulative voting for election of directors.
Firm permits voting by mail.
Firm discloses director candidates to shareholders in advance of shareholder meeting.
Board approval required for related party transactions (*required 2000 for top 10 chaebol, mid-2001 for all chaebol, 2001 on for large and chaebol firms*)
Board Structure Subindex (B)
Firm has at least 50% outside directors *(rule adopted 1999 required beginning mid-2001 for large firms)*
Firm has more than 50% outside directors (director database except as indicated)
Firm has outside director nominating committee (*rule adopted 1999, required from mid-2001 for large firms*).
Audit committee of the board of directors exists (*rule adopted 1999, required from mid-2001 for large firm*)
Firm has compensation committee
Board Procedure Subindex (C)
Directors' positions on board meeting agenda items are recorded in board minutes.
Board chairman is an outside director or (from 2003) firm has outside director as lead director.
A system for evaluating directors exists.
A bylaw to govern board meetings exists.
Firm holds four or more regular board meetings per year.
Firm has one or more foreign outside directors.
Shareholders approve outside directors' aggregate pay (separate from all directors' pay).
Outside directors attend at least 70% of meetings, on average
Board meeting solely for outside directors exists.
100% outside directors on audit committee
Bylaws governing audit committee (or internal auditor) exist.
Audit committee includes person with expertise in accounting
Audit committee (or internal auditor) approves the appointment of the internal audit head.
Audit committee meets ≥ 4 times per year
Disclosure Subindex (E)
Firm conducted investor relations activity in last year
Firm website includes resumes of board members
English financial disclosure exists
Ownership Parity (P)
Ownership Parity = (1 - ownership disparity); disparity = ownership by all affiliated shareholders - ownership by controlling shareholder and family members

Table 2. Summary Statistics and Correlations

Panel A. Summary Statistics for KCGI and Subindices

	Year	N	mean			median	sd	min	max
			all	large	small				
KCGI	1998	476	23.3	27.8	22.6	23	5.1	10.3	55.6
	2000	505	28.7	43.2	26.7	26.7	8.6	7	67.2
	2002	434	40.6	61.1	37	37.2	12.1	13.8	87.9
	2004	498	42.9	68	39.4	40.3	12.7	18.4	98.6
	Total	3551	33.7	50.4	31.2	31.2	12.4	7	98.6
Board Structure	1998	476	0.1	0	0	0	0.9	0	10
	2004	498	3.5	15.1	1.8	0	5.5	0	20
	Total	3551	2	9.2	1	0	4.2	0	20
Ownership Parity	1998	476	17.6	16.3	17.7	18.9	3	3.6	20
	2004	498	17	17	17.1	18.5	3.6	4.2	20
	Total	3551	17.4	16.6	17.5	18.8	3.2	2	20
Disclosure Subindex	1998	476	0.6	1.7	0.4	0	2	0	20
	2004	498	6.2	13.2	5.1	6.7	5.8	0	20
	Total	3551	2.7	6.6	2.1	0	4.6	0	20
Board Procedures	1998	476	4.4	5.8	4.2	4	2.4	0	10
	2004	498	7.9	12.1	7.3	8	3	0	18.6
	Total	3551	6.6	10	6.1	6.7	3	0	18.6
Sharehoder Rights	1998	476	0.7	3.9	0.3	0	2.6	0	20
	2004	498	8.4	10.7	8	6.7	2.9	5	20
	Total	3551	5	7.8	4.5	5	4.4	0	20

Panel B. Pearson Correlation Coefficients for KCGI and Subindices

*, **, and *** indicate significance at 10%, 5%, and 1% levels, respectively.

	KCGI	Board Structure	Ownership Parity	Disclosure	Board Procedure	Shareholder Rights
KCGI	1					
Board Structure	0.75***	1				
Ownership Parity	0.17***	-0.04	1			
Disclosure	0.75***	0.44***	-0.07**	1		
Board Procedure	0.69***	0.54***	-0.13***	0.40***	1	
Shareholder Rights	0.72***	0.39***	-0.06***	0.42***	0.43***	1

Table 3. Variable Definitions

Definitions and outlier treatment description of principal variables. Amounts are in billion Korean won. Balance sheet amounts are measured at fiscal year-end and income and cash flow amounts are for fiscal year, unless otherwise specified. Indicated variables are winsorized within each year at the 1% level, 99% level, or both.

