Discussion of “CEO compensation and corporate risk-taking: Evidence from a natural experiment”∗

Christopher S. Armstrong
The Wharton School, University of Pennsylvania, 1300 Steinberg-Dietrich Hall, Philadelphia, PA 19104-6365, USA

A R T I C L E  I N F O
Available online 12 November 2013
JEL classification:
C18
C21
C26
G30
G32
G34
Keywords:
Legal liability
Regulatory risk
Tail risk
Stock options
Compensation
Managerial incentives

A B S T R A C T
Gormley, Matsa, and Milbourn (in this issue) examine the design and causal effects of CEOs’ equity portfolio incentives on firm risk in a novel research setting in which certain firms experience a large exogenous shock that increases their left-tail risk and reduces their investment opportunities. Gormley et al. find that boards and CEOs both make adjustments to CEOs’ equity portfolios following the shock. They also find that CEOs with more convex equity portfolios (i.e., Vega) prior to the shock reduce risk less following the shock. Despite certain measurement and identification concerns, Gormley et al. is an innovative attempt to address an important and challenging research question. Partial identification and sensitivity analysis an important class of techniques that are well-suited for providing causal inferences about Gormley et al.’s and other important research questions that are impeded by endogeneity concerns.

1. Introduction

Gormley, Matsa, and Milbourn (in this issue and hereafter referred to as “GMM”) examine a novel research setting in which certain firms experience a large and exogenous shock to their left-tail risk. They address two important and related research questions about the relation between CEOs’ equity incentives and firm risk. GMM first examine whether boards and CEOs adjust CEOs’ equity portfolios in response to the shock. As GMM discuss in their paper, this research question is important because it speaks to the efficiency with which CEOs’ incentive-compensation contracts are designed and subsequently adjusted—both by boards and by CEOs themselves—in response to changes in the contracting environment.

GMM’s second and more prominent research question asks how the convexity of CEOs’ equity portfolios affects their risk-taking decisions. The intuition that underlies GMM’s theoretical predictions and empirical tests of this question is that CEOs with more convex equity holdings should be less sensitive to changes in firm risk. Following an unexpected increase in left-tail risk, CEOs with more convex equity holdings should be more tolerant of their firm’s new, higher level of risk and, consequently, make fewer adjustments to firm risk. Thus, in contrast to the traditional empirical approach that asks whether convex equity contracts cause CEOs to influence their firm’s risk, GMM examine a shock to firm risk and ask whether convex equity contracts make CEOs more tolerant of the change in firm risk and therefore cause them to not make changes to firm

∗This discussion has benefitted from discussions with John Core (editor), Ian Gow, Wayne Guay, David Larcker, Daniel Taylor, David Tsui, and Rahul Vashishtha, as well as from comments and questions from the participants at the 2012 Journal of Accounting and Economics Conference. I thank John Lin and Tanya Paul for their excellent research assistance. I gratefully acknowledge financial support from the Wharton School.

E-mail address: carms@wharton.upenn.edu

0165-4101/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jacceco.2013.11.003
risk. This research question is important because it speaks to the extent to which incentive-compensation contracts can be designed to mitigate risk-related agency problems, whereby risk averse and undiversified CEOs forego risky, positive net present value (NPV) projects. This particular agency problem is thought to be one of the largest sources of shareholder value destruction in publicly-traded corporations (Jensen and Ruback, 1983), so it is important to understand the extent to which it can be mitigated.

Although both research questions are important, as GMM correctly point out, the endogenous relation between incentive-compensation contracts and the features of executives’ contracting environments—including current and anticipated firm risk—make it challenging to identify the causal effect of one on the other. As I discuss in more detail in Section 2, GMM’s research setting is well-suited for providing insight regarding the design and adjustment of CEOs’ equity incentives. In particular, if the shock to firms’ left-tail risk and investment opportunities is exogenous, it can be used to identify how CEOs’ contracts are redesigned—by actions of both boards and CEOs themselves—to accommodate the new contracting environment.

For a variety of reasons that I discuss in Section 2, GMM’s research setting is less amenable for providing inferences about the causal effect of CEOs’ equity incentives on firm risk. In essence, GMM’s research design represents an innovative and clever attempt to overcome the difficult task of finding valid instruments for CEO equity incentives (i.e., Vega and Delta). In particular, they use a shock that increases firms’ left-tail risk and reduces their investment opportunities to examine how CEOs’ equity incentives affect their risk-taking decisions following the shock. However, because GMM cannot observe CEOs’ equity incentives after the shock, they use CEOs’ equity incentives measured before the shock as proxies or instruments, which involves considerable measurement error. Although CEOs’ contracts (e.g., the number of options) may be slow to change following a shock to firm risk, the incentives provided by their contracts (e.g., Vega, or the sensitivity of these options to risk) will change because they are a function of, among other things, firm risk. Moreover, since the direction of the change in CEOs’ risk-taking incentives is theoretically ambiguous, not all CEOs will necessarily want to decrease firm risk following the shock. In addition, since the shock also affects firms’ investment opportunities, it will alter the risk-return tradeoff faced by CEOs, which further addles predictions about CEOs’ risk-taking decisions following the shock. Consequently, for these and other reasons that I discuss, GMM’s results related to the causal effect of CEOs’ equity incentives on firm risk are less persuasive.

