## MACHINES, ANALYSTS, AND FINANCIAL MARKETS

By

Xinyu Wang

## A DISSERTATION

Submitted to Michigan State University in partial fulfillment of the requirements for the degree of

Business Administration Finance—Doctor of Philosophy

2021

#### ABSTRACT

#### MACHINES, ANALYSTS, AND FINANCIAL MARKETS

#### By

### Xinyu Wang

In the first essay of this dissertation, I provide evidence on the usefulness of machine learning techniques to predict a firm's future earnings and its implied cost of capital. These techniques have the potential to offer significant marginal explanatory power over prior approaches which rely on either linear models or analyst forecasts. I adopt a deep neural network approach that incorporates lagged and contemporaneous accounting variables to predict future earnings. My evidence demonstrates that this forecasting approach offers significant explanatory power that can improve on analyst forecasts. In addition, the deep learning approach outperforms linear models and displays substantially less bias than the human analyst forecasts. When I turn to the implied cost of capital from my earnings forecast model, I find that my deep-learning-derived estimates significantly outperform popular linear-model-based estimates. I argue that these findings have interesting implications for a variety of questions in finance and accounting.

In the second study composing this dissertation, I turn from machines to people and consider reputational spillovers from buy-side analysts to sell-side analysts after a financial misconduct event. I detect evidence of negative reputational spillovers in the form of diminished market reliance on recommendations by sell-side analysts after buy-side analysts from the same brokerage firm are associated with financial misconduct. This penalty is significantly related to the buy-side analyst's gender, suggesting that market participants condition their expectations regarding analyst behavior on analyst-specific characteristics.

This dissertation is dedicated to my parents, Dr. Gang Wang and Ms. Shuming Yu.

#### ACKNOWLEDGEMENTS

I would like to express the depth of my gratitude to my adviser and dissertation committee chair Professor Charles Hadlock, who has provided me kind helps and supports during my PhD life. Dr. Charles Hadlock not only teaches me how to do serious research, but also provides me guidance on study, career, and more importantly, life. His charm of personality lights up my way. His optimistic attitude towards life impresses me a lot. His encouragement motivates me to go further no matter what difficulties I will meet in my life. It is my great honor to have him as my adviser.

I would like to thank my committee members, Professor Andrei Simonov, Professor John (Xuefeng) Jiang, and Professor Nuri Ersahin. Without their persistent help, I cannot finish the dissertation and graduate from the PhD program.

I would like to thank my parents Dr. Gang Wang and Ms. Shuming Yu. They firmly and selflessly support me through my PhD life, my study life, and my whole life. As parents, they build a healthy family environment for me, spare no effort to foster me, and teach me philosophy of life and conduct of life, which also influence my academic career. Without my parents, I cannot make the first achievement of my scholarly life. I feel very lucky that I born in this family!

# **TABLE OF CONTENTS**

LIST OF TABLES	. viii
LIST OF FIGURES	viiii
CHAPTER 1. The Implied Cost of Capital: A Deep Learning Approach 1. INTRODUCTION	1 2
2.1. Implied Cost of Capital	6
<ul><li>2.2. Deep Learning</li><li>3. DATA AND EMPIRICAL METHODOLOGY</li></ul>	7
3.1. Data 3.2. Deep Learning Model	11 11
<ul><li>3.3. The Benchmark: Cross-sectional Linear Regression</li><li>3.4. Computing the ICCs</li></ul>	18 18
4. RESULTS	20
Earnings Forecasts	20
4.2. Comparison of the Deep-Learning-Based and Linear-Model-Based Earnings Foreca	.sts 23
4.3. Performance of Deep Learning Model and Linear Regression on Estimating Compo	site 26
4.4. Comparison of the Individual Deep Learning Model-based ICC and the Individual Linear Model-based ICCs.	28
4.5.Realized Returns, ICCs, and Firm Characteristics	32
5.1. Extra Predictive Power of Deep Learning on Earnings	38
<ul><li>5.2. Comparison of Deep Learning and Analysts on Earnings Forecasts</li><li>5.3. Deep Learning Model with the Common Variables in the Linear Model</li></ul>	40 43
6. CONCLUSION	46 48
CHAPTER 2. Reputational Spillovers and Information Production by Analysts	54
1. INTRODUCTION 2. LITERATURE REVIEW	55 59
<ul><li>2.1. Financial Misconduct</li><li>2.2. Gender Difference</li></ul>	59 60
3. DATA AND EMPIRICAL METHODOLOGY	62
3.2. Empirical Methodology	64
<ol> <li>RESULTS</li></ol>	67 acts
- 	67