outcomes	**description**
ln(market value of assets)	natural logarithm of (book value of debt + book value of preferred stock + market value of common stock, in thousand won)
Tobin's *q*	market value of assets / book value of assets; winsorized at 1/99%
ln(Tobin's *q*)	natural logarithm of Tobin's *q*
Disclosure Subindex	See Table 1 (construction of KCGI)
Shareholder Rights Subindex	See Table 1 (construction of KCGI)
Board Procedure Subindex	See Table 1 (construction of KCGI)
ln(MD&A word count)	*ln*(number of words in the MD&A section of annual reports)
ROA	net income / book value of assets; winsorized at 1/ 99%
Book leverage	book value of debt / book value of assets; winsorized at 99%
Market leverage	book value of debt / market value of common stock; winsorized at 99%
One-Year Sales Growth	sales growth = $[(sales_t/sales_{t-1})-1]$; winsorized at 1/99%
(capex/assets)×100	(capital expenditures/assets) × 100; winsorized at 99%
(R&D/sales)×100	(research and development expenditures/sales)×100; missing treated as 0; winsorized at 99%
Abs(AA)×100	absolute value of abnormal accruals×100; winsorized at 99%
Signed(AA)×100	signed abnormal accruals×100; winsorized at 1/ 99%
main independent variable	**description**
KCGI	Korean Corporate Governance Index
other variables	**description**
large	large firm dummy equals 1 if book value of assets>2 trillion won at end of 2000 and zero otherwise (or as specified in the text)
post	post reform dummy equals 1 if fiscal year >=2001 and zero otherwise (or as specified in the text)
ln(assets)	natural logarithm of assets, in thousand won
ln(years listed)	natural logarithm of number of years listed on Korean Stock Exchange
debt/mvce	book value of debt/market value of common equity; winsorized at 99%
five-year sales growth	geometric mean growth during past five fiscal years (or available period if shorter); winsorized at 1/99%
R&D/sales	research and development expenditures/sales; missing treated as 0
advertising/sales	advertising expenditures/sales; missing treated as 0
exports/sales	export revenue/sales; missing treated as 0
ppe/sales	property, plant and equipment/sales
capex/ppe	capital expenditures/property, plant and equipment
ebit/sales	earnings before interest and taxes/sales; winsorized at 1/99%
market share	firm's share of sales by all firms in the same 4-digit industry listed on KSE.
turnover	common shares traded during the year / common shares held by public shareholders; winsorized at 99%
foreign ownership	common shares held by foreign investors/common shares outstanding
sole ownership	common shares held by controlling shareholder and family members/common shares outstanding
Chaebol	1 if a member firm of top 30 chaebols; 0 otherwise
ADR (level 1)	1 if a firm issued level 1 ADR; 0 otherwise
ADR (level 2 or 3)	1 if a firm issued level 2 or 3 ADR; 0 otherwise
MSCI	1 if a member of MSCI Emerging Markets Index; 0 otherwise
ln(mvce)	natural logarithm of market value of common equity
btm	(book value of equity/market value of equity); winsorized at 99%
cash flow/lagged assets	$cash\ flow_t/(assets_{t-1})$; winsorized at 1/ 99%

Table 4. Descriptive Statistics

Table shows *t*-test for differences in means in 1999 for covariates x_j (indexed by *j*), $|t_j| = |\bar{x}_{jt} - \bar{x}_{jc}| / [(s_{jt}^2 / N_t + s_{jc}^2 / N_c)]^{1/2}$ where s_{jt} and s_{jc} are standard deviations for treated and control groups. Tobin's *q* means are as of mid-1999. *, **, *** indicate significance at the 10%, 5%, and 1% levels. Significant differences (at 5% level or better) are in **boldface.** Table also shows absolute values of "normalized differences" (Imbens and Rubin, 2015), defined as $ND_j = |\bar{x}_{jt} - \bar{x}_{jc}| / [(s_{jt}^2 + s_{jc}^2)/2]^{1/2}$.