Nevertheless, GMM’s research setting and the intuition behind their theoretical predictions represent a novel attempt to provide convincing causal evidence about an important but challenging research question. GMM’s study is also commendable because it explicitly recognizes that providing credible causal evidence about this important research question calls for an innovative research design. I build on these themes in Section 3 by discussing how causal inferences depend crucially on the specific causal effect that is identified by a particular instrumental variable or, in GMM’s case, by a particular “natural experiment.” I also discuss how causal inference in non-experimental settings is a product of both the data coupled with—ideally explicit and theoretically justified—identifying assumptions. I conclude with a brief discussion of partial identification and sensitivity analysis, which are a class of techniques that explicitly acknowledge this tradeoff between data and identifying assumptions that is inherent in causal inference. These techniques provide a promising set of tools that may be useful for providing credible causal evidence about GMM’s and other important, yet challenging research questions.

2. GMM’s research design and identification strategy

2.1. CEOs’ incentive-compensation contracts as a function of firm risk

GMM’s first research question relates to the design of incentive-compensation contracts and, in particular, how these contracts and the risk-taking incentives they provide are changed in response to changes in the contracting environment.¹ GMM argue that “after-tail risk increases, there are two primary ways for a manager’s financial exposure to the firm’s stock price and volatility to change—changes initiated by the company’s board and changes initiated by the manager” (p. 19). GMM test for changes instigated by the board by examining the Vega and Delta provided by annual equity grants and changes instigated by managers by examining their stock sales and option exercises.²

In Table 2, GMM report that CEOs at firms that have a sudden increase in their left-tail risk because of exposure to a carcinogenic substance (i.e., “exposed firms”) receive, on average, between $7,600 and $10,000 less Vega and between $12,900 and $25,200 less Delta from their annual equity grants than their counterparts at non-exposed firms. They also find that CEOs of exposed firms, on average, exercise between $790,800 and $1,184,000 more of their vested options following

¹ Note that “are changed” is italicized to emphasize how GMM’s tests of their first research question examine deliberate changes that are made by boards and CEOs. However, as I discuss in more detail below, it is important to note that the incentives provided by an existing contract will change when firm risk (or other parameters) change.

² GMM adopt the standard definitions of Vega and Delta from the empirical equity incentives literature (e.g., Guay, 1999; Core and Guay, 1999; and others). A CEO’s equity portfolio vega, or Vega, is defined as the change in the risk-neutral (i.e., Black-Scholes) value of the CEO’s option portfolio for a 0.01 change in the standard deviation of the underlying stock returns. Following Guay (1999) and others, GMM assume that Vega from a CEO’s stockholdings is negligible. A CEO’s equity portfolio delta, or Delta, is defined as the change in the risk-neutral value of the CEO’s equity portfolio—both stock and options—for a 1% change in the price of the underlying stock.
exposure than their counterparts at non-exposed firms. Based on this evidence, GMM conclude that “boards do in fact modify the incentive structure of managers’ annual compensation after tail risk increases” and that “managers of exposed firms also take actions to directly reduce their own financial exposure to their firms’ risk” (pp. 19 and 20). GMM also examine how the portfolio Vega and Delta of CEOs at exposed firms change following exposure. In Table 3 they report that the average portfolio Vega and Delta of CEOs at exposed firms declines nearly monotonically in the years following exposure. In particular, equity portfolio Vega (Delta) falls by an average of $24,600 and $55,310 ($181,700 and $483,800) two and five years following exposure, respectively.

GMM’s conclusion that CEOs’ equity incentives are slow to adjust following a relatively large increase in firm risk is at odds with their evidence in Table 2, which shows that boards and CEOs begin adjusting CEOs’ equity portfolios relatively soon after the shock. Intuitively, it is unclear how boards provide fewer equity incentives through annual stock and option grants and, at the same time, CEOs also work to reduce their equity incentives by exercising options and selling shares of stock, yet there is no (statistically) discernable effect of their collective actions on CEOs’ equity incentives. GMM attempt to reconcile this inconsistency between their two sets of results by suggesting that “vesting schedules restrict the changes available to a manager” and that boards may be similarly constrained “because it is arguably costly for firms to modify securities that have already been awarded.” Although both explanations are plausible and may explain why it takes time to adjust CEOs’ equity incentives in general, these explanations are less satisfying in GMM’s research setting, where the CEOs of interest experience a relatively large and sudden change to their contracting environment. Even though it may be costly for boards to modify the terms of previously granted securities to better suit the new contracting environment, it is also likely very costly for boards not to do so. If boards allow sub-optimal risk-taking incentives to persist following a large change in the contracting environment, they should expect CEOs to make risk choices that are inconsistent with shareholders’ interests. Thus, although the basis for GMM’s test of their second research question is an explicit assumption that incentives adjust slowly, the assumption calls into question whether these CEOs’ incentives were optimal before the shock to firm risk.