4.2. Comparison in One Year Before and After the Misconduct Dates	73
4.3. Comparison of Non-Misconduct Related Recommendations	79
5. CONCLUSION AND FUTURE POTENTIAL DIRECTIONS	
BIBLIOGRAPHY	

# LIST OF TABLES

# **LIST OF FIGURES**

Figure 1: An Example of the Operation Process of a Single Neuron.	13
Figure 2: The Basic Structure of a Deep Learning Model	14
Figure 3: An Example of the Financial Misconduct Report	63

CHAPTER 1. The Implied Cost of Capital: A Deep Learning Approach

### 1. INTRODUCTION

Estimating a firm's expected returns is a key issue in finance and accounting, and implied cost of capital (ICC), as a popular proxy of expected returns, has been widely applied in literature.<sup>1</sup> Prior researchers have largely relied on time-series estimates (e.g., Ball and Watts, 1972; Brooks and Buckmaster, 1976; Brown and Rozeff, 1978; Myers, Drake, Bradshaw, and Myers, 2012) and/or analyst forecasts on earnings as inputs into the implied cost of capital (or expected returns) calculation. However, there are some important limitations to these approaches, including limited data availability on analyst forecasts and substantial noise when relying on time-series estimates. In a major step forward, Hou, van Dijk, and Zhang (2012) offer a model-based framework to estimate future earnings and a firm's corresponding implied cost of capital (or expected returns). This approach has proven quite popular and has been relied on in several recent studies.

In this paper, I follow the principles behind the Hou et al. (2012) approach by considering a model-based method for predicting earnings. Instead of relying on the classic linear regression model, I use deep learning techniques trained on common accounting information to offer a substantial improvement to the earnings prediction model. Since both simple regression models and the complicated thought processes of analysts appear to uncover important elements of a firm's earnings dynamics, my hope is that a computer trained to think in some ways like the unstructured brain of an analyst but with more discipline (e.g., no optimism bias) may offer the best of both

<sup>&</sup>lt;sup>1</sup>Previous literature about implied cost of capital (ICC) estimation includes Gordon and Gordon (1997), Gebhardt, Lee, and Swaminathan (2001), Claus and Thomas (2001), Easton (2004), Ohlson and Juettner-Nauroth (2005), etc. Implied cost of capital has been treated as a good proxy of expected returns (Pástor, Sinha, and Swaminathan, 2008; Li, Ng, and Swaminathan, 2013), and it has been applied to study the relevance of expected returns to various types of risk (Lee, Ng, and Swaminathan, 2009; Chava and Purnanandam, 2010; Hwang, Lee, Lim, and Park, 2013; Dhaliwal, Judd, Serfling, and Shaikh, 2016; Kalev, Saxena, and Zolotoy, 2019), labor mobility (Donangelo, 2014), information asymmetry (El Ghoul, Guedhami, Ni, Pittman, and Saadi, 2013), political factors (Boubakri, Guedhami, Mishra, and Saffar, 2012; Boubakri, El Ghoul, and Saffar, 2014), option trading (Naiker, Navissi, and Truong, 2013), financial constraints (Campbell, Dhaliwal, and Schwartz, 2012), etc. A list of top finance and accounting journal papers using ICC as a dependent variable of analysis can be found in Lee, So, and Wang (2020).

worlds and therefore a substantially improved prediction ability.

