Strategic Management Journal
Strat. Mgmt. J., **38**: 1721–1740 (2017)
Published online EarlyView 9 March 2017 in Wiley Online Library (wileyonlinelibrary.com) DOI: 10.1002/smj.2632
Received 5 October 2015; *Final revision received 15 December 2016*

SKEW AND HEAVY-TAIL EFFECTS ON FIRM PERFORMANCE

SHIGE MAKINO[1]* and CHRISTINE M. CHAN[2]
[1] *Department of Management, The Chinese University of Hong Kong, Shatin, Hong Kong*
[2] *School of Business, The University of Hong Kong, Pokfulam, Hong Kong*

Research summary*: Most strategic management studies adopt an average-centered view that uses the central tendency to explain between-group variation in performance (i.e., performance differences between business units, firms, industries, and countries). In this study, we explain within-group variation using a variance-centered view that focuses on the peripheral characteristics of performance distributions as defined by skew and heavy tails (i.e., variance and kurtosis). Drawing on performance feedback theory, we hypothesize that successful firms tend to develop a positive skew in their performance distributions, which we call a "positive skew effect" in this study, and that heavy tails moderate this effect. Our analysis of the performance of a group of foreign affiliates provides general support for our hypotheses at both the firm and segment (industry and country) levels.*

Managerial summary*: Managers of multi-business firms use various approaches to improve the aggregate performance of their business units. Some expand the range of upper performance outliers (exploration) or reduce the range of lower outliers (downsizing); others improve the performance of current business units (exploitation). We find that firms with superior performance tend to have a balanced mix of the three approaches. We also find that segments (countries and industries) with higher mean performances provide environments that facilitate the entry of productive firms and the exit of unproductive firms and provide environments in which incumbents can further improve their performance by learning from others. We observe that successful firms and segments have a positive skew in their performance distributions, which we call a "positive skew effect."*

INTRODUCTION

The primary purpose of strategic management research is to identify the sources of variation in firm performance. Drawing on industrial organization theory and resource- and institution-based views, previous studies have identified a variety of "conditions" (e.g., industry structure, firm resources and capabilities, and institutions) that explain performance variation. However, these "conditions" do not sufficiently explain performance variation; studies examining the decomposition of the effects of business units, firms, industries, sub-national regions, and countries on performance variation have revealed that these effects explain at most 50% of the observed variation (Chan, Makino, and Isobe, 2010; Ma, Tong, and Fitza, 2013; Makino, Isobe, and Chan, 2004; McGahan and Porter, 1997; McGahan and Victer, 2010; Rumelt, 1991).

Keywords: analysis of outliers; heavy-tail effects; non-normal distribution; skew effect; variance-centered view
*Correspondence to: Shige Makino, Department of Management, Chinese University of Hong Kong, Shatin, N.T., Hong Kong. E-mail: makino@baf.cuhk.edu.hk

Underlying the conventional condition-based explanations is the assumption that once a theoretically defined condition is met, all of the firms will behave in a coherent manner (i.e., similar to average firms) and achieve a similar level of firm performance. We call this conventional perspective the *average-centered view*. This view implicitly assumes that average events (behavior or performance) in a given condition represent the essential characteristics of all of the events in the condition and that events that deviate from the average are "errors in theory." Researchers have attempted to decrease such errors by identifying omitted conditions (i.e., developing new theories or combining current theories). Although such attempts have certainly advanced our knowledge of strategic management, they often result in "an undue emphasis on the development of theory at the expense of research which observes and reports actual facts" (Hambrick, 2007: 1346).

The average-centered view, which is deeply embedded in strategic management studies, also makes the assumption that events are independent and will generate Gaussian (normal) distribution (McKelvey and Andriani, 2005). As the average-centered view focuses on the central (average) tendencies of the observed events, it considers outliers or residual values that deviate from the average as unimportant sources of information or even as noise in the descriptions of the characteristics of the observed events (Andriani and McKelvey, 2007, 2009). For example, a typical OLS imposes assumptions on the errors, including *homoscedasticity* (variance of the errors is the same across all levels of explanatory variables), *no autocorrelation* (the errors are independent of one another), and *normality* (the errors are normally distributed) with the zero average value. However, according to the behavioral tradition of management research, these assumptions are often not appropriate for strategic management practices, as the behavior of firms is typically characterized as interdependent, self-selected, and biased. There are three reasons for this characterization. First, a firm's strategic behavior can hardly be considered random or independent as it is (a) influenced by the firm's past behavior and the choices made by other firms (Cyert and March, 1963), (b) inherently "endogenous" to the firm's expected performance outcomes (Hamilton and Nickerson, 2003; Shaver, 1998), and (c) subject to bias and bounded rationality (Kahneman and Tversky, 1979; Simon, 1965) (i.e., a violation of the assumption of independence). Second, firms do not only choose behavior that follows the behavior of average firms; their behavior may deviate from that of average firms. Firms have different orientations in learning (i.e., exploitation and exploration) that result in variations in strategic behavior and hence in performance outcomes, even between firms in the same situations (March, 1991) (i.e., a violation of the assumption of homoscedasticity). Third, firms are biased toward behavior that produces superior performance. Firms have a natural incentive to retain high performing business units and improve or terminate underperforming business units (Porter, 1991), resulting in a non-symmetrical distribution of performance among the retained business units around the mean (i.e., a violation of the assumption of normality).

Although the average-centered view can explain how mean performance varies between conditions and can identify the conditions that lead to differences in the mean performance, which we call the *effects on the mean*, it fails to explain how the shape of performance distribution varies within conditions (i.e., the range and configuration of performance variation within conditions), which we call the *effects on the variance*. To achieve a precise understanding of the sources of strategic behavior and performance variation, the conventional average-centered view should be complemented by a view that explains *effects on the variance*, which we call the *variance-centered view*.

The variance-centered view can be extended in two directions to examine two different aspects of strategic management that the average-centered view cannot fully explain. First, it can be extended to identify factors that explain *effects on the variance*. This approach has been widely adopted by researchers examining inequities or disparities in social and economic behavior and outcomes (i.e., unequal regional economic growth within and between countries; and income, consumption, and expense disparities between individuals).[1] Second, it can be extended to explain the possible associations between *effects on the mean* and *effects on the variance*. The average-centered view

[1] There has been some methodological development in this approach. Western and Bloome (2009), for example, developed a method of examining $V(w)$, called "variance function regression," which estimates the effects of covariates on both the mean and variance (i.e., conditional variance measured by residual dispersion) of a dependent variable.

implicitly assumes that *effects on the variance* are symmetric at, constant across, and independent of *effects on the mean*. However, the behavioral tradition of management research suggests that this assumption is less likely to hold for strategic management practices. Although the analysis of associations of moments (mean, variance, skewness, and kurtosis[2]) has a long tradition in financial studies, it has received scant attention in strategic management research, aside from seminal studies of risk-return relationships in the industrial sectors (Andersen and Bettis, 2015; Andersen, Denrell, and Bettis, 2007; Bowman, 1980; Bromiley, 1991; Henkel, 2009; Miller and Leiblein, 1996; Wiggins and Ruefli, 2002).

This study, focusing on the second approach, proposes a performance distribution model that draws on performance feedback theory (Cyert and March, 1963). The key thrusts of the model are the hypotheses that the shape of the performance distributions of successful firms tends to be characterized by a positive skew at the firm and segment levels over the long run and that the tail characteristics, which are defined by variance and kurtosis in a performance distribution, have specific associations with the mean performance in skew situations. We call the former the *skew effect* and the latter the *heavy-tail effect* on the mean performance. To examine these hypotheses, we focus on two research settings: the performance distributions of business units (foreign affiliates) within firms (parent firms) and within segments (industries and countries). Specifically, we examine how the aggregate performance of foreign affiliates is associated with the shape of the performance distributions at the parent firm, industry, and host country levels.

In the next section, we introduce the key concepts of the variance-centered view. We then propose our performance distribution model and develop hypotheses to explain the relationship between *effects on the variance* and *effects on the mean*. Using performance data from 7,224 foreign affiliates for the 1997–2009 period, we examine how the mean and shape of the distributions of the annual performance of foreign affiliates co-vary among parent firms, industries, and host countries. Finally, we discuss the results and their implications for future research.

[2] Descriptions of the key terms are provided in the File S1.

PERFORMANCE DISTRIBUTION MODEL

Variance-centered view

An analysis of performance variation can be decomposed into two parts: mean performance that varies between conditions (i.e., *effects on the mean*) and performance distribution that varies within conditions (i.e., *effects on the variance*). This relationship is described in the following formula:

$$V(T) = V(b) + V(w),$$

where $V(T)$ represents the total variance of performance; $V(b)$ represents the between-condition variation (i.e., *effects on the mean*); and $V(w)$ represents the within-condition variation (i.e., *effects on the variance*). The average-centered view focuses on the conditions that explain $V(b)$ and assumes that the average performance represents the central tendency of the performance outcomes of the firms in a given condition and that those outcomes that deviate from the average are peripheral to the condition. In essence, the average-centered view assumes that $V(b)$ is sufficiently large compared to $V(w)$.