Moreover, although it is “costly for firms to modify securities that have already been awarded,” if the benefits are great enough, there are ways for boards to do so. One way is to reprice “underwater” stock options that are far out-of-the-money (e.g., Carter and Lynch, 2001). Another more extreme mechanism is to terminate CEOs whose incentives are not appropriate for the new contracting environment. If it is costly to adjust the incumbent CEO’s incentives and if there are other benefits to terminating the CEO (e.g., the CEO is performing poorly), it may be efficient to replace the CEO and provide the incoming CEO with appropriate incentives. While terminating a CEO is very costly, one benefit is that the board can write a new incentive-compensation contract with the incoming CEO.

In Table 4, GMM present evidence that suggests that some CEOs are, in fact, replaced following a large increase in firm risk and that their replacements have significantly less portfolio Vega than both their predecessors and replacement CEOs at firms that did not experience an increase in risk. These results are consistent with the idea that it is more costly to adjust an incumbent CEO’s incentive-compensation contract to accommodate the firm’s increased risk and diminished investment opportunities. In these cases, it is more efficient for the board to terminate the CEO and provide the replacement CEO with appropriate incentives.3

Recognizing that CEO termination is an incentive-compacting mechanism also highlights the importance of incentives other than equity and considering how they relate to the firm’s contracting environment. For example, if a change in risk alters a CEO’s probability of termination, this can have a powerful effect on the CEO’s risk-taking incentives going forward. Although GMM, as well as most other studies of equity-based risk-taking incentives, ignore termination and other non-equity sources of incentives, a more comprehensive research design would account for the joint effect of all of the relevant mechanisms that provide CEOs with risk-taking incentives.

2.2. Firm risk as a function of CEOs’ incentive-compensation contracts

GMM adopt a novel research design to examine whether CEOs’ equity incentives have an effect on firm risk. Before discussing GMM’s research design, I briefly review the research design that has been used in most empirical risk-taking studies, and is based on the following specification.

\[
\text{FirmRisk}_{it} = \beta_0 + \beta_1 \text{Delta}_{it-1} + \beta_2 \text{Vega}_{it-1} + \text{Controls}_i + \epsilon_{it}
\]  

\(FirmRisk\) is either a proxy for the riskiness of specific managerial decisions (e.g., research and development expenditures, firm leverage, and diversifying acquisitions) or a comprehensive measure of firm risk (e.g., stock return volatility), \(Delta\) and \(Vega\) are CEOs’ equity portfolio delta and vega, respectively, \(Controls\) includes variables that are also thought to be related to firm risk (e.g., size and growth options), and the subscripts \(i\) and \(t\) index the firm and year, respectively. Studies that adopt this specification include Guay (1999), Rajgopal and Shevlin (2002), Coles et al. (2006), Lewellen (2006), Low (2009), and

---

3 It is possible that some CEOs recognized the downside of using the ultimately carcinogenic substance, but continued to use it, perhaps because of their equity risk-taking incentives (i.e., the substance being labeled as a carcinogen represents a negative realization of a risky choice). If this risky choice were at least in part encouraged by these CEOs’ equity incentives, the shock to firm risk would not be exogenous. This concern is somewhat mitigated in GMM since they do not include CEOs who are terminated soon after exposure to the carcinogen becomes known. This concern could also be addressed with sensitivity analysis discussed in Section 3.
Armstrong and Vashishtha (2012). These studies primarily focus on Vega as the source of CEOs’ equity risk-taking incentives, but they also consider Delta, which also influences risk-averse CEOs’ attitude toward risk.

Most of these studies—as well as GMM—acknowledge that there is likely to be an endogenous relation between CEOs’ equity incentives and firm risk. To identify the causal effect of Vega and Delta on FirmRisk, these studies typically use instrumental variables as an attempt to address these endogeneity concerns. An ideal instrumental variable (1) induces variation in Vega and Delta and (2) has no direct effect on FirmRisk. An instrument that satisfies both conditions can be thought of as mimicking a randomized experiment by inducing “controlled variation” in the endogenous variable (i.e., Vega), which is then used to identify its causal effect on the outcome of interest. However, as GMM and others note, the primary challenge with this approach is the inherent difficulty in knowing whether a potential instrument satisfies the second condition, which is known as the exclusion restriction. Accordingly, it is useful to consider alternative research designs that may also be capable of identifying the causal effect of Vega on firm risk.

2.2.1. GMM’s theoretical predictions and research design

In contrast to the traditional approach, GMM seek to estimate the following theoretical model.

$$\text{Change in FirmRisk}_{it} = b_0 + b_1 \text{Exposure}_{it} + b_2 \text{Post}_{it} \cdot \Delta \text{Exposure}_{it} + b_3 \text{Post}_{it} \cdot \text{Vega}_{it} + \text{Controls}_{it} + \epsilon_{it}$$

(2)

where Post-Exposure Delta and Vega are CEOs’ portfolio delta and vega immediately after it becomes known that their firms were exposed to carcinogens. An ideal research design would identify the dates on which exposure became known and isolate the accompanying changes in price and volatility. These changes in price and volatility, which represent exogenous variation in two of the underlying parameters of Vega, would be used to calculate CEOs’ Vega (and Delta) immediately after exposure. These values would serve as an instrument for Vega along the lines of the traditional research design given by Eq. (1).

