ISSN:1946-6404

International Review of Accounting, Banking and Finance

CONTENTS

International Review of Accounting, Banking and Finance

IRABF **Summer** **2025** **Volume 17** **No. 2**

Sponsored by: Shih Chien University and Louisiana State University

Journal Website: www.irabf.org

Published by: Shih Chien University

Print copy ISSN: 1946-6404 Online version ISSN:1946-6412

International Review of Accounting, Banking and Finance

IRABF Summer 2025 Volume 17 No. 2

CONTENTS

International Review of Accounting, Banking and Finance
Vol 17, No. 2, Summer, 2025, Pages 1-19

IRABF

Market Volatility and US-China Trade Policy: Evidence from the First Trump Administration

Nathan Mauck *[1], Stephen Pruitt [2], Wenjia Zhang [3]

1. *University of Missouri – Kansas City*
2. *University of Missouri – Kansas City*
3. *China Foreign Affairs University*

Accepted June 2025

A B S T R A C T

Prior work on the impact of the first Trump administration on the intersection between financial markets and international relations—especially with People's Republic of China—suggest that Mr. Trump's decidedly less "measured" tone compared with his predecessors (George W. Bush and Barack Obama) resulted in striking increases in financial market volatility and investor interest around trade-related announcements. Now, with the re-election of Mr. Trump as President of the United States for a second non-consecutive term, we re-examine financial market reactions to the Trump administration trade events. The present study significantly extends prior work by investigating the impacts of a sample of 90 key US/China trade announcements on VIX and VXFXI volatility metrics made over the course of the first Trump administration. Interestingly, trade announcements classified as likely to lead to increases in trade tensions had no impact on changes in the two studied indexes, whereas events thought likely to lead to decreases in trade tensions were associated with rather dramatic decreases of 2.5% in VIX volatility. Although conjecture, the results are consistent with the hypothesis that financial market participants did not respond to antagonistic pronouncements simply because they may have been fully anticipated, whereas announcements of a more conciliatory tone were perceived as unexpected. Tests of volatility spillovers between the US and Chinese markets document highly significant spillovers from the VIX index to the VXFXI index but not from the VXFXI to the VIX.

Keywords: U.S. and China trade; volatility; volatility spillover; trade war

[1] Corresponding Author, Email : mauckna@umkc.edu

[2] Email : pruittst@umkc.edu

[3] Email : wzhang@cfau.edu.cn

1. Introduction

With the November 5, 2024 re-election of Donald J. Trump as President of the United States—the first US president elected to non-consecutive terms since Grover Cleveland in 1893—and Republican control of both the US House and Senate, the world entered a new and potentially much more volatile political and economic environment. Compounding the considerable domestic uncertainties associated with Mr. Trump's oft-repeated pledge to deport millions of undocumented immigrants and billionaire Elon Musk's newly-created Department of Government Efficiency's (DOGE) efforts to dramatically shrink the size of the U.S. government, the ultimate ramifications of geopolitical events of an unusually serious nature—the wars in Gaza and Lebanon (precipitated by a deadly surprise attack against Israel by Gaza-based Hamas fighters), the increasingly hostile relationship between Israel and Iran, the Trump administration's shift away from NATO and Ukraine in favor of closer ties with Russia, and, perhaps most seriously for the global economy, the growing threat of a trade war triggered by Mr. Trump's promise of dramatically increased U.S. tariffs[4] and the increasingly belligerent rhetoric from China regarding its intentions in the South China Sea (especially involving the threat of forced unification with Taiwan[5])—remain unknown, but suggest an obvious need for serious and careful diplomacy.[6]

It is within the context of the words "serious" and "diplomacy" that recent work by Mauck, Pruitt, and Zhang (2022) regarding the observed positive correlation between US and Chinese investor attention and market-wide share-price volatility in both nations in response to trade-related announcements—especially the measurably "less diplomatic" statements issued over the course of the Trump administration—is particularly relevant. Concluding that "words matter," the authors' findings imply that financial market participants, no less than diplomats and forward-thinking politicians and bureaucrats, carefully consider both the substance and tone of various economic pronouncements.

The present study significantly extends the Mauck, Pruitt, and Zhang study by presenting the first-ever event-specific analysis of the responses of market-wide US and Chinese equity proxies of expected future volatility in the context of the recent US-China trade war by exploiting a remarkably comprehensive catalog of US-Chinese trade-related events over much of the Trump administration (from 5/2/2016 to 1/15/2020).[7] In addition, the study also examines US and Chinese equity proxies for the presence of volatility spillover effects around these events—both from the US to China and from Chinese markets to the US. Given the extraordinary importance of US and Chinese trade to the world economy (valued at $575 billion in 2023 alone[8]) and the aforementioned current deterioration of US and Chinese political relations, information concerning the sensitivity of US and Chinese—and, by extension, world—equities markets to trade-related stimuli over the 2016 to 2020 time period is likely to prove of significant interest to many constituencies, including politicians, bureaucrats, corporate managers, and financial market participants in the hotter still present political and military environment.

---

[4] See, e.g., https://www.nytimes.com/2024/11/25/business/economy/trump-tariffs-canada-mexico-china.html.

[5] In a March 3, 2024, speech at the opening of the National People's Congress (NPC), Premier Li Qiang for the first time officially dropped the previously employed terms "peaceful reunification" regarding China's relationship with Taiwan. See, e.g., https://www.reuters.com/world/china/china-drops-peaceful-reunification-reference-taiwan-raises-defence-spending-by-2024-03-05/

[6] As noted by famed British MP Tony Benn (1925-2014), "All war represents a failure of diplomacy."

[7] As discussed below, the beginning of the *global* COVID-19 pandemic in mid-January 2020 significantly altered the diffusion of information and market pricing dynamics in US and Chinese markets. For a detailed timeline of the COVID pandemic, see, e.g., https://www.cdc.gov/museum/timeline/covid19.html#:~:text=January%2020%2C%202020,respond%20to%20the%20emerging%20outbreak.

[8] See, e.g., https://www.statista.com/statistics/277679/total-value-of-us-trade-in-goods-with-china-since-2006/.

2. Previous Results

Generally considered to have officially begun on January 17, 2018, when then-US President Donald Trump began setting tariffs and other trade restrictions on the import of washing machines and solar panels—the vast majority of which were (and remain) made in China—the genesis of the US/China "trade war" actually began on May 2, 2016, when then-candidate Trump used the "r-word" (as in "rape") in an Allen County (Ohio) campaign speech: "We can't continue to allow China to rape our country and that's what they're doing. It's the greatest theft in the history of the world."[9] Not surprisingly, since that time, scholars of various academic disciplines have sought to evaluate the impact of the conflict on trade, financial markets, and, more broadly, political relations. For example, Liu, Sun, Xu, and Zhang (2023), Feng, Li, Peng, and Tan (2021), and Cheng, Hua, and Wang (2023) have studied the impact of US and Chinese trade discord on trade contraction, the cost of debt, and the influence of corporate culture on firm resilience, respectively.

Not surprisingly, several prior studies have examined the influence of trade-related information on cross-border equities markets. For example, Chen, Lui, Lu, and Tang (2016) analyze the impact of regularly scheduled official Chinese trade announcements on both equity market price levels and volatility. The authors employ the Baidu Search Index[10] as a proxy for investor attention and document the expected positive correlation between investor interest and equity price reactions around the time of the announcements. Related studies by Bank, Larch, and Peter (2011) and Takeda and Wakao (2014) present virtually identical results within the context of German and Japanese equities markets, respectively.

As noted above, Mauck, Pruitt, and Zhang (2022) exploited innovations in Google Trends' Search Volume Index (SVI) for the query "U.S. China trade"—an explicit proxy for investor interest in US/China trade news—to assess volatility changes in both US and Chinese equity market indexes to trade-related news under different US presidential administrations. Using data collected over the 2004 to 2018 period, the authors' findings demonstrate a clear distinction between the changes in volatility observed in response to the more measured pronouncements made during the George W. Bush (a decrease) and Obama (no change) administrations and the statistically significant increases in volatility associated with the more truculent language of those issued during the first Trump administration. Tests of volatility spillovers suggest the primary direction of contagion was from the US to Chinese markets, with little evidence of spillovers from the Shanghai Stock Exchange to the S&P 500.

In a study that proved influential in the collection of data for the present work, Yang, Luo, and Jiang (2021) created a daily economic policy uncertainty (EPU) index and employed a series of complex network analyses to ascertain the relationships between various Pacific Basin markets. The authors concluded i) that China was the clear center of the larger Asia-Pacific network, ii) that the US and China were the most important sources of cross-nation spillover effects in the studied financial networks, and iii) that correlations between the constructed EPU and financial networks significantly changed (i.e., increased) during the COVID-19 outbreak as compared with prior experience. Lei and Song (2022) similarly examines economic policy uncertainty in China and finds that stock price crash risk for Chinese firms increased during the US/China trade war. While Wang and Wang (2010) and Zhou, Zhang, and Zhang (2012) find that a “spillover” of volatility between U.S. and Chinese equities that runs in both directions, Vuong, Nguyen, and Huynh (2022) also present compelling evidence of a statistically significant "breakpoint" in equity market spillovers between the US and China due to the COVID-19 pandemic. Combined, these findings lend substantial empirical support for the decision to truncate the present analysis to the beginning of the pandemic.

[9] See, e.g., https://www.politico.com/blogs/2016-gop-primary-live-updates-and-results/2016/05/trump-china-rape-america-222689

[10] Baidu, founded in 2000, is the second largest internet search engine in the world and is used almost exclusively by Chinese citizens and nationals.

3. Data and Methodology

3.1 Data

As noted above, the basic research sample of trade-related events for this study begins on May 2, 2016, the day before the Republican Party's two remaining presidential candidates—Ted Cruz and John Kasich—suspended their campaigns for the presidency and the Republican National Committee's then-chairman (Reince Priebus) declared Mr. Trump to be "the presumptive Republican nominee."[11] The sample ends on January 15, 2020, in deference to the findings of LI, et al. (2021) that the advent of the COVID-19 pandemic at that time significantly altered prior marketplace correlations (by essentially closing large swaths of the global economy). As such, the data encompass the final six months of the 2016 US presidential election campaign and all the Trump administration prior to the beginning of the COVID-19 pandemic.

Consistent with Frino et al. (2011), the purpose of the study was to examine market-wide volatility around specific events classified as likely to either increase or decrease the overall temperature of the US/China trade relationship in place at that time. For example, Mr. Trump's widely reported January 1, 2018, threat to impose "a big fine" on China over alleged intellectual property theft was classified as increasing the trade temperature, while his May 13, 2018, Twitter[12] tweet promising to help Chinese telecom company ZTE compete for US business was classified as decreasing trade temperatures. Naturally, information viewed as increasing trade tensions would be expected to be associated with increases in overall market risk (that is, financial market volatility) and vice versa. Chen, Jiang, Li, and Xu (2016) similarly examine Chinese futures markets and volatility around specific US events, although their focus is on US. Consumer Price Index (CPI) announcements.

The trade-related events included in the sample were obtained by melding the informational content of multiple published news reports. For example, an October 2019 report from the Reuters news service provided a detailed timeline of 33 major events involved in "the U.S.-China trade war."[13] An additional and even more comprehensive report from Dezan Shira and Associates' China Briefing presented a timeline of 79 trade-related events.[14] Finally, the Peterson Institute for International Economics—an independent nonprofit, nonpartisan research organization dedicated to international trade issues—summarized the various events into five broad "battle" categories: solar panels and washing machines, steel and aluminum, and technology and intellectual property.[15] Each of the studied announcements was classified by the members of the research team as likely to be interpreted by financial market participants as either increasing or decreasing trade tensions. The appendix presents a summary of all 90 of the events included in the study.

The volatility of the U.S. stock market around the 90 studied U.S.-China trade-related events was assessed via daily changes in the Chicago Board Options Exchange (CBOE) well-known Market Volatility Index (VIX), commonly known as "the fear index." This index, which is created by value-weighting all out-of-the-money call and put prices on options expiring between 16 and 44 days into the future (two calendar weeks on each side of one 30-day month) outstanding on the S&P 500 market index, where the weights employed are based on the number of minutes to the expiration of each individual option contract relative to the total. As constructed, the index is designed to quantify expected market volatility over the following thirty days.[16] VIX is used in other spillover research such as Smales (2022) who finds that US market uncertainty (measured by VIX) spreads to many

[11] See, e.g., https://en.wikipedia.org/wiki/Political_career_of_Donald_Trump#:~:text=By%20March%202016%2C%20Trump%20was,Trump%20the%20presumptive%20Republican%20nominee.

[12] Now known simply as "X."

[13] https://www.reuters.com/article/us-usa-trade-china-timeline/timeline-key-dates-in-the-us-china-trade-war-idUSKBN1WP23B/

[14] https://www.china-briefing.com/news/the-us-china-trade-war-a-timeline/

[15] https://www.piie.com/blogs/trade-and-investment-policy-watch/2018/trumps-trade-war-timeline-date-guide

[16] A detailed discussion of the computational procedures involved in the creation of the VIX index is available from sources such as Investopia. See, e.g., https://www.investopedia.com/articles/active-trading/070213/tracking-volatility-how-vix-calculated.asp.

global markets including China. Similarly, the China ETF Volatility Index (VXFXI) is a metric of the expected volatility of the Chinese stock market, calculated based on the information obtained from over-the-counter options listed on the Stock Exchange of Hong Kong Ltd (SEHK), also published by CBOE. Historical data for both the VIX and VXFXI indexes were obtained directly from the CBOE website.

3.2 Methodology

As noted above, during the period under scrutiny (933 trading days), 90 significant trade events were recorded.[17] Our initial analysis delved into the variations in VIX and VXFXI levels on the specific dates of each U.S.-China trade event, with the correlation between these dates and implied volatility structured in the form of a VAR(1) regression. The specified model closely follows Jiang, Konstantinidi, and Skiadopoulos (2012) and Krieger, Mauck, and Vasquez (2015), both of which extensively examined volatility spillovers between U.S. and foreign markets.

We first estimated the following equation to investigate the existence of volatility spillovers:

$$\Delta IVt = C + \varphi \Delta IVt - 1 + \mu t, \quad (1)$$

where $\Delta IVt = IVt - IVt{-}1$ is a (2 x 1) vector of changes in the implied volatility indices for the U.S. (VIX) and China (VXFXI), C is a vector of constants, and φ is a (2 x 2) matrix of regression coefficients. To capture the volatility effects of trade friction announcements, we generalized equation (1) by incorporating indicator variables for each trade announcement. This augmented model enabled us to examine the direct linkage between the trade announcements and coincident spillovers in implied volatility between the U.S. and Chinese stock markets. In addition, we also investigated whether the responses of the volatility indices to the studied announcements were shaped by the market's anticipation of trade fears. This endeavor aimed to ascertain if events characterized by increases or decreases in trade tensions resulted in differences in market dynamics and, if so, how these differences modulated volatility spillovers between the two markets.

4. Empirical Results

Table 1 presents the results of the mean and median percentage volatility changes observed in the VIX (Panel A) and VXFXI (Panel B) indexes in response to the 90 aggregated trade-related events analyzed over the May 2016 to January 2020 interval. Overall, there is little evidence that the announcements led to significant changes in the volatility of either index. Indeed, what evidence exists is mainly contradictory. Whereas the non-parametric sign-test (Z) of the simple fraction of announcements that resulted in VIX volatility reductions is statistically significant at the 5 percent level, the mean and median volatility metrics actually moved in opposite directions—with the means rising and the medians falling—underscoring the general lack of any meaningful volatility impact in the complete sample. Interestingly, in no case is there any evidence that the studied trade-related announcements elicited reactions in the VXFXI index (Panel B), although, again, the directional change of the means and medians are reversed. However, since, as noted above, the 90 studied trade announcements represent events likely to be interpreted by financial market participants as either leading to increasing or decreasing US-China trade tensions (but not both), bifurcation of the full sample into separate "increase" and "decrease" sub-samples is necessary. Tables 2 and 3 present the results of these tests.

[17] Nine trade-specific events were eliminated from the sample due to issues related to "clustering," where one event announcement classified as likely to increase U.S./China trade tensions occurred on the very same day as one classified as likely to decrease trade tensions. It should be noted that there were no qualitative differences observed between the two samples.

Table 1 VIX and VXFXI Changes on Major Events During the Trade Frictions

	Mean % Change	Median % Change
Panel A: Reactions from VIX		
VIX	1.27%	-1.30%
t-statistics	[0.786]	
VIX increases	36	
VIX decreases	54	
Z-statistics for sign-rank test	-2.003	
Panel B: Reactions from VXFXI		
VXFXI	0.19%	-0.11%
t-statistics	[0.281]	
VXFXI increases	43	
VXFXI decreases	47	
Z-statistics for sign-rank test	-0.527	

This table presents changes in the VIX and VXFXI levels on major trade events during the trade frictions. The percentage changes relative to the previous days, are reported. The sample period is May 2nd, 2016–Jan 15th, 2020. T-test results of the mean change and sign-rank tests of the median change are reported. ***, **, and * denote statistical significance at the 1%, 5%, and 10%, respectively.

Panels A and B of Table 2 present the results of the studied trade-friction announcements on the VIX volatility index, with Panel A (Panel B) including events assumed to represent a decrease (increase) in US-China tensions. The employed methodology and test statistics are identical to those presented in Table 1 above.

Unlike the general lack of statistical significance for the aggregated sample, Panel A of the table shows strong evidence of decreases in overall VIX volatility in response to trade-related events thought to lead to decreases in U.S.-China trade tensions. In both cases the results are significant at the 1 percent level or less. Indeed, of the 42 events in the likely decrease sample, 33 (78.6%) exhibited declines in volatility at the time of the announcements. Further, the directional changes in volatility (decreases) were consistent across the board. Interestingly, unlike the case of the likely decrease group, the impact of events thought likely to lead to increases in US-China trade tensions were not statistically different from zero, with just over half (54.2%) of the 48 events leading to increases in the volatility of the VIX index (although both the means and medians moved in the same direction).

Although impossible to test empirically, the results of the Mauck, Pruitt, and Zhang (2022) study suggest a plausible explanation for the clear asymmetry of the volatility impacts of the likely increase and decrease samples. As noted by the authors, President Trump's ". . . belligerent statements regarding China and U.S./Chinese trade seemed almost a staple of daily news reports." Indeed, even a casual reading of the 90 events included in the present study suggests a consistent personal and political animosity (if not outright vitriol) toward China unlikely to be in any way "enhanced" short of a formal declaration of war. In other words, financial market participants may not have responded to antagonistic trade-related pronouncements by the Trump administration for the simple reason that they were probably fully anticipated. Conversely, those (rarer) instances in which Mr. Trump "changed his stripes" and established a more positive, conciliatory tone were likely perceived by financial markets as unexpected surprises, thus leading to a general reduction in trade fears and, hence, overall lower levels of market volatility. As the authors conclude, "In the final analysis, the results of the study strongly suggest that—at least in the case of U.S./China trade—words matter."