	0-2 T vs. > 2T				0.5-2T vs. 2-8T			
	Mean Treated	Mean Controls	Norm. Diff.	t-test	Mean Treated	Mean Controls	Norm. Diff	t-test
outcome								
ln(mkt value of assets)	22.18	18.99	3.51	**30.47*** **	21.96	20.57	3.01	**17.00*** **
ln(Tobin's *q*)	-0.02	0.02	0.12	1.03	-0.02	0.09	0.28	1.60
Tobin's *q*	1.01	1.08	0.18	1.47	1.01	0.95	0.20	1.13
Disclosure Subindex	7.10	1.96	0.46	**2.44** **	6.15	3.13	0.14	0.74
Shareholder Rights Subindex	7.67	4.47	0.91	**5.13*** **	7.64	5.20	0.69	**3.45*** **
Board Procedure Subindex	9.65	4.47	1.28	**8.39*** **	9.43	5.91	0.78	**4.11*** **
ln(MD&A word count)	8.27	7.92	0.72	**4.11*** **	8.22	8.05	0.27	1.29
ROA	1.27	2.22	0.10	078	0.30	2.45	0.28	1.47
Book Leverage	0.66	0.54	0.71	5.60	0.69	0.64	0.31	1.73*
Market Leverage	6.89	4.75	0.24	1.55	7.96	8.73	0.07	0.37
One-Year Sales Growth	0.30	0.20	0.20	1.25	0.31	0.16	0.30	1.50
(capex/assets)×100	5.20	4.44	0.15	0.91	5.10	4.80	0.06	0.27
(R&D/sales)×100	1.33	1.02	0.16	0.97	0.94	0.71	0.20	0.90
Abs(AA)×100	3.80	5.09	0.05	0.21	3.56	4.04	0.21	0.83
Signed(AA)×100	0.24	-0.53	-0.10	-0.47	0.25	0.06	-0.01	-0.05
main indep. variable								
KCGI	30.02	24.96	0.78	**4.65*** **	29.02	26.69	0.38	**1.99** **
covariates								
ln(assets)	8.39	5.32	3.49	**30.89*** **	8.18	6.81	3.39	**18.51*** **
ln(years listed)	2.82	2.45	0.50	**3.60*** **	2.85	2.75	0.16	0.89
debt/mvce	7.70	4.61	0.34	**2.18** **	8.63	7.56	0.10	0.55
five-year sales growth	0.19	0.11	0.54	**3.80*** **	0.19	0.17	0.17	0.97
R&D/sales	0.01	0.03	0.07	0.96	0.01	0.01	0.02	0.13
advertising/sales	0.01	0.01	0.22	**2.12** **	0.01	0.01	0.12	0.71
exports/sales	0.35	0.28	0.24	1.64	0.30	0.25	0.18	0.96
ppe/sales	0.68	0.49	0.36	**2.29** **	0.69	0.55	0.23	1.21
capex/ppe	0.14	0.15	0.10	0.83	0.14	0.15	0.05	0.30
ebit/sales	0.05	0.06	0.11	0.65	0.04	0.09	0.35	1.78*
market share	0.17	0.05	0.69	**4.32*** **	0.13	0.09	0.22	1.23
turnover	5.08	6.00	0.25	**2.18** **	5.20	4.68	0.18	0.99
foreign ownership	15.25	5.82	0.78	**5.12*** **	14.85	8.67	0.47	**2.58*** **
sole ownership	8.20	20.79	0.98	**8.27*** **	8.30	15.12	0.56	**3.20*** **
Chaebol	0.86	0.15	1.99	**13.97*** **	0.85	0.33	1.23	**7.02*** **
ADR (level 1)	0.16	0.00	0.60	**3.19*** **	0.09	0.01	0.34	1.70
ADR (level 2 or 3)	0.02	0.00	0.15	0.85	0.02	0.00	0.21	1.00
MSCI	0.63	0.10	1.28	**7.75*** **	0.62	0.33	0.59	**3.21*** **
ln(mvce)	7.05	4.68	1.57	**7.63*** **	6.68	5.31	1.11	**4.46*** **
btm	1.13	1.42	0.26	**1.67***	2.08	3.15	-0.41	**-2.00** **
cash flow/lagged assets	0.07	0.05	0.05	0.28	0.067	0.073	-0.21	-0.87

Table 5. Comparison of Results across Research Designs

Table summarizes evidence for relation between *KCGI* or large firm shock and indicated outcomes, across indicated research designs. "Yes" and **boldface** means statistically significant relationship (at 5% level or better), in predicted direction; "Marg" "and italics means marginally significant relationship (at 10% level), in predicted direction; "No" means insignificant result. AA = abnormal accruals. See text for methods details. Results for causal designs are with limited covariates. RD results use "broad" bandwidth ([0.5T, 8T]); RD-only regressions do not control for *ln*(assets).

Method	Benchmark	Classic panel		Simpler causal				Combined causal	
	Annual DiD /RD	RE	FE	Simple DiD	Shock-IV	Post-shock RD	Annual DiD	Simple DiD/ RD	IV/RD
Bandwidth	Broad	All	All	All	All	Broad	All	Broad	Broad
Disclosure and subindices									
Disclosure subindex	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Shareholder Rights subindex	No	Yes	No	Marg	No	No	No	No	No
Board Procedure subindex	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Ln(MD&A word count)	No	Yes	Yes	Yes	No	No	No	No	No
Financial outcomes									
Ln(Tobin's q)	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	No
ROA	No	*Marg*	No	No	Yes	No	*Marg*Some	No	Yes
Book leverage	No	*Marg*	No	Yes	No	No	No	No	No
Market leverage	No	Yes	Marg	Yes	No	No	No	No	No
One-year sales growth	*Marg*	Yes	Marg	Yes	Yes	No	Yes	*Marg*	Yes
Capex/assets	No	No	No	Yes	No	No	No	No	No
R&D/sales	No	No	No	No	No	No	No	No	No
Earnings management									
Abs(AA)	No	Yes	Yes	No	No	No	No	No	No
Signed AA	No	No	No	No	No	No	No	No	No

Note: Black, Jang and Kim (2006) and Black and Kim (2012) find evidence, across a variety causal methods, including an event study of the reform period, that the reform shock predicts higher share price and Tobin's q. Black, Kim, Jang and Park (2015) report evidence that control of related party transactions is one channel which can explain the effect of governance on Tobin's q.