Since GMM are unable to directly observe post-exposure Vega and Delta, they use pre-exposure Vega and Delta as proxies, or instruments, and instead estimate the following model (which corresponds to their Eq. (5)).

$$\text{FirmRisk}_{it} = \gamma_0 + \gamma_1 \text{Exposure}_{it} + \gamma_2 \text{Exposure}_{it} \cdot \Delta \text{Exposure}_{it-1} + \gamma_3 \text{Exposure}_{it} \cdot \text{Vega}_{it-1} + \text{Controls}_{it} + \epsilon_{it}$$

(3)

GMM also include firm-cohort fixed effects in this specification, in part, as an attempt to control for time-invariant unobservable factors that may also influence CEOs’ risk-taking decisions (e.g., CEO risk aversion).

This research design, which differs from the traditional specification given by Eq. (1), focuses “on how vega is related to responses to unanticipated increases in risk, rather than the absolute levels of risk-taking” (p. 13, emphasis original). GMM’s implicit identification strategy is to use “ex ante vega as a proxy for the (unobserved) vega managers face after risk increases— but before they respond” (p. 25, emphasis original). Similarly, ex ante, or pre-exposure delta is used as a proxy, or instrument, for post-exposure delta, which also influences CEOs’ risk-taking decisions following exposure to a carcinogen.

GMM’s research design represents an innovative and clever attempt to overcome the difficult task of finding valid instruments for Vega and Delta that is inherent in the traditional research design. Nevertheless, for reasons that I now discuss, GMM’s research design also entails a number of potentially limiting concerns, many of which are common to virtually every empirical risk-taking study.

2.2.2. Concerns with GMM’s research design

A basic concern with GMM’s research design is their use of CEOs' pre-exposure Vega as a proxy for their post-exposure Vega. By using pre-exposure Vega as a proxy for post-exposure Vega, GMM’s research design ignores differences in the strength of the shock (i.e., cross-sectional differences in the change in volatility and stock price), to the extent that exposure to a carcinogen has differential effects on firms’ stock price and volatility, these differences, or “heterogeneous treatments,” could be used to facilitate sharper theoretical predictions and more powerful empirical tests (e.g., risk-averse CEOs who experience larger increases in volatility may respond more aggressively in their subsequent risk-taking decisions).

Second, although GMM’s research design does not formally rely on instrumental variables estimation, their identification strategy relies on essentially the same assumptions (see pp. 13 and 14), which effectively allows them to use pre-exposure Vega and Delta as instruments for their post-exposure counterparts. Intuitively, however, it is unclear whether GMM’s research design can separately identify the causal effects of two endogenous variables (i.e., Vega and Delta) from a single source of exogenous variation (i.e., Exposure). Separating Exposure into its effects on stock price and volatility—as well as incorporating heterogeneity in these effects—would yield exogenous variation along two channels that, in tandem, would facilitate identifying the separate effects of Vega and Delta on CEOs’ risk-taking decisions.

Third, pre-exposure is a noisy proxy. As GMM discuss, even if a CEO’s existing incentive-compensation contract does not change (or does not change quickly) following a shock to firm risk, the incentives provided by the contract will change because both Vega and Delta are functions of firm risk. Fig. 1 illustrates how the sign of the partial derivative of Vega with respect to firm risk is ambiguous and, consequently, why a CEO’s Vega prior to an unexpected change in risk is not necessarily a reliable proxy for the CEO’s Vega after the change in risk. In particular, Panels A and B plot the risk-neutral Black-Scholes Vega and risk-averse Certainty Equivalent Vega of an option as a function of stock price and stock return volatility. These panels correspond to GMM’s empirical measure and theoretical construct of risk-taking incentives, respectively. Both panels clearly show that, regardless of how it is measured, an option’s Vega depends on stock price and volatility. Moreover, the partial derivative of Vega is not even always monotonic.
GMM acknowledge the relation between risk and stock price and Vega when they note that “the partial derivatives of Vega with respect to volatility and stock price … can be positive or negative depending on other parameters” (p. 11). However, GMM explain that they rely on pre-exposure Vega as a proxy for post-exposure Vega because they do not know exactly when exposure becomes known during the year. GMM argue that pre-exposure Vega is a “reasonable proxy” for post-exposure Vega because “the convexity of the manager’s payoff prior to the unexpected change in risk … is a positive predictor of convexity immediately after the change in risk,” as evidenced by the high correlation between the two (0.77).