Table 2 VIX Changes on Different Types of Major Trade Events

	Mean % Change	Median % Change
Panel A: Likely Decrease group		
VIX	-2.50%***	-2.39%
t-statistics	[-2.947]	
VIX increases	9	
VIX decreases	33	
Z- statistics for sign-rank test	-3.549	
Panel B: Likely Increase group		
VIX	4.57%	0.80%
t- statistics	[1.595]	
VIX increases	26	
VIX decreases	22	
Z- statistics for sign-rank test	0.433	

This table presents changes in the VIX levels on different types of major trade events. The percentage changes relative to the previous days, are reported. The sample period is May, 2016–Jan, 2020. T-test results of the mean change and sign-rank tests of the median change are reported. ***, **, and * denote statistical significance at the 1%, 5%, and 10%, respectively.

Table 3 presents identical tests to those in Table 2 but, in this case, employs the volatility of the VXFXI index. Not only do the results not approach conventional levels of statistical significance, but the mean and median changes in volatility of the VXFXI index and the simple fraction of events registering volatility increases and decreases are inconsistent, a finding which strongly suggests VXFXI traders may have been less attuned to trade-related developments than VIX traders—at least over this interval of calendar time.

Table 3 VXFXI Changes on Different Types of Major Events

	Mean % Change	Median % Change
Panel A: Likely Decrease Group		
VXFXI	0.19%	0.53%
t- statistics	[0.203]	
VXFXI increases	22	
VXFXI decreases	20	
Z- statistics for sign-rank test	0.463	
Panel B: Likely Increase Group		
VXFXI	-0.28%	-0.51%
t-statistics	[-0.279]	
VXFXI increases	21	
VXFXI decreases	27	
Z-statistics for sign-rank test	-1.01	

This table presents changes in the VXFXI levels on different types of major trade events. Changes in the absolute level in VXFXI, as well as the percentage changes relative to the previous days, are reported. The sample period is May, 2016–Jan, 2020. T-test results of the mean change and sign-rank tests of the median change are reported. ***, **, and * denote statistical significance at the 1%, 5%, and 10%, respectively.

Table 4 presents the results of a series of multivariate regression tests on volatility spillovers between the US and China employing daily percentage changes in the VXFXI as the dependent variable. As shown in Panel A, although there is evidence of first-order negative autocorrelation between daily changes in the VXFXI—that is, increases in the VXFXI volatility index on, say, Tuesday preceded decreases in the index on Wednesday (and vice versa)—there is no statistically significant evidence of first-order autocorrelation (either negative or positive) between the VIX index and the VXFXI.

Panel B of Table 4 replicates the Panel A analysis discussed above but with the addition of an indicator variable on the date of each of the aggregated sample of trade events (first column) and both the event date and an interaction variable between the prior day's change in the VXFXI and the date of each trade event. As shown, there is no evidence that the aggregated event sample and changes in the VXFXI index are in any way related. However, again, since the aggregated sample includes trade-related events thought likely to lead to increases or decreases (but not both simultaneously) in the volatility of the VXFXI index, the overall impact of the studied events cannot be assessed without bifurcation of the sample. The results of these tests are presented in Panel C of the table.

As shown in Panel C, trade-related events classified as likely to lead to increases in the daily volatility of the VXFXI index were associated with both economically and statistically significant increases in volatility. Indeed, adjusted for first-order autocorrelations, the studied trade tension increase events led to a mean 2.5 percent increase in the volatility of the VXFXI index (a change significant at the 1 percent level). Interestingly, although consistent in sign, events classified as likely to decrease US/China trade tensions did not elicit statistically significant decreases in VXFXI volatility at conventional levels, as the overall percentage impact of the decrease events was just 44 percent of the magnitude of the increased sample. There is no evidence that the interaction terms (the change in the VIX multiplied by the increase or decrease dummies) are significant at conventional levels.

Table 5 continues the analysis by studying the impact of the same variables included in Table 4 above on the percentage of daily changes in the CBOE's VIX index. Although there is no evidence of first-order autocorrelation between the change in the VIX controlled for the VIX change the prior trading day, there is extreme evidence of volatility spillovers between the VIX index and "same day" movements of the VXFXI (Panel B) due to the lack of synchronicity between the trading locations (with China being fourteen hours ahead of Chicago).

Panel C of Table 5 repeats the Panel C of Table 4 analysis with the likely increase and likely decrease events and interaction terms. Unlike the lack of significance of the interaction terms of the Table 4 results discussed above, the interaction between the VXFXI, the change in the VIX index, and the likely volatility-increasing events is highly significant. As expected, there is no evidence that the likely decrease events and changes in the VXFXI index are in any way correlated with changes in the VIX.

Table 4 Spillover Effects on the Percentage Changes in Volatility Levels: from the U.S. to China

	ΔVXFXI%	
Panel A: Volatility spillover		
C	0.001	
	[0.554]	
ΔVIX%(t-1)	0.015	
	[0.553]	
ΔVXFXI%(t-1)	-0.175***	
	[-3.879]	
Adj. R^2	0.023	
N	933	
Panel B: Volatility spillover with trade event days		
C	0.0002	0.000
	[0.102]	[0.116]
ΔVIX%(t-1)	0.016	0.008
	[0.579]	[0.282]
ΔVXFXI%(t-1)	-0.178***	-0.183***
	[-3.932]	[-4.048]
Trade Events	0.009	0.008
	[1.439]	[1.365]
ΔVXFXI%(t-1)* Trade Events		0.083
		[1.403]
Adj. R^2	0.024	0.025
N	933	933
Panel C: Volatility spillover with different types of trade event days		
C	0.000	0.000
	[0.105]	[0.117]
ΔVIX%(t-1)	0.014	0.008
	[0.529]	[0.284]
ΔVXFXI%(t-1)	-0.178***	-0.184***
	[-3.959]	[-4.064]
Increase	0.026***	0.024***
	[3.153]	[3.008]
Decrease	-0.011	-0.011
	[-1.285]	[-1.280]
ΔVIX(t-1)*Increase		0.088
		[1.296]
ΔVIX(t-1)*Decrease		0.023
		[0.211]
Adj. R^2	0.033	0.033
N	933	933

This table tests the spillover effects from the U.S. to China, using multivariate regressions on the percentage (daily) changes of VXFXI for China market. The coefficient estimates, t-statistics (in brackets), and adjusted R2 are reported. ***, **, and * denote statistical significance at the 1%, 5%, and 10%, respectively.

Table 5 Spillover Effects on the Percentage Changes in Volatility Levels: from China to U.S.

	ΔVIX%	
Panel A: Volatility spillover		
C	0.002	
	[1.091]	
ΔVXFXI%	1.160***	
	[29.579]	
ΔVIX%(t-1)	0.012	
	[0.528]	
Adj. R^2	0.485	
N	933	
Panel B: Volatility spillover with trade event days		
C	0.002	0.002
	[1.036]	[1.053]
ΔVXFXI%	1.160***	1.096***
	[29.536]	[25.569]
ΔVIX%(t-1)	0.012	0.009
	[0.527]	[0.381]
Trade Events	0.000	-0.002
	[0.018]	[-0.317]
ΔVXFXI%(t-1)* Trade Events		0.377***
		[3.635]
Adj. R^2	0.485	0.491
N	933	933
Panel C: Volatility spillover with different types of trade event days		
C	0.002	0.002
	[1.041]	[1.062]
ΔVXFXI%	1.152***	1.095***
	[29.227]	[25.701]
ΔVIX%(t-1)	0.011	0.007
	[0.455]	[0.282]
Increase	0.013	0.003
	[1.330]	[0.317]
Decrease	-0.015	-0.022*
	[-1.438]	[-1.968]
ΔVXFXI*Increase		0.469***
		[4.197]
ΔVXFXI*Decrease		-0.484*
		[-1.766]
Adj. R^2	0.486	0.497
N	933	933

This table tests the spillover effects from China to the U.S., using multivariate regressions on the percentage (daily) changes of CBOE VIX for the US market. The coefficient estimates, t-statistics (in brackets), and adjusted R2 are reported. ***, **, and * denote statistical significance at the 1%, 5%, and 10%, respectively.

5. Conclusion

Without question, Donald Trump's first US presidential term (2017-2021) was characterized by a very different geopolitical emphasis than that of his immediate predecessors (George W. Bush and Barack Obama). Prior research by Mauck, Pruitt, and Zhang (2022) demonstrated this difference quantitatively by examining changes in internet search frequencies and the volatility of both US and Chinese stock markets associated with US/China trade announcements made during the Bush (associated with a volatility decrease), Obama (no impact), and Trump (a volatility increase) administrations. The present study significantly extends this analysis by investigating the impacts of a sample of 90 key US/China trade-related events on daily changes in VIX and VXFXI volatility metrics over the course of the majority of the first Trump administration.[18]

Viewed in totality, there is no evidence that the 90 studied events had any measurable impacts on the volatility of the VIX or VXFXI indexes. However, once the full sample was bifurcated between events likely to lead to increases or decreases in market volatility, a very different picture emerged. Specifically, while events classified a priori as likely to lead to increases in US/China trade tensions elicited no economic or statistical changes in the VIX index, events thought likely to lead to decreases in trade tensions were associated with rather dramatic decreases in VIX volatility (-2.5%; significant at the 1% level). Although the likely increase trade tensions sample was associated with a rather dramatic increase in VIX volatility of over 4.5%, the large standard errors associated with this sample meant that this increase was not significant at conventional levels. Although conjecture, these results are consistent with the hypothesis that financial market participants may not have responded to antagonistic trade-related pronouncements by the Trump administration for the simple reason that they may have been fully anticipated, whereas announcements of a more conciliatory tone were viewed as unexpected surprises. Interestingly, the 90 studied trade announcements were not associated with any statistically significant changes in the VXFXI index. Tests for volatility spillovers between US (VIX) and Chinese (VXFXI) markets present evidence of the highly substantial spillovers between the VIX index and "same day" movements of the VXFXI expected due to the lack of synchronicity between the two trading locations.

With the November 2024 re-election of Donald Trump as the 47th President of the United States and his subsequent (and almost immediate) launching of what appears to be a new worldwide trade war based upon "reciprocal" increases in tariffs with Mexico, Canada, and China[19], the course of future US trade relations with its many worldwide partners is presently in flux. Future historians and economists may well find ample additional evidence that, as noted by Mauck, Pruitt, and Zhang (2022), when it comes to US/China trade, "words (really do) matter."

[18] As noted above, the present sample was truncated to the beginning of the COVID-19 pandemic in deference to research by LI, Luo, and Jiang (2021) and Vuong, Nguyen, and Huynh (2022), who present compelling evidence of a statistically significant "breakpoint" in market spillovers between US and Chinese financial markets due to the pandemic.

[19] See, e.g., https://www.whitehouse.gov/fact-sheets/2025/02/fact-sheet-president-donald-j-trump-imposes-tariffs-on-imports-from-canada-mexico-and-china/

References

Bank, M., Larch, M., & Peter, G. (2011). Google search volume and its influence on liquidity and returns of German stocks. Financial markets and portfolio management, 25, 239-264.

Chen, J., Jiang, F., Li, H., and Xu, W. (2016). Chinese stock market volatility and the role of U.S. economic variables. Pacific-Basin Finance Journal, 39, 70-83.

Chen, J., Liu, Y. J., Lu, L., & Tang, Y. (2016). Investor attention and macroeconomic news announcements: Evidence from stock index futures. Journal of Futures Markets, 36(3), 240-266.

Cheng, S., Hua, X., & Wang, Q. (2023). Corporate culture and firm resilience in China: Evidence from the Sino-US trade war. Pacific-Basin Finance Journal, 79, 102039.

Feng, X., Li, W., Peng, Y., & Tan, Y. (2021). International trade friction and the cost of debt: Evidence from China. Pacific-Basin Finance Journal, 67, 101550.

Frino, A., Lecce, S., & Segara, R. (2011). The impact of trading halts on liquidity and price volatility: Evidence from the Australian Stock Exchange. Pacific-Basin Finance Journal, 19(3), 298-307.

Jiang, G. J., Konstantinidi, E., & Skiadopoulos, G. (2012). Volatility spillovers and the effect of news announcements. Journal of Banking & Finance, 36(8), 2260-2273.

Krieger, K., Mauck, N., & Vazquez, J. (2015). Comparing US and European market volatility responses to interest rate policy announcements. International Review of Financial Analysis, 39, 127-136.

Lei, A. C., & Song, C. (2022). Economic policy uncertainty and stock market activity: Evidence from China. Global Finance Journal, 52, 100581.

Liu, D., Sun, W., Xu, L., & Zhang, X. (2023). Time-frequency relationship between economic policy uncertainty and financial cycle in China: Evidence from wavelet analysis. Pacific-Basin Finance Journal, 77, 101915.

Mauck, N., Pruitt, S., Zhang W., 2022. Words matter: Market responses to changes in U.S. and Chinese trade-related internet search frequency under different U.S. administrations. Global Finance Journal. 53. https://doi.org/10.1016/j.gfj.2022.100742.

Smales, L. A. (2022). Spreading the fear: The central role of CBOE VIX in global stock market uncertainty. Global Finance Journal, 51, 100679.

Takeda, F., & Wakao, T. (2014). Google search intensity and its relationship with returns and trading volume of Japanese stocks. Pacific-Basin Finance Journal, 27, 1-18.

Vuong, G. T. H., Nguyen, M. H., & Huynh, A. N. Q. (2022). Volatility spillovers from the Chinese stock market to the US stock market: The role of the COVID-19 pandemic. The Journal of Economic Asymmetries, 26, e00276.

Yang, L. I., Luo, J., & Jiang, Y. (2021). Policy uncertainty spillovers and financial risk contagion in the Asia-Pacific network. Pacific-Basin Finance Journal, 67, 101554.

Wang, P., and Wang, P. (2010). Price and volatility spillovers between the greater China markets and the developed markets of U.S. and Japan. Global Finance Journal, 21, 304-317.

Zhou, X., Zhang, W., and Zhang, J. (2012). Volatility spillovers between the Chinese and world equity markets. Pacific-Basin Finance Journal, 20, 247-270.

Appendix: Major Events during the U.S.-China Trade War

This table lists the 90 important events happened during the trade war[20]. These events are classified into five groups: (1) news good for both; (2) news bad for both; (3) news bad for China/good for US; (4) news good for US/bad for China; (5) news unclear for both parties. For US, group (1) and (3) are good news for them, whereas group (2) and (4) are bad news for them. In our empirical analysis, for US, news are divided into two types: "good news for US", which includes (1) and (3), and "bad news for US", which includes (2) and (4). Similarly, for China, news are also divided into two types: "good news for China", which includes (1) and (4), and "bad news for China", which includes (2) and (3).

	Date	Event	Brief Description/Minor trade issue	
1	3/31/2017	Two executive orders signed by Trump.	One calls for tighter tariff enforcement in anti-subsidy and anti-dumping trade cases. The other orders a review of U.S. trade deficits and their causes.	Increase
2	4/7/2017	Xi visits Trump's Mar-a-Lago estate in Florida.	Trump and Chinese President Xi Jinping agree to set up a 100 Day Action Plan to resolve trade differences.	Decrease
3	5/22/2017	Trade deal reached.	US and China agree to a trade deal that would give US firms greater access to China's agriculture, energy, and financial markets, while China gains access to sell cooked poultry to the US.	Decrease
4	7/19/2017	The two sides fail to agree on new steps to reduce the U.S. deficit with China after the 100 days of talks.		Increase
5	8/14/2017	"Section 301" case against China initiated.	Trump orders "Section 301" probe into alleged Chinese intellectual property theft, described as his first direct trade measure against Beijing.	Increase
6	11/10/2017	Trump pays a "state visit plus" to China.	Relations were considered to have warmed.	Decrease
7	1/17/2018	Trump threatens a big "fine" on China	Trump, in a Reuters interview, threatens a big "fine" on China over alleged IP theft, without providing details.	Increase
8	1/22/2018	Trump imposes tariffs on all imported washing machines and solar panels - not just those from China.		Increase
9	2/5/2018	China Investigates US Exports of Sorghum		Increase
10	2/7/2018	'Global safeguard tariffs' implemented.	The US implements 'global safeguard tariffs' – placing a 30 percent tariff on all solar panel imports, except for those from Canada, (worth US$8.5 billion) and a 20 percent tariff on washing machine imports (worth US$1.8 billion).	Increase

[20] We take reference of several sources. "Timeline: Key dates in the U.S.-China trade war", see, https://www.reuters.com/article/us-usa-trade-china-timeline/timeline-key-dates-in-the-us-china-trade-war-idUSKBN1WP23B; "The US-China Trade War: A Timeline", see, https://www.china-briefing.com/news/the-us-china-trade-war-a-timeline/; "Trump's Trade War Timeline: An Up-to-Date Guide", see, https://www.piie.com/blogs/trade-investment-policy-watch/trump-trade-war-china-date-guide.