Recent advances in deep learning, a subbranch of machine learning techniques in artificial intelligence, appear ideally suited to this task. In particular, the deep neural network approach simulates the activity of neurons in the human brain when reacting to informational inputs such as quantitative data, text, or imagery. While current models of this type are far less sophisticated than the human brain, they offer a potential compensating advantage when it comes to prediction in that various types of human subjectivity bias can be eliminated. Thus, the performance of these types of models in predicting earnings and estimating the associated implied cost of capital is ultimately an empirical question.

To conduct this analysis, I generate deep-learning-based earnings forecasts for up to five future years and calculate the implied cost of capital using these estimates of earnings. I show that my derived earnings forecasts do offer marginal explanatory power above and beyond analyst forecasts. Moreover, the deep-learning forecasts display substantially less bias than the human analyst forecasts. When I turn to predicting returns and the implied cost of capital from my earnings forecast model, I find that my deep-learning-derived estimates significantly outperform the linear-model-based estimates of Hou et al. (2012).

This study contributes to the literature in multiple ways. First, this is the first study that generates an implied cost of capital by an approach that has both model-based features and human brain-based features. Prior studies do not allow these two types of features to be combined. Deep learning is a machine learning approach that is ideally suited to this task, as it is an artificial intelligence method for combining the powerful yet unstructured creative nature of human thinking with the discipline of computer-directed and structured optimization. Thus, one contribution of this study is to illustrate in one specific context the promise of this approach more generally to

address important issues in finance and accounting.

A second contribution of this study is to offer a substantial step forward in more accurately estimating a firm's ICCs. In particular, I show that 38 of the most common accounting items can be used by a deep neural network to generate earnings predictions that lead to an ICC that predicts returns substantially better than prior model-based approaches. Thus, any researcher interested in estimating an implied cost of capital would be well served by using the techniques explored in this study.

An additional contribution of this study is that I establish that behavioral factors in human thinking that lead to biases in predictions by analysts can be solved by deep-learning methods. While Hou et al. (2012) have established that linear regression models can also lead to bias-free predictions, I show that the data can be used much more richly and in a human-like way while still eliminating this flaw in true human thinking. Thus, machine learning has the potential to richly and creatively process information while placing structure and discipline on the information aggregation process.

Finally, my study shows that machine learning and deep learning can be used in prediction contexts related to real activities by firms rather than solely financial market predictions. While prior authors have explored deep learning models in the areas of asset pricing, portfolio management, and mortgage risk, the promise of this approach in predicting the outcome of a firm's real activities (earnings, investment, growth) has not been widely recognized. Hopefully this study will lead to additional work in which predicting the values of real variables is an important underlying economic issue.

The rest part of the article is organized in the following structure. Section 2 reviews the literature, while Section 3 introduces the data, models, and empirical methodology. In Section 4, I

present the main results regarding the performance of the deep learning models, both in an absolute and relative sense, in forecasting earnings and in estimating the implied cost of capital. Section 5 reports a set of additional evidences and robustness checks, while Section 6 will reach to conclusions.

#### 2. LITERATURE REVIEW

#### 2.1. Implied Cost of Capital

The expected return is a major topic in research related to firm resources allocations or decision making. Since the ex ante expected returns are unknown, in the past, researchers use historical realized returns to forecast expected returns based on time-series models, but academic researches (e.g., Froot and Frankel, 1989; Elton, 1999) indicate that the expected returns by time-series models do not work well, majorly since ex post realized returns are noisy. Besides, for the new firms with limited historical information, it is hard to get a proper estimate on expected return based on non-sufficient ex post realized returns.