Table 6. Panel Regressions: Does KCGI Predict Tobin's *q* and other outcomes?

Panel A. *ln*(Tobin's q)

Each row reports coefficients from separate regressions, with industry × year FE, and either firm RE or FE as indicated, of *ln*(Tobin's q) on KCGI, indicated covariates, and constant term, over 1998-2004. Variables are defined in Table 3. *t*-statistics, with standard errors clustered on firm, are in parentheses. *, **, *** indicate significance at the 10%, 5% and 1% level; significant results at 5% or better are in **boldface**. R^2 is overall for RE and within for FE. Median RE λ indicates whether RE results are closer to OLS (λ ~ 0) or FE (λ ~ 1) is 0.71 for col. (1) and 0.66 for col. (3).

Regression	(1)	(2)	(3)	(4)
industry×year FE	yes	yes	yes	yes
firm FE or RE	RE	FE	RE	FE
$KCGI_t$	**0.0054*** (6.58)**	**0.0036*** (4.01)**	**0.0043*** (5.40)**	**0.0030*** (3.37)**
ln(assets)	**0.9693*** (84.99)**	**0.9141*** (35.04)**	**0.9501*** (81.95)**	**0.8884*** (32.75)**
ln(years listed)	**-0.0892*** (-5.53)**	**-0.1169*** (-3.00)**	**-0.1006*** (-7.43)**	**-0.1542*** (-4.26)**
debt/mvce			**0.0033*** (4.10)**	**0.0028*** (3.14)**
5-year sales growth			0.0426 (1.13)	0.1094** (2.29)
R&D/sales			**0.0468** (2.48)**	**0.0269*** (2.95)**
advertising/sales			**1.5160** (2.52)**	1.4521* (1.73)
exports/sales			**-0.0635** (-2.14)**	**-0.1262*** (-2.91)**
ppe/sales			**-0.0577*** (-2.59)**	-0.0445* (-1.88)
$(ppe/sales)^2$			**0.0016** (2.10)**	0.0012 (1.48)
capex/ppe			**0.1187*** (4.44)**	**0.0726*** (3.14)**
ebit/sales			**-0.1184** (-1.98)**	-0.0401 (-0.65)
market share			**0.3352*** (3.85)**	**0.3713*** (3.48)**
turnover			**0.0055*** (5.35)**	**0.0040*** (3.68)**
foreign ownership			**0.0044*** (5.57)**	**0.0047*** (5.42)**
sole ownership			**-0.0040*** (-2.61)**	-0.0021 (-1.08)
sole $ownership^2$			0.0000 (0.48)	0.0000 (0.35)
Chaebol			0.0141 (0.59)	
ADR (level 1)			0.0761 (1.49)	
ADR (level 2 or 3)			0.3070 (1.30)	
MSCI			0.0417 (1.35)	
constant	**13.97*** (215.64)**	**14.40*** (87.21)**	**14.13*** (222.58)**	**14.62*** (90.59)**
No of firms (observations)	672 (3,580)	672 (3,580)	668 (3,459)	668 (3,459)
R^2	0.9525	0.6009	0.9632	0.6393

Panel B. Other outcomes

Each row reports coefficients from separate regressions, with industry × year FE, and either firm RE or FE as indicated, of indicated outcome variables on indicated KCGI variant, either limited or full covariates, and constant term. Limited covariates are ln(assets) and ln(years listed). Specifications and covariates are same as in Panel A except: Full covariates for abs(AA)×100 and signed(AA)×100 are cash flow/lagged assets; (cash flow/lagged assets)^2, btm, *ln*(mvce). Coefficients on covariates and constant term are suppressed. Sample period and number of observations are shown in Table 10. *t*-statistics, with standard errors clustered on firm, in parentheses. *, **, *** indicate significance at the 10%, 5% and 1% level; significant results at 5% or better are in **boldface**.