11	3/8/2018	Trump orders 25% tariffs on steel imports and 10% on aluminum from all suppliers - not just China.		Increase
12	3/22/2018	Trump signs a memorandum directing some acts.	To file a WTO case against China for their discriminatory licensing practices; To restrict investment in key technology sectors; and To impose tariffs on Chinese products (such as aerospace, information communication technology and machinery).	Increase
13	3/23/2018	Tariffs on steel and aluminum imports imposed.	US imposes a 25 percent tariff on all steel imports (except from Argentina, Australia, Brazil, and South Korea) and a 10 percent tariff on all aluminum imports (except from Argentina and Australia).	Increase
14	4/2/2018	Tariffs on US goods imposed by China.	China imposes tariffs (ranging 15-25 percent) on 128 products (worth US$3 billion) including fruit, wine, seamless steel pipes, pork and recycled aluminum in retaliation to the US' steel and aluminum tariffs.	Increase
15	4/3/2018	Initial list released by US.	The USTR releases an initial list of 1,334 proposed products (worth US$50 billion) subject to a potential 25 percent tariff (list revised June 15).	Increase
16	4/4/2018	Tariffs on US goods imposed by China.	China reacts to USTR's initial list, and proposes 25 percent tariffs to be applied on 106 products (worth US$50 billion) on goods such as soybeans, automobile, chemicals (list revised on June 16).	Increase
17	4/5/2018	Additional tariffs proposed by Trump.	Trump instructs trade officials to consider whether an additional $100 billion of US imports from China should be imposed.	Increase
18	4/16/2018	US Department of Commerce concludes that Chinese telecom company ZTE violated US sanctions.	US companies are banned from doing business with ZTE for seven years.	Increase
19	4/17/2018	China announces antidumping duties of 178.6 percent on imports of sorghum from the US.		Increase
20	5/7/2018	US-China engage in trade talks in Beijing.	The US demands that China reduce the trade gap by US$200 billion within two years. Talks end with no resolution.	Decrease
21	5/13/2018	Trump promises to help ZTE in a tweet.		Decrease
22	5/18/2018	China's Commerce Ministry announces that it will stop tariffs on US sorghum at negotiations.		Decrease
23	5/20/2018	The trade war is put on hold.	US and China agree to put the trade war on hold after China reportedly agrees to buy more US goods.	Decrease
24	5/29/2018	US reinstates tariff plans after brief truce		Increase

25	6/5/2018	Two days of trade talks between US and China held in Beijing.		Decrease
26	6/7/2018	US and ZTE agree to deal that will allow ZTE to resume business.		Decrease
27	6/15/2018	(US) Initial list of products reduced and finalized.	List 1 now implements a 25 percent tariff on a reduced 818 products (from 1,334) and is set to take effect on July 6, 2018. List 2 of 284 new products is also announced and under consideration.	Decrease
28	7/6/2018	US implements first China-specific tariffs US		Increase
29	7/10/2018	US releases second tariff list US (The United States unveils plans for 10% tariffs on $200 billion of Chinese imports).	The USTR releases a third list of tariffs (List 3) of over 6,000 commodities originating in China (worth US$200 billion), which will be subject to a 10 percent tariff.	Increase
30	7/16/2018	Trump Administration Files WTO Challenges	The US Trade Representative files separate disputes at the World Trade Organization against Canada, China, the	Increase
31	7/20/2018	Trump Threatens Tariffs on All Imports from China		Increase
32	8/1/2018	Trump orders USTR to increase the tariffs on $200 billion of Chinese imports to 25% from the originally proposed 10%.		Increase
33	8/2/2018	US tariffs revisions (US$200 billion)	The US Department of Commerce also adds 44 Chinese entities to its export control list that pose a “significant risk” to US national security.	Increase
34	8/3/2018	China announces second round of tariffs on US products		Increase
35	8/7/2018	Second round of tariffs finalized and released by US	US releases a revised version of tariffs on a final list of US$16 billion worth of imports from China	Increase
36	8/8/2018	China revises its $50 billion tariff list, removing crude oil.		Decrease
37	8/14/2018	China files WTO claim against US	The Chinese Ministry of Commerce announces that a formal case has been lodged at the WTO against the US for its tariffs on solar panels, alleging that US tariffs have damaged China’s trade interests.	Increase
38	8/23/2018	US and China implement second round of tariffs, China files second WTO complaint US		Increase
39	9/7/2018	Trump threatens new tariffs	Trump threatens to impose tariffs on US$267 billion more.	Increase
40	9/12/2018	US invites China to re-open negotiations	The White House’s top economic advisor, Larry Kudlow, says that the US has invited China to restart trade negotiations before tariffs on US$200 billion worth of Chinese goods (List 3) go into effect.	Decrease

41	9/17/2018	US finalizes tariffs on US$200 billion of Chinese goods		Increase
42	9/18/2018	China announces retaliation for US tariffs	China announces that it will implement tariffs on US$60 billion worth of US goods (List 3) after the latest round of tariffs from the US (worth US$200 billion) go into effect on September 24.	Increase
43	9/24/2018	US and China implement third round of tariffs US	On Sep 22nc, China cancels trade talks with US	Increase
44	10/25/2018	US and China officials resume contact.	US and China working-level officials reportedly resume contact after weeks of silence.	Decrease
45	10/30/2018	US reportedly prepared to announce tariffs on remaining Chinese products	The US is reportedly prepared to announce tariffs on all remaining Chinese products by early December if talks between Trump and Xi at the G20 in Argentina are not successful.	Increase
46	11/9/2018	US and China resume trade talks	According to the report, the two sides discussed a framework for a trade deal, or at least a "ceasefire" to reduce tensions.	Decrease
47	11/19/2018	US releases list of proposed export controls on emerging technologies	The rules do not specify China, but are widely considered by observers to be related to US efforts to prevent China from acquiring sensitive technologies.	Increase
48	12/1/2018	The United States and China agree on a 90-day halt to new tariffs.		Decrease
49	12/14/2018	China to temporarily lower tariffs on US autos; resumes buying US soybean exports		Decrease
50	1/9/2019	US and China engage in 3-day trade talks in Beijing	After the talks, China's Commerce Ministry issued a statement that the talks were "extensive and established a foundation for the resolution of each other's concerns."	Decrease
51	1/22/2019	US cancels preparatory talks with China	US officials cited disagreements over the enforcement of IP rules as the reason for the cancellation.	Increase
52	1/31/2019	US and China hold 2-day trade talks in Washington D.C	China offers to buy five million tons of US soybeans. Trump announces that he will meet with Xi in-person in February.	Decrease
53	2/15/2019	US and China hold trade talks in Beijing	the US and China continue to have differences, but agree to keep talking in Washington the following week.	Decrease
54	2/24/2019	Deadline extended by US.	Trump extends the March 1 deadline, leaving the tariffs on $200 billion of Chinese goods at 10% on an open-ended basis.	Decrease
55	3/29/2019	US and China hold trade talks in Beijing after one month break	Officials call the trade talks constructive, with an enforcement mechanism to monitor China's commitment to trade concessions reportedly a sticking point.	Decrease
56	4/1/2019	China bans all types of fentanyl on April 1st; China extends the suspension of additional tariffs on US autos and auto parts on Mar 31th.	China announces that it will ban all variants of the synthetic opioid fentanyl, effective May 1, 2019, in what is considered a concession to the US amid trade talks	Decrease

57	4/5/2019	US and China hold trade talks in Washington	On Thursday, April 4, Trump meets with Liu He, and says that the two sides will know "over the next four weeks" whether they can strike a deal. US and Chinese negotiators agree to continue talks the following week.	Decrease
58	4/10/2019	US and China agree to establish trade deal enforcement offices		Decrease
59	5/5/2019	Trump tweets that he intends to raise the tariffs rate on $200 billion of Chinese goods to 25% on May 10.		Increase
60	5/8/2019	The Trump administration gives formal notice of its intent to raise tariffs on $200 billion of Chinese imports to 25% from 10%, effective May 10.		Increase
61	5/13/2019	China announces tariff hikes on US products, China launches tariff exemption system	China announces that it will increase tariffs on US$60 billion worth of US goods from June 1, 2019, in response to the tariff increases imposed by the US on May 10.	Increase
62	5/16/2019	US places Huawei on its 'entity list', banning it from purchasing from US companies		Increase
63	5/31/2019	China establishes its very own 'unreliable entities' list	China announces that it will establish its very own unreliable entities list in retaliation to the US' entity list.	Increase
64	6/1/2019	China increases tariffs on US$60 billion worth of products	Tariffs of 25 percent, 20 percent, and 10 percent, which were first announced on May 13, 2019 are now in effect on US$60 billion worth of American goods exported to China.	Increase
65	6/18/2019	Xi and Trump rekindle trade talks ahead of G20 meeting		Decrease
66	6/19/2019	US Tariff Exemption Process for Chinese Imports	The Office of the US Trade Representative (USTR) announces a process by which US interested parties could request the exclusion of certain Chinese products – subject to additional tariffs – as per the September 2018 list (List 3).	Decrease
67	6/21/2019	US adds another five Chinese entities to its 'entity list'		Increase
68	6/26/2019	Tentative truce reached days before G20 Summit		Decrease
69	6/29/2019	Trade talks to restart, ban on Huawei relaxed		Decrease
70	7/9/2019	US exempts 110 Chinese products from 25 percent tariffs, issues licenses to American Huawei suppliers		Decrease
71	7/16/2019	Trump threatens tariffs on US$325 billion of Chinese goods, new member on China's negotiating team		Increase
72	8/1/2019	Trump says US will impose 10 percent tariffs on another US$300 billion of Chinese goods starting September 1		Increase

73	8/6/2019	Chinese companies suspend new US agricultural product purchases; US declares China is a currency manipulator.		Increase
74	8/13/2019	US delays tariffs on certain products and removes items from the list	US and China agree to talk again in two weeks	Decrease
75	8/23/2019	China announces US$75 billion in tariffs on US goods, Trump threatens tariff increases on Chinese goods		Increase
76	8/26/2019	Liu calls for calm, Trump says talks will proceed		Decrease
77	9/2/2019	China lodges WTO tariff case against the US	On Sep 1st, tariffs come in force as scheduled (from both sides).	Increase
78	9/5/2019	China and US agree to 13th round of trade talks		Decrease
79	9/11/2019	China unveils tariff exemption list for US imports		Decrease
80	9/13/2019	China exempts various agricultural products from additional tariffs		Decrease
81	9/20/2019	US releases new tariff exemption lists, which exempt over 400 Chinese goods from tariffs	US-China mid-level trade talks in Washington. The two countries agreed to keep communicating on related trade issues and discussed the details of the 13th round of bilateral high-level economic and trade consultations scheduled for October as reported by state media.	Decrease
82	9/23/2019	Purchase of US goods.	Chinese companies the following monday buy about 600,000 tonnes of U.S. soybeans, resuming modest purchases started earlier in September that would reach 3.5 million tonnes by early October — about 10% of China's annual pre-trade war volumes.	Decrease
83	10/7/2019	The U.S. Commerce Department puts 28 Chinese companies on its "entity list".	Largely banning U.S. firms from selling to them, over their alleged involvement in human rights abuses against Uighur Muslims in Xinjiang.	Increase
84	10/10/2019	High level talks held.	High level delegates from China and the U.S. meet in Washington for two days of talks.	Decrease
85	10/11/2019	US announces "Phase 1" deal, delays tariff increase for Chinese goods	As part of the Phase 1 agreement, China will reportedly purchase US$40-50 billion in US agricultural products annually, strengthen intellectual property provisions, and issue new guidelines on how it manages its currency.	Decrease
86	10/18/2019	US tariff exclusion process for US$300 billion of Chinese imports		Decrease
87	11/1/2019	China wins WTO case, able to sanction US$3.6 billion worth US imports		Decrease

88	11/8/2019	US and China Talk Tariff Rollback	US, China negotiators talk over phone, agree on trade points “in principle”	Decrease
89	11/26/2019	US releases new regulatory guidelines for its telecom networks procedure to protect telecom networks from national security threats	While the document makes no mention of Huawei or ZTE equipment, it might impact the two Chinese companies as they were placed on the US entity “blacklist”, earlier in May, and on Friday, November 22, were voted unanimously as national security risks by the US Federal Communications Commissions.	Increase
90	12/13/2019	US, China agree to ‘phase one deal’ just before next tariff hike	China releases second set of US products to be excluded from additional tariffs	Decrease

International Review of Accounting, Banking and Finance

Vol 17, No. 2, Summer, 2025, Pages 20-38

IRABF

Evaluating Escalating Managers' Performance with Outcome Knowledge: Experimental Evidence and Implications for Project Management

Young-Won Her [1]

David Nazarian College of Business & Economics Department of Accounting
California State University, Northridge

Accepted June 2025

A B S T R A C T

This study investigates how the escalating behavior by project managers and the subsequent outcome affect evaluations of their decision performance, which is an important but unanswered question in prior escalation research. A decision-making experiment was conducted to examine the role of these factors in performance appraisal by asking 117 student participants to evaluate the decision performance of a hypothetical project manager. The results indicate that a manager's decision to escalate his commitment to a failing project has a negative effect on performance evaluation. The evaluations are also affected by the outcome valence (successful or unsuccessful) perceived by evaluators although this ex post information is not indicative of the decision quality. Additionally, there is a significant interaction effect found between the escalation decision and the outcome valence, suggesting that evaluators tend to take a more cautious attitude toward the decision quality in their appraisal process when they receive negative outcome than positive outcome information. The implications of findings for practice and for future research are discussed.

Keywords: Escalation of commitment; outcome effect; decision performance evaluation; sunk cost
JEL Classification: M41

[1] Email: youngwon.her@csun.edu

1. Introduction

A significant body of research has documented the anecdotal evidence on escalation of commitment by project managers (Yang et al., 2023; Sleesman, 2019; Devigne et al., 2016; Hsieh et al., 2015; Drummond, 2014; Gomez & Sanchez, 2013; Denison, 2009; Whyte & Saks, 2007). The typical finding from this stream of research is that managers sometimes become overly committed to their projects and this over-commitment, surprisingly, appears more apparent when they receive negative rather than positive feedback about the future of their projects. This paradoxical phenomenon where a failing project continues to be funded instead of being abandoned or redirected has been known as the escalation effect in the project management context (Sleesman, 2019; Behrens & Ernst, 2014; He & Mittal, 2007).

A considerable amount of research has been undertaken to explore what causes such irrational resource allocation decisions and how this dysfunctional behavior can be mitigated (Ohlert & Weißenberger, 2020; Brüggen & Luft, 2016; Sarangee et al., 2014; Behrens & Ernst, 2014; Sleesman et al., 2012; Kadous & Sedor, 2004). It should, however, be noted that the primary focus of prior escalation research has been on the part of a decision maker as an "actor" and no attention has been paid to the part of a performance evaluator as an "observer." This is understandable because the nature of the central issues raised with this topic directs researchers' attention toward the decision maker who is subject to escalation and how such costly behavior can be suppressed. The present study, however, proposes that escalation research should extend its boundary to the point where the role of a performance evaluator can also be examined since the actor and observer are both important players in this unique phenomenon.

Staw (1981), although he hoped to reject his conjecture, may be correct to suggest that the escalation phenomenon might represent a post-hoc reconstruction of events by observers (Bowen, 1987). In fact, in many cases, the ultimate judgment on whether a decision maker's action was appropriate or not is made by observers, typically after the associated outcome information is available. Therefore, this type of ex post judgment by observers may be inefficient or even fallible since the issue of escalation, by the definition, will be raised only when the eventual outcome turns out to be negative. Further, it has been suggested that the tendency of managers to escalate their commitment may be influenced by the nature of performance evaluation systems by which their actions and decisions are observed and evaluated (Liang, 2019; Kadous & Sedor, 2004; Frederickson et al., 1999). This indicates that it may also be important to consider the judgmental role of observers in the escalation context for a fuller understanding of the phenomenon. As such, the research in this field may not be complete until this additional issue is addressed: how evaluators perceive the manager's escalating behavior? The current study seeks to answer this question which has been largely ignored in prior literature.

Bowen (1987), in his critical review of earlier escalation research, argued that the post-hoc practice of labeling certain decisions as "errors" (presumably because of the negative outcomes resulted) may be misleading in some cases. Although his major intention in this argument is to emphasize the uncertainty that might have existed in the decision situations faced by managers who are blamed for their escalation, it also implies that performance evaluations by observers are inherently vulnerable to the "outcome effect."

The outcome effect refers to the phenomenon whereby performance evaluators systematically overweight their outcome knowledge in assessing a manager's decision performance, and thereby developing their evaluative judgement in the direction of the outcome known (Chen et al., 2021; Mertins et al., 2013). Thus, when the outcome turns out to be positive (negative), evaluators tend to evaluate the manager more positively (negatively), regardless of the actual appropriateness of the decision which resulted in that outcome. Since it is possible that the escalation of commitment by managers may bring a turnaround of the situation and a positive as well as a negative consequence (although less likely), evaluating managers' decision performance in the escalation context may also be susceptible to the outcome effect. To explore this issue in the context of project escalation, the study incorporates outcome information as another important research variable. Accordingly, the

main objective of this study is to investigate how managers' project management behavior (either escalating or non-escalating) and the information about its resultant outcomes (either good or bad) affect evaluators' assessments of the managers' decision-making performance.

A case study-based experiment, in which participants were asked to evaluate a hypothetical project manager's decision performance, was performed to test the hypotheses that managers who discontinued their existing project in favor of a better alternative project will be more favorably viewed by evaluators than managers who continued their current project despite its anticipated lower profitability, and that a successful outcome will lead to a more favorable evaluation than will an unsuccessful outcome. The experimental findings of this study supported the predicted relationships. A manager who decided to escalate his commitment to a failing project was less favorably evaluated than a manager who did not exhibit such escalating behavior. The evaluations were also influenced by the outcome valence (successful or unsuccessful) even though this ex post information could not be an indicative of the decision quality. As such, decision appropriateness and outcome information were both important determinants of assessing managers' decision quality, although the outcome knowledge turned out to be apparently a dominant factor in the current study.

The major contribution of this study is to integrate two independent research streams in the areas of escalation of commitment and outcome effects to consider an important research question which is otherwise difficult to be addressed by separate research endeavors. Another important contribution is that the study provides implications for the way performance evaluation systems are designed and implemented by management. The results of this study show that a manager's escalating behavior may be appraised in a significantly different manner depending upon the subsequent outcomes. The consequence of such evaluations may have costs for organizations: managers who made an improper escalating decision may be praised and rewarded because of an unexpected positive outcome. Conversely, managers who made an appropriate decision may be blamed and punished because of a negative outcome which was expected to be less probable at the time of their decisions. Both instances have negative impacts on the generalization of valid organizational learning as they may subvert rules, policies, and procedures maintained by organizations to guide their members' professional judgement and decision-making. It is thus important for companies to be aware of the findings of this study and to consider its implications for the design of their evaluation and feedback systems.

The remainder of this paper is organized as follows. A brief review of extant literature on both escalation of commitment and outcome effects follows this section. A series of hypotheses focusing on the research question raised by this study is then developed. In the following section, the procedures used to perform the experiment which tests the hypotheses are described. The analysis procedures and the results of the experiment are then summarized. Finally, implications of the findings, limitations of the study, and some suggestions for future research are discussed.

2. Literature Review and Hypotheses

2.1 Research on the Escalation of Commitment

Due to its costly behavioral implications, the escalation of commitment by managers in a firm has been one of the most intensively researched topics in the area of organizational behavior and management accounting. A substantial body of research in this area suggests that managers persist in committing resources to their projects even after receiving negative feedback that the initial investment has not reached its goals (Sleesman et al., 2012). A need to justify their previous decision, which is thought to be driven by personal responsibility for the negative consequence, was proposed as the main motivation for this seemingly irrational decision behavior (Gomez & Sanchez, 2013; Schultze et al., 2009; Staw, 1981). That is, in committing more resources to their prior decision, managers are described to hope to prove that they were not erroneous in their initial judgment or choice.