However, research suggests that, due to the effects of outliers, $V(w)$ is often too large to ignore for explaining in $V(T)$ and has a highly skewed distribution. Skewed distributions are frequently observed in business, economic, and social phenomena (Simon, 1955). Some recent studies of economic growth have shown that data from many emerging economies do not conform to the normal distribution model (Bekaert and Harvey, 2002; Chan, Isobe, and Makino, 2008; Harvey, 1995). Other studies of entrepreneurship have examined the 49 input and outcome variables used in theories of entrepreneurship and find that all of the variables except one are more accurately described as having a power law distribution, which takes the form of a lognormal distribution that is positively skewed with a wide range of outliers (Crawford et al., 2015). Financial researchers have reported that outliers have a substantial effect on asset return and allocation beyond the conventional mean-variance measures (e.g., Harvey et al., 2010; Harvey and Siddique, 2000; Kon, 1984; Patton, 2004; Xiong and Idzorek, 2011). Innovation research has also shown that the size distributions of the returns on inventions and innovations are highly skewed, with the top 10% of the sample representing over 90% of the total returns (Scherer and Harhoff, 2000).

DOI: 10.1002/smj

In a similar vein, recent strategic management research has cautioned against dependence on the average-centered view. These studies highlight the importance of examining extreme values, because firms' actions do not always regress toward the mean, but may also deviate from it (e.g., Andriani and McKelvey, 2007, 2009; Crawford et al., 2015; Hansen, Perry, and Reese, 2004; Henkel, 2009; McKelvey and Andriani, 2005; Wiggins and Ruefli, 2002). Baum and McKelvey (2006) argue that "statistical methods employed to address such problems are concerned primarily with what goes on at the center of a statistical distribution and do not pay particular attention to its tails, or in other words, the most extreme values at either the high or the low end" (pp. 124–125). The neglect of outliers inevitably leads to systematically biased theories of organizations and to a discrepancy between research and practice (Denrell, 2003),[3] because practitioners pay more attention to actions that lead to extremes than to actions that lead to average performance (Andriani and McKelvey, 2007). Although researchers have recently attempted to develop alternative methodologies for evaluating extremes by incorporating the insights of the power law distribution into their analyses (e.g., Baum and McKelvey, 2006; Crawford et al., 2015; Hansen et al., 2004; Johnson, Bellman, and Lohse, 2003; Zanini, 2008),[4] these methodologies have received limited attention in the mainstream strategic management research, which has been dominated by the average-centered view.

As a complement to the average-centered view, the proposed variance-centered view explains the portion of $V(T)$ that cannot be fully explained by $V(b)$. The variance-centered view provides two approaches for understanding the property of $V(w)$. The first approach identifies the conditions that lead to deviations from a normal distribution and explains why the shapes of the distribution vary within boundary conditions. The second approach, adopted in this study, examines the possible associations between $V(b)$ and $V(w)$ in a given distribution. In our study of performance distribution, our underlying premise is that $V(b)$ is inherently endogenous to $V(w)$, because $V(w)$ is an outcome of a firm's self-selecting behavior that is intended to increase $V(b)$. Self-selecting behavior influences $V(w)$ in two ways. First, it leads to variation in strategic choices. Second, the choice is biased towards creating successful operations. The former leads to a varying-range of outliers and the latter leads to a skew in the performance distribution, which together affect $V(w)$ and hence the shape of the distribution.

The following section presents the performance distribution model that explains how the shape of performance distribution, as defined by skewness, kurtosis, and variance, is associated with the mean performance at the firm and segment levels. The descriptions of the key terms used in our conceptual discussions are provided in the File S1.

Modeling a performance distribution

A firm's performance represents the aggregate performance of its multiple business units. In terms of strategic management practice, a firm can improve its performance by using any one of the following methods.

1. Improve all business units.
2. Create upward outliers.
3. Remove downward outliers.
4. Improve some of the business units within a given performance range.

The first method leads to a greater mean performance by moving the performance distribution to the right without changing the shape of the performance distribution; this is called the *structural effect* in this study. The structural effect is explained by the average-centered view and describes a situation in which certain conditions uniformly influence the performance of all of the business units of a firm. The structural effect can occur at the firm level because each firm possesses a unique set of routines or other firm-level resources and capabilities that systematically improve the performance of all of its current business units. These firm-specific conditions create the between-firm variation in mean performance, whereas the within-firm variation is assumed to be constant across firms. The structural effect also arises at the segment level (i.e., industries, and countries),

[3] See some typical examples in Baum and McKelvey's study (2006, pp. 125–126).

[4] Power law distributions refer to the Pareto rank/frequency distributions that have potentially infinite ranges of extreme values (and hence long and fat tails), (potentially) infinite variance, unstable means, and unstable confidence intervals (Andriani and McKelvey, 2007, 2009). A good introduction to the concept of power law can be found in economics and finance research (e.g., Gabaix, 2016) and in entrepreneurship research (e.g., Crawford et al., 2015).

DOI: 10.1002/smj

because segment-specific conditions (e.g., competitive industry structure, institutional environment, and naturally inherited factors of production) systematically influence the performance of all of the firms within a segment. These segment-specific conditions create the between-segment variation in the mean performance, whereas the within-segment variation is assumed to be constant across the segments.

Skew effect

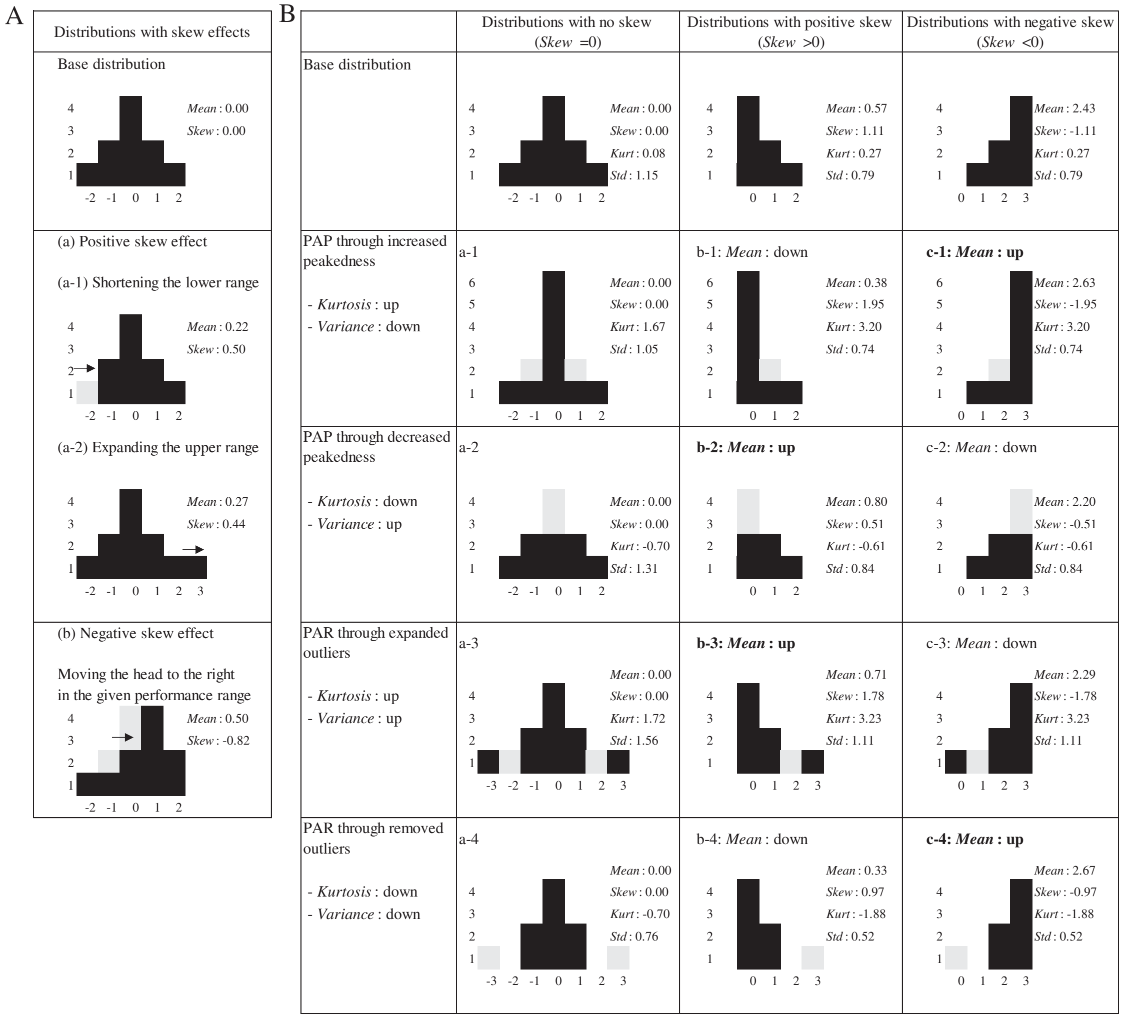
The remaining three methods for improving performance lead to a greater mean performance by changing the shape of the performance distribution. This is called the *skew effect* in this study, and is explained by the variance-centered view. The *skew effect* describes a situation in which the skewness in a performance distribution has a positive association with the mean performance. Specifically, the second and the third methods create a *positive skew effect* that involves a *positive adjustment to the range* of a performance distribution (PAR) (i.e., expanding the upper range and/or shortening the lower range of a performance distribution), which results in a greater mean with a positive skew in the performance distribution. A positive skew in a performance distribution is characterized by the highest point of the distribution, i.e., the head, being located on the left side of the distribution and a longer tail on the right side of the distribution. The fourth method, which creates a *negative skew effect*, involves *a positive adjustment of the proportion* of a performance distribution (PAP) (i.e., the creation of a greater [smaller] concentration in the right [left] side of the distribution in a given range of the performance distribution), resulting in a greater mean with a negative skew in the performance distribution. A negative skew is characterized by a head located on the right side and a longer tail on the left side of the distribution. Panel A in Figure 1 illustrates how these adjustments change the shape of the performance distribution and improve the mean performance.[5]