regression		(1)	(2)	(3)	(4)
firm RE or FE covariates		RE limited	FE limited	RE full	FE full
outcome	predictor variable				
ln(Tobin's q), from Panel A	$KCGI_t$	**0.0054*** (6.58)**	**0.0036*** (4.01)**	**0.0043*** (5.40)**	**0.0030*** (3.37)**
Tobin's *q*	$KCGI_t$	**0.0050*** (5.57)**	**0.0022** (1.97)**	**0.0037*** (4.37)**	0.0016 (1.44)
Disclosure Subindex	$KCGI_{t-1}$ less Disclosure	**0.0921*** (5.59)**	**0.0773*** (4.05)**	**0.0735*** (4.74)**	**0.0778*** (4.15)**
Shareholder Rights Subindex	$KCGI_{t-1}$ less Shareholder Rights	**0.0288** (2.37)**	-0.0041 (-0.25)	**0.0421*** (3.31)**	0.0024 (0.15)
Board Procedure Subindex	$KCGI_{t-1}$ less Board Procedure	**0.0606*** (5.63)**	**0.0389** (2.56)**	**0.0464*** (5.10)**	**0.0302** (2.49)**
ln(MD&A word count)	$KCGI_{t-1}$	**0.0026*** (4.07)**	**0.0021*** (2.84)**	**0.0025*** (3.62)**	**0.0023*** (2.89)**
ROA	$KCGI_{t-1}$	-0.0708* (-1.78)	0.0296 (0.69)	**-0.0807** (-2.29)**	0.0119 (0.28)
Book Leverage	$KCGI_{t-1}$	0.0010* (1.83)	0.0003 (0.52)	**0.0011** (2.29)**	0.0007 (1.26)
Market Leverage	$KCGI_{t-1}$	**-0.0530*** (-2.77)**	-0.0469* (-1.89)	-0.0300 (-1.52)	-0.0337 (-1.36)
One-Year Sales Growth	$KCGI_{t-1}$	**-0.0022*** (-2.92)**	-0.0022* (-1.69)	**-0.0027*** (-3.09)**	**-0.0034** (-2.58)**
(capex/assets)× 100	$KCGI_{t-1}$	0.0014 (0.12)	-0.0165 (-1.14)	-0.0090 (-0.75)	-0.0062 (-0.43)
R&D/sales	$KCGI_{t-1}$	0.0079 (1.28)	0.0012 (0.18)	0.0115* (1.75)	0.0036 (0.54)
abs(AA)×100	$KCGI_{t-1}$	**-0.0336** (-2.33)**	**-0.0551** (-2.36)**	**-0.0342** (-2.28)**	**-0.0563*** (-2.85)**
signed(AA)× 100	$KCGI_{t-1}$	0.0222 (1.44)	0.0091 (0.32)	0.0023 (0.13)	-0.0008 (-0.03)

Table 7. Simple DiD Regressions (Full and Within-Bandwidth Samples)

Each row reports coefficient on post dummy (=1 starting in 1999 for share-price-based outcomes and in 2001 for other outcomes) interacted with large firm dummy for indicated dependent variable. Large-firm dummy is measured at year-end 1999 for Tobin's q and at year-end 2000 for other outcomes. Coefficients are from separate DiD regressions with firm and industry × year FE of dependent variable on post dummy × large firm dummy, covariates, and constant term. Covariates are same as in Table 6, coefficients on covariates and constant term are suppressed. Post dummy and large dummy are absorbed by the fixed effects. See Table 10 for sample periods and sample sizes. *t*-statistics, with standard errors clustered on firm, are in parentheses. *, **, *** indicate significance at the 10%, 5% and 1% level; significant results (at 5% or better) are in **boldface**.

regression sample covariates	(1) full limited	(2) broad limited	(3) narrow limited	(4) full full	(5) broad full	(6) narrow full
ln(Tobin's *q*)	**0.2268*** (6.16)**	**0.1121*** (2.84)**	**0.1550*** (2.79)**	**0.1403*** (4.79)**	**0.0808** (2.27)**	0.0665 (1.53)
Tobin's *q*	**0.2358*** (5.28)**	**0.1365*** (2.71)**	**0.1714*** (2.69)**	**0.1286*** (3.59)**	**0.0908** (2.22)**	0.0616 (1.33)
Disclosure Subindex	**5.3444*** (9.36)**	**3.3866*** (5.13)**	**2.5179*** (2.82)**	**4.4797*** (7.70)**	**2.6082*** (3.77)**	1.6427* (1.69)
Shareholder Rights Subindex	**-0.8670** (-1.99)**	-0.7173 (-1.10)	-1.0829 (-1.31)	-0.7980* (-1.83)	-0.6026 (-0.93)	-1.3694* (-1.81)
Board Procedure Subindex	**1.2315*** (4.33)**	**1.6989*** (4.21)**	0.8039 (1.37)	**1.1259*** (3.98)**	**1.1148*** (2.77)**	0.0936 (0.17)
ln(MD&A word count)	**0.0791*** (2.69)**	0.0376 (1.07)	-0.0061 (-0.13)	**0.0719** (2.55)**	0.0219 (0.62)	-0.0457 (-0.89)
ROA	0.8480 (0.61)	0.3855 (0.24)	0.7360 (0.35)	1.2306 (1.16)	0.6694 (0.65)	1.3684 (0.97)
Book Leverage	**0.0768*** (3.23)**	0.0364 (1.33)	0.0130 (0.35)	0.0250 (1.45)	0.0383* (1.86)	0.0213 (1.04)
Market Leverage	**-2.5645*** (-2.70)**	-0.2983 (-0.22)	0.5799 (0.29)	**-2.1558** (-2.30)**	0.4229 (0.33)	0.9908 (0.55)
One-Year Sales Growth	**-0.1035*** (-2.71)**	-0.0916* (-1.93)	-0.0735 (-1.06)	**-0.0928** (-2.34)**	-0.1021* (-1.92)	-0.1061 (-1.34)
(capex/assets)×100	**-1.2948** (-2.43)**	-0.9504 (-1.43)	-0.9409 (-1.04)	-0.8241* (-1.68)	-0.5573 (-0.91)	-0.5333 (-0.70)
R&D/sales	-0.2045 (-1.02)	-0.0887 (-0.38)	-0.1564 (-0.63)	-0.1129 (-0.50)	0.1600 (0.64)	0.0694 (0.15)
abs(AA)×100	0.1741 (0.22)	0.8994 (0.92)	0.5906 (0.47)	0.0043 (0.01)	0.9414 (1.10)	0.4228 (0.40)
signed(AA)×100	0.5262 (0.63)	1.1241 (0.98)	0.7992 (0.54)	0.4757 (0.60)	1.3265 (1.34)	1.0127 (0.78)