This type of affective explanations based on the psychological perspective, however, was not successful in covering all kinds of empirical evidence reported because some findings were not in accordance with the prediction based on such explanations. For example, contrary to the justification theory, some studies found that escalation still occurred even when managers were not personally

responsible for the prior decisions (Arkes & Blumer, 1985) and when negative feedback could be attributable to external events, which should have lessened managers' feelings of responsibility (Staw & Ross, 1978). In some instances, managers demonstrated rational decision behavior even though the conditions faced by these managers were conducive to escalation (Whyte & Saks, 2007; Leatherwood & Conlon, 1988).

Accordingly, subsequent research efforts in this area were directed at finding new approaches which may either complement the existing psychological model or provide better explanations for the irregularities detected in prior studies. A good example is Harrell and Harrison's study (1994). They suggested that the conflicting results in prior escalation research may be resolved by applying the expanded view of rational economic decision-making assumed by agency theory. To test this idea, they assessed the viability of two agency theory concepts, an incentive to shirk and privately held information, in addressing the escalation issue. Their results indicated that managers' escalation behavior is conditional on the existence of these two agency theory constructs, confirming the potential usefulness of this alternative approach.

Another example can be found in a group of studies that have attempted to improve the existing models by rectifying either conceptual or methodological flaws identified in earlier experiments (Victoravich, 2010; Denison, 2009; Schulz-Hardt et al., 2009; He & Mittal, 2007; Harrison & Harrell, 1995). These studies typically argue that past research on escalation often failed to establish decision making situations in which escalating decision is clearly economically inadvisable. Since no credible criteria or standards against which to compare the manipulated negative feedback were provided, the subjects in the earlier studies are prone to have difficulties in perceiving the given feedback as truly negative, which makes their decisional choice highly ambiguous. As a result, it is difficult to determine whether the escalation effects reported by previous experiments truly reflect a decisional error or simply indicate a preference for consistency as a way to respond to uncertainty when there is no systematic decision rule. A commonly suggested solution to this methodological deficiency was to provide explicit and relevant prospective information as a decision criterion, which may help decision makers determine the exact nature of the feedback received. Once this correction has been made, researchers found that the strong responsibility effect found in earlier studies either disappears or significantly weakens, which implies that prospective information is more important than either retrospective information or personal responsibility in managers' project evaluation decision.

More recent studies tend to focus on exploring factors that may moderate the level of escalation in various decisional contexts. The main assumption taken by this stream of research is that the escalation phenomenon should be viewed as contingent on a number of factors affecting specific situations. Factors that have been examined as relevant include the magnitude of prior resource commitments (Devigne et al., 2016), the frequency of failures associated with a project (Sarangee et al., 2019; Lant & Hurley, 1999), the availability of decision aids or consultant advices (Ohlert & Weißenberger, 2020; Loh et al., 2019; Behrens & Ernst, 2014), and other relevant personality and psychological variables, such as self-efficacy and anticipated regret (Liang, 2019; Sarangee et al., 2019). All of these variables were found influential in determining a decision maker's commitment level.

As can be seen from the literature review provided above, commitment is a complex process, influenced by multiple and sometimes conflicting factors. As noted earlier, however, there has been no research attempt to understand how the managers' escalating behavior is perceived by performance evaluators who may have an important feedback or learning effect on managers' project-related decisions. Addressing this issue requires a review of the outcome effect literature since the relevant research has shown that evaluators are typically susceptible to the outcome effect.

2.2 Research on the Outcome Effects

The hindsight (or outcome) bias paradigm[2] was established in the probabilistic judgment literature mainly by Fischhoff and his colleagues' contributions (Fischhoff, 1975; Fischhoff & Beyth, 1975;

[2] Previous literature has made a distinction between the effect of outcome knowledge on remembered probabilities and on evaluations. The former is generally called the hindsight bias and the latter the outcome effect (Lipe, 1993).

Slovic & Fischhoff, 1977). Since then, many researchers have attempted to apply the main implications of this paradigm to the evaluative contexts. Such expanded applications were undertaken in a variety of experimental settings (Chen et al., 2021; Mertins et al., 2013). For instance, Mitchell and Kalb (1981) examined the outcome effect on supervisors' evaluations of subordinates in a health care setting. The study found that supervisors with outcome knowledge, especially in the case of a negative outcome, rated the outcome as more probable, held the subordinate more responsible for the outcome, and made more internal attributions for the outcome than did supervisors with no outcome knowledge. In a series of similar experimental studies, Baron and Hershey (1988) provided subjects with a set of 12-16 medical and gambling decisions to evaluate and the outcome of each decision as well. The results showed that the valence of outcome (good or bad) systematically influenced subjects' judgement of the quality of each decision. Similarly, Lipshitz (1989) who employed a military setting to test the outcome effect reported that decision makers and their decisions were perceived more favorably when favorable outcomes were disclosed. Decisions that resulted in a successful outcome were also considered more justifiable and perceived to follow from a superior decision process.

Like these examples, studies that test for the outcome effect in the evaluative context have generally found that the effect is quite pervasive in its occurrence. As a result, research interest in this area has shifted away from whether or not evaluators exhibit the outcome effect, and toward other topics such as identifying the possible causes of the effect and testing various debiasing schemes that are designed to eliminate the outcome effect (Mertins et al., 2013; Peecher & Piercey, 2008). For example, Brown and Solomon (1993) experimentally investigated the viability of three possible competing explanations for the outcome effect: cognitive reconstruction, self-enhancing motive, and an escalation-of-commitment analogue. Their experimental results revealed that the cognitive interpretation provides the most complete account of the outcome effect. They also attempted to attenuate the outcome effect associated with the capital investment decision appraisal, by enhancing the involvement of evaluators in a manager's (i.e., an evaluatee's) decision process. As they predicted, prior advisory involvement was found to be effective in making the manager's decision environment more visible to evaluators, and thereby facilitating their evidence recall (other than the outcome) during the evaluation process. Likewise, Fisher and Selling (1993) observed that an ex ante agreement on outcome prediction between the evaluator and the evaluatee significantly reduced the outcome effect. Other studies also reported several different factors as potential moderating variables. Such variables include mental framing (Jones & Chen, 2005; Lipe, 1993), outcome controllability (Ghosh, 2005; de Villiers, 2002; Tan & Lipe, 1997), decision uncertainty (Peecher & Piercey, 2008; Ghosh & Ray, 2000), the degree of surprise associated with outcomes (Charron & Lowe, 2008) and the type of performance evaluation systems (Mertins, 2010; Frederickson et al., 1999).

Despite these numerous efforts to mitigate the outcome effect on performance evaluation, the general conclusion reached by researchers is that the outcome effect is fairly robust as they found that most debiasing procedures they tested were only marginally successful. The persistence of this outcome effect, therefore, validates the expectation of this study that managers' escalating behavior may be viewed differently depending on how the associated outcomes turn out. In the next section, the effects of escalation of commitment and outcome information on performance evaluation are considered, and testable hypotheses are developed.

2.3 Effects of Escalation of Commitment on Performance Evaluation

In the process of capital investment decisions by which a firm's limited resources are allocated to one decisional alternative over others, it is essential to compare competing projects on the basis of their economic merits. This type of comparative analysis is not necessarily confined to only the initial investment selection stage, but it is also applied to the post investment stage to verify the continued economic viability of the selected projects. For example, the profitability of ongoing projects is often periodically reviewed by comparing their economic performances with certain criteria. Such criteria may include the predetermined hurdle rate (the minimum acceptable return on investment), the profitability of alternative investment opportunities, or the salvage value (the opportunity cost of continuing the current project). Whichever criterion is used, managers' decision whether to continue

or discontinue an existing project typically requires both performance and criterion information. Such information can be either historical (past-oriented) or prospective (future-oriented) in nature. However, it is important to note that the only relevant information in managers' project evaluation decisions is prospective information because their choice should be based on the predicted difference in future performance under each alternative (Horngren et al., 2022). Historical data, such as sunk costs, on the other hand, do not provide any relevant information to the current decision although they may have indirect bearing on the decision by helping in forecasting the future. Critics of past escalation research (Victoravich, 2010; Denison, 2009; He & Mittal, 2007), as already noted in the preceding literature review, clarify this point by arguing that many prior studies failed to provide necessary future information, resulting in highly indeterminable situations where essentially, any decisional choice can be right and wrong. The absence of decision-relevant information also makes it impossible on the part of a performance evaluator to judge the decision quality of a manager without referring to the associated outcome because there are no credible criteria other than outcome available for evaluation.

In order to avoid this problem, the present study employs a life-cycle model of resource allocation based on expected values, which was proposed by Northcraft and Wolf (1984) and adopted by several recent studies on escalation (Liang, 2019; Brüggen & Luft, 2016; Denison, 2009). Northcraft and Wolf suggested that the time-adjusted-rate-of-return be used as a criterion for the allocation of resources to projects throughout their useful lives in situations where calculations of such profitability measures can be realistically made. This approach allows managers to compare the expected rate of return for the remaining life of their projects with that of competing investment alternatives. The major benefit of using this model, therefore, is that it provides managers with explicit decision-relevant prospective information, and thereby clarifies when a financial setback is likely to constitute a rational reason to terminate or abandon their current project. In addition to this conceptual superiority, the use of this model is consistent with the current practice of capital investment appraisal where the adoption of such discounting models has dramatically increased over the years (Horngren et al., 2022).[3]

As described above, if decision information that is both relevant to and necessary for project evaluation is available, decision theories based on the rational choice paradigm assume that a firm's managers will reach decisions that maximize the profitability of their firm. In other words, if managers find that the future performance of their project is not likely to meet certain criteria (e.g., a hurdle rate, the expected ROI of alternative projects, etc.), they are expected to discontinue the project so that its resources can be redirected to a better alternative. However, as evidenced in numerous escalation studies, managers are often overly committed to their projects, and thereby sometimes making a decision which is largely divergent from what normative decision theories suggest. From a viewpoint of performance evaluators, managers' escalating commitment will be perceived dysfunctional only when its associated consequences are expected to be negative. Accordingly, the escalating behavior perceived by evaluators to be undesirable is by nature against a normatively suggested solution. In other words, a decision which does not conform to the normative decision rules, such as the expected return rule discussed above, should be viewed by evaluators as an indication of inappropriate escalation.

Several studies investigated the impact of normativeness of decisions on performance evaluation.[4] For example, Lipshitz (1989) found that those taking normatively appropriate actions were evaluated more favorably than others. Similarly, Mowen and Stone (1992) observed significantly higher performance ratings when a manager's decision matched with expected value calculations. Lipe's (1993) experiment, in which a manager's variance investigation decision was assessed by student subjects, also reported a significant effect for the normativeness of the decision reviewed. Such evidence leads to a prediction that, in the context of this study, managers' escalation

[3] It has been reported that discounted-cash-flow (DCF) models, which explicitly consider the time value of money, are used by more than 85% of the large industrial firms in the US (Horngren *et al*., 2022).

[4] These studies typically use the term, "decision quality," operationalizing it as the congruence of decisions with normative decision models such as the expected value model (Tan & Lipe, 1997).

of commitment to a project which is not recommended by a normative decision model will have a negative impact on performance evaluation. Since the normative decision rule in the current study is based on the expected value of profitability, the normatively correct decision is defined as continuation of a project which has a higher expected profitability and cancellation of a project which has a lower expected profitability than those of alternatives. Accordingly, the prediction proposed above is restated within this study's framework as follows:

H1: Managers who decided to cancel the current project in favor of an alternative project which was expected to be more profitable will be more favorably evaluated than managers who decided to continue the current project.

2.4 Effects of Outcome Information on Performance Evaluation

From the perspective of normative decision theories, the distinction between a good decision and a good outcome is crucial in decision analysis since it is the decision process and not the decision outcome that is a relevant criterion for judging decision quality (Mertins et al., 2013; Ghosh, 2005; Lipshitz & Barak, 1995). Information that can only be available after a decision is made is irrelevant to assessing the quality of that decision. In certain instances, however, outcomes may be valid, although imperfect, inputs to assessments of decision quality (Peecher & Piercey, 2008). Specifically, when a manager has extensive ex ante information (and more information than the evaluator), outcomes can serve as diagnostic cues to the decision process used by the manager since it is reasonable to assume that bad (good) outcomes are more likely to result from poor (good) decisions (Mertins et al., 2013; Tan & Lipe, 1997). If there is no information asymmetry problem, however, the evaluator should only use the information about potential outcomes and the probabilities and utilities of those outcomes that existed at the time of the decision made in assessing the decision quality since the actual outcomes in this case are uninformative (Peecher & Piercey, 2008; Hershey & Baron, 1995).[5] Thus, whether or not it is appropriate to use outcome information in performance appraisals depends on the observability of the manager's decision process and the causal relationship between the decision and its associated outcome (e.g., outcome controllability). For this reason, the effects of outcome information on performance evaluation, unlike in the hindsight bias paradigm in the psychology literature, cannot be unambiguously labeled a bias (Chen et al., 2021; Mertins et al., 2013).

Since a manager's project continuation decision is made ex ante (i.e., before the results of his/her decision are known), the manager should be evaluated based on the ex ante information if that information is shared by the manager and the evaluator. Research on the outcome effect, however, has shown that ex post information also affects performance evaluation even when the available outcome information cannot be indicative of the decision quality. For instance, Baron and Hershey (1988) allowed their student subjects to have all the relevant information known to the decision maker, plus the outcome knowledge. Nevertheless, the students still appeared to take their outcome knowledge into account in rating the quality of the decision they reviewed. Likewise, Fisher and Selling (1993) investigated whether the outcome effect can be eliminated when the decision process used by an evaluatee is observable to evaluators. Their experiment results showed that the outcome effect still existed even under conditions of perfect knowledge of the decision process algorithm. They conjectured that the reason of failure in complete elimination of the outcome effect by the observability might be due to the absence or inadequacy of knowledge about the optimal transformation process of information into the decision. This uncertainty as to what is the definitely appropriate decision may have caused evaluators to still rely on the outcome information even though they had the perfect knowledge of the decision process.

Based on the theoretical discussion and the existing empirical evidence provided above, it is predicted that evaluators of this study will also consider the ex post outcome information (i.e., the degree of success in the chosen project), as well as the ex ante forecasted profitability data which

[5] This discussion is consistent with the agency theory perspective (Holmstrom, 1979). That is, when the agent's effort is observable, the first best solution is to reward the agent based upon those efforts. On the other hand, if the agent's effort is either unobservable or imperfectly observable, the best solution is to reward the agent upon outcomes alone or some combination of outcomes and information that is incrementally revealing of effort.

were used by managers to make their project continuation decisions. As a result, when the outcome turns out successful, the manager's decision may appear more appropriate ex post, and thereby leading to a higher performance rating than when the outcome turns out unsuccessful. The following hypothesis formulates this prediction:

H2: Managers who chose a project that turns out to be successful will be more favorably evaluated than managers who chose a project that turns out to be unsuccessful.

2.5 Interactive Effects of Escalation of Commitment and Outcome Information

Another noteworthy finding of prior outcome effect research is that the valence of outcome (whether it is positive or negative) may have differential impacts on evaluators' cognitive efforts in their assessment task (Mertins et al., 2013). In general, research found that negative information is more heavily weighted than positive information in performance evaluation. For example, Mitchell and Kalb (1981) found that a poorly performing subordinate is blamed more harshly when the resulting outcome is negative than when it is positive. In the context of audit litigation, Peecher & Piercey (2008) observed that the possession of adverse outcome information by evaluators leads to harsher judgements of auditor negligence. Similarly, Tan and Lipe (1997) investigated whether the performance evaluation is affected by the outcome controllability by managers, and reported that their predicted relationships are found only under the negative outcome condition. That is, with a failed outcome, the performance evaluation varied depending on whether the outcome was controllable or uncontrollable by managers, whereas with a successful outcome, there was no such variation. They explained this conditional effect of controllability by referring to the salience of negative outcomes in performance evaluation. It was speculated that the justification or consideration of reasons as to why a particular outcome occurred may become more important when the outcome is negative than when it is positive. Accordingly, under the circumstance where justifications are more essential for poor outcomes, it is natural for evaluators to perform more intensive decision analysis when outcomes are negative than when they are positive.

Some prior research provides support for this idea. In a study of corporate annual reports' letter to shareholders, Bettman and Weitz (1983) found that more detailed causal analysis is provided for unfavorable than favorable company performance. Wong and Weiner (1981) reported that spontaneous attributional thinking is more prevalent in the face of negative outcomes than for positive outcomes. Empirical evidence like this has an implication for the current study as it suggests that a negative outcome may lead to a more engaged and careful analysis of causes for the outcome. In the context of this study, the valence of a project outcome (either successful or unsuccessful) could make evaluators take very different attitudes in analyzing the manager's decision process. In other words, evaluators told of an unsuccessful outcome may exert greater cognitive efforts to analyze the manager's decision process, whereas those told of a successful outcome may either put inadequate efforts into their analysis or become less sensitive to the quality of the decision made. Accordingly, it is expected that the distinction made by evaluators between the good (non-escalating) and bad (escalating) decisions will be more significant under the condition of an unsuccessful outcome than under the condition of a successful outcome. The following interaction hypothesis formulates this expectation:

H3: There will be a greater difference between the performance evaluations of escalating managers and non-escalating managers when the project outcome is unsuccessful than when it is successful.

3. Method

3.1 Participants

A behavioral decision-making experiment was conducted to examine the hypothesized relationships. Participants were 128 MBA students enrolled in an advanced managerial accounting course at a large public university (61 females and 67 males). The typical subject was about 27 years old and had 4-5 years of work experience. All of the participants provided usable responses. The questionnaires were

distributed and returned during a regular class session, taking about 30 minutes for their completion. Participation was voluntary and consent was obtained. A small amount of course credit was offered to encourage participation, and participants were assured that their responses were confidential and anonymous. Since most participants either had just completed or were currently taking courses in which the basic knowledge and skills for various types of decision-making in business are taught (e.g., economics, finance and managerial accounting), they appeared to be academically prepared for the current study's experimental task.