Although skew effects have not been the focus of strategic management studies, some researchers have highlighted the importance of understanding these effects in relation to firm performance. One stream of research based on the behavioral tradition of management research suggests that previous downside risks tend to be positively related to subsequent returns (e.g., Miller and Reuer, 1996). The rationale is that when the realized performance falls below the aspirational level, firms are likely to engage in risk-taking behavior as a means of improving their subsequent performance. In other words, when there are many underperforming business units, firms can improve their performance by creating a positive skew effect (i.e., expanding the upper range of performance outliers while shortening the lower range of performance outliers). Another stream of research built on the real option theory suggests that firms can increase the upside opportunities and decrease the downside risks of investment by maximizing the value of an option to wait (e.g., McGrath, 1999). The real option effect is essentially the same as the positive skew effect in that by exercising the option to wait, firms can reduce downside risks while retaining the option to explore new opportunities, thereby increasing their chances of performance improvement in the long run. In contrast, total quality management (TQM) improves a firm's performance by creating a negative skew effect. TQM represents a firm's organization-wide commitment to improving its current practices (management procedures and processes). TQM's primary goal is to improve the performance of the current practices closer to the upper possible limit rather than to create new practices that expand the limit.

Although the preceding discussion of the skew effect is in the context of a firm with multiple business units, the same pattern holds at the segment level. As the economic performance of a segment depends on the aggregate performance of the individual firms operating in the segment, a positive skew effect is created by the entry of productive firms and the exit of unproductive firms (i.e., PAR), whereas a negative skew effect is created by the improvement of existing firms in a given performance range (i.e., PAP) in the segment. These adjustments in the performance distribution of a segment correspond to two separate views on competition (Mueller, 2005). One view regards competition as a process for maximizing the efficient (the Pareto optimal) allocation of resources and the other regards competition as a dynamic process for creating resources. In a segment that has a higher level of mean performance,

[5] These represent the prototypes of skew effects in a unimodal distribution. In addition, many firms attempt to simultaneously acquire high-performing operations, remove underperforming operations, and improve current operations. The actual observed skew in the performance distribution reflects the outcomes of the combined effects.

DOI: 10.1002/smj



C

		Skewness in the Distribution: Positive	Skewness in the Distribution: Negative
Type of adjustment	PAP	(a) Exploitation/Learning: Performance improvement due to a decreased concentration in the lower range of the distribution (b-2 in Panel B). Kurtosis: down; Variance: up. Weakens a positive skew effect	(b) Exploitation/Learning: Performance improvement due to an increased concentration in the higher range of the distribution (c-1 in Panel B). Kurtosis: up; Variance: down. Strengthens a negative skew effect
	PAR	(c) Exploration/Entry: Performance improvement due to the expanded upward outliers (b-3 in Panel B). Kurtosis: up; Variance: up. Strengthens a positive skew effect	(d) Downsizing/Exit: Performance improvement due to the removal of downward outliers (c-4 in Panel B). Kurtosis: down; Variance: down. Weakens a negative skew effect

PAR: Positive adjustment to the range of a performance distribution
PAP: Positive adjustment to the proportionsof a performance distribution

Figure 1. Performance improvement through adjusting distribution: prototypes. (A) Skew effects, (B) heavy-tail effects. Heavy-tail effects involve the cases of b-2 and b-3 in a positive skew situation and c-1 and c-4 in a negative skew situation. In these cases, the mean performance improves with changes in kurtosis and variance. (C) Heavy-tail effects on skew effects.

DOI: 10.1002/smj

resources are mobilized from unproductive firms to productive ones through the exit of inefficient firms and the entry of efficient ones; resources are created as a consequence of learning among the existing firms.[6]

Despite the limited attention to the skew effect at the segment level, some studies provide evidence of the association between skewness and mean performance. At the country level, Harvey (1995) finds that emerging economies tend to have higher average stock returns and volatility than developed countries and to have non-normal distributions of the stock returns with a varying degrees of skew, generally with a positive value (positive skew). Makino and Yiu's (2014) study reveals that the skewness in the performance distributions of listed firms in Japan, China, and India have greater negative values during recession periods, during which the aggregate economic performance declines, than in post-recession periods. At the industry level, Henkel (2009) observes that individual firms' performance distributions have a considerable spurious effect of skewness on the estimated mean and variance across industries, indicating that performance distributions of poorly performing firms are likely to be negatively skewed. Based on the analysis of the mean-variance with different levels of skewness, Harvey and Siddique (2000) suggest that negatively skewed stock investment portfolios should have high expected returns. Although these studies show that the extent of the association between skewness and mean performance differs across segments, we argue that the differences in the extent of the association across segments is a result of the differences in the levels of PAR and PAP across segments. These discussions lead to the following propositions.

Proposition 1a: A positive skew effect is created by a positive adjustment to the range of a performance distribution.

Proposition 1b: A negative skew effect is created by a positive adjustment to the proportions of a performance distribution.

Heavy-tail effect

In addition to creating skew effects, firms can improve their mean performance by creating a *heavy-tail effect*, which describes a situation in which changes in the values of *kurtosis* and *variance*, the two related yet distinct elements in a distribution, lead to an improvement in the performance mean.[7] High kurtosis is characterized by a high peak in the head or long tails with more outliers and thinner shoulders relative to the normal distribution.[8] High variance is characterized by a low peak in the head and heavy tails with many values deviating from the mean.

Kurtosis and variance are conceptually independent measures of a heavy tail. They are positively related when the kurtosis is created by outliers, because outliers increase the variability, and are negatively related when kurtosis is created by the peakedness of the head, because increased peakedness implies a greater concentration around the head and reduces the variability in the distribution. Furthermore, kurtosis and variance have no association with the mean when there is no skew in a distribution. The first column in Panel B in Figure 1 illustrates these relations.

Kurtosis and variance are only associated with the mean performance in the presence of a skew effect. In a positive skew situation, the mean performance improves when either the peakedness decreases (see b-2 in Panel B in Figure 1) or the range of upward outliers becomes wider (see b-3 in Panel B in Figure 1). The first process reduces the kurtosis (due to the decreased peakedness) and increases the variance (due to the increased variability caused by the decreased peakedness), leading to PAP. The second process increases the kurtosis (due to the expanded upper range of outliers) and increases the variance (due to the increased upward variability), leading to PAR.[9] In a negative skew situation, the mean performance improves when either the peakedness increases (see

[6] Some empirical studies support this view. For example, Liu (1991) finds that competitive pressures force less efficient producers to fail and that the gap in productivity between existing and exiting plants, and between entering and exiting plants, has widened over time, whereas the gap between existing and entering plants has shrunk over time.

[7] In this study, heavy tails refer to tails that are heavier than the tails in a normal distribution.

[8] A "thin shoulder" arises in a distribution that has "density crossing" over the normal distribution, meaning that the probability densities cross the normal curve twice on each side of the mean when the distribution has excess kurtosis (DeCarlo, 1997).

[9] The peakedness of the distribution should be described in relation to the range of outliers (Westfall, 2014), because the peakedness can increase (decrease) as the range of outliers narrows (widens). Kurtosis, as a measurement of peakedness, also reflects the range of outliers in a distribution.

c-1 in Panel B in Figure 1) or the range of downward outliers becomes narrower (see c-4 in Panel B in Figure 1). The first process increases the kurtosis (due to the increased peakedness) and reduces the variance (due to the decreased variability caused by an increased peakedness), leading to PAP. The second process reduces the kurtosis (due to the narrowed lower range of outliers) and reduces the variance (due to the decreased downward variability), leading to PAR. The remaining cases (b-1, b-4, c-2, and c-3 in Panel B in Figure 1) are not considered heavy-tail effects as they do not lead to an improvement in the performance mean. Panel C in Figure 1 illustrates the four heavy-tail effects discussed above.

The conventional notions of exploration, exploitation, and downsizing at the firm level can be related to the concepts of kurtosis and variance effects, as shown in Panel C in Figure 1. First, *exploitation*, or the improvement of the existing business units, either decreases a concentration of underperforming business units in a positive skew situation, which leads to reduced kurtosis and increased variance (cell-a), or increases a concentration of high performing business units in a negative skew situation, which leads to increased kurtosis and reduced variance (cell-b) in the given range of the distribution. Second, *exploration* of exceptionally high performing business units expands the upper range of the distribution in a positive skew situation, which leads to increased kurtosis and variance (cell-c). Third, *downsizing* underperforming business units shortens the lower range of the distribution in a negative skew situation, which leads to reduced kurtosis and variance (cell-d). The first two processes involve PAP and further weaken the positive skew effect (and strengthen the negative skew effect). The third and the fourth processes involve PAR and further strengthen the positive skew effect (and weaken the negative skew effect).