Although pre-exposure Vega may be highly correlated with post-exposure Vega, using pre-exposure Black-Scholes Vega as a proxy for post-exposure risk-adjusted (Certainty Equivalent) Vega introduces a second source of measurement error: the first source discussed above comes from using values measured pre- rather than post-exposure, and the second source comes from using risk-neutral (i.e., Black-Scholes) rather than risk-adjusted (i.e., Certainty Equivalent) values. The reported correlation between pre- and post-exposure Vega (i.e., 0.77) relates to the risk-neutral Black-Scholes Vega, which is the empirical measure, and not the Certainty Equivalent Vega, which is the underlying theoretical construct of interest. Depending on CEOs’ unobservable risk-aversion and other parameters (e.g., CEO wealth), the Black-Scholes Vega and the Certainty Equivalent Vega can have different sensitivities to stock price and volatility. Although a high correlation between CEOs’ pre- and post-exposure Black-Scholes Vega is comforting, it only speaks to measurement error in Vega as a proxy for post-exposure Vega, and does not necessarily imply that CEOs’ pre- and post-exposure Certainty Equivalent Vega are highly correlated. It is important to note, however, that most studies also use risk-neutral (i.e., Black-Scholes) Vega and Delta as empirical proxies for CEOs’ Certainty Equivalent Vega and Delta, which are the theoretical constructs of interest, but are very difficult to measure.

Second, although pre- and post-exposure (Black-Scholes) Vega may be highly correlated in the cross-section, this does not seem to correspond to the source of identifying variation in GMM’s research design. In particular, GMM’s primary specification (i.e., Eq. (3) above) includes the interactions Exposure × Vega, and Exposure × Delta, along with firm-cohort fixed effects. It seems as if much of the cross-sectional variation in Vega (or, more accurately, Exposure × Vega) is captured by the firm-cohort fixed effects and Exposure × Vega only captures the extent to which the Vega of CEOs at exposed firms differs from their firm-cohort average.

2.2.3. Concerns with GMM’s theoretical predictions

There are also concerns related to GMM’s theoretical predictions and the intuition that underlies their notion of “risk-curvature.” Recall that a risk averse and undiversified CEO may pass up certain positive NPV projects that entail “too much” idiosyncratic risk even though diversified shareholders, who are indifferent towards idiosyncratic risk, would benefit from these projects. The board, acting on behalf of shareholders, may be able to mitigate this risk-related agency problem by designing an incentive-compensation contract that makes the CEO’s induced utility function “less concave” and more linear like that of shareholders. In other words, by “convexifying” the CEO’s inherently concave utility function, the incentive-compensation contract will make the CEO less averse to risk when evaluating the risk-return tradeoff associated with the

---

Fig. 1. Vega as a function of stock price and stock return volatility. This figure plots the risk-neutral Black-Scholes Vega (Panel A) and the Certainty Equivalent Vega (Panel B) of a portfolio of 10,000 options for different values of the underlying stock price and stock return volatility. In both panels, the stock option has an exercise price of $50 and five years until expiry and the risk-free rate is 5%. The Black-Scholes Option Vega in Panel A is calculated as the partial derivative of the Black-Scholes option value with respect to volatility (i.e., sigma). The Certainty Equivalent Option Vega in Panel B is calculated as the change in a risk-averse agent’s Certainty Equivalent for a 0.01 change in volatility (i.e., sigma). The risk-averse agent has power utility with a coefficient of relative risk aversion of two and fixed (i.e., non-stochastic) wealth of $20,000,000.
firm’s investment opportunities. In the limit, it may be possible for the board to design an incentive-compensation contract that makes the CEO indifferent towards risk, thereby resolving the risk-related agency problem.

Since GMM’s notion of “risk-curvature” relates to a CEO’s induced utility function, it is a product of both the CEO’s inherent utility function and the CEO’s incentive-compensation contract. In GMM’s numerical examples, the attributes of the CEO (e.g., risk-aversion and wealth) are held constant, so differences in CEOs’ risk-curvature come entirely from differences in their equity holdings. GMM’s implicit assumption that other characteristics (e.g., utility functions, risk aversion, and wealth) are identical across CEOs is imported into their empirical tests. However, assuming that CEOs are identical in the cross-section and that any differences in their aversion to risk are due entirely to differences in their equity holdings implicitly ignores the endogenous matching of CEOs and incentive-compensation contracts on the basis of certain CEO attributes (e.g., risk aversion). To the extent this assumption is violated, GMM’s theoretical predictions represent an approximation of the cross-section. GMM acknowledge this concern and develop robustness tests to account for cross-sectional differences in unobservable CEO characteristics. Although these tests attempt to control for unobservable cross-sectional differences in CEO attributes that also influence their risk-taking decisions (e.g., wealth and risk aversion), developing reliable proxies of these largely unobservable constructs poses a difficult empirical challenge. Importantly, this concern is not unique to GMM, but is pervasive in the empirical contracting literature, and is one for which no reliable and widely accepted solution yet exists.

The final concern with GMM’s theoretical predictions is that, although the “risk-curvature” of a specific incentive-compensation contract is not a function of the firm’s investment opportunities, the optimal level of risk-curvature should be largely determined by the firm’s investment opportunities. If, as GMM argue, exposure to a carcinogenic substance alters both a firm’s left-tail risk and its investment opportunities, it will change a CEO’s menu of available projects (i.e., the risk-return tradeoff). Thus, a change in the firm’s investment opportunities introduces an additional complication that is not apparent from GMM’s numerical examples, which hold the firm’s risk-return tradeoff constant. Without knowing the expected payoffs that accompany the various levels of risk from which CEOs can choose, the direction in which CEOs would want to change their firm’s risk following the shock is ambiguous.  