Starting from Botosan (1997), researchers widely use the implied cost of capital (ICC) as a new proxy of the expected return. ICC is the discount rate equating the share price and the present value of expected future cash flows. It is appeal to people because of the present value relation involved as well as the use of forecasts of firms' fundamentals in the future. According to Wang (2017), more than 70 papers from 1997 to 2016 have been published in the top journals of finance and accounting. In recent studies, researchers use ICCs as the proxy of expected returns to study the relevance of expected returns to regulations or acts (Ashbaugh-Skaife, Collins, Kinney, and Lafond, 2009; Dhaliwal, Krull, and Li, 2007), disclosure (Botosan, 1997; Botosan and Plumlee, 2002; Francis, Khurana, and Pereira, 2005b; Francis, Nanda, and Olsson, 2008), option trading (Naiker et al., 2013), risk (Chava and Purnanandam, 2010; Hwang et al., 2013; Dhaliwal et al., 2016), financial constraints (Campbell et al., 2012), information asymmetry (El Ghoul et al., 2013), auditor's characteristics (Chen, Chen, Lobo, and Wang, 2011; Khurana and Raman, 2006; Krishnan, Li, and Wang, 2013), conservatism (García Lara, García Osma, and Penalva, 2011), tax (Dhaliwal, Krull, Li, and Moser, 2005; Dhaliwal, Heitzman, and Li, 2006; Goh, Lee, Lim, and Shevlin, 2016), and internal control weakness (Ogneva, Subramanyam, and Raghunandan, 2007).

Since ICC is an important proxy of expected returns in academic research, a good way of estimating ICC may affect the results of research a lot. Majorly, the studies treat forecasted earnings of financial analysts as expected future cash flows and calculate ICC. However, analysts' forecasts contain not only objective information, but also the subjective judgments from financial analysts, which may lead to biased results on the estimation of ICC. Also, since financial analysts will only do forecasting for a limited number of firms, using analysts' forecasts may limit the coverage of firms. Due to considerations such as bias, firms coverage, and mixed results on the performance of forecasting future realized returns, Hou et al. (2012) introduce an innovative method on estimating ICC. They use a cross-sectional model to get estimated earnings first and then calculate ICC using the average of ICCs from five commonly use ICC estimation methods. By doing the cross-sectional estimation, they reduce the level of bias on earnings estimates comparing to analysts' forecasts, increase the range of firms covered in the analysis, and improve the performance of estimated ICCs on forecasting the future realized returns. This cross-sectional linear model-based method has been widely used in recent studies.

### 2.2. Deep Learning

Machine learning is a set of algorithms that can provide machines ability to learn by themselves without being explicitly instructed or programmed. The phrase "Machine learning" was first came up by Arthur Samuel in 1952. In the long history of development, there are lots of machine learning techniques applied in the real world, including business, health industry, and security, etc. Traditional machine learning techniques include Bayesian model averaging, random forest, nearest neighbors, support vector machine, and neural networks, etc. For different purposes, people rely on either linear machine learning algorithms (for instance, LASSO and ridge regression), or non-linear machine learning algorithms (for example, decision trees) for more complex tasks. Based on the characteristics of each machine learning algorithms, people use machine learning algorithms for pattern forecasting, regression, or classification purposes.

In recent years, artificial intelligence and machines are applied in various areas, and automated valuations start playing an important role in the capital market. People use machine learning and artificial intelligence techniques in areas such as asset pricing (Bianchi, Büchner, and Tamoni, 2019; Brogaard and Zareei, 2018; Ye and Zhang, 2019; Bloch, 2019), credit risk management (Altman, Marco, and Varetto, 1994; Kim and Sohn, 2004; Abdou, Alam, and Mulkeen, 2014; Khandani, Kim, and Lo, 2010; Bonelli, Figini, and Giovannini, 2017), exchange rate prediction (Peng and Albuquerque, 2019) and fraud detection (Bertomeu, Cheynel, Floyd, and Pan, 2019; Bao, Ke, Li, Yu, and Zhang, 2019), etc. For example, Ding, Lev, Peng, Sun, and Vasarhelyi (2019) show that machine learning can improve managerial estimates. They use machine learning techniques to generate loss estimates using the data of insurance companies and conclude that the estimates by machines are better than the managers' actual estimates. Though machine learning techniques share common characteristics like self-learning and updating ability, the operating principles behind each of them are various.