The above processes correspond to the notions of entry, exit, and learning at the segment level. First, *learning*, or the improvement of existing firms, either decreases a concentration of unproductive firms in a positive skew situation (cell-a) or increases a concentration of productive firms in a negative skew situation (cell-b) in the given range of the distribution. Third, the *entry* of exceptionally productive firms expands the upper range of the distribution in a segment (cell-c). Fourth, the *exit* of unproductive firms shortens the lower range of the distribution in a segment (cell-d). As in the firm level, the first two processes involve PAP and further strengthen the negative skew effect (and weaken the positive skew effect). The third and the fourth processes involve PAR and further strengthen the positive skew effect (and weaken the negative skew effect).

In sum, heavy-tail effects can improve the mean performance in the presence of skew effects, by reinforcing or attenuating the skew effects. The above discussions lead to the following propositions.

Proposition 2a: In the presence of a positive skew effect, aggregate-performance is further improved by either of the following two processes: (a) decreasing the concentration of underperforming business units/ firms (Exploitation/Learning) or (b) expanding the range of upward outliers (Exploration/Entry).

Proposition 2b: In Proposition 2a, the Exploitation/Learning process leads to a PAP with decreased kurtosis and increased variance that weakens the positive skew effect, and the Exploration/Entry process leads to a PAR with increased kurtosis and variance that strengthens the positive skew effect in the performance distribution.

Proposition 3a: In the presence of a negative skew effect, aggregate performance is further improved by either of the following two processes: (a) increasing the concentration of high performing business units/firms (Exploitation/Learning) or (b) removing downward outliers (Downsizing/Exit).

Proposition 3b: In Proposition 3a, the Exploitation/Learning process leads to a PAP with increased kurtosis and decreased variance that strengthens the negative skew effect, and the Downsizing/Exit process leads to a PAR with decreased kurtosis and variance that weakens the negative skew effect in the performance distribution.

PERFORMANCE DISTRIBUTION MODEL: HYPOTHESES

The propositions outlined above suggest that in addition to leveraging structural effects, firms can improve their performance by creating skew and heavy-tail effects.[10] Drawing on performance feedback theory (Cyert and March, 1963; Greve, 2003), we argue that performance distributions tend to exhibit a positive skew in the long run and that successful firms are more likely to benefit from positive skew effects.

Firms' performance feedback is defined as "organizational learning from their experience by collecting performance measures, creating aspiration levels based on their own past performance or that of other organizations, and changing organizational activities if the performance is lower than the aspiration level" (Greve, 2003: 1). Central to performance feedback theory is the idea that firms engage in problemistic searches when they are doing poorly but not when they are doing well (Cyert and March, 1963; Greve, 2003). This idea is consistent with that of prospect theory (Kahneman and Tversky, 1979), which argues that firms take more risks when they are doing poorly (March, 1991). In contrast, the key thrust of the threat-rigid hypothesis is that individuals and firms tend to behave rigidly in threatening situations and are thus less likely to engage in search behavior when they are doing poorly (Staw, Sandelands, and Dutton, 1981).

Although both types of firm behavior are likely to occur, we argue that performance distributions will eventually exhibit a positive skew because firms have a general tendency to self-select behavior that creates or maintains successful business units. Under the condition of performance decline, problemistic searches drive firms to explore new opportunities to compensate for realized losses, which is, if successful, likely to result in a creation of upward performance outliers (i.e., exploration). In contrast, organizational rigidity drives firms to reexamine the current (inefficient) practices to minimizes expected losses, which is likely to result in the abandonment of underperforming operations (i.e., downsizing). Both processes lead to a positive adjustment of the range of a performance distribution (PAR), creating a positive skew.

In contrast, under the condition of incremental performance change, firms are more likely to make incremental improvements in current operations (i.e., exploitation) or even to maintain the status quo, rather than make major strategic actions such as exploration or downsizing. The process of incremental improvement leads to a positive adjustment of the proportion in a performance distribution (PAP), creating a negative skew. However, the negative skew effect will eventually be attenuated when the performance improvement reaches an upper limit (Henkel, 2009: 288). To make further performance improvements beyond the limit of what is possible with the negative skew effect, firms need to engage in PAR, namely, either create performance outliers in the upper range or remove underperforming units in the lower range. This process, while improving the aggregate performance, eventually weakens a negative skew and creates a positive skew in the performance distribution. As a result, firms that are successful at achieving high aggregate performance of their business units tend to have a greater positive skew in their performance distributions than less successful firms.

Hypothesis 1: Firms with a higher aggregate performance of business units tend to have a more positive skew in their performance distributions.

Firms' performance feedback at the segment level can be explained by the "Red Queen" (Barnett and Hansen, 1996) and "Blue Ocean" (Kim and Mauborgne, 2005) ideas. The "Red Queen" idea (Barnett and Hansen, 1996) describes the process by which firms learn from and adapt to competition. The key argument is that competition prompts rival firms to conduct problemistic searches, which both makes them stronger competitors and forces weaker competitors to either fail or adapt. In essence, the success or failure of firms is a consequence of both adaptive and maladaptive processes of learning through competitive interactions. Although the "Blue Ocean" idea (Kim and Mauborgne, 2005) presents a somewhat different perspective, it also suggests that firms increase their chances of success through searching and learning. The key argument is that firms should stay away from intense competition (or the "Red Ocean") and explore untapped markets to create demand (Kim and Mauborgne, 2005). Both approaches suggest that competition facilitates the entry of productive firms, the exit of unproductive firms, and searching and learning among all firms.

As competition sets in, high performing firms enter the market, whereas underperforming firms either exit the market or learn to improve their

[10] It would be interesting to explore the type of non-normal probability distribution that can best describe the proposed associations between skew, kurtosis, variance, and the mean, but this is beyond the scope of this study.

performance. These processes leads to PAR at the segment level due to the creation of upward outliers and the removal of downward outliers, leading to a positive skew. This lead to the following hypothesis.

Hypothesis 2: Segments with a higher aggregate firm performance tend to have a more positive skew in their performance distributions than segments with a lower aggregate performance.

In the presence of a positive skew effect, a firm consists of business units including a small number of upward outliers and a large number of average performers. To further improve its performance, a firm eventually needs to make PAP because it cannot continue to create upward outliers or to downsize downward ones to make PAR. This strategic shift resembles a punctuated equilibrium process of organizational learning (Lant and Mezias, 1992) in which organizations experience long periods of incremental changes and adaptation punctuated by fundamental reorientations. In our framework, "fundamental reorientations" represent PAR (i.e., exploration and downsizing) that results in drastic changes in a firm's current portfolio of business units. In the subsequent periods, a firm engages in "local searches" within the boundary of current expertise or knowledge (Rosenkopf and Nerkar, 2001; Stuart and Podolny, 1996) and exploitation to update and refine the current operations (March, 1991). This process creates PAP, as the current operations improve their performance without making PAR.

PAP at the segment level, like PAP at the firm level, eventually replaces PAR in the presence of a positive skew effect. At the segment level, potential entrants have the option of either entering the segment (industry or country) or not, and incumbent firms have the option of either exiting or staying. In the presence of a positive skew effect, the segment consists of a small number of exceptionally productive competitors and a large number of average firms. In this situation, potential entrants may see limited room to develop a position that exceeds that of the top performers, whereas incumbents see the opportunity to learn from these outliers, or the "skewed few," to emulate their best practices (Baum, 2011; March and Sutton, 1997). These processes facilitate PAP, as the incumbents improve the performance in the segment, while simultaneously creating a limit to PAR.

According to Proposition 2b, PAP in the presence of a positive skew effect improves the aggregate performance by creating decreased kurtosis and increased variance. In summary, we set forth the following hypothesis.

Hypothesis 3: The positive skew effect is negatively moderated by kurtosis and positively moderated by variance.

METHODOLOGY

Sample

Our data are drawn from the *Trend Survey of Overseas Business Activities* (hereafter the *Trend Survey*). The *Trend Survey*, which is conducted annually by the Japanese Ministry of Economics, Trade, and Industry, surveys all of the foreign affiliates in which any Japanese corporation has invested capital of 10% or more, except for those in the real estate, financial, and insurance industries. We use the *Trend Survey* for the 1997–2009 period (fiscal years 1996–2010), which gives a sample of 13,944 foreign affiliates. The average response rate of the survey is 60.9%.

We select foreign affiliates that reported their performance information (i.e., profit before taxes and sales) for every year in the observation period. The foreign affiliates established after the beginning of the observation period (i.e., 1997), or terminated before the end of the observation period (i.e., 2009), are excluded from the sample. We also exclude cases that exhibit a misalignment in the units used in the performance calculation. For example, some foreign affiliates report 1,000 as a unit of profit but 1 million as a unit of sales. To exclude such miscalculated values, we check whether there are any obvious discrepancies in the values in the reported return on sales (ROS) across years for each foreign affiliate. If the ROS in a certain year is abnormally high or low compared to the ROS in other years, we consider it a suspicious case. The suspicious cases are removed from the sample when they indicate a more than 10-fold growth in either sales or returns (profit before taxes) compared to their average for the period.