Table 8. Shock-IV Regressions (Full and Within-Bandwidth Samples)

Table reports first- and second-stage 2SLS regressions, with firm and industry × year FE, of indicated dependent variables on instrumented KCGI, with post-reform dummy × large firm dummy (measured annually) as the instrumental variable, controlling for indicated covariates. Coefficients on covariates, and constant term are suppressed. See Table 10 for sample periods and samples. First stage coefficient on post-reform dummy × large-firm dummy is shown for ln(Tobin's q) (first-stage is similar for other share-price-related outcomes) and for ROA (similar for other non-share-price outcomes). For abs(AA) we use ln(market value of equity) as the size measure, following prior literature, instead of ln(assets). Limited covariates are ln(assets) and ln(years listed). See Table 6 for full covariates. *t*-statistics, with standard errors clustered on firm, are in parentheses. *, **, *** indicate significance at the 10%, 5% and 1% level; significant results (at 5% or better) are in **boldface**.

	(1)	(2)	(3)	(4)	(5)	(6)
sample **First stage**	full	broad	narrow	full	broad	narrow
Share-price related outcomes						
Ln(Tobin's q)	**9.7237*** (7.42)**	**7.7976*** (6.66)**	**5.6552*** (4.71)**	**9.0308*** (7.05)**	**7.9916*** (6.48)**	**5.7018*** (4.61)**
Other outcomes						
ROA	**13.1031*** (9.96)**	**9.6899*** (6.93)**	**6.9163*** (4.27)**	**13.2506*** (10.13)**	**9.6602*** (6.84)**	**6.9314*** (4.21)**
rest of KCGI covariates	yes limited	yes limited	yes limited	yes full	yes full	yes full
Second stage outcome	coefficient on instrumented *KCGI*			coefficient on instrumented *KCGI*		
ln(Tobin's *q*)	**0.0151*** (5.11)**	0.0067* (1.77)	0.0063 (0.89)	**0.0127*** (3.81)**	0.0051 (1.39)	0.0009 (0.16)
Tobin's *q*	**0.0144*** (4.23)**	0.0073 (1.62)	0.0034 (0.59)	**0.0123*** (3.20)**	0.0067 (1.45)	-0.0007 (-0.12)
Disclosure Subindex	**0.4173*** (7.17)**	**0.3285*** (4.10)**	**0.2535** (2.09)**	**0.3883*** (6.35)**	**0.2998*** (3.62)**	**0.2783** (1.96)**
Shareholder Rights Subindex	-0.0102 (-0.30)	-0.0198 (-0.40)	-0.0210 (-0.25)	-0.0047 (-0.13)	-0.0012 (-0.02)	-0.0283 (-0.34)
Board Procedure Subindex	**0.1325*** (4.61)**	**0.1563*** (3.56)**	**0.1315** (1.99)**	**0.0937*** (3.93)**	**0.1084*** (2.93)**	0.0755 (1.15)
ln(MD&A word count)	**0.0038** (2.45)**	0.0012 (0.48)	**-0.0080** (-1.98)**	0.0033* (1.92)	0.0010 (0.39)	-0.0078 (-1.63)
ROA	**0.2255** (2.10)**	0.3310* (1.91)	0.1905 (0.87)	0.1733* (1.88)	0.2801* (1.89)	0.1350 (0.71)
Book Leverage	-0.0007 (-0.34)	0.0000 (0.01)	0.0003 (0.06)	-0.0003 (-0.17)	-0.0009 (-0.46)	-0.0010 (-0.30)
Market Leverage	-0.1203 (-1.34)	-0.0467 (-0.32)	0.1036 (0.40)	-0.0897 (-1.02)	0.0484 (0.37)	0.2546 (1.20)
One-Year Sales Growth	**-0.0094*** (-3.53)**	-0.0061 (-1.43)	-0.0108 (-1.41)	**-0.0110*** (-3.67)**	**-0.0100** (-2.24)**	-0.0166 (-1.84)
(capex/assets)×100	-0.0181 (-0.41)	-0.0081 (-0.12)	0.1179 (0.92)	-0.0292 (-0.62)	0.0417 (0.60)	0.2390* (1.75)
R&D/sales	0.0154 (0.61)	0.0376 (1.11)	0.0732 (1.23)	0.0119 (0.54)	0.0283 (1.02)	0.0629 (1.12)
abs(AA)×100	-0.0718 (-1.23)	-0.0352 (-0.52)	-0.1074 (-1.03)	-0.0911* (-1.73)	-0.0427 (-0.79)	-0.1220 (-1.39)
signed(AA)×100	0.0319 (0.53)	0.0116 (0.14)	0.0137 (0.10)	0.0260 (0.42)	0.0213 (0.30)	0.0208 (0.17)