However, moving into the era of big data, people need to deal with tasks with huge amount of data and more complex analysis structures, and traditional machine learning algorithms are facing some obstacles in application. For example, most of the time, people need to manually determine which features (or variables) to be used before the construction of traditional machine learning models. This feature creation process, or feature engineering, generally depends on the knowledge in the certain area that we are applying the machine learning model. For example, a doctor diagnoses a patient as a certain illness based on symptoms, and symptoms here are features used to make a judgement. But, professional knowledge based on the past experience may not always be a good source of finding the best features for the machine learning model, and human beings have the limitation on time, ability and energy to perfectly deal with the huge amount of data. Thus, people start to consider a new algorithm which can not only complete the learning process as traditional machine learning algorithms, but also efficiently do feature engineering works instead of human beings.

To mimic the complicated human actions like feature engineering, people may think about an algorithm which can closely simulate the operation of brain. Neural network is a machine learning technique borrowing the features of neurons of creatures, and deep neural network, the deep learning algorithm discussed in this paper, is a more complicated version of neural networks and an advanced machine learning technique mimicking the operating principles of creature neurons and human brains. Different from traditional neural networks, deep neural network generally includes multiple hidden layers between an input layer and an output layer, which may analyze input data with much more dimensions. With more layers and neurons, deep neural network can transform the information through a complicated network structure in a way imitating the decision making process of human beings. According to Zhou and Feng (2017), the layer-bylayer processing principle and representation learning ability are the keys of success of deep neural networks in different areas. These two characteristics allow deep learning models to generate new features in the learning process without the involvement of human beings and ensure the feature transformation within the model structures, which is more efficient than traditional machine learning algorithms. Besides, the multilayer structures and the choices of activation functions at neuron level provide stronger ability to analyze big data. These advantages make deep learning algorithm a promising machine learning technique to deal with big data, reduce human

involvement, and provide better estimation once the model is well trained. Thus, in recent years, deep learning becomes one of the hottest machine learning algorithms applied in the real world, including its application in creating Google's AlphaGo and detecting Covid-19 cases (Zheng, Deng, Fu, Zhou, Feng, Ma, Liu, and Wang, 2020).

But in business area, based on current literature, this technique imitating neurons and human brains is employed only on portfolio management (Heaton, Polson, and Witte, 2017), mortgage risk management (Sirignano, Sadhwani, and Giesecke, 2018) and asset pricing (Chen, Pelger, and Zhu, 2019; Messmer, 2017). My study in this paper will be the first one applying deep learning techniques to earnings forecasts and ICC estimation.

### 3. DATA AND EMPIRICAL METHODOLOGY

### 3.1. Data

The data related to firms comes from Compustat and CRSP. The study includes firms from NYSE, Amex, and Nasdaq, and sharecodes should be 10 or 11. All firms in the sample should be available in both Compustat and CRSP, and the range of data is from 1962 to 2018. When constructing the accounting variables for the deep learning model, I first rank the frequency of all accounting items in Compustat across the whole time period and then pick up the most frequent seventy items. Next, I calculate the correlation between each pair of the seventy variables and rule out the variable with a lower level of frequency if the two variables in the pair have a correlation higher than 0.9. In the end, there are 38 variables from Compustat left in the sample. For the variables used in estimating the linear model-based earnings and ICCs, I follow the definitions in Hou et al. (2012) and construct the variables for the pooled cross-sectional regressions. I also obtain forecasted earnings of financial analysts and actual earnings of firms from I/B/E/S.

#### 3.2. Deep Learning Model

In this paper, the deep learning model I employ is a deep neural network. A deep neural network is a type of machine learning technique that imitates the operating process of human brains. Different from general neural networks, deep neural networks have more than two layers, meaning that there will be at least one layer between input layer and output layer, leading to a more sophisticated estimation process. Once the input data comes available, the information will be passed through the "neurons" on different layers and finally reach to the output layer, showing the output of the whole analysis process. With enough training data, deep neural networks can adjust by itself and provide judgments on the final output results like human beings.