The industry classification of the foreign affiliates is based on the list of industries reported in the *Trend Survey*, which follows the three- or four-digit SIC classifications. A small number of foreign affiliates

DOI: 10.1002/smj

in our sample change their industry classifications during the observation period. In these cases, we use the industry in which they remained the longest. The parent firms comprise both listed and non-listed companies. Some of the parent firms in our sample merged or closed their operations during the observation period. The corresponding foreign affiliate cases are excluded from the analysis. The country classifications are based on the list of countries used by the Statistics Division of the United Nations. To obtain a consistent inference for the analyses at the host country level, we exclude those foreign affiliates that moved to another host country during the observation period. The final sample is comprised of 7,224 foreign affiliates.

The usable cases are compiled in a panel dataset that contains repeated observations for the same foreign affiliates during the 1997–2009 period. Using this base-panel dataset, we develop three separate datasets, corresponding to the parent firm, industry, and host country levels. The datasets for the analyses at the parent firm, industry, and host country levels comprise a balanced panel of 10 or more foreign affiliate cases for each parent firm, industry, and host country in each year. They represent a 13-year balanced panel of 86 parent firms, 46 industries, and 42 host countries during the 1997–2009 period.

Variables and analysis

Our dependent variable, *Mean*, is measured by the mean ROS of the foreign affiliates within each parent firm, industry, and host country per year over the observation period. ROS is a profit measure defined by the ratio of profit before tax over total sales. Unlike conventional strategic management research, which often uses return on assets (ROA) or return on invested capital (ROIC), we use ROS to measure the profit return for three reasons. First, sales (the denominator of ROS) may be a better measure of the outcomes of ongoing businesses in the global environment, as they are a better reflection of changes in performance across business climates than fixed assets or invested capital (Chan et al., 2008; Lincoln, Gerlach, and Ahmadjian, 1996). Second, asset turnover and interest-bearing liabilities may vary significantly between host countries due to differences in their market value and the cost of capital. Third, Japanese corporations have long used ROS as the primary performance measure.

The independent variable, *skewness* (*Skew*), and the two moderating variables, *kurtosis* (*Kurt*) and *standard deviation* (*Std*), are measured according to the following formulas:

$$Skew = \frac{n}{(n-1)(n-2)} \sum \left(\frac{ROS - \overline{ROS}}{Std} \right)^3,$$

$$Std = \sqrt{\frac{1}{n-1} \sum \left(ROS - \overline{ROS} \right)^2}, \text{ and}$$

$$Kurt = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left(\frac{ROS - \overline{ROS}}{Std} \right)^4 - 3.$$

These variables are calculated both at the firm (parent firm) level and at the segment (industry and host country) levels per year.

Data analysis

Firm performance is a multi-faceted construct that can be influenced by factors that are internal or external to the firm. To test our hypotheses, we estimate two empirical models: OLS with fixed-effects and generalized method of moments (GMM). The fixed-effects model rules out unobserved group effects (parent firm, industry, and country) on the dependent variable. GMM uses an instruments approach for estimating the models to rule out the unobserved effects.

First, we estimate the following OLS model with fixed-effects:

$$Y_{it} = X_{it}\beta + \lambda_t + \eta_i + \nu_{it}. \tag{1}$$

Here, Y_{it} represents the mean ROS (*Mean*) across all of the foreign affiliates in the *i*th group (parent firm, industry, and country) at the end of year t; and the vector X_{it} represents a set of explanatory variables (*Skew*, *Kurt*, *Std*, $Skew \times Kurt$, and $Skew \times Std$) in the *i*th group at the end of year t. The model also contains a year-effect λ_t, an unobserved group-effect η_i, and an error term ν_{it}. Year dummy variables are added to the analyses to capture the year-effect λ_t. The unobserved group-effect η_i represents the differences between groups (parent firms, industries, or host countries) that are stable

DOI: 10.1002/smj

over the years. In the fixed-effects model, η_i is regarded as a fixed-parameter and its effect on Y_{it} is ruled out in the analysis. The results of the analyses are reported in the File S1.

Second, we add a lagged value of the dependent variable ($Y_{i(t-1)}$) to capture possible persistence (autocorrelation), heteroscedasticity, and mean-reversion in ROS. This is important because firm performance may follow time trends under certain external economic conditions (e.g., the "Lost Decades"—a long recession period beginning in the early 1990s in Japan) and can be affected by organizational routines and inertia that evolve over time and have persistent effects on performance across long periods of time. As the dependent variable $Y_{i(t-1)}$ is correlated with the error term v_{it}, the standard estimation in a fixed-effects model is inconsistent. To address this problem, we use the generalized method of moments (GMM) proposed by Arellano and Bond (1991). We estimate the following time differencing equation to eliminate the unobserved group effects η_i:

$$\Delta Y_{it} = \alpha \Delta Y_{i(t-1,)} + \Delta X_{it}\beta' + \Delta\lambda_t + \Delta v_{it}. \quad (2)$$

In the above equation, $\Delta Y_{i(t-1)} = Y_{i(t-1)} - Y_{i(t-2)}$ and $\Delta v_{it} = v_{it} - v_{i(t-1)}$. $Y_{i(t-2)}$ is correlated with $\Delta Y_{i(t-1)}$, but uncorrelated with Δv_{it} (in the absence of serial correlation in v_{it}) and hence is used as a valid instrument for $\Delta Y_{i(t-1)}$ to obtain consistent estimations.

RESULTS

Table 1 presents the descriptive statistics and correlations for the variables. Table 2 presents the results of two sets of GMM regressions (Panels A and B) for the parent firm, industry, and host country samples.[11] The results show that skewness (*Skew*) has a positive and significant association with mean performance (*Mean*) at all three levels in the GMM models. The Arellano-Bond test shows no serial correlation in first-difference errors in any of the GMM models. Overall, the evidence provides strong support for Hypotheses 1 and 2. We also analyze the historical patterns of the associations between *Mean* and *Skew* in the performance distributions of each parent firm, industry, and host country using a 5-year rolling window to estimate an OLS model in which mean performance (*Mean*) is a dependent variable, skewness (*Skew*) is an explanatory variable, and kurtosis and variance are control variables. The results of the analysis show that with the exception of a few firms, industries, and countries with abnormal patterns, the patterns are stable across year-windows. The percentages of positive coefficients of *Skew* are 82.6% for the parent firm sample, 86.0% for industry sample, and 87.0% for the host country sample, with 27.9% of firms, 43.5% of industries, and 45.2% of countries having a positive coefficient for the entire 9-year window. None of the cases have a negative coefficient for the entire period.

To test the moderating effects of kurtosis (*Kurt*) and variance (*Std*) on the skew effect, we include two moderating variables in the analyses. The results of the fixed-effects regression show that the interaction between *Skew* and *Kurt* has a significant and negative association with the mean performance for the parent firm and host country samples but not for the industry sample. Similarly, the results show that the interaction between *Skew* and *Std* has a significant and positive association with the mean performance for the parent firm and host country samples but not for the industry sample. The analyses of the GMM regression models also provide consistent results. These findings suggest that heavy-tail effects exist at the parent firm and host country levels, but these effects are not clear at the industry level.

We also examine whether the relative effect of skew and the moderating effects of kurtosis and variance may vary between positive and negative performance situations. We think this is a valid issue, as research has suggested that realized performance outcomes create bias in and serve as a reference for the choice of subsequent behavior, which in turn affects future outcomes (Cyert and March, 1963; Kahneman and Tversky, 1979). To examine this possibility, we estimate the following GMM model:

$$\Delta Y_{it} = \alpha \Delta Y_{i(t-1)} + \tau_{it}\Delta X_{it}\beta'^{POS} + \left(1 - \tau_{it}\right)$$
$$\Delta X_{it}\beta'^{NEG} + \Delta\lambda_t + \Delta v_{it},$$

where τ_{it} is a dummy variable that equals 1 if ROS is greater than the sample mean, and 0 otherwise. β'^{POS} represents the main effect of *Skew* and the moderating effects of *Kurt* and *Std*

[11] As a base analysis, we conduct an OLS with fixed effects. The results are reported in the File S1.

Table 1. Descriptive statistics and correlations

	Mean	Std	Min	Max	*Mean*	*Skew*	*Kurt*	*Std*
Parent firm sample ($N = 1118$)								
Mean	0.00	1.00	−9.03	11.23	1.00			
Skew	0.00	1.00	−4.18	4.15	0.49	1.00		
Kurt	0.00	1.00	−0.63	8.35	0.00	−0.17	1.00	
Std	0.00	1.00	−0.24	19.97	0.07	0.04	0.34	1.00
Industry sample ($N = 598$)								
Mean	0.00	1.00	−11.13	3.14	1.00			
Skew	0.00	1.00	−4.12	4.90	0.31	1.00		
Kurt	0.00	1.00	−0.72	9.42	−0.07	−0.40	1.00	
Std	0.00	1.00	−0.42	9.84	−0.79	−0.14	0.29	1.00
Host country sample ($N = 546$)								
Mean	0.00	1.00	−11.49	7.63	1.00			
Skew	0.00	1.00	−4.22	3.68	0.34	1.00		
Kurt	0.00	1.00	−0.54	8.68	−0.07	−0.31	1.00	
Std	0.00	1.00	−0.32	17.23	−0.20	−0.11	0.41	1.00

that are conditional on a sample having positive values of mean ROS, capturing the extent to which these effects explain the mean ROS in a positive performance situation. Similarly, β'^{NEG} represents the main effect of *Skew* and the moderating effects of *Kurt* and *Std* that are conditional on a sample having negative values of ROS mean, capturing the extent to which these effects explain the mean ROS in a negative performance situation. The results, presented in Panel C in Table 2, show that skewness has a significant and positive effect on the mean performance in the both positive and negative performance situations at all three levels. In the industry and host country samples, the moderating effects of kurtosis and variance on skew are significant in a positive performance situation but not in a negative performance situation, whereas in the parent firm sample, they are significant in both situations. The latter result suggests that heavy-tail effects work primarily in positive performance situations. With this evidence, we conclude that Hypothesis 3 is generally supported.