Table 9. Post-Shock Regression Discontinuity

Table shows coefficient on above-threshold dummy (threshold = 2T; equals1 in year *t* if firm has assets > 2T in that year) for indicated outcome variables from regressions over post-shock period (1999-2004 for share-price-related outcomes; 2001-2004 for all others) with indicated controls for the running variable (*ln*(assets)). Broad bandwidth is [0.5T, 8T]; narrow bandwidth is [1T, 4T]. Within each bandwidth, the first column does not control for the running variable. The middle column includes a linear control for ln(assets), but allows for a jump at the threshold. The right column allows different coefficients (different slopes) below and above the threshold. The regressions are: without slopes: y = a + b × large firm dummy. With single slope: y = a + b × large firm dummy + c × ln(assets). With different slopes: y = a + b × large firm dummy + c × [ln(assets)] + d × [large firm dummy × ln[(assets)]. *t*-statistics, with standard errors clustered on firm, in parentheses. *, **, *** indicate significance at the 10%, 5% and 1% level; significant results (at 5% or better) are in **boldface.**

Bandwidth	Broad			Narrow		
Regression	(1)	(2)	(3)	(4)	(5)	(6)
Ln(assets)	No	Yes	× threshold dummy	No	Yes	× threshold dummy
ln(Tobin's *q*)	**0.1383***** **(3.93)**	0.0266 (0.38)	0.0154 (0.24)	0.0699 (1.56)	0.0836 (0.99)	0.0750 (0.95)
Tobin's *q*	**0.1121***** **(2.89)**	0.0304 (0.44)	0.0089 (0.13)	0.0577 (1.19)	0.0515 (0.63)	0.0475 (0.60)
Disclosure Subindex	**4.4712***** **(6.37)**	1.6338 (1.39)	1.9304 (1.58)	**3.5893***** **(3.94)**	-0.3635 (-0.24)	-0.2484 (-0.17)
Shareholder Rights Subindex	**2.1753***** **(6.04)**	0.3974 (0.60)	0.7089 (1.13)	**1.4199***** **(3.12)**	-0.1649 (-0.17)	-0.0253 (-0.03)
Board Procedure Subindex	**3.6080***** **(9.74)**	**1.6135**** **(2.10)**	**2.0390***** **(2.74)**	**2.8465***** **(5.34)**	0.8315 (0.86)	1.1592 (1.28)
ln(MD&A word count)	**0.2040***** **(3.21)**	0.0510 (0.50)	0.0610 (0.57)	**0.1844**** **(2.32)**	-0.1371 (-1.07)	-0.1521 (-1.18)
ROA	0.6443 (0.87)	1.9331* (1.70)	1.3676 (1.14)	1.0842 (1.25)	-0.2153 (-0.13)	-0.1569 (-0.10)
Leverage	**0.0770**** **(2.47)**	-0.0066 (-0.12)	0.0109 (0.21)	0.0356 (0.96)	0.0418 (0.58)	0.0458 (0.69)
Market Leverage	0.5375 (0.70)	0.0002 (0.00)	0.7869 (0.53)	0.1513 (0.13)	2.8572 (1.24)	2.7705 (1.11)
One-Year Sales Growth	0.0107 (0.44)	-0.0188 (-0.36)	-0.0384 (-0.66)	-0.0197 (-0.51)	-0.0674 (-0.91)	-0.0622 (-0.81)
(capex/assets)×100	0.4231 (0.74)	-0.6074 (-0.77)	-0.5918 (-0.72)	0.0579 (0.10)	-0.3778 (-0.31)	-0.3475 (-0.29)
R&D/sales	0.4652 (1.41)	0.5690 (1.08)	0.7861 (1.25)	0.6381 (1.47)	0.5899 (0.77)	0.6789 (0.88)
abs(AA)×100	-0.6133 (-1.45)	-0.3480 (-0.50)	-0.6383 (-1.04)	-0.5132 (-1.01)	-0.4165 (-0.43)	-0.7940 (-0.87)
signed(AA)×100	0.2978 (0.47)	1.2712 (1.16)	1.0109 (1.00)	0.5479 (0.71)	0.6548 (0.45)	0.4664 (0.33)