3. Suggestions for future empirical research

GMM’s second research question regarding whether CEOs’ equity incentives cause them to take risks is an important and extensively examined question in corporate finance. As GMM point out, recent studies tend to rely on relatively sophisticated econometric techniques since establishing a contribution generally requires providing causal evidence. Because of the relatively stringent threshold imposed by GMM’s research question, it is well-suited for discussing some of the subtle but increasingly important aspects of causal inference. Moreover, the joint endogenous relation between incentive-compensation contracts and firm risk that is inherent in GMM’s two research questions is useful for illustrating how partial identification techniques may be useful for advancing this and other important research questions when perfectly exogenous instruments—or the appropriate natural experiments—are not available.

3.1. Graphical representation of an ideal instrument

Although instrumental variables are usually characterized by their covariance with the endogenous variable(s) and outcome variable of interest, a graphical representation is useful for illustrating the distinction between various causal effects that instruments may be able to identify.  

Fig. 2, Panel A illustrates the case of an ideal instrument, Z, that is both strong in the sense that it has a large covariance with endogenous variable of interest, X, and “valid” in the sense that it perfectly satisfies the exclusion restriction, which requires no direct effect of the instrument on the outcome of interest, Y. The strength of the instrument is illustrated by its large area of overlap with the endogenous variable. The validity of the instrument is illustrated by the only area of overlap between the instrument and the outcome of interest coming through the endogenous variable.

Even though the instrument in Panel A is both strong and perfectly exogenous, it is important to note that it only identifies the effect of the endogenous variable on the outcome within the area of overlap (i.e., the covariation between the endogenous variable and the instrument). As Murnane and Willett (2011, 274) note, “... IV estimates capitalize only on variation in the [endogenous variable] that ‘falls within’ or is ‘sensitive to’ variation in the instrument.” When there are so-called heterogeneous treatment effects (i.e., where the effect of the endogenous variable on the outcome is different across

---

(footnote continued)

incentive-compensation contract that relates a certain performance measure (e.g., stock price) to the agent’s payoff (and therefore expected utility). See Ross (2004) for a more detailed discussion of induced utility functions.

6 For example, a CEO may want to increase firm risk (following a shock that increased the firm’s risk and reduced its investment opportunities) if the shock pushes the firm close to financial distress. In that case, even a CEO with only shares of stock may actually prefer an increase in firm risk because of the option value of equity.

7 A number of authors have used Venn diagrams as a way to depict the intuition behind instrumental variables. The Venn diagrams and the accompanying discussion are based on those presented by Murnane and Willett (2011).
observations), the effect estimated by a particular instrument should be regarded as a local average treatment effect, or LATE (Imbens and Angrist, 1994).

Recognizing that estimated treatment effects are typically local means that any estimate may not be representative of a global causal effect that prevails in the broader population. For example, a particular research question might call for an estimate of the average treatment effect, or ATE, which is the average causal effect that would obtain if the treatment were applied randomly throughout the population, rather than only to a particular subsample. The average treatment effect can be very different from any particular local average treatment effect, so generalizability is a major concern if the research question relates to a treatment effect other than the local one that is identified by a particular instrument. This concern arises in GMM’s research setting if their sample of firms that experience a shock to their left-tail risk is not representative of the broader population of firms. In this case, since GMM’s estimate of the causal effect of Vega on firm risk represents a local average treatment effect, it may not generalize to support broader inferences about more general casual effects of Vega on firm risk.

Another important and related consequence of distinguishing between various treatment effects is that different instruments (or “natural experiments”) estimate different treatment effects (i.e., different LATEs). Thus, any single study is unlikely to prove definitive about the existence and magnitude of a particular causal effect. Instead, causal inferences about a particular research question are developed by accumulating and jointly evaluating the collective evidence across multiple studies that identify different treatment effects. Thus, a contribution of GMM is to add to the collective evidence about the causal effect of CEOs’ equity incentives on firm risk.

---

**Fig. 2.** Graphical representation of alternative instrumental variables. This figure presents a graphical depiction of the properties of alternative instrumental variables. In all four panels, the outcome of interest, labeled Y, and the endogenous variable of interest, labeled X, have the same relation as illustrated by the identical area of overlap. Panel A presents an instrumental variable, labeled Z, that is strong by virtue of its relatively large overlap with X and perfectly exogenous by virtue of having no direct overlap with Y. Panel B presents an instrumental variable that is weak by virtue of its relatively small overlap with X and perfectly exogenous by virtue of having no direct overlap with Y. Panel C presents another instrumental variable that is weak by virtue of its relatively small overlap with X and perfectly exogenous by virtue of having no direct overlap with Y. Panel D presents an instrumental variable that is strong by virtue of its relatively large overlap with X, but is not perfectly exogenous by virtue of having an area of direct overlap with Y that is not shared by X.