Robustness check

As performance distribution can be skewed with outliers, it is possible that conventional parametric estimation methods may not be appropriate for analyzing performance data. We therefore apply a non-parametric method to examine the hypotheses. Specifically, we perform a quantile regression using a 50% quantile, which estimates the median of the dependent variable conditional on the values of the independent variables. Quantile regression is a fitting mechanism that minimizes the sum of the absolute residuals rather than the sum of the squares of the residuals, as in ordinary regression. It has some useful features compared to OLS; for example, the estimated coefficient vector is less sensitive to outliers in the median regression and the median regression estimators can be more efficient when the error term is non-normal (Koenker and Bassett, 1978). We estimate Equation (1) with a quantile (median) regression. We find consistent results at all three levels.

DISCUSSION

Building on the variance-centered view and drawing on performance feedback theory, this study proposes a model of performance distribution. The model illustrates that mean performance is associated with the shape of the distribution curve, as defined by skewness, kurtosis, and variance. The key point is that the aggregate performance of subsets (i.e., business units within a firm; firms in a segment) does not precisely reflect the individual performance of these subsets. Simple comparisons of the mean performance across firms or segments or of conditions underlying their differences provide incomplete and often misleading implications for both strategic management theory and practice, especially when the observed distribution is characterized by a high skew with heavy tails. Our findings reveal that successful firms tend to have

DOI: 10.1002/smj

Table 2. Results of the regression analyses

	Firm								Industry								Country							
DV: *Mean* $_{it}$	B	p	95% CI		B	p	95% CI		B	p	95% CI		B	p	95% CI		B	p	95% CI		B	p	95% CI	
Panel A. GMM																								
Constant	0.07	0.38	−0.09	0.24	0.01	0.75	−0.05	0.07	0.11	0.36	−0.13	0.35	0.11	0.35	−0.12	0.34	0.01	0.91	−0.19	0.21	0.07	0.69	−0.25	0.38
	(0.08)				(0.03)				(0.12)				(0.12)				(0.10)				(0.16)			
Mean $_{it-1}$	−0.02	0.51	−0.10	0.05	−0.02	0.56	−0.07	0.04	−0.01	0.83	−0.10	0.08	−0.01	0.82	−0.09	0.07	0.13	0.05	0.00	0.26	0.14	0.01	0.03	0.26
	(0.04)				(0.03)				(0.04)				(0.04)				(0.07)				(0.06)			
Skew $_{it}$	0.54	0.00	0.42	0.67	0.48	0.00	0.35	0.61	0.31	0.00	0.21	0.41	0.29	0.00	0.17	0.41	0.35	0.00	0.19	0.51	0.48	0.00	0.26	0.70
	(0.07)				(0.07)				(0.05)				(0.06)				(0.08)				(0.11)			
Kurtosis $_{it}$	−0.02	0.85	−0.21	0.17	0.11	0.21	−0.06	0.28	0.28	0.00	0.20	0.36	0.18	0.17	−0.07	0.43	0.04	0.82	−0.33	0.42	−0.06	0.62	−0.31	0.19
	(0.10)				(0.09)				(0.04)				(0.13)				(0.19)				(0.13)			
Std $_{it}$	0.03	0.79	−0.18	0.23	−0.26	0.12	−0.58	0.06	−0.87	0.00	−1.06	−0.68	−0.72	0.00	−1.13	−0.31	−0.12	0.59	−0.58	0.33	0.10	0.73	−0.45	0.64
	(0.10)				(0.16)				(0.10)				(0.21)				(0.23)				(0.28)			
Skew $_{it}$ × *Kurtsis* $_{it}$					−0.49	0.00	−0.74	−0.25					−0.15	0.44	−0.53	0.23					−0.51	0.00	−0.84	−0.18
					(0.13)								(0.19)								(0.17)			
Skew $_{it}$ × *Std* $_{it}$					0.82	0.00	0.52	1.11					0.22	0.34	−0.22	0.66					0.49	0.01	0.14	0.84
					(0.15)								(0.22)								(0.18)			
Year dummy	Included				Included				Included				Included				Included				Included			
N	946				946				506				506				462				462			
N (group)	86				86				46				46				42				42			
N (instruments)	81				83				81				83				81				83			
R^2	0.24				0.59				0.73				0.74				0.20				0.31			
Arellano-bond test	−0.07	0.94			0.79	0.43			−0.92	0.36			−1.13	0.26			0.18	0.86			0.11	0.91		
Panel B. GMM: Comparisons of positive and negative performance situations																								
Constant	0.09	0.27	−0.07	0.26	−0.04	0.54	−0.16	0.08	−0.02	0.66	−0.13	0.08	−0.34	0.01	−0.58	−0.09	0.19	0.38	−0.23	0.60	−0.09	0.68	−0.51	0.33
	(0.08)				(0.06)				(0.05)				(0.12)				(0.21)				(0.21)			
Mean $_{it-1}$	−0.02	0.53	−0.09	0.05	−0.01	0.60	−0.07	0.04	−0.01	0.83	−0.10	0.08	−0.01	0.74	−0.08	0.06	0.13	0.04	0.01	0.26	0.08	0.00	0.05	0.11
	(0.04)				(0.03)				(0.04)				(0.04)				(0.07)				(0.02)			
Kurtosis $_{it}$	0.03	0.79	−0.21	0.27	0.05	0.41	−0.08	0.19	0.28	0.00	0.20	0.36	−0.13	0.35	−0.41	0.15	0.09	0.61	−0.27	0.46	0.08	0.58	−0.19	0.35
	(0.12)				(0.07)				(0.04)				(0.14)				(0.19)				(0.14)			
Std $_{it}$	0.03	0.77	−0.17	0.23	−0.04	0.86	−0.47	0.39	−0.87	0.00	−1.07	−0.67	−0.96	0.00	−1.42	−0.49	−0.12	0.62	−0.58	0.34	−1.04	0.10	−2.30	0.21
	(0.10)				(0.22)				(0.10)				(0.24)				(0.24)				(0.64)			

Table 2. Continued

	Firm								Industry								Country							
DV: *Mean* $_{it}$	B	p	95% CI		B	p	95% CI		B	p	95% CI		B	p	95% CI		B	p	95% CI		B	p	95% CI	
Positive performance situation																								
Skew $_{it}$	0.48	0.00	0.34	0.63	0.38	0.00	0.26	0.50	0.31	0.00	0.21	0.42	0.27	0.00	0.14	0.41	0.31	0.00	0.15	0.46	0.33	0.00	0.13	0.53
	(0.08)				(0.06)				(0.06)				(0.07)				(0.08)				(0.10)			
Skew $_{it}$ × *Kurtsis* $_{it}$					−0.26	0.06	−0.53	0.01					−0.72	0.01	−1.28	−0.16					−1.11	0.00	−1.79	−0.44
					(0.14)								(0.29)								(0.35)			
Skew $_{it}$ × *Std* $_{it}$					0.59	0.01	0.18	1.00					1.36	0.00	0.52	2.20					3.62	0.00	1.14	6.09
					(0.21)								(0.43)								(1.26)			
Negative performance situation																								
Skew $_{it}$	0.62	0.00	0.42	0.83	0.53	0.00	0.38	0.67	0.29	0.00	0.11	0.48	−0.02	0.92	−0.39	0.35	0.42	0.00	0.16	0.68	0.14	0.38	−0.17	0.46
	(0.10)				(0.07)				(0.10)				(0.19)				(0.13)				(0.16)			
Skew $_{it}$ × *Kurtsis* $_{it}$					−0.66	0.00	−0.98	−0.33					−0.23	0.20	−0.57	0.12					−0.01	0.97	−0.40	0.38
					(0.17)								(0.17)								(0.20)			
Skew $_{it}$ × *Std* $_{it}$					1.12	0.00	0.48	1.75					−0.11	0.64	−0.60	0.37					−0.56	0.27	−1.55	0.43
					(0.32)								(0.25)								(0.51)			
Year dummy	Included				Included				Included				Included				Included							
N	946				946				506				506				462				462			
N (group)	86				86				46				46				42				42			
N (instruments)	82				86				82				86				82				86			
R^2	0.25				0.60				0.73				0.78				0.20				0.51			
Arellano-bond test	0.06	0.95			0.51	0.61			−0.92	0.37			−0.86	0.39			0.18	0.86			0.48	0.63		

Note. Robust standard errors in parenthesis. All of the variables are standardized.

a greater positive skew in their performance distributions and that the heavy-tail effects moderate the relationship between skewness and mean performance. These findings generally remain consistent at the parent firm, industry, and host country levels.