3.2 Experimental Task

The participants were projected into the role of a senior manager who has been asked to evaluate the decision-making performance of a hypothetical junior project manager.[6] It was described in the case scenario (see Appendix) that Patrick, the hypothetical junior project manager, who possessed independent decision-making authority for his project selection and management, currently launched his third project (Project Q) after successful completion of the initial two projects that he had initiated and managed. The new project had an expected lifetime of five years and its overall expected net present value was estimated as $7,000,000 at that time. After 3 years, however, the project was behind schedule with cash flows about 50% less than originally forecasted. Accordingly, he reevaluated his project's future prospect to decide whether the project should be continued for the remaining two years of its lifetime, or terminated early so its resources could be used for another project (Project Z) which had the same years of expected lifetime as the remaining period of the current project. The explicit future performance data associated with these two alternative projects were available for his project continuation decision as presented in Table 1.

Table 1. Expected Future Performance Information of the Projects

Project Q **(Current Project):** Expected net present value for the remaining two years:

20% chance of a net present value of $6,000,000; .20 × $6,000,000 = $1,200,000
80% chance of a net present value of $3,000,000; .80 × $3,000,000 = $2,400,000
Expected net present value …… **$3,600,000**

Project Z **(Alternative Project):** Expected overall net present value for its two year lifetime:

50% chance of a net present value of $5,000,000; .50 × $5,000,000 = $2,500,000
50% chance of a net present value of $4,000,000; .50 × $4,000,000 = $2,000,000
Expected net present value …… **$4,500,000**

As shown in the table, since the expected net present value of the current project for the remaining two years of its lifespan is less than the expected net present value of the alternative project during the same period, a normatively correct decision is to discontinue the current project and transfer its resources to the alternative project. Thus, if Patrick is persistent in continuing his current project despite its lower profitability projected, this should be viewed as an indication of escalating

[6] The selection of a junior project manager as an evaluatee was intended to provide the participants with an implication that this type of managers may have a stronger incentive to escalate their commitment even in a failing project. Prior research (Harrell & Harrison, 1994) indicates that a junior project manager with a growing reputation for successfully managing projects (like Patrick in this study's case scenario) could be more vulnerable to the sunk cost or escalation effects than a senior project manager with a relatively solid reputation gained over a period of years. This is because the relative impacts of their performances in a single project (particularly when it was unsuccessful) on their reputation, job security and/or marketability would be much different between the junior and senior project managers. To make this point clearer to the participants, it was stated that generally, managers gain a reputation as being talented when their managed projects are successful while unsuccessful project can damage their reputation and career potential.

his commitment in the failing project.

After reading the case scenario described above, the participants received information about what decision Patrick eventually made and how the subsequent outcome related to his decision turned out. The participants then were asked to evaluate his decision performance. The response was elicited on a Likert-type scale that ranged from 1 to 10 in which the end points were labeled "unsatisfactory decision-making performance" and "satisfactory decision-making performance," respectively.[7] Accordingly, larger numerical responses indicate more positive performance evaluation.

3.3 Research Design

The study employed a 2 × 2 between-groups factorial design as depicted in Table 2. The two independent variables are escalation of commitment (yes/no) and project outcome (good/bad). The participants were randomly assigned to one of the four experimental groups. The escalation of commitment was manipulated at two levels by informing the participants that the project manager (Patrick) decided to either continue (escalating) or discontinue (non-escalating) a project which was not recommended by the normative decision rule (i.e., the expected net present value approach). Similarly, the project outcome was manipulated at two levels by providing the participants with the actual performance information about the project chosen, as either better (successful) or worse (unsuccessful) than was originally estimated.

Table 2. Research Design Used in the Experiment*

Project Outcome***	Escalation of Commitment**	
	Yes	No
Good	***Group 1*** (n = 30)	***Group 2*** (n = 31)
Bad	***Group 3*** (n = 29)	***Group 4*** (n = 27)

* Dependent variable: evaluations of the decision performance of a hypothetical project manager (1 = Unsatisfactory; 10 = Satisfactory)

** Escalation of commitment was manipulated at two levels by informing that the hypothetical project manager decided to either continue (Yes) or discontinue (No) a project which was not recommended by the expected net present value approach.

*** Project outcome was manipulated at two levels by informing that the actual performance of the project chosen by the hypothetical project manager was either better (Good) or worse (Bad) than was originally estimated.

Two manipulation check questions were used to assess how well the participants understood the specific treatments given to them. For the escalation of commitment treatment, the manipulation question asked them to answer whether the project manager decided to continue his project in the case scenario they just reviewed. A 9-point response scale was used to measure the degree of participants' agreements with this statement so that larger numbers indicated stronger agreements while smaller numbers stronger disagreement with the statement. For the project outcome treatment,

[7] In order to minimize the centralizing tendency of responses which often occurs when there is overly strong wording such as "definitely" or "extremely", the end points were labeled using relatively mild words.

it was asked whether the eventual outcome of the manager's chosen project was successful. The same response scale described above was used to measure the agreements with this second manipulation check statement. Lastly, the participants were asked to provide some demographic information such as gender, age, education level, and work experience.

4. Analysis and Result

4.1 Preliminary Analyses

Prior to hypothesis testing, various preliminary analyses were performed to assure the effectiveness of the randomization process and the satisfaction of model requirements. The chi-square test results indicated that the participants' gender and education level were not significantly different across the four experimental groups ($\chi2 = 0.89$, $p = 0.83$ and $\chi2 = 10.80$, $p = 0.29$, respectively). The results of ANOVA which was performed for the other quantitative demographic variables (age and work experience) also indicated no significant differences for these variables among groups ($F = 0.28$, $p = 0.84$ and $F = 1.04$, $p = 0.38$, respectively). Thus, the randomization procedure appeared to be successful. In addition, no systematic relationships were found between the demographic variables and the participants' responses, from a regression analysis in which all demographic variables listed above were incorporated as independent variables together with the two experimental factors. Accordingly, differences in demographic variables do not appear to influence the results of this study. Other univariate tests performed on the response variable suggested that the basic assumptions for the analysis of variance (e.g., normality, equal variances) were reasonably well met by the data.

An examination of the manipulation check data revealed that the manipulation of this study was successful. As expected, the mean response in the escalation condition (7.7) was significantly larger than the mean response in the non-escalation condition (2.3) for the first manipulation check question which asked how strongly the participants agreed with the statement that the project manager decided to continue his current project ($t = 22.33$, $p < .0001$). Similarly, for the outcome manipulation check question in which the participants were asked whether the actual outcome information they received indicated success, a significantly greater mean agreement was found for the good outcome condition (7.9) than for the bad outcome condition (2.4) as intended ($t = 24.55$, $p < .0001$). Since eleven of the 128 participants (about 9%) made obviously wrong answers to one or both of the manipulation check questions, the 3-step ANOVA procedure previously used by Harrell and Harrison (1994) was employed to determine whether these participants' incorrect answers on the manipulation check influenced their evaluation responses.[8] While there was no evidence of such influences, it was decided to include only those participants who passed the manipulation check (117 respondents) for hypothesis testing. The study results, however, remained unchanged when the analyses described below were performed with all 128 participants.

4.2 Hypothesis Testing

A 2 × 2 analysis of variance (ANOVA) was conducted to test the hypotheses. Table 3 summarizes the results of this analysis. As shown in Panel A of this table, the overall model is statistically significant ($F = 32.74$, $p < .0001$). The results also indicate significant main effects for both escalation of commitment ($F = 28.36$, $p < .0001$) and project outcome ($F = 59.62$, $p < .0001$). Of more importance to this study is that the interaction between these two variables was also found significant ($F = 8.88$, $p = 0.0035$), as will be explained in greater detail below. The omega-square statistics (ω^2), which measure the relative impact size of each variable, indicate that the outcome effect (28%) dominates the escalation effect (13%) and its interactive effect with escalation (4%) on participants' evaluation responses, which is similar to the typical findings of prior outcome effect research (e.g., Ghosh & Ray, 2000; Tan & Lipe, 1997; Lipshitz, 1989).

[8] This procedure incorporates in an ANOVA model a categorical variable which indicates whether the participants' manipulation check responses were correct or incorrect. In the first step, this manipulation check variable is added as a main effect. Next, if it is not found significant, an interaction term between this variable and one of the experimental variables replace the main effect term. This process continues until all possible interactions are tested.

Table 3. Results of Hypothesis Testing (N = 117)

Panel A: Results of Analysis of Variance*

Source	F	p-value	ω^2
Model	32.74	< .0001	
Escalation of Commitment (E)	28.36	< .0001	0.13
Project Outcome (O)	59.62	< .0001	0.28
Interaction (E × O)	8.88	.0035	0.04

Panel B: Mean Performance Evaluation by Groups (standard deviations in parentheses)

Project Outcome***	Escalation of Commitment**		Overall
	Yes	No	
Good	***Group 1*** 6.83 (1.46) *n* = 30	***Group 2*** 7.61 (2.19) *n* = 31	7.23 (1.89) *n* = 61
Bad	***Group 3*** 3.28 (1.67) *n* = 29	***Group 4*** 6.04 (1.76) *n* = 27	4.61 (2.20) *n* = 56
Overall	5.09 (2.37) *n* = 59	6.88 (2.14) *n* = 58	6.07 (2.45) *N* = 117

Panel C: Planned Comparisons

	t	*p*-value
Escalation *vs.* no escalation decisions under good outcome:		
(Group 1 vs. Group 2)	1.64	0.107
Escalation *vs.* no escalation decisions under bad outcome:		
(Group 3 vs. Group 4)	6.02	< 0.001

* Dependent variable: evaluations of the decision performance of a hypothetical project manager (1 = Unsatisfactory; 10 = Satisfactory)

** Escalation of commitment was manipulated at two levels by informing that the hypothetical project manager decided to either continue (Yes) or discontinue (No) a project which was not recommended by the expected net present value approach.

*** Project outcome was manipulated at two levels by informing that the actual performance of the project chosen by the hypothetical project manager was either better (Good) or worse (Bad) than was originally estimated.

Further analyses were proceeded to examine the nature of the observed main effects as well as interaction effects. Panel B contains the mean level of performance ratings given by participants in each experimental group. As shown in the table, the significant differences are in the predicted directions.

Specifically, the participants evaluated the manager's decision performance more positively when the manager discarded his current project in favor of the alternative project which was

recommended as a better option by the expected return rule (M = 6.88) than when he was continuously committed in his current project (M = 5.09). These results are consistent with the first hypothesis predicting that the performance evaluation of managers will be affected by the normativeness of decision-making reflected in their project continuation decisions. Accordingly, H1 was supported. Similarly, the mean performance evaluation was significantly higher when the project outcome was reported as successful (M = 7.23) than when it was reported as unsuccessful (M = 4.61). This is consistent with the second hypothesis predicting that the performance evaluation of managers will be influenced by the outcome information as well. Thus, H2 was also supported.

In terms of extremes within the four groups, Group 2 (no escalation/good outcome) provided the highest level of performance rating (M = 7.61) and Group 3 (escalation/bad outcome) reported the lowest level of performance rating (M = 3.28) as expected. However, it is important to note that only this group (Group 3) among the four groups indicated virtually negative performance evaluation.[9] In fact, the mean performance evaluation of Group 1 (escalation/good outcome, M = 6.83) is positive and in the second highest position although this group, like Group 3, was informed that the manager escalated his commitment in a less profitable project (i.e., a rationally incorrect choice against the normative decision model). Thus, the manager assessed by this group was actually praised rather than blamed for his undesirable escalating behavior because of the successful outcome. On the other hand, the mean performance rating given to the manager of Group 4 (no escalation/bad outcome, M= 6.04) does not indicate a clearly favorable level of assessment even though he made a rationally correct choice based on the normative decision rule.[10] Instead, the manager of this group who made a right decision (no escalation) but unfortunately experienced a bad outcome was actually less favorably evaluated than the manager of Group 1 who made a wrong decision (escalation) but luckily obtained a good outcome.[11] This implies that the participants in the present study were unable to ignore ex post outcome information, which was not available at the time of project managers' decision, but in fact more heavily utilized this information in their assessment task than they used the escalation information.

H3 predicted that there would be a greater negative effect of escalation of commitment on performance evaluations by respondents under the bad outcome condition than under the good outcome condition, implying a significant interaction between the escalation and outcome factors. Since this interaction turned out significant (F = 8.88, p = 0.0035) as shown in Panel A of Table 3, additional analyses were carried out to clarify the nature of the interaction effect found. Panel C of Table 3 reports the results of planned comparisons for testing the third hypothesis. As displayed in the table, when the reported project outcome was successful, there was no significant difference between the performance rating on the escalating manager and on the non-escalating manager (t = 1.64, p = 0.107). In contrast, when the reported project outcome was unsuccessful, there was a significant difference in evaluating the escalating manager and the non-escalating manager (t = 6.02, p < 0.001). Accordingly, the significant main effect observed for escalation in the ANOVA model appears to be mostly due to the difference found in the bad outcome condition. The results of these pairwise comparisons therefore suggest that the participants were more vigilant in discerning the manager's decision quality when the outcome was unsuccessful than when it was successful, as implied in H3.

5. Discussion

Before discussing the results of this study and their implications, some limitations and strengths of this study should be noted. One limitation is that the participants were all students with, perhaps, no or little prior experience or formal training in doing performance ratings. Experienced managers or

[9] The mean performance evaluation of this group (3.28) was significantly lower than the neutral point of 5.5 ($t = 7.19$, $p < 0.001$).

[10] The mean performance evaluation of this group (6.04) was not significantly different from the uncertain value of 5.5 ($t = 1.58$, $p = 0.126$).

[11] The mean performance evaluation of this group (6.04) was significantly lower than the mean evaluation of Group 1 (6.83) at the 10% significance level ($t = 1.86$, $p = 0.068$).

practitioners in the field may have reacted differently to the performance data presented in the current experiment. Accordingly, the study needs to be expanded using participants with professional experience in performance evaluation to see if the findings of this study are replicable. In addition, other evaluation-relevant information (e.g., manager's past performance history) which is perhaps available in a more realistic case was not considered in this study for the purpose of simplicity. The inclusion of this additional relevant information, however, could have produced different results. Accordingly, caution must be taken in extending the results of this study to other groups or settings. A strength of this study is that it employed an experimental research design that is generally considered to have the highest level of internal validity required to establish stronger causal relationships among the variables of interest.

The present study investigated how a project manager's escalation of commitment in a failing project is perceived by evaluators when the eventual outcome is already known. Since a manager's escalating behavior which is not in accordance with a normatively appropriate action should be viewed as dysfunctional, it was hypothesized that the escalation of commitment by a manager will have a negative impact on performance evaluation if the manager's decision process is observable (H1). In addition, since manager's escalating decision may produce a favorable outcome (though less likely) as well as an unfavorable outcome and the evaluators are typically vulnerable to the outcome effect, it was hypothesized that the project outcome information will also affect performance ratings by evaluators (H2). Finally, the salience of negative outcomes in performance appraisal, which was often observed in prior research, resulted in an additional hypothesis that the outcome information will have differential impacts on the evaluators' decision analysis (H3).

The experimental results of this study provided strong support for both the main effect hypotheses (H1 and H2) and the interaction hypothesis (H3). As predicted, a manager who displayed normatively irrational behavior in his project continuation decision (the escalating manager) was less favorably evaluated than a manager who abstained from such behavior (the non-escalating manager). Additionally, the results show that evaluation by outcomes is also evident in the context of evaluating escalating managers. An identical decision was appraised differently depending on its resultant outcome. When the manager decided to continue his commitment in a less profitable project and the subsequent outcome was successful, the manager was in fact praised for his escalating behavior, while the same behavior was criticized, as it ought to be, when the outcome was unsuccessful. These results empower Staw's conjecture (1981) that the escalation phenomenon may represent a post-hoc reconstruction of events by observers. Finally, the results of planned pairwise comparisons suggest that the salience of negative outcomes may have triggered and intensified evaluators' decision analysis efforts in their evaluation process.

The results of this study have implications for both practice and research. From the practical standpoint, the study has implications for the design of systems used to evaluate managers' decision performance. In the present study, ex ante decision relevant information available for a manager' project continuation decision was fully conveyed to the evaluators, and communicating such information did influence their evaluation process. However, as can be seen from the effect size analysis in the results section, the outcome effect was so dominant that most participants did not adequately consider the decision quality information which is more important than the outcome information from the normative perspective and so should be more heavily weighted in their evaluation task. Thus, simply supplying information regarding managers' decision process may not necessarily insure against reaction to the outcome information. This implies that outcome effects may be best dealt with through the design of better performance evaluation systems. The systems may be designed by management, intentionally or unintentionally, such that either decision process quality or outcome valence is the major basis for appraisal. The results of this study suggest that a major emphasis be placed on the decision process and its judgmental criteria (e.g., formal decision rules based on professional experience or expertise, standard operating procedures, etc.) to avoid potentially detrimental consequences from both escalation and outcome effects.

The study also has implications for research examining the effect of feedback on learning. Given that performance evaluation in an organization can be a learning mechanism, ill-structured evaluation

systems may restrict managers' abilities to learn from experience (Brown & Solomon, 1993). If evaluators were subject to outcome effects when evaluating the decisions made by managers, over time such effects could shift the managers' attention away from rational and prospective decision-making and toward risky and persistent commitment to their past decision. For instance, if project managers learn by experience that their performance will be assessed mainly based on outcomes rather than their decision quality and that neither project audits nor information gathering by their company will be undertaken until the completion of the project they initiated, managers are more likely and more frequently to escalate their commitment. The management should consider this issue when designing their evaluation systems.

Further research is needed to determine whether the findings from this study can be generalizable to other contexts than the specific one studied here. There are a wide range of investment situations where the escalation of commitment by actors and the outcome-based evaluation by observers may interplay to preclude optimal investment choices and effective organizational learning. Among those potential research contexts, it is of particular interest to examine an investment setting in which clear prescriptive decision rules or feedback may not be readily available, such as a research and development funding case. Also, prior research indicates that subjective performance evaluation like the one used in this study may be influenced by a variety of personality and organizational variables, such as risk propensity, uncertainty avoidance, locus of control, interpersonal similarity, budgetary participation, and incentive structure. Thus, future research could examine the role of these additional factors in performance evaluation along with the information about managers' decision process.

References

Arkes, H. R., & Blumer, C. (1985). The psychology of sunk cost. *Organizational Behavior and Human Decision Processes*, 35, 124-140.

Baron, J., & Hershey, J. C. (1988). Outcome bias in decision evaluation. *Journal of Personality and Social Psychology*, 54, 569-579.

Behrens, J., & Ernst, H. (2014). What keeps managers away from a losing course of action? Go/stop decisions in new product development. *Journal of Product Innovation Management*. 31(2), 361–374.

Bettman, J., & Weitz, B. (1983). Attributions in the board room: causal reasoning in corporate annual reports, *administrative Science Quarterly*, 28, 165-183.