---

8 Imbens and Rosenbaum (2005) note that since different instrumental variables produce different manipulation in the endogenous variable of interest, even if none of the instruments across a collection of studies is certain to be completely exogenous, but they are at least plausible and not biased in the same direction, the collective evidence across these studies can be quite persuasive.
3.2. Graphical representation of a “weak” and perfectly exogenous instrument

Although a strong and perfectly exogenous instrument will deliver a consistent estimate of a particular local average treatment effect, instruments that simultaneously satisfy both conditions are rare in most corporate finance settings. Moreover, even if such an instrument is available, the local average treatment effect that it identifies may not speak to the research question of interest. Fig. 2, Panel B illustrates a more realistic scenario in which the instrument is perfectly exogenous, but is “weak” in the sense that it has only a modest covariance with the endogenous variable of interest. As in Panel A, the lack of any direct overlap between the instrument and the outcome indicates that the former has no direct effect on the latter, and is therefore exogenous. However, unlike in Panel A, there is only a small area of overlap between the instrument and the endogenous variable, which indicates that the instrument induces only modest variation in the endogenous variable.

Many of the concerns related to weak instruments are relatively well known and have been discussed by a number of authors, including Bound et al. (1993, 1995), Staiger and Stock (1997), and Stock et al. (2002). A major concern with weak instruments is that the small sample bias that is inherent in instrumental variables estimators is exacerbated when the instrument is only weakly related to the endogenous variable of interest (e.g., Leamer, 2010). Although GMM do not formulate their empirical tests using instrumental variables, as discussed above their identification strategy relies on assumptions that are largely the same as those required for instrumental variables estimation. Accordingly, the concern about small sample bias seems to apply to GMM based on the relatively small sizes in their tests.

Panel B illustrates another concern with weak instruments that can be especially problematic when there are heterogeneous treatment effects. Because the area of overlap between a weak instrument and the endogenous variable is relatively small, its common overlap with the outcome is also necessarily small. Thus, the aforementioned concern about the generalizability of a particular local average treatment effect can be particularly heightened in the presence of a weak instrument. Panel C presents a similar, but even more extreme case in which the instrument induces variation in the endogenous variable of interest, but there is no associated effect on the outcome in the particular subpopulation that is affected by the instrument. In this case, it would be incorrect to infer that the endogenous variable has no effect on the outcome based on the lack of a local effect estimated from this particular instrument.

3.3. Partial identification with a “strong” but not perfectly exogenous instrument

Strong and perfectly exogenous instruments are rare in most corporate finance settings, due in part to the presumption of rational optimizing behavior by the relevant economic actors (e.g., boards, shareholders, and CEOs). Manski (2003), Tamer (2010), and others discuss methods that are collectively referred to as “partial identification” techniques, which can be used to draw inferences from instruments that are not perfectly exogenous, but are sufficiently strong. The traditional econometric approach characterizes identification as a binary event where a potential instrument either satisfies the exclusion restriction (i.e., \( \text{cov}(Z, \varepsilon) = 0 \)) or does not (i.e., \( \text{cov}(Z, \varepsilon) \neq 0 \)), and the parameter that describes the causal effect is either “identified” or “not identified,” respectively. Partial identification techniques generalize this framework to accommodate potential instruments that are not perfectly exogenous, but where something may be known about the direction and/or magnitude of the correlation between the instrument and the structural error.

Panel D of Fig. 2 presents an example of a strong, but not perfectly exogenous instrument that is a candidate for partial identification techniques. The area of direct overlap between the instrument and the outcome that is not shared by the endogenous variable indicates that the instrument is not completely exogenous. The standard econometric approach would either discard this instrument or assume that the instrument is perfectly exogenous and proceed to draw inferences based on this (invalid) assumption. In contrast to these two extremes, partial identification techniques recognize that if something is known about the direct overlap between the instrument and the outcome, this knowledge can be used to isolate and utilize the variation in the instrument that is exogenous.

One way that partial identification techniques might be applied in GMM’s research setting would be to use Exposure as an instrument for Vega in the traditional research design shown in Eq. (1). As previously discussed in the context of Fig. 1, Exposure induces variation in Vega through its effect on stock price and volatility, which indicates that it is relevant (i.e., \( \text{cov}(Z, X) \neq 0 \)). In order to use Exposure as an instrument, the traditional approach also requires invoking the untestable assumption that it is perfectly exogenous with respect to future firm risk (i.e., \( \text{cov}(Z, \varepsilon) = 0 \)). Rather than invoke this strong assumption, partial identification may instead start with the weaker assumption that Exposure and future firm risk are positively correlated (i.e., \( \text{cov}(Z, \varepsilon) > 0 \)). Because this assumption is relatively uninformative, the resulting inferences would likely come in the form of a set of values for the parameter of interest (i.e., “set identification”) rather than a conventional point estimate (i.e., “point identification”). The range of values in which the parameter lies could be narrowed by invoking stronger assumptions about the relation between Exposure and future firm risk (e.g., \( 0 < \text{cov}(Z, \varepsilon) < 0.20 \)). In the limit, partial...
identification encompasses the special case of the traditional instrumental variables estimator, which delivers a point estimate of the parameter by invoking the strongest possible assumption (i.e., \( \text{cov}(Z,e) = 0 \)). Armstrong et al. (2013) rely on partial identification techniques when examining the effect of shareholder voting on CEO pay because, although strong, their instrument for shareholder voting (i.e., voting recommendations by proxy advisors) may not be perfectly exogenous.