One possible avenue for future research is to combine the average- and variance-centered views and identify common firm- or segment-specific conditions that simultaneously affect both the mean performance and the shape of distributions. The average-centered view explains how certain internal or external conditions affect the central tendencies of the strategic behavior and performance of firms, assuming that the errors (the values deviating from the mean) are not important sources of information and that the distribution is normal and the shape and scale of the distribution are constant across conditions. The variance-centered view suggests that the strategic behavior of a firm often operates in opposition to a regression toward the mean even when the same internal or external conditions apply, assuming that the distribution can be non-normal and vary under the same conditions. Neither view alone can fully explain the observed variations in strategic behavior and performance. Research that simultaneously examines these two approaches may provide a better illustration of the real process. Although such attempts have recently been made in social studies of inequity (e.g., Western and Bloome, 2009; Zheng, Yang, and Land, 2011), to the best of our knowledge, little progress has been made on this front in the strategic management field.

Our evidence shows that the positive skew effect is negatively moderated by kurtosis and positively moderated by variance. The evidence suggests that successful firms tend to improve the mean performance by first creating a positive skew (i.e., exploration and/or downsizing). However, to further improve their mean performance, firms tend to pursue exploitation to upgrade their existing business units, rather than continuing to rely on further exploration or downsizing. One managerial implication is that to sustain their competitive advantages, firms need to adopt a balanced mix (i.e., ambidexterity) of exploration, downsizing, and exploitation. Future studies may extend the variance-centered view to the literature of organizational learning to examine how organizational learning changes the shape of performance distribution as it enhances the sustainable competitive advantages of firms. Our evidence also shows that segments with higher mean performances provide environments that facilitate the entry of high performing firms and the exit of underperforming firms and provide environments in which existing firms can further improve their performance by learning from other firms. This evidence is consistent with the conventional Industrial Organization view that excessive regulations on market entry and the protection of weak firms can weaken the competitiveness of the economy in the long term, and the evolutionary view of economic growth that emphasizes the importance of learning (Nelson and Winter, 1982). Future research can extend the variance-centered view to research on the micro-foundations of economic growth; in particular, it can be used to explain how economic policies on the strategic behavior of the firms can affect the sustainable development and growth of an economy.

Another interesting extension of the current research would be to explore the factors that deter some firms from creating positive skew effects. Our evidence shows that performance improvement through negative skew effects may be a tentative process. Our main observation is that negative skew effects tend to become less influential when the performance improvement reaches an upper limit, and firms will eventually need to create positive skew effects to further improve their performance. In fact, the strategic management literature has identified a number of factors that may deter firms from creating positive skew effects, including the *competency trap* (psychological barriers to searching for new alternatives due to successful experiences with the current options), *sequential dependence* (poor-performing operations acting as a source of other, successful operations), *irreversibility* (poor-performing operations and their supporting resources cannot be sold or diverted for alternative uses), *organizational imprinting* and *organizational inertia* (a persistent organizational resistance to change), and the *agency problem* (decision makers intentionally distorting or overestimating the value of certain operations for their own interests). Although these studies have focused on certain elements of or specific functions in management practices, future research should explain how problems that arise in certain organizational areas affect the shape of a firm's performance distribution and thereby the aggregate performance of a firm.

Our findings also show that the performance distributions of some segments (industries or host countries) exhibit a negative skew and a lower aggregate performance. There are at least four possible reasons for this finding. First, extremely underperforming firms, or "zombie firms," can survive due to institutional factors such as government policies that protect troubled firms (Hoshi and Kashyap, 2004). Second, institutional voids create a space in which underperforming firms can survive due to the poor functioning of market mechanisms that fail to screen out less efficient firms (Carroll, 1985). Third, some countries are slow to recover from external shocks (e.g., financial crises) due to weak political and economic institutions (Makino and Yiu, 2014). Fourth, the interdependence of economies across segments can increase economic disparity, making some weaker economies even weaker (Kremer and Maskin, 2003). Future research may integrate these insights into the variance-centered view and examine the circumstances under which they persistently block the development of a positive skew in the performance distribution and thereby affect firm performance at an aggregate level across segments.

Our study argues that firms can create a positive-skew effect by engaging in explorative behavior or downsizing (i.e., creating the upper range and/or shortening the lower range of performance outliers). In practice, the creation of positive skew effects can be achieved either through a portfolio management approach (acquiring exceptionally successful operations from outside and divesting or selling-off unsuccessful ones) or a capability building approach (upgrading the currently unsuccessful operations to the average ones; and the average operations to successful ones). An interesting extension of this study would involve examining how these two approaches influence a firm's long-term competitive advantage. The portfolio investment approach is usually performed through capital investment, i.e., buyouts or spinouts of operations, and may be a faster way to improve performance than the capability building approach. However, if a firm relies too much on returns from capital management, it may lose the chance to develop core capabilities, which require a path-dependent, accumulative, and organization-wide commitment to learning (Prahalad and Hamel, 1990). Renowned economist Thomas Piketty recently argued that when the rate of return on capital (r) is greater than the rate of economic growth (g) over the long term, the result is an unequal distribution of wealth, which in turn causes social and economic instability (Piketty, 2014). The implication of Piketty's thesis on strategic management is that firms that achieve a greater rate of return on capital investment tend to outperform those that focus on capability building, which ensures the growth of real businesses. However, we could argue the opposite: when the rate of return from capital investment exceeds the rate of return from capability building, firms may lose their sustainable competitive advantage in the long run despite the superior short-term financial returns. Future research might extend our study by examining how the capability building and portfolio management approaches affect the process of creating skew effects, that is, the shape of performance distributions, and thereby the sustainable competitive advantages of firms.

Limitations

This study has several limitations. First, the models proposed for the skew and heavy-tail effects are prototypes. In reality, firms use these effects simultaneously with varying weights. For example, firms can strengthen their positive skew effects by simultaneously adding upward outliers and removing downward outliers. Our evidence merely demonstrates the combined effect of both skew and heavy-tail effects at the aggregate level. Future studies should examine the ways in which these effects independently and jointly affect performance improvement. Furthermore, we assume that performance distributions have continuous and unimodal densities. Although we believe that the two approaches to performance improvement can apply to any structures of performance distribution, our model may not explain certain cases in which performance distributions have discrete values or multimodal structures.

Second, the scope of our analysis is limited to Japanese firms with multi-business-unit structures (i.e., multinational corporations) in which the relations between the business units are constant. Future studies should examine whether the observed associations also occur in the performance distributions of firms with different national origins and organizational forms that have varying degrees of interdependence between the business units. Finally, to constitute the balanced panel, we measure the key variables over a short period (13 years) and use only units that have complete information of 10 or more observations (i.e., affiliates within a firm or firms in a segment) per year during the observation

DOI: 10.1002/smj

period. The results may therefore be affected by an under-sampling bias. Although our study provides strong and robust evidence, future studies could minimize the bias in the dataset.

CONCLUSIONS

We propose and examine a performance distribution model based on the variance-centered view. Although we recognize that the ideas supporting the proposed model must be refined in future studies, we believe that our study provides a new approach to the investigation of the practices and outcomes of strategic management. We take the view that non-normality is the default state for social behavior and make statistical inferences by comparing basic but often overlooked descriptive statistics such as skewness and kurtosis. The key thrust of this view is that factors affecting the shape of distributions provide far more useful implications for advancing our understanding of social behavior and its outcomes than those factors that merely explain the average. Our study suggests that creating positive skew is essential for improving the aggregate performance at both the firm and segment levels. A lack of positive skew in a firm's performance distribution indicates room for improvement, and managers and policy makers should examine the factors impeding the creation of the positive skew effect and further improve their performance using the heavy tail effect. We hope that future studies will consider the variance-centered approach when examining variations in behavior and their outcomes.

ACKNOWLEDGEMENTS

We thank Associate Editor Glenn Hoetker and two anonymous reviewers for their helpful comments and suggestions. The authors thank Ying-Foon Chow, Takehiko Isobe, Akitoshi Ito, Ching-Chyi Lee, Amanda Phalin, Haipeng Shen, Hugh Thomas, and Xin Wang for helpful comments on the paper, and Chong He, Jingyi Wang, and Xiaoteng Wu for research assistance. The work described in this paper was partially supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region (Project No. CUHK14519516 & Project No. HKU731412)

REFERENCES

Andersen, T. J., & Bettis, R. A. (2015). Exploring longitudinal risk–return relationships. *Strategic Management Journal*, **36**(8), 1135–1145.

Andersen, T. J., Denrell, J., & Bettis, R. A. (2007). Strategic responsiveness and Bowman's risk–return paradox. *Strategic Management Journal*, **28**(4), 407–429.

Andriani, P., & McKelvey, B. (2007). Beyond Gaussian averages: Redirecting international business and management research toward extreme events and power laws. *Journal of International Business Studies*, **38**(7), 1212–1230.

Andriani, P., & McKelvey, B. (2009). From Gaussian to Paretian thinking: Causes and implications of power laws in organizations. *Organization Science*, **20**(6), 1053–1071.

Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies*, **58**(2), 277–297.

Barnett, W. P., & Hansen, M. T. (1996). The red queen in organizational evolution [Special Issue]. *Strategic Management Journal*, **17**, 139–157.

Baum, J. A. (2011). Free-riding on power laws: Questioning the validity of the impact factor as a measure of research quality in organization studies. *Organization*, **18**(4), 449–466.