Bowen, M. (1987). The escalation phenomenon reconsidered: Decision dilemmas or decision errors? *Academy of management Review*, 12, 52-66.

Brown, C. E., & Solomon, I. (1993). An experimental investigation of explanations for outcome effects on appraisals of capital budgeting decisions. *Contemporary Accounting Review*, 10, 83-111.

Brüggen, A., & Luft, J. L. (2016). Cost Estimates, Cost Overruns, and Project Continuation Decisions. *The Accounting Review*, 91(3), 793–810.

Charron, K. F., & Lowe, D. J. (2008). An examination of the influence of surprise on judges and jurors' outcome effects. *Critical Perspectives on Accounting*, 19, 1020–1033.

Chen, J., Kwan, L. C., Ma, L. Y., Choi, H. Y., Lo, Y. C., Au, S. Y., Tsang, C. H., Cheng, B. L., & Feldman, G. (2021). Retrospective and prospective hindsight bias: Replications and extensions of Fischhoff (1975) and Slovic and Fischhoff (1977). *Journal of Experimental Social Psychology*, 96, Article 104154.

de Villiers, C. J. (2002). The effect of attribution on perceptions of managers' performance. *Meditari Accountancy Research*, 10, 53–70.

Denison, C. A. (2009). Real options and escalation of commitment: a behavioral analysis of capital investment decisions. *The Accounting Review*, 84(1), 133–155.

Devigne, D., Manigart, S., & Wright, M. (2016). Escalation of commitment in venture capital decision making: Differentiating between domestic and international investors. *Journal of Business Venturing*, 31(3), 253–271.

Drummond, H. (2014). Escalation of commitment: when to stay the course? *The Academy of*

Management Perspectives, 28(4), 430-446.
Fischhoff, B. (1975). Hindsight ≠ Foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of experimental psychology: human perception and performance*, 104, 288-299.
Fischhoff, B., & Beyth, R. (1975). I know it would happen: Remembered probabilities of once—Future things. *Organizational Behavior and Human Performance*, 13, 1-16.
Fisher, J., & Selling, T. I. (1993). The outcome effect in performance evaluation: decision process observability and consensus. *Behavioral Research in Accounting*, 5, 58-77.
Frederickson, J. R., Peffer, S. A., & Pratt, J. (1999). Performance evaluation judgments: Effects of prior experience under different performance evaluation schemes and feedback frequencies. *Journal of Accounting Research*, 37, 151-165.
Ghosh, D. (2005). Alternative measures of managers' performance, controllability, and the outcome effect. *Behavioral Research in Accounting*, 17, 55–70.
Ghosh, D., & Ray, M. R. (2000). Evaluating managerial performance: mitigating the "outcome effect." *Journal of Managerial Issues*, 12, 247-260.
Gomez, C., & Sanchez, J. I. (2013). Cultural boundaries of self-justification and prospect theories in escalation of commitment: A US–Mexico comparison. *The International Journal of Human Resource Management*, 24(13), 2646–2656.
Harrell, A., & Harrison, P. (1994). An incentive to shirk, privately held information, and managers' project evaluation decisions. *Accounting, Organizations and Society*, 10, 569-577.
Harrison, P., & Harrell, A. (1995). Initial responsibility, prospective information, and managers' project evaluation decisions. *Advances in Management Accounting*, 4, 127-146.
He, X., & Mittal, V. (2007). The effect of decision risk and project stage on escalation of commitment. *Organizational Behavior and Human Decision Processes*, 103, 225–237.
Hershey, J. C., & Baron, J. (1995). REJOINDER Judgment by outcomes: When is it warranted? *Organizational Behavior and Human Decision Processes*, 62(1), 127.
Holmstrom, B. (1979). Moral hazard and observability. *The Bell Journal of Economics*, 10(1), 74-91.
Horngren, C. T., Sundem, G. L., Burgstahler, D., & Schatzberg, J. O. (2022). *Introduction to management accounting*. Pearson.
Hsieh, K., Tsai, W., & Chen, M. (2015). If they can do it, why not us? Competitors as reference points for justifying escalation of commitment. *Academy of Management Journal*, 58(1), 38-58.
Jones, K. T., & Chen, C. C. (2005). The effect of audit outcomes on evaluators' perceptions. *Managerial Auditing Journal*, 20(1), 5–18.
Kadous, K., & Sedor, L. M. (2004). The efficacy of third-party consultation in preventing managerial escalation of commitment: The role of mental representations. *Contemporary Accounting Research*, 21(1), 55–88.
Lant, T. K., & Hurley, A. E. (1999). A contingency model of response to performance feedback. *Group & Organization Management*, 24, 421-437.
Leatherwood, M. L., & Conlon, E. J. (1988). The impact of prospectively relevant information on persistence in a project following setbacks. *Advances in Information Processing in Organizations*, 3, 207-219.
Liang, B. (2019). The effects of self-efficacy, process feedback, and task complexity on escalation of commitment in new product development. *Journal of Business & Industrial Marketing*, 34(8), 1641–1653.
Lipe, M. G. (1993). Analyzing the variance investigation decision: The effects of outcomes, mental accounting, and framing. *The Accounting Review*, 68, 748-764.
Lipshitz, R. (1989). Either a medal or a corporal: The effects of success and failure on the evaluation of decision making and decision makers. *Organizational Behavior and Human Decision Processes*, 44, 380-395.
Lipshitz, R., & Barak, D. (1995). Hindsight wisdom: Outcome knowledge and the evaluation of decisions. Acta Psychologica, 88(2), 105–125.
Loh, C.-Y., Coyte, R., & Cheng, M. M. (2019). Is a fresh pair of eyes always better? The effect of

consultant type and assigned task purpose on communicating project escalation concerns. *Management Accounting Research*, 43, 1–14.

Mertins, L. (2010). The impact of the choice of performance evaluation system on the magnitude of the outcome effect. Working paper. Presented at 2010 Annual American Accounting Association Meeting.

Mertins, L., Salbador, D., & Long, J. H. (2013). The outcome effect - A review and implications for future research. *Journal of Accounting Literature*, 31, 2-30.

Mitchell, T. R., & Kalb, L. S. (1981). Effects of outcome knowledge and outcome valence on supervisors' evaluations. *Journal of Applied Psychology*, 66, 604-612.

Mowen, J. C., & Stone, T. H. (1992). An empirical analysis of outcome biases in constituent evaluations of public policy decision makers. *Journal of Public Policy and Marketing*, 11, 24-30.

Northcraft, G. B., & Wolf, G. (1984). Dollars, sense, and sunk costs: A life-cycle model of resource allocation decisions. *Academy of Management Review*, 9, 225-234.

Ohlert, C. R., & Weißenberger, B. E. (2020). Debiasing escalation of commitment: The effectiveness of decision aids to enhance de-escalation. *Journal of Management Control*, 30(4), 405–438.

Peecher, M. E., & Piercey, M. D. (2008). Judging audit quality in light of adverse outcomes: Evidence of outcome bias and reverse outcome bias. Contemporary Accounting Research, 25(1), 243–274.

Sarangee, K. R., Schmidt, J. B., & Calantone, R. J. (2019). Anticipated regret and escalation of commitment to failing, new product development projects in business markets. *Industrial Marketing Management*, 76, 157–168.

Sarangee, K. R., Woolley, J. L., Schmidt, J. B., & Long, E. (2014). De-escalation mechanisms in high-technology product innovation. *Journal of Product Innovation Management*, 31(5), 1023–1038.

Schulz-Hardt, S., Thurow-Kröning, B., & Frey, D. (2009). Preference-based escalation: A new interpretation for the responsibility effect in escalating commitment and entrapment. *Organizational Behavior and Human Decision Processes*, 108, 175–186.

Sleesman, D. J. (2019). Pushing through the tension while stuck in the mud: Paradox mindset and escalation of commitment. *Organizational Behavior and Human Decision Processes*, 155, 83-96.

Sleesman, D. J., Conlon, D. E., McNamara, G., & Miles, J. E. (2012). Cleaning up the big muddy: a meta-analytic review of the determinants of escalation of commitment. *Academy of Management Journal*, 55(3), 541-562.

Slovic, P., & Fischhoff, B. (1977). On the psychology of experimental surprises. *Journal of Experimental Psychology: Human Perception and Performance*, 3, 544-551.

Staw, B. M. (1981). The escalation of commitment to a course of action. *Academy of Management Review*, 6, 577-587.

Staw, B. M., & Ross, J. (1978). Commitment to a policy decision: A multitheoretical perspective. *Administrative Science Quarterly*, 23, 40-64.

Tan, H., & Lipe, M. G. (1997). Outcome effects: The impact of decision process and outcome controllability. *Journal of Behavioral Decision Making*, 10, 315-325.

Victoravich, L. M. (2010). When do opportunity costs count? The impact of vagueness, project completion stage, and management accounting experience. *Behavioral Research in Accounting*, 22(1), 85–108.

Whyte, G., & Saks, A. M. (2007). The effects of self-efficacy on behavior in escalation situations. *Human Performance*, 20(1), 23–42.

Wong, P., & Weiner, B. (1981). When people ask "why" questions, and the heuristics of attributional search. *Journal of Personality and Social Psychology*, 40, 650-663.

Yang, M. M., Yang, F., & Li, X. (2023). Open innovation and de-escalation of commitment in underperforming new product development projects. *Journal of Business Research*, 164, 1-15.

Appendix: Case Materials

Project Management Case
(Yes-Escalation / Good-Outcome Condition)

As a senior manager of Harvest Project, Inc., you supervise Patrick Lopez, who is a junior project manager. Patrick is employed at another location and you cannot directly observe his work activities. Your company uses a highly decentralized management approach that provides each project manager with a high level of independent decision-making authority. When the projects they initiate and manage are successful, project managers gain a reputation as being talented. When project managers initiate and manage a project that fails, this can damage their reputation and career potential. The initial two projects that Patrick initiated and managed were successful. Project Q, the third project that Patrick initiated and currently manages, has an expected lifetime of five years. Initially, there was a 75% chance Project Q would provide a net present value of $8,000,000 and a 25% chance it would provide a net present value of $4,000,000. Thus Project Q had an overall expected net present value of $7,000,000 (.75 × $8,000,000 = $6,000,000; .25 × $4,000,000 = 1,000,000; $6,000,000 + $1,000,000 = $7,000,000). After three years, Project Q has fallen significantly behind schedule with cash flows that were about 50% less than originally estimated. ***At this point, Patrick evaluated Project Q's future expectations to decide whether Project Q should be continued for the remaining two years of its lifetime, or cancelled so its resources could be used for an alternative project, Project Z.*** Project Z had an expected lifetime of two years and would provide benefits similar to those provided by Project Q. The expected net present value approach is usually used for such decisions in the Company. The two projects are described below.

Project Q**:** Expected net present value for the remaining two years:

20% chance of a net present value of $6,000,000; .20 × $6,000,000 = $1,200,000
80% chance of a net present value of $3,000,000; .80 × $3,000,000 = $2,400,000
Expected net present value …… **$3,600,000**

Project Z**:** Expected overall net present value for its two year lifetime:

50% chance of a net present value of $5,000,000; .50 × $5,000,000 = $2,500,000
50% chance of a net present value of $4,000,000; .50 × $4,000,000 = $2,000,000
Expected net present value …… **$4,500,000**

Patrick has independent decision-making authority for this decision. He was, however, required to either (1) continue Project Q or (2) cancel Project Q and use its resources for Project Z. Patrick believed that he could make Project Q successful, so ***he decided to continue Project Q***.

After Project Q's completion, an internal project audit revealed the results for the last two years of its lifetime. During this time period, Project Q's results were ***better*** than was predicted for either Project Q or Project Z. Project Q was, therefore, a ***successful*** project.

As Patrick's supervisor, you are required to provide an evaluation of his managerial **decision-making performance**. Project Q was the only project Patrick managed during this period. Circle a number below to indicate your evaluation.

Unsatisfactory Decision-Making Performance									**Satisfactory Decision-Making Performance**
1	2	3	4	5	6	7	8	9	10

DO NOT LOOK BACK OR CHANGE YOUR PREVIOUS RESPONSE WHILE COMPLETING THIS PAGE!

I. Respond to the following two questions based on the information presented to you in the case you just completed.

1. When Patrick was required to choose between the two projects, Projects Q and Z, he believed that he could make ***Project Q*** successful, so he decided to ***continue Project Q***.

1	2	3	4	5	6	7	8	9
Strongly Disagree								Strongly Agree

2. After Patrick completed the project that he had chosen, an internal project audit revealed that the actual results of his chosen project were ***better*** than was predicted for either Project Q or Project Z. Thus, the project that he had chosen was a ***successful*** project.

1	2	3	4	5	6	7	8	9
Strongly Disagree								Strongly Agree

II. Participant Information

Please provide the following information about yourself. As mentioned earlier, your replies are not associated with your name; so all replies are completely anonymous.

1. Gender: Male_____ Female_____

2. Age: _____

3. Education level:

Ph.D. or equivalent_____ Masters Degree (in progress)_____

Bachelors Degree (in progress)_____ Other (specify)_____

4. Number of years of work experience (if any)_____

5. Number of years of manufacturing experience (if any)_____

6. Number of individuals you have ever supervised (if any)_____

7. Total approximate annual compensation (if any)_____

THANK YOU VERY MUCH FOR YOUR PARTICIPATION.

International Review of Accounting, Banking and Finance
Vol 17, No. 2, Summer, 2025, Pages 39-52

IRABF

The Impact of Female Directors and ESG on Bank Risk: Evidence from the Asian Banks

Tsui-Jung Lin, Lien-Wen Liang*, Pei-Chun Lin

1. *Chinese Culture University*
2. *Chinese Culture University*
3. *Chinese Culture University*

Accepted June 2025

ABSTRACT

This study aims to investigate the impact of female directors and environmental, social, and governance (ESG) practices on bank risk. The empirical quantitative study covers a sample of 64 Asia banks for the period of 2016 to 2022 and analyzes using the System Generalized Method of Moments (GMM), we find that the presence of female directors significantly lowers risk. Their diverse perspectives foster richer board deliberations and less-overconfident decision-making, leading to more prudent risk assessments and greater resilience. By contrast, ESG practices alone do not appear to reduce risk. An interaction analysis shows, however, that when female directors and robust ESG engagement coexist, bank risk declines in emerging economies-an effect not observed in developed markets. These findings suggest that expanding female representation on bank boards can curb risk exposure and that empowering these directors to drive ESG initiatives further strengthens risk management, especially in less-developed countries. Regulators might therefore consider incentives that encourage banks to increase female board participation and deepen directors' sustainability expertise.

Keywords: ESG, Women Directors, Bank Risk, GMM
JEL classification: G20, G28

* Corresponding author: llw@ulive.pccu.edu.tw

1. Introduction

Sustainable development, along with the environmental (E), social (S), and governance (G) dimensions of ESG, plays a crucial role in mitigating the damage to human life and property caused by increasingly frequent global extreme weather events. This is especially true for the financial industry, where compliance with growing ESG-related regulations and the integration of ESG principles into investment and lending strategies have become essential for long-term profitability and sustainability. Within the ESG framework, gender equality is not only a core component of international human rights law but also a critical management issue. While gender equality has been extensively discussed, key challenges remain, such as overcoming traditional cultural constraints on gender roles, narrowing the gender gap across various sectors, and encouraging women to actively participate in society, assume leadership positions, and fully realize their potential and influence. Appointing women to corporate boards, in particular, represents a concrete expression of the values of gender equality and social equity.

In recent years, numerous studies have analyzed the impact of appointing women as outside directors in the financial sector. Female directors introduce diverse perspectives that strengthen board governance and mitigate agency conflicts (Chatjuthamard, Jiraporn, and Lee, 2021), while their presence also signals lower risk exposure during financial crises (Mohsni, Otchere, and Shahriar, 2021). Banks led by women tend to exhibit higher capital adequacy and equity-to-asset ratios, alongside lower relative risk (Menicucci and Paolucci, 2022a). They likewise benefit from stronger oversight and more rigorous risk management (Gulamhussen and Santa, 2015; Birindelli, Chiappini, and Savioli, 2020). Emphasizing gender diversity, therefore, not only reduces operational risk but also enhances a bank's image and credibility with investors and other stakeholders (Olsen, Awuah-Offei, and Bumblauskas, 2021).

Rising awareness of gender equality has further heightened interest in women's roles within corporate leadership. Empirical evidence shows that female leaders contribute broader strategic perspectives (Glass and Cook, 2018), help curb corporate risk (Nadeem, Suleman, and Ahmed, 2019), and typically devote greater attention and resources to ESG initiatives (Zhang, 2023).

This paper focuses on Asian banks because corporate-governance reforms were introduced in Asia later than in Europe and the United States, and their effectiveness remains underexplored. Although ESG reporting has become widespread in advanced Western economies, three of the world's four largest economies- Japan, China, and India are in Asia (Abdul Rahman and Alsayegh, 2021). Consequently, ESG implementation and disclosure have become increasingly important across the region (Alsayegh, Abdul Rahman, and Homayoun, 2020). Against this backdrop, the study investigates how the interaction between female board participation and ESG practices influences risk in the Asian banking industry.

Overall, this study makes three contributions. First, it advances the literature on the role of female directors in risk oversight within banks in emerging markets. Corporate success fundamentally hinges on the board of directors, which is charged with monitoring the execution of strategic objectives, upholding sound governance structures, and shaping organizational culture. Effective directors also prioritise business ethics and corporate responsibility (Birindelli, Dell'Atti, Iannuzzi, and Savioli, 2018). Prior research indicates that female directors enhance board governance through more rigorous monitoring (Atif, Hossain, Alam, and Goergen, 2021; Gull, Saeed, Suleman, and Mushtaq, 2022) and greater independence (Adams and Ferreira, 2009). A higher proportion of women on the board increases organisational sensitivity to social and environmental considerations, thereby fostering sustainable development (Hafsi and Turgut, 2013; Yasser, Al Mamun, and Ahmed, 2017; Veltri, Mazzotta, and Rubino, 2021). Consistent with these findings, the study shows that the presence of female directors is associated with reduced bank risk. By contributing diverse perspectives and experiences, they facilitate more comprehensive deliberations and prudent decision-making, which temper managerial overconfidence and enhance institutional resilience. Second, this study deepens the understanding of how ESG implementation influences risk in the banking sector. As the linchpin of the global financial system, banks' risk-management practices are critical to economic stability. By directing capital flows, financial institutions also possess the leverage to propel a worldwide

transition toward sustainability and a low-carbon economy. Prior research attributes the banking sector's vulnerability during the 2008 crisis to the excessive accumulation of risk (Brunnermeier, 2009; DeYoung, Peng, and Yan, 2013). In the post-crisis landscape, leading banks have increasingly acknowledged that ESG-related exposures embedded in their financing portfolios can translate into substantial long-run financial losses. Simultaneously, the emergence of green finance and related instruments has enabled banks to incorporate sustainability considerations into their core operations, thereby promoting sustainable development and mitigating risk. Empirically, this study finds that the environmental pillar (E) is negatively associated with bank risk, whereas the social (S) and governance (G) pillars are positively associated. These divergent effects suggest that ESG dimensions do not uniformly influence risk; rather, each pillar exerts a distinct risk-modulating impact. Finally, this paper contributes to the literature by examining the interaction between female board representation and ESG practices and their influence on bank risk. Gender diversity has been associated with enhanced corporate innovation, improved performance, and reduced risk exposure. However, the interaction between female directors and ESG does not exhibit a significant risk-reducing effect in banks operating in developed countries, whereas a statistically significant association is observed in banks from developing countries. This suggests the presence of entrenched gender biases and cultural stereotypes in certain Asian countries, which may constrain the ability of female directors to fully exercise their professional and managerial influence. As a result, the effectiveness of ESG-related policy implementation in mitigating bank risk appears limited in such contexts.