Although there are some subtle differences, it is important to note that partial identification is closely related to sensitivity or “bounding” analyses that have been developed by Frank (2000), Rosenbaum (2002) and others, and have been used in the accounting literature by Larcker and Rusticus (2010) and Armstrong et al. (2010). Manski (2003) describes sensitivity analysis as “mathematically complementary approach” to partial identification that “begin[s] with some point-identifying assumption and examine[s] how identification decays as this assumption is weakened in specific ways.” Sensitivity analysis acknowledges that the independent variable of interest (e.g., Vega in Eq. (2) above) may not be completely exogenous (i.e., \( \text{cov}(X,e) \neq 0 \)) and, consequently, the point estimate of its effect may be biased. Sensitivity analysis then assesses how identification change from progressively relaxing the initial assumption that the variable is perfectly exogenous. In the case of GMM, sensitivity analysis would acknowledge that Vega in Eq. (2) above may not be completely exogenous, and would ask how correlated Vega would have to be with the unobservable structural error to alter GMM’s inference that Vega causes CEOs to take risk. In the context of Fig. 2, this essentially amounts to increasing the area of direct overlap between the instrument \((Z)\) and the outcome of interest \((Y)\). Core (2010) discusses how sensitivity analysis is relatively simple to perform and should be a standard part of accounting research.

This example illustrates two of the more appealing aspects of partial identification techniques vis-à-vis traditional identification strategies. First, as Manski (2005) notes, partial identification techniques explicitly separate the inferential problem into its statistical and identification components. He further notes that although the former can typically be alleviated by drawing more observations from the sampling process, the latter “cannot be solved by gathering more of the same kind of data.” Instead, overcoming identification problems requires “invoking stronger assumptions” or “initiating new sampling processes that yield different kinds of data.”

Second, because partial identification techniques admit instruments that are less than perfectly exogenous, these methods greatly expand the set of feasible instruments that can be used to provide causal inferences when data from “natural experiments” is not available. Moreover, partial identification techniques explicitly highlight the cost–benefit tradeoff that is inherent in invoking the identifying assumptions that are required for causal inferences. The benefit is that stronger assumptions deliver more precise inferences (i.e., a smaller identification set). These inferences, however, come at the cost of what Manski (2003) refers to as “the law of decreasing credibility: The credibility of inference decreases with the strength of the assumptions maintained.” To the extent that identifying assumptions are based on sound economic theory, partial identification provides a framework for combining data and theory as substitutes for producing causal inferences.

4. Conclusion

Gormley, Matsa, and Milbourn (in this issue) examine two related research questions that are interesting and important, but at the same time challenging because of the endogenous relation between CEOs’ equity incentives and firm risk. GMM attempt to overcome these challenges by studying a novel research setting in which certain firms experience a large change in (i.e., “shock” to) their left-tail risk and their investment opportunities. For the reasons discussed above, to the extent this shock produced material changes in the contracting environment, it is well suited for providing insight into GMM’s first research question regarding how features of the contracting environment influence the design and subsequent adjustment of CEOs’ incentive-compensation contracts.

In contrast, GMM’s evidence about their second, more prominent research question regarding whether CEOs’ equity incentives cause them to take risk is less persuasive. GMM attempt to overcome the difficult task of finding valid instruments for CEOs’ equity incentives by using the shock to firm risk. Because they cannot observe CEOs’ equity incentives (i.e., Vega and Delta) following the shock, they use CEOs’ equity incentives prior to the shock as proxies or instruments. However, this approximation entails measurement error. The efficacy of their innovative, but somewhat unconventional research design depends on the direction and magnitude of this measurement error. As I discuss, there are a number of reasons that CEOs’ equity incentives change following an unexpected change in firm risk. Not only do boards make adjustments via equity grants, but CEOs also make adjustments through their option exercises and stock sales. In addition to these two deliberate choices that alter equity incentives, a shock to firm risk (and stock price) causes changes in the incentives provided by CEOs’ existing equity portfolios. The collective effect of these various changes, coupled with the
additional change in firms’ investment opportunities—and therefore, the risk-return tradeoff faced by CEOs—add measurement error and complicate GMM’s theoretical predictions about CEOs’ risk-taking decisions following a shock to firm risk.

Although GMM’s results related to their second research question are less persuasive, their underlying research design represents an innovative attempt to advance our understanding of an important but challenging research question. As suggested by GMM’s novel research design, further progress in answering this and related research questions that are impeded by endogeneity concerns is likely to require new and innovative research designs and methods, such as partial identification and sensitivity analysis. GMM’s research setting is also useful for highlighting the importance of clearly articulating the specific causal effect (i.e., the local average treatment effect) that is identified by a particular instrument or natural experiment, and ensuring that the estimated effect speaks to the research question of interest. Overall, GMM is an important study for illustrating the challenges and subtleties that are inherent in providing credible causal—rather than simply descriptive—inferences about many of the important but unanswered corporate finance research questions.

References

Bound, J., Jaeger, D., Baker, R., 1993. The cure can be worse than the disease: a cautionary tale regarding instrumental variables. NBER technical working paper no. 137.
Lewellen, K., 2006. Financing decisions when managers are risk averse. J. Financ. Econ. 82, 551–589.