Baum, J. A., & McKelvey, B. (2006). Analysis of extremes in management studies. In D. Ketchen & D. Bergh (Eds.), *Research methodology in strategy and management* (pp. 123–196). New York, NY: JAI Press.

Bekaert, G., & Harvey, C. R. (2002). Research in emerging markets finance: Looking to the future. *Emerging Markets Review*, **3**(4), 429–448.

Bowman, E. H. (1980). A risk/return paradox for strategic management. *Sloan Management Review*, **21**, 17–33.

Bromiley, P. (1991). Testing a causal model of corporate risk taking and performance. *Academy of Management Journal*, **34**(1), 37–59.

Carroll, G. R. (1985). Concentration and specialization: Dynamics of niche width in populations of organizations. *American Journal of Sociology*, **90**(6), 1262–1283.

Chan, C. M., Isobe, T., & Makino, S. (2008). Which country matters? Institutional development and foreign affiliate performance. *Strategic Management Journal*, **29**(11), 1179–1205.

Chan, C. M., Makino, S., & Isobe, T. (2010). Does subnational region matter? Foreign affiliate performance in the United States and China. *Strategic Management Journal*, **31**, 1226–1243.

Crawford, G. C., Aguinis, H., Lichtenstein, B., Davidsson, P., & McKelvey, B. (2015). Power law distributions in entrepreneurship: Implications for theory and research. *Journal of Business Venturing*, **30**, 696–713.

Cyert, R. M., & March, J. G. (1963). *A behavioral theory of the firm*. Englewood Cliffs, NJ: Prentice-Hall.

DeCarlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods*, **2**(3), 292–307.

Denrell, J. (2003). Vicarious learning, undersampling of failure, and the myths of management. *Organization Science*, **14**(3), 227–243.

Gabaix, X. (2016). Power laws in economics: An introduction. *Journal of Economic Perspectives*, **30**(1), 185–206.

Greve, H. R. (2003). *Organizational learning from performance feedback: A behavioral perspective on innovation and change*. Cambridge, England: Cambridge University Press.

Hambrick, D. C. (2007). The field of management's devotion to theory: Too much of a good thing? *Academy of Management Journal*, **50**(6), 1346–1352.

Hamilton, B. H., & Nickerson, J. A. (2003). Correcting for endogeneity in strategic management research. *Strategic Organization*, **1**(1), 51–78.

Hansen, M. H., Perry, L. T., & Reese, C. S. (2004). A Bayesian operationalization of the resource-based view. *Strategic Management Journal*, **25**(13), 1279–1295.

Harvey, C. R. (1995). Predictable risk and returns in emerging markets. *Review of Financial Studies*, **8**(3), 773–816.

Harvey, C. R., & Siddique, A. (2000). Conditional skewness in asset pricing tests. *Journal of Finance*, **55**(3), 1263–1295.

Harvey, C. R., Liechty, J. C., Liechty, M. W., & Müller, P. (2010). Portfolio selection with higher moments. *Quantitative Finance*, **10**(5), 469–485.

Henkel, J. (2009). The risk–return paradox for strategic management: Disentangling true and spurious effects. *Strategic Management Journal*, **30**(3), 287–303.

Hoshi, T., & Kashyap, A. K. (2004). Japan's financial crisis and economic stagnation. *Journal of Economic Perspectives*, **18**(1), 3–26.

Johnson, E. J., Bellman, S., & Lohse, G. L. (2003). Cognitive lock-in and the power law of practice. *Journal of Marketing*, **67**(2), 62–75.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, **47**(2), 263–292.

Kim, W. C., & Mauborgne, R. (2005). *Blue ocean strategy: How to create uncontested market space and make competition irrelevant*. Cambridge, MA: Harvard Business Press.

Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica*, **46**(1), 33–50.

Kon, S. J. (1984). Models of stock returns-a comparison. *Journal of Finance*, **39**(1), 147–165.

Kremer, M., & Maskin, E. (2003). *Globalization and inequality* (Working paper), Department of Economics, Harvard University. Retrieved from http://dev.wcfia.harvard.edu/sites/default/files/Kremer_GlobalizationInequality.pdf

Lant, T. K., & Mezias, S. J. (1992). An organizational learning model of convergence and reorientation. *Organization Science*, **3**(1), 47–71.

Lincoln, J. R., Gerlach, M. L., & Ahmadjian, C. L. (1996). Keiretsu networks and corporate performance in Japan. *American Sociological Review*, **61**(1), 67–88.

Liu, L. (1991). *Entry-exit, learning, and productivity change: Evidence from Chile* (Working Paper WPS769). Washington, DC: The World Bank.

Ma, X., Tong, T. W., & Fitza, M. (2013). How much does subnational region matter to foreign subsidiary performance? Evidence from fortune global 500 corporations' investment in China. *Journal of International Business Studies*, **44**, 66–87.

Makino, S., Isobe, T., & Chan, C. M. (2004). Does country matter? *Strategic Management Journal*, **25**(10), 1027–1043.

Makino, S., & Yiu, D. (2014). Strategy and performance of Asian enterprises. In G. Redding & M. Witt (Eds.), *Oxford handbook of Asian business systems* (pp. 561–586). New York, NY: Oxford University Press.

March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science*, **2**(1), 71–87.

March, J. G., & Sutton, R. I. (1997). Crossroads—organizational performance as a dependent variable. *Organization Science*, **8**(6), 698–706.

McGahan, A. M., & Porter, M. E. (1997). How much does industry matter, really? [Summer Special Issue] *Strategic Management Journal*, **18**, 15–30.

McGahan, A. M., & Victer, R. (2010). How much does home country matter to corporate profitability? *Journal of International Business Studies*, **41**(1), 142–165.

McGrath, R. G. (1999). Falling forward: Real options reasoning and entrepreneurial failure. *Academy of Management Review*, **24**(1), 13–30.

McKelvey, B., & Andriani, P. (2005). Why Gaussian statistics are mostly wrong for strategic organization. *Strategic Organization*, **3**, 219–228.

Miller, K. D., & Leiblein, M. J. (1996). Corporate risk–return relations: Returns variability versus downside risk. *Academy of Management Journal*, **39**(1), 91–122.

Miller, K. D., & Reuer, J. J. (1996). Measuring organizational downside risk. *Strategic Management Journal*, **17**(9), 671–691.

Mueller, D. C. (2005). *The dynamics of company profits*. Cambridge, MA: Cambridge University Press.

Nelson, R. R., & Winter, S. G. (1982). *An evolutionary theory of economic change*. Cambridge, MA: Belknap Press/Harvard University Press. (Reprinted in 1996).

Patton, A. J. (2004). On the out-of-sample importance of skewness and asymmetric dependence for asset allocation. *Journal of Financial Econometrics*, **2**(1), 130–168.

Piketty, T. (2014). *Capital in the twenty-first century*(A. Goldhammer, Trans.). Cambridge, MA: Harvard University Press.

Porter, M. E. (1991). Towards a dynamic theory of strategy. *Strategic Management Journal*, **12**(S2), 95–117.

Prahalad, C. K., & Hamel, G. (1990 (May–June)). The core competence of the corporation. *Harvard Business Review*, **68**, 79–91.

Rosenkopf, L., & Nerkar, A. (2001). Beyond local search: Boundary-spanning, exploration, and impact in the optical disk industry. *Strategic Management Journal*, **22**(4), 287–306.

Rumelt, R. P. (1991). How much does industry matter? *Strategic Management Journal*, **12**(3), 167–185.

Scherer, F. M., & Harhoff, D. (2000). Technology policy for a world of skew-distributed outcomes. *Research Policy*, **29**, 559–566.

Shaver, J. M. (1998). Accounting for endogeneity when assessing strategy performance: Does entry mode choice affect FDI survival? *Management Science*, **44**(4), 571–585.

Simon, H. A. (1955). On a class of skew distribution functions. *Biometrika*, **42**(3/4), 425–440.

Simon, H. A. (1965). *Administrative behavior*. New York, NY: Free Press.

Staw, B. M., Sandelands, L. E., & Dutton, J. E. (1981). Threat rigidity effects in organizational behavior: A multi-level analysis. *Administrative Science Quarterly*, **26**, 501–524.

Stuart, T. E., & Podolny, J. M. (1996). Local search and the evolution of technological capabilities [Summer]. *Strategic Management Journal*, **17**, 21–38.

Western, B., & Bloome, D. (2009). Variance function regressions for studying inequity. *Sociological Methodology*, **36**, 75–98.

Westfall, P. H. (2014). Kurtosis as peakedness. *American Statistician*, **68**(3), 191–195.

Wiggins, R. R., & Ruefli, T. W. (2002). Sustained competitive advantage: Temporal dynamics and the incidence and persistence of superior economic performance. *Organization Science*, **13**(1), 81–105.

Xiong, J., & Idzorek, T. (2011). The impact of skewness and fat tails on the asset allocation decision. *Financial Analysts Journal*, **67**(2), 23–35.

Zanini, M. (2008, November 15). Using "power curves" to assess industry dynamics. *McKinsey Quarterly*, pp. 1–5.

Zheng, H., Yang, Y., & Land, K. C. (2011). Variance function regression in hierarchical age-period-cohort models: Applications to the study of self-reported health. *American Sociological Review*, **76**(6), 955–983.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

File S1. I. Descriptions of key terms. II. Additional analyses.

DOI: 10.1002/smj