Table 10. Overview of Sample Time Periods for Outcome-Methods Combinations

This table provides an overview of the time periods used for each regression. KCGI, Board Procedure Subindex and Shareholder Rights Subindex are available from 1998-2004. Disclosure Subindex and all other variables are available from 1998-2005. Data on share prices, outcomes other than governance indices, and covariates, is available from 1998-2005. We chose to start the pre-treatment period in 1998 because 1997 was a core year for the East Asian financial crisis. **Panel regressions**. For panel regressions, with KCGI as the core predictor variable, and outcomes for which we expect a lagged effect, the time period for the regressions starts in 1999 because KCGI is available starting in 1998. For outcomes for which we do not expect a lag (those derived from share prices), we regress the outcome on contemporaneous KCGI, and the regression time period ends in 2004 because the last year KCGI is available is 2004. **Panel DiD and IV**. We use the same time periods as for the panel regressions. **RD**. For outcomes for which we expect a lagged effect, the time period for RD starts in 2001 because the principal reforms began in 2000. For outcomes for which we do not expect a lag (those derived from share prices), the RD period starts in 1999, using year-end values, because the large firm reforms were adopted in 1999. **Yearly DiD**. We begin the sample period in 1998 in order to show this year as a pre-treatment year.

Panel A. Sample Periods

	Method				
Outcome	**panel regressions**	**panel DiD**	**RD**	**IV**	**yearly DiD**
ln(Tobin's *q*) Tobin's *q*	1998-2004	1998-2005	1999-2005	1998-2004	1998-2005
Shareholder Rights Subindex Board Procedure Subindex	1999-2004	1999-2004	2001-2004	1999-2004	1998-2004
Disclosure Subindex *ln*(MD&A word count) ROA Leverage One-Year Sales Growth (capex/assets)×100 R&D/sales abs(AA)×100 signed(AA)×100	1999-2005	1999-2005	2001-2005	1999-2004	1998-2005

Panel B. Sample Sizes

Table shows sample sizes for panel DiD regressions with full covariates. Each cell shows number of large firms, number of small firms (in parentheses), and number of observations. Sample sizes are smaller for capex/asset and R&D/sales because we exclude non-bank financial firms, and are smaller for accruals measures because the abnormal accruals estimation requires additional variables and a minimum number of observations per industry-year. Samples are slightly larger with limited covariates. **IV.** Sample sizes are same as for panel DiD. **Panel regressions.** Sample sizes are same as for panel DiD except there is no distinction between large and small firms. **RD.** Number of firms is similar to panel DiD but number of observations is smaller for samples that begin in 2001. **Yearly DiD**. Number of firms is similar to that for panel DiD.

	full sample	**broad sample**	**narrow sample**
ln(Tobin's *q*)	57 (568); 3,783	48 (97); 983	31 (36); 462
Tobin's *q*	57 (568); 3,783	48 (97); 983	31 (36); 462
Disclosure Subindex	**58 (576); 3,328**	**53 (134); 1,023**	**41 (60); 496**
Shareholder Rights Subindex	58 (576); 2,895	52 (122); 778	38 (53); 380
Board Procedure Subindex	58 (576); 2,907	52 (123); 781	38 (53); 381
ln(MD&A word count)	57 (574); 3,318	51 (130); 887	38 (56); 425
ROA	56 (579); 3,820	46 (98); 980	32 (38); 488
Leverage	56 (579); 3,820	46 (98); 980	32 (38); 488
Market Leverage	56 (579); 3,820	46 (98); 980	32 (38); 488
One-Year Sales Growth	56 (579); 3,819	46 (98); 980	32 (38); 488
(capex/assets)×100	46 (544); 3,532	36 (81); 803	24 (31); 382
R&D/sales	46 (544); 3,532	36 (81); 803	24 (31); 382
Abs(AA)×100	39 (530); 3,438	35 (110); 737	26 (46); 347
Signed(AA)×100	39 (530); 3,438	35 (110); 737	26 (46); 347