The rest of the paper is organized as follows: Section 2 reviews the relevant literature and outlines the research hypotheses. Section 3 describes the data, variables, and research methodology. Section 4 presents the empirical findings, including the main regression results. Section 5 concludes the study and offers policy implications.

2. Literature Review and Hypotheses Development

2.1 The relationship between female directors and bank risks

Women's participation in corporate leadership has received increasing scholarly and practitioner attention. Empirical evidence shows that, relative to their male counterparts, female directors display heightened sensitivity to environmental issues (Nadeem et al., 2020), adhere to more stringent ethical standards (Moreno-Ureba, Bravo-Urquiza, and Reguera-Alvarado, 2022), and actively advance environmental strategies and investments (Atif et al., 2021; Gull et al., 2022; Issa and Bensalem, 2023). Because board composition is a cornerstone of the governance mechanism, greater female representation enriches decision-making, shapes corporate strategy and performance (Bennouri et al., 2018; Nielsen and Huse, 2010a), and mitigates intra-board conflict through superior interpersonal skills (Nielsen and Huse, 2010b). It also enhances discussion quality and reporting effectiveness (Gul, Srinidhi, and Ng, 2011), is associated with higher meeting attendance and stronger oversight (Adams and Ferreira, 2009), and improves the company's reputation (Bear, Rahman, and Post, 2010). Collectively, these outcomes signal a commitment to diversity and legitimacy, thereby strengthening support from key stakeholders, particularly customers, suppliers, and investors (Hillman, Shropshire, and Cannella, 2007).

From an agency theory perspective, female directors are regarded as effective monitors of management and contribute a wide range of ideas, perspectives, skills, and experiences to the company's board of directors (McGuinness, Vieito, and Wang, 2017). Their presence also serves as a balancing influence within male-dominated boards. Faccio, Marchica, and Mura (2016) demonstrate that women generally exhibit greater risk aversion than men, rendering them less susceptible to financial distress (Mittal and Lavina, 2018) and contributing to lower levels of risk in the banking sector (Menicucci and Paolucci, 2022a). Furthermore, Belaounia, Tao, and Zhao (2020) suggest that female directors enhance corporate value in at least three ways: by improving overall firm performance, mitigating earnings management, and curbing excessive risk-taking behavior. Based on the aforementioned literature, the following hypothesis is proposed:

H1: There is a negative relationship between female directors and bank risks.

2.2 The relationship between ESG and bank risks

The banking sector is inherently exposed to high levels of moral hazard and therefore demands more stringent regulatory oversight (Wu and Shen, 2013). The 2008 global financial crisis, in particular, highlighted the dangers of institutional complacency among banks and regulators, severely undermining public confidence in financial institutions (Larosière et al., 2009). In response, increased attention has been directed toward corporate social responsibility investment and business philosophy and the integration of environmental, social, and governance (ESG) principles, which are believed to enhance operational resilience and restore stakeholder trust. Empirical evidence suggests that firms engaging in ESG strategies tend to outperform those that neglect their social responsibilities (Flammer and Kacperczyk, 2019).

However, the benefits of ESG implementation remain subject to debate. Critics argue that CSR activities fall outside the core objectives of firms, which should primarily focus on maximizing shareholder value (Friedman, 1970). Some studies further suggest that while ESG engagement positively affects cash flow and operational efficiency, it may simultaneously exert a negative influence on the cost of equity (Azmi, Hassan, Houston, and Karim, 2021). Among ESG components, environmental initiatives are found to have the most pronounced impact on bank valuation. Nevertheless, ESG practices not only help mitigate environmental and social risks but also contribute to improved corporate governance and long-term firm value. Based on the foregoing discussion, this study posits the following hypothesis:

H2: There is a negative relationship between ESG performance and bank risk.

With respect to the influence of female directors, prior research suggests that a higher proportion of women on corporate boards is associated with increased attention to sustainability and social responsibility, thereby enhancing firms' ESG performance (Menicucci and Paolucci, 2022b). Moreover, firms with greater female board representation tend to disclose more environmental information, as female directors are generally more attuned to environmental concerns. The presence of women on boards also contributes to a broader range of perspectives and decision-making styles, which can enhance the overall quality of governance and strategic innovation, further reinforcing ESG outcomes. In summary, the inclusion of female directors may mitigate corporate risk through diversified viewpoints and heightened social accountability. Based on this rationale, the following research hypothesis is proposed:

H3: Female directors and ESG performance are negatively associated with bank risk.

3. Research methods

3.1. Source and sample screening

This chapter outlines the research methodology employed in this study and is organized into three sections. The first section describes the data sources and the selection criteria for the research sample. The second section defines the variables used in the analysis, including both key independent and dependent variables, as well as control variables. The third section presents the regression model specification, detailing the empirical strategy and model formulation adopted for hypothesis testing.

The study investigates the impact of female board representation and environmental, social, and governance (ESG) practices on bank risk. The data are obtained from Datastream, BankFocus, and the World Bank database. Given the improved availability and consistency of ESG-related data in recent years, the study period spans from 2016 to 2022, covering a total of seven years.

In accordance with the International Monetary Fund (IMF) classification, Asian banks in the sample are divided into those from developed and developing countries. The developed economies include China, Hong Kong, Japan, South Korea, and Taiwan, while the developing economies comprise India, Indonesia, Malaysia, Pakistan, the Philippines, and Thailand. After excluding observations with missing or incomplete data, the final sample consists of 64 banks.

3.2 Research methods and models

To address potential endogeneity concerns, this study employs the System Generalized Method of Moments (System GMM) estimator, which effectively mitigates endogeneity bias and accounts for unobserved heterogeneity by incorporating lagged dependent variables and controlling for fixed effects. In the bank risk regression model, the Z-score is utilized as a proxy for overall bank risk. A higher Z-score indicates greater financial stability and a lower probability of insolvency (Laeven and Levine, 2009).

The Z-score is widely adopted in the banking literature as a standard measure of risk exposure (e.g., Delis, Hasan, and Tsionas, 2014; Houston, Lin, Lin, and Ma, 2010) and is computed using the following equation:

$$Zscore = \frac{Return\ on\ Asset + \left(\frac{Equity}{Asset}\right)}{Standard\ Deviation\ of\ Return\ on\ Asset}$$

Where Return on Assets (ROA) represents the bank's profitability relative to its total assets, Equity-to-Asset Ratio denotes the proportion of equity capital to total assets, and Standard Deviation of ROA captures the volatility of the bank's return on assets, serving as a measure of income variability.

To empirically examine the effects of ESG performance and the interaction between female board representation and ESG on bank risk, this study specifies the following regression model:

$$\begin{aligned} Zscore_{it} = {} & \alpha_0 + \alpha_1 WOB_{it-1} + \alpha_2 ESG_{it-1} + \alpha_3 (WOB_{it-1} \times ESG_{it-1}) + \alpha_4 CAR_{it} + \alpha_5 NPL_{it} \\ & + \alpha_6 CURR_{it} + \alpha_7 DBR_{it} + \alpha_8 SIZE_{it} + \alpha_9 GDP_t + \alpha_{10} IRS_t + \alpha_{11} INF_t \\ & + \alpha_{12} Covid19_t + \eta_t + \varepsilon_{it} \end{aligned}$$

where WOB_{it-1}denotes the laggard of female directors, ESG_{it-1}represents the laggard of total ESG score. The interaction term $(WOB_{it-1} \times ESG_{it-1})$ captures the moderating effect of female board representation on ESG performance in relation to bank risk. The model also includes a set of control variables. Specifically, CAR_{it} represents the capital adequacy ratio, NPL_{it} represents the delinquency ratio, $CURR_{it}$ represents the liquidity ratio, DBR_{it} is debt ratio, $SIZE_{it}$ is bank size, GDP_t represents the economic growth rate, IRS_t is deposit spread, INF_t represents for inflation rate and $Covid19_t$ is for the dummy variable of epidemic. η_t is the time effect, and ε_{it} is the residual.

WOB_{it-1} is calculated as the number of females as a proportion of the total number of board members. The ESG score is based on definitions provided by the Datastream database, where it represents a composite score calculated as a weighted average of environmental, social, and governance indicators disclosed by each bank, along with corresponding category-specific scores (Shakil, Mahmood, Tasnia, and Munim, 2019). CAR_{it} is defined as the ratio of a bank's capital to its risk-weighted assets. NPL_{it} is measured as the total non-performing loan divided by total loan. $CURR_{it}$ is defined as the liquid assets divided by deposits and short-term funding. DBR_{it} is calculated as total liabilities divided by total assets. $SIZE_{it}$ is measured by taking the natural log of total assets. The macroeconomic control variables include Gross Domestic Product (GDP), Inflation (INF), Interest Rate Spread (IRS), and a dummy variable for the COVID-19 pandemic (COVID-19), representing systemic shocks during the sample period.

4. Empirical Results

This chapter presents the empirical analysis based on the data and research methodology outlined in the preceding sections. Section 1 provides the descriptive statistics of the variables, while Section 2 reports and interprets the empirical findings derived from the regression model.

4.1 Descriptive Statistics

As shown in Table 1, the average proportion of women on boards of directors (WOB) is 15.05%,

indicating that, on average, less than 16% of board members are female. The WOB variable ranges from a minimum of 0 to a maximum of 69.23%, suggesting that the appointment of female directors is not uniformly mandated across Asian countries. The average ESG score is 52.27, reflecting a moderate level of environmental, social, and governance performance among the sampled banks.

The average capital adequacy ratio (CAR) is 15.91%, which exceeds the Basel III regulatory minimum of 10.5%, implying a relatively strong capital position and lower risk exposure. The average non-performing loan (NPL) ratio is 2.67%, significantly below the commonly accepted threshold, indicating good loan portfolio quality. The average debt-to-asset ratio (DBR) is 0.1005, suggesting low reliance on debt financing and, consequently, lower financial risk. The average bank size (SIZE), measured as the natural logarithm of total assets, is 11.23.

With regard to macroeconomic indicators, the average GDP growth rate is 3.06%, the average interest rate spread (IRS) is 2.13%, and the average inflation rate (INF) is 2.38%. The variable Covid19 is a dummy variable representing the presence of the COVID-19 pandemic, where a value of 0 is assigned to observations prior to 2019 and a value of 1 to those from 2019 onward.

To mitigate the potential bias caused by multicollinearity, this study examines the correlation coefficients among the independent variables. As presented in Table 2, the correlation coefficients range from -0.26 to 0.63, suggesting that no serious multicollinearity issues are present within the model.

Table 1. Descriptive Statistics

Variables	Mean	Median	Std.	Min.	Max.	Obs.
WOB	15.0471	13.8095	11.3439	0	69.2308	448
ESG	52.2696	54.1800	19.4142	9.5400	89.4300	448
CAR	15.9083	14.9850	5.7825	8.4300	56.2100	448
NPL	2.6747	1.4996	3.0344	0.1408	18.4581	448
CURR	26.5716	21.2417	18.1761	4.5014	122.5728	448
DBR	0.1005	0.0741	0.0872	0.0041	0.5091	448
SIZE	11.2269	11.2526	1.2697	8.8306	15.5023	448
GDP	3.0555	3.2286	3.8277	-9.5183	9.0503	448
IRS	2.1312	1.9917	1.1434	0.6733	5.3100	448
INF	2.3843	1.9680	2.6555	-1.1387	19.8739	448
Covid19	0.5714	1	0.4954	0	1	448

Note: WOB refers to the variable for female directors; ESG represents the total score comprising Environmental (E), Social (S), and Governance (G) dimensions; CAR denotes the capital adequacy ratio; NPL is the non-performing loan ratio; CURR represents the liquidity ratio; DBR is the debt-to-asset ratio; SIZE indicates bank size; GDP refers to the economic growth rate; IRS denotes the deposit interest rate spread; INF represents the inflation rate; and Covid19 is a dummy variable indicating the presence of the COVID-19 pandemic.

Table 2. Pearson Correlation

	WOB	ESG	CAR	NPL	CURR	DBR	SIZE	GDP	IRS	INF	Covid19
WOB	1										
ESG	0.3752	1									
CAR	0.0546	0.1723	1								
NPL	-0.0378	-0.1373	0.0084	1							
CURR	-0.0345	0.1019	0.6250	0.0161	1						
DBR	-0.2000	-0.1234	-0.0167	0.0219	0.0727	1					
SIZE	-0.0120	0.1019	-0.2308	-0.1072	-0.1127	0.0397	1				
GDP	0.0722	0.1137	-0.0669	0.1082	-0.1555	-0.0314	0.0520	1			
IRS	0.1561	0.2183	0.0669	0.3607	0.0042	-0.1465	-0.1952	0.2474	1		
INF	0.0102	0.0659	0.0748	0.4772	0.0285	0.0677	-0.1517	0.2809	0.4184	1	
Covid19	0.1153	0.2056	0.0766	-0.0091	0.0580	-0.0488	0.0763	-0.2638	0.0051	0.1459	1

Note: WOB refers to the variable for female directors; ESG represents the total score comprising Environmental (E), Social (S), and Governance (G) dimensions; CAR denotes the capital adequacy ratio; NPL is the non-performing loan ratio; CURR represents the liquidity ratio; DBR is the debt-to-asset ratio; SIZE indicates bank size; GDP refers to the economic growth rate; IRS denotes the deposit interest rate spread; INF represents the inflation rate; and Covid19 is a dummy variable indicating the presence of the COVID-19 pandemic.

4.2 Regression analysis results of bank risk

This study employs the System Generalized Method of Moments (System GMM) to examine the impact of female board representation and ESG performance on bank risk, incorporating an interaction term to explore the joint effect of the two factors. The empirical results are presented below.

Column I of Table 3 reports the baseline System-GMM estimates for the effect of female board representation (WOB) on bank risk. The coefficient on WOB is 0.0051 and is statistically significant at the 10 percent level, thereby lending empirical support to Hypothesis 1, which posits that greater female participation on the board is associated with lower bank risk. In practical terms, a higher proportion of female directors corresponds to enhanced stability and reduced risk-taking at the bank level. This finding aligns with Arayssi, Dah, and Jizi (2016), who demonstrate that female board representation mitigates firms' equity risk. Moreover, it supports the conclusions of Ingersoll, Cook, and Glass (2023), who find that female directors tend to exhibit greater prudence and reduced overconfidence in risk assessment decisions.

The empirical results regarding the relationship between ESG performance and bank risk are presented in Column II of Table 3. The estimated coefficient for ESG is negative and statistically significant at the 5% level (coefficient = -0.0040, $p < 0.05$). However, contrary to the expectations of this study, the result does not support the proposed hypothesis that ESG performance is negatively associated with bank risk. In other words, increased ESG engagement does not appear to significantly reduce bank risk within the sample. One possible explanation for this finding is that the implementation of ESG initiatives entails substantial costs. While the integration of Environmental, Social, and Governance (ESG) principles, particularly through CSR investments and sustainability-driven business strategies, has gained considerable attention in recent years and may enhance investor confidence, the cost-effectiveness of such efforts remains uncertain and warrants further investigation.

4.3 The Impact of the Interaction between Female Directors and ESG on Bank Risks

The empirical results regarding the interaction between female board representation (WOB) and ESG performance on bank risk are presented in Column III of Table 3. The interaction term is statistically insignificant, indicating that the joint effect of WOB and ESG does not exhibit a meaningful association with bank risk. Consequently, this finding fails to support Hypothesis 3, which posits a negative relationship between the interaction of female directors and ESG engagement on bank risk. A plausible explanation for this result is that the effectiveness of female directors and ESG practices may be contingent upon the broader institutional and cultural context. In environments where structural or cultural resistance to gender diversity and sustainability practices exists, the implementation of ESG strategies and the influence of female board members may be constrained. Such resistance may result in misalignment, internal conflict, or inefficiencies in governance and decision-making, potentially exacerbating operational risks within banks (Peng and Chandarasupsang, 2023).

In light of this, the study proceeds to conduct a sub-sample analysis by separating the data into developed and emerging market countries to further investigate contextual differences.

The sample includes 34 banks from developed countries and 30 banks from developing countries. The empirical results for developed countries are presented in Table 4. Column I shows that the coefficient for female board representation is positive and statistically significant at the 10% level (coefficient = 0.0063, $p < 0.1$), indicating that the presence of female directors is associated with lower bank risk. This finding supports the argument that female directors contribute to more effective board governance and reduced agency conflicts by introducing diverse perspectives (Chatjuthamard et al., 2021) and is consistent with the main empirical findings of this study.

Table 3. Regression Analysis of Female Directors and ESG on Bank Risk (Z-score)

Zscore	I	II	III
WOB	0.0051*		0.0115
	(1.84)		(0.99)
ESG		-0.0040**	-0.0049
		(-1.98)	(-1.50)
WOB×ESG			-0.0001
			(-0.30)
CAR	-0.0218***	-0.0187**	-0.0197**
	(-2.82)	(-2.32)	(-2.35)
NPL	-0.1420***	-0.1500***	-0.1510***
	(-11.99)	(-12.89)	(-12.93)
CURR	0.0001	0.0001	0.0005
	(0.06)	(0.03)	(0.20)
DBR	-2.0780***	-2.2440***	-2.1070***
	(-4.23)	(-4.54)	(-4.35)
SIZE	-0.0220	-0.0122	-0.0105
	(-0.77)	(-0.41)	(-0.34)
GDP	-0.0070	-0.0032	-0.0045
	(-0.76)	(-0.34)	(-0.48)
IRS	-0.1830***	-0.1560***	-0.1620***
	(-6.09)	(-4.70)	(-4.89)
INF	-0.0148	-0.0156	-0.0133
	(-0.90)	(-0.94)	(-0.81)
Covid19	-0.0424	0.0051	-0.0081
	(-0.61)	(0.07)	(-0.12)
_cons	5.7440***	5.8180***	5.7280***
	(14.90)	(15.24)	(15.27)
N	448	448	448
Wald chi^2	337.23	366.28	402.11
R-squared	0.4482	0.4497	0.4577
Root MSE	0.6748	0.6738	0.6689

Note: 1. *,**,***represent significance at the 10%,5%, and 1% levels (2-tailed), respectively.
2.WOB refers to the variable for female directors; ESG denotes the total ESG score; WOB × ESG represents the interaction term between female directors and the ESG score; CAR is the capital adequacy ratio; NPL is the non-performing loan ratio; CURR is the liquidity ratio; DBR is the debt-to-asset ratio; SIZE indicates bank size; GDP refers to the economic growth rate; IRS is the deposit interest rate spread; INF denotes the inflation rate; and Covid19 is a dummy variable indicating the presence of the COVID-19 pandemic.
3.The study uses the Z-score as a proxy for bank risk. A higher Z-score indicates greater bank stability and lower risk, while a lower Z-score reflects weaker stability and higher risk exposure.

As shown in Column II of Table 4, the coefficient for ESG is negative but not statistically significant, suggesting that ESG practices alone do not have a meaningful impact on bank risk in developed countries. However, Column III reports a statistically significant negative coefficient for the interaction term between female directors and ESG (coefficient = -0.0004, $p < 0.05$), implying that the joint effect of gender diversity and ESG engagement contributes to risk reduction.

This result suggests that, while ESG implementation or female board presence alone may have a limited effect, their interaction is more effective in mitigating bank risk. Nonetheless, the effectiveness of such governance mechanisms may vary across institutional and cultural contexts. In settings where resistance to ESG practices or gender diversity persists, implementation challenges

may arise, leading to internal conflicts, inefficiencies, and, ultimately, elevated operational risk (Peng and Chandarasupsang, 2023).

Table 4. The Impact of Female Directors and ESG on Bank Risk (Z-score) in Developed Countries

Zscore	I	II	III
WOB	0.0063*		0.0315**
	(1.76)		(2.47)
ESG		-0.0003	0.0038
		(-0.13)	(1.20)
WOB×ESG			-0.0004**
			(-2.16)
CAR	0.0171**	0.0169**	0.0142*
	(2.31)	(2.29)	(1.88)
NPL	0.0198	-0.0094	-0.00167
	(0.20)	(-0.11)	(-0.02)
CURR	-0.0202***	-0.0201***	-0.0199***
	(-5.85)	(-5.52)	(-5.37)
DBR	0.1900	0.0269	0.0931
	(0.58)	(0.07)	(0.26)
SIZE	-0.0100	-0.0100	-0.0176
	(-0.38)	(-0.34)	(-0.59)
GDP	0.0311*	0.0310*	0.0258
	(1.92)	(1.88)	(1.52)
IRS	-0.1250***	-0.1110***	-0.1190***
	(-3.14)	(-2.86)	(-3.00)
INF	-0.0164	-0.0054	-0.0260
	(-0.37)	(-0.12)	(-0.58)
Covid19	-0.0430	-0.0217	-0.0542
	(-0.54)	(-0.27)	(-0.68)
_cons	5.0240***	5.1270***	4.9720***
	(11.56)	(12.73)	(12.44)
N	238	238	238
Wald chi^2	339.87	319.86	352.98
R-squared	0.2811	0.2742	0.2975
Root MSE	0.6056	0.6085	0.5987

Note: 1. *,**,***represent significance at the 10%,5%, and 1% levels (2-tailed), respectively.
2. WOB refers to the variable for female directors; ESG denotes the total ESG score; WOB × ESG represents the interaction term between female directors and the ESG score; CAR is the capital adequacy ratio; NPL is the non-performing loan ratio; CURR is the liquidity ratio; DBR is the debt-to-asset ratio; SIZE indicates bank size; GDP refers to the economic growth rate; IRS is the deposit interest rate spread; INF denotes the inflation rate; and Covid19 is a dummy variable indicating the presence of the COVID-19 pandemic.
3. The study uses the Z-score as a proxy for bank risk. A higher Z-score indicates greater bank stability and lower risk, while a lower Z-score reflects weaker stability and higher risk exposure.

Table 5. The Impact of Female Directors and ESG on Bank Risk (Z-score) in Developing Countries

Zscore	(1)	(2)	(3)
WOB	0.0186***		-0.0378*
	(4.18)		(-1.67)
ESG		0.0045	-0.0183**
		(1.26)	(-2.07)
WOB×ESG			0.0009**
			(2.35)
CAR	0.0566***	0.0444**	0.0659***
	(4.16)	(3.04)	(4.41)
NPL	-0.1350***	-0.1360***	-0.1290***
	(-14.06)	(-13.17)	(-12.29)
CURR	0.0130***	0.0133***	0.0091***
	(4.41)	(4.33)	(2.92)
DBR	-4.0900***	-4.2900***	-4.1200***
	(-7.46)	(-6.95)	(-7.39)
SIZE	-0.0589	-0.0625	-0.0570
	(-1.25)	(-1.30)	(-1.26)
GDP	-0.0036	-0.0032	-0.0040
	(-0.39)	(-0.34)	(-0.42)
IRS	-0.1400**	-0.2160***	-0.1060*
	(-2.51)	(-4.05)	(-1.74)
INF	0.0026	-0.0084	0.0013
	(0.19)	(-0.56)	(0.10)
Covid19	-0.2130***	-0.1710	-0.2090**
	(-2.37)	(-1.80)	(-2.27)
_cons	4.2850***	4.8480***	5.1230***
	(6.42)	(7.02)	(7.96)
N	210	210	210
Wald chi^2	1132.15	981.33	1172.38
R-squared	0.6593	0.6355	0.6705
Root MSE	0.5341	0.5525	0.5253

Note: 1. *,**,***represent significance at the 10%,5%, and 1% levels (2-tailed), respectively.
2. WOB refers to the variable for female directors; ESG denotes the total ESG score; WOB × ESG represents the interaction term between female directors and the ESG score; CAR is the capital adequacy ratio; NPL is the non-performing loan ratio; CURR is the liquidity ratio; DBR is the debt-to-asset ratio; SIZE indicates bank size; GDP refers to the economic growth rate; IRS is the deposit interest rate spread; INF denotes the inflation rate; and Covid19 is a dummy variable indicating the presence of the COVID-19 pandemic.
3. The study uses the Z-score as a proxy for bank risk. A higher Z-score indicates greater bank stability and lower risk, while a lower Z-score reflects weaker stability and higher risk exposure.

With respect to the impact of female directors and ESG on bank risk in developing countries, the empirical results are presented in Table 5. Column I indicates a positive and statistically significant relationship between female board representation and bank risk (coefficient = 0.0186, $p < 0.01$). This suggests that the presence of female directors contributes to reduced risk, likely due to the introduction of diverse perspectives, enhanced board effectiveness, and diminished agency conflicts (Chatjuthamard et al., 2021). This finding is consistent with the main empirical conclusions of the study.

Column II shows that the coefficient for ESG is positive but not statistically significant, although the direction of the effect aligns with the theoretical expectations of this paper. Notably, the interaction term between female directors and ESG, reported in Column III, is positive and

statistically significant (coefficient = 0.0009, $p < 0.05$). This result implies that in developing countries, female board members play a moderating role in strengthening the relationship between ESG engagement and reduced bank risk. The significant interaction supports Hypothesis 3 of this study, which posits that the combined effect of female directorship and ESG practices contributes to risk mitigation in the banking sector.

5. Conclusion

This study investigates the extent to which female board representation, ESG performance, and their interaction influence bank risk in Asian countries. Prior literature has suggested that women contribute diverse perspectives and insights to corporate boards and that female leadership is often more attentive to ESG-related concerns, allocating greater resources to sustainability initiatives. However, the extent to which female directors directly contribute to risk reduction remains an open empirical question.

Drawing on the existing literature, this study formulates three hypotheses: (1) female board representation is negatively associated with bank risk; (2) ESG performance is negatively associated with bank risk; and (3) the interaction between female board representation and ESG performance is negatively associated with bank risk.

The empirical findings reveal that, in the context of Asian banks, female directors are associated with lower levels of risk and enhanced bank stability. This supports the notion that women tend to exhibit greater prudence and lower overconfidence in risk assessment (Ingersoll et al., 2023). However, the independent effects of ESG performance and its interaction with female board representation are not consistently significant across all country contexts. Specifically, in developed Asian countries, the interaction between female directors and ESG performance does not significantly reduce bank risk. In contrast, in developing Asian countries, the interaction term is both negative and statistically significant, suggesting that the joint influence of female directors and ESG initiatives is more effective in risk mitigation.

These findings underscore the importance of institutional and developmental contexts in shaping the effectiveness of gender diversity and ESG strategies. From a managerial perspective, the results suggest that banks should consider increasing female representation on their boards and actively promote ESG-related policies to enhance stability. Furthermore, banks are encouraged to support female directors in pursuing ESG-related training to strengthen board-level expertise in sustainability. Regulatory authorities may also consider introducing incentive mechanisms to promote sustainable governance and professional development among bank directors.

References

Abdul Rahman, R. and M. F. Alsayegh. 2021. Determinants of corporate environment, social and governance (ESG) reporting among Asian firms. ***Journal of Risk and Financial Management***, 14(4): 167.

Adams, R. B. and D. Ferreira. 2009. Women in the boardroom and their impact on governance and performance. ***Journal of Financial Economics***, 94(2): 291-309.

Alsayegh, M. F., R. Abdul Rahman, and S. Homayoun. 2020. Corporate economic, environmental, and social sustainability performance transformation through ESG disclosure. ***Sustainability***, 12(9): 3910.

Arayssi, M., M. Dah, and M. Jizi. 2016. Women on boards, sustainability reporting and firm performance. ***Sustainability Accounting, Management and Policy Journal***, 7(3): 376-401.

Atif, M., M. Hossain, M. S. Alam, and M. Goergen. 2021. Does board gender diversity affect renewable energy consumption? ***Journal of Corporate Finance***, 66: 101665.

Azmi, W., M. K. Hassan, R. Houston, and M. S. Karim. 2021. ESG activities and banking performance: International evidence from emerging economies. ***Journal of International Financial Markets, Institutions and Money***, 70: 101277.

Bear, S., N. Rahman, and C. Post. 2010. The impact of board diversity and gender composition on

corporate social responsibility and firm reputation. ***Journal of Business Ethics***, 97: 207-221.

Belaounia, S., R. Tao, and H. Zhao. 2020. Gender equality's impact on female directors' efficacy: A multi-country study. ***International Business Review***, 29(5): 101737.

Bennouri, M., T. Chtioui, H. Nagati, and M. Nekhili. 2018. Female board directorship and firm performance: What really matters? ***Journal of Banking & Finance***, 88: 267-291.

Birindelli, G., H. Chiappini, and M. Savioli. 2020. When do women on board of directors reduce bank risk? ***Corporate Governance: The International Journal of Business in Society***, 20(7): 1307-1327.

Birindelli, G., S. Dell'Atti, A. P. Iannuzzi, and M, Savioli. 2018. Composition and activity of the board of directors: Impact on ESG performance in the banking system. ***Sustainability***, 10(12): 4699.

Brunnermeier, M. K. 2009. Deciphering the liquidity and credit crunch 2007-2008. ***Journal of Economic Perspectives***, 23(1): 77-100.

Chatjuthamard, P., P. Jiraporn, and S. M. Lee. 2021. Does board gender diversity weaken or strengthen executive risk-taking incentives? ***PloS one***, 16(10): e0258163.

Delis, . D., I. Hasan, and E. G. Tsionas. 2014. The risk of financial intermediaries. ***Journal of Banking & Finance***, 44: 1-12.

DeYoung, R., E. Y. Peng, and M. Yan. 2013. Executive compensation and business policy choices at US commercial banks. ***Journal of Financial and Quantitative Analysis***, 48(1): 165-196.

Faccio, M., M. T. Marchica, and R. Mura. 2016. CEO gender, corporate risk-taking, and the efficiency of capital allocation. ***Journal of Corporate Finance***, 39: 193-209.

Flammer, C. and A. Kacperczyk. 2019. Corporate social responsibility as a defense against knowledge spillovers: Evidence from the inevitable disclosure doctrine. ***Strategic Management Journal***, 40(8): 1243-1267.

Friedman, M. (1970) The Social Responsibility of Business Is to Increase Its Profits. *New York Times Magazine, 13 September 1970*, 122-126.

Glass, C. and A. Cook. 2018. Do women leaders promote positive change? Analyzing the effect of gender on business practices and diversity initiatives. ***Human Resource Management***, 57(4): 823-837.

Gul, F. A., B. Srinidhi, and A. C. Ng. 2011. Does board gender diversity improve the informativeness of stock prices? ***Journal of Accounting and Economics***, 51(3): 314-338.

Gulamhussen, M. A. and S. F. Santa. 2015. Female directors in bank boardrooms and their influence on performance and risk-taking. ***Global Finance Journal***, 28: 10-23.

Gull, A. A., A. Saeed, M. T. Suleman, and R. Mushtaq. 2022. Revisiting the association between environmental performance and financial performance: Does the level of environmental orientation matter? ***Corporate Social Responsibility and Environmental Management***, 29(5): 1647-1662.

Hafsi, T. and G. Turgut, 2013. Boardroom diversity and its effect on social performance: Conceptualization and empirical evidence. ***Journal of Business Ethics***, 112: 463-479.

Hillman, A. J., C. Shropshire, and A. A. Cannella Jr. 2007. Organizational predictors of women on corporate boards. ***Academy of Management Journal***, 50(4): 941-952.

Houston, J. F., C. Lin, P. Lin, and Y. Ma. 2010. Creditor rights, information sharing, and bank risk taking. ***Journal of Financial Economics***, 96(3): 485-512.

Ingersoll, A. R., A. Cook, and C. Glass. 2023. A free solo in heels: Corporate risk taking among women executives and directors. ***Journal of Business Research***, 157: 113651.

Issa, A. and N. Bensalem. 2023. Are gender-diverse boards eco-innovative? The mediating role of corporate social responsibility strategy. ***Corporate Social Responsibility and Environmental Management***, 30(2): 742-754.

Laeven, L. and R. Levine. 2009. Bank governance, regulation and risk taking. ***Journal of Financial Economics***, 93(2): 259-275.

De Larosière, J., Balcerowicz, L., Issing, O., Masera, R., Mc Carthy, C., Nyberg, L., and Perez Onno Ruding, J. (2009). *The High-Level Group on financial supervision in the EU, Brussels.*

European Commission.
McGuinness, P. B., J. P. Vieito, and M. Wang. 2017. The role of board gender and foreign ownership in the CSR performance of Chinese listed firms. ***Journal of Corporate Finance***, 42: 75-99.
Menicucci, E. and G. Paolucci. 2022a. Gender diversity and bank risk-taking: an empirical investigation in Italy. ***Corporate Governance: The International Journal of Business in Society***, 22(2): 317-339.
Menicucci, E. and G. Paolucci. 2022b. Board diversity and ESG performance: evidence from the Italian banking sector. ***Sustainability***, 14(20): 13447.
Mittal, S., and Lavina. 2018. Females' representation in the boardroom and their impact on financial distress: An evidence from family businesses in India. ***Indian Journal of Corporate Governance***, 11(1): 35-44.
Mohsni, S., I. Otchere, and S. Shahriar. 2021. Board gender diversity, firm performance and risk-taking in developing countries: The moderating effect of culture. ***Journal of International Financial Markets, Institutions and Money***, 73: 101360.
Moreno-Ureba, E., F. Bravo-Urquiza, and N. Reguera-Alvarado. 2022. An analysis of the influence of female directors on environmental innovation: When are women greener? ***Journal of Cleaner Production***, 374: 133871.
Nadeem, M., S. Bahadar, A. A. Gull, and U. Iqbal. 2020. Are women eco-friendly? Board gender diversity and environmental innovation. ***Business Strategy and the Environment***, 29(8): 3146-3161.
Nadeem, M., T. Suleman, and A. Ahmed. 2019. Women on boards, firm risk and the profitability nexus: Does gender diversity moderate the risk and return relationship? ***International Review of Economics & Finance***, 64: 427-442.
Nielsen, S. and M. Huse. 2010a. Women directors' contribution to board decision-making and strategic involvement: The role of equality perception. ***European Management Review***, 7(1): 16-29.
Nielsen, S. and M. Huse. 2010b. The contribution of women on boards of directors: Going beyond the surface. ***Corporate Governance: An International Review***, 18(2): 136-148.
Olsen, B. C., K. Awuah-Offei, and D. Bumblauskas. 2021. Setting materiality thresholds for ESG disclosures: A case study of US mine safety disclosures. ***Resources Policy***, 70: 101914.

Peng, H. and T. Chandarasupsang. 2023. The effect of female directors on ESG practice: Evidence from China. ***International Journal of Financial Studies***, 11(2): 66.
Shakil, M. H., N. Mahmood, M. Tasnia, and Z. H. Munim. 2019. Do environmental, social and governance performance affect the financial performance of banks? A cross-country study of emerging market banks. ***Management of Environmental Quality: An International Journal***, 30(6): 1331-1344.
Veltri, S., R. Mazzotta, and F. E. Rubino. 2021. Board diversity and corporate social performance: Does the family firm status matter? ***Corporate Social Responsibility and Environmental Management***, 28(6): 1664-1679.
Wu, M. W., and C. H. Shen. 2013. Corporate social responsibility in the banking industry: Motives and financial performance. ***Journal of Banking & Finance***, 37(9): 3529-3547.
Yasser, Q. R., A. Al Mamun, and I. Ahmed. 2017. Corporate social responsibility and gender diversity: Insights from Asia Pacific. ***Corporate Social Responsibility and Environmental Management***, 24(3): 210-221.
Zhang, Z. 2023, September. Influence of Female Executives on the ESG Rating Performance of Enterprises. ***In 2023 9th International Conference on Humanities and Social Science Research (ICHSSR 2023),*** Atlantis Press.

ISSN:1946-6404

2025 Volume 17 Number 2