As indicated in Amel-Zadeh, Calliess, Kaiser, and Roberts (2020), the deep learning

approach is a model that maps independent variables to a predicted value with a set of parameters:  $\hat{y} = f(x; \Theta)$ . In particular,  $\Theta$  represents a set of weights and biases, and the formula can be rewritten as  $\hat{y} = f(x; w, b)$ , in which w represents weights and b represents biases. Learning means the process to find the optimal values of w and b so that  $\hat{y} = f(x; w, b)$  can be the best estimation comparing to the actual value y. Thus, the general process is using the training data to find the optimal values of w and b first and utilizing the input data to calculate the forecasts.

In this paper, the number of inputs of a single unit (or neuron) is determined by the number of neurons on the previous layer. For instance, the single neuron on the first hidden layer will have 38 input variables from the input layer of the model. Each single neuron has a non-linear activation function, which will combine the inputs and conduct an output of this neuron for next layer. The process is shown in Figure 1. There are many common forms of activation functions (e.g., Sigmoid, Tanh, Relu, etc., according to Zhou (2019)). I use "Relu" as the form of the activation function, which is the commonly used activation function, and the activation function used here is:

$$a = max\left(0, z\right) \tag{1}$$

Thus, the output value from a single neuron *i* equals to:

$$a_{i} = max (0, \sum_{j=1}^{n} w_{ij} x_{j} + b_{i})$$
(2)

where  $a_i$  is the output value of neuron *i*,  $w_{ij}$  is the weight between input *j* and neuron *i*,  $x_j$  is the value of input *j*, and  $b_i$  is the bias.





Once a neuron conducts an output value, the output becomes one of the inputs for the neurons on the next layer. After discussing the principles behind single neurons, we focus on the full picture of the model shown in Figure 2. The deep learning model I applied in this research include two hidden layers between input layer and output layer. The input layer consists of the 38 accounting items from financial statements. The information is passed through the "links" between each neuron following the principles introduced in Figure 1, while each link is related to a specific weight. An estimated value  $\hat{y}$  can be calculated at the output layer. The above procedure from inputs to  $\hat{y}$  is called "forward propagation." At the very beginning, when the deep neural network is an initial one or "a blank brain," the weights are assigned randomly (e.g., all weights can be an equal value). During the training process, once  $\hat{y}$  is calculated using inputs from training data, the machine will apply mean squared error (mse) as the loss function to compare the estimated value  $\hat{y}$  with the actual value y, and will use optimizer to update weights and biases in  $f(x; \Theta)$ . The optimizer used in this research is RMSprop, an optimizer similar to the gradient descent algorithm

with momentum. After repeating the training process above, a deep learning model with optimal parameters can be reached, and earnings forecasts can be calculated based on the model and new input data.



Figure 2: The Basic Structure of a Deep Learning Model

Starting from 1968 to 2018, for each year, I utilize the data of the past 10 years as training data and construct five deep learning models in five different earnings horizons (from one year ahead to five years ahead) for the year *t*. Taking the information availability issue into consideration, I apply a 3-month lag on the financial reporting, meaning that I treat the accounting data (including actual earnings in training data set) of firms having fiscal years located in the range from April of previous year to March of current year as the data for current year. Thus, during the training process, to estimate the models for year *t*, both accounting items and actual earnings from

financial statements should be available before the March of year *t*. I continue this process to construct deep learning models for each year *t* in a 10-year rolling base. After the training of models for each year *t*, still considering the information availability issue, I put the accounting item numbers of companies with fiscal years within the range from April of previous year *t*-1 to March of current year *t* into the deep neural network models for year *t* to compute earnings forecasts of firms in year  $t + \lambda$  ( $\lambda$ =1 to 5). For each firm, in each year, I generate the estimated earnings for up to five future years. Since deep neural network models in this study require nonmissing observations for all the 38 independent variables, to maximize the coverage of firms in the estimation, I choose the 38 independent variables based on the frequency of available observations and the correlations among variables. Table 1 presents a list of accounting items used in the deep learning model and descriptive statistics.

Variables	Obs.	Mean	Median
Assets - Total	241222	3411.01	144.01
Cash and Short-Term Investments	238598	383.88	11.38
Liabilities - Other - Total	238841	570.31	1.02
Current Assets - Other - Sundry	215323	43.90	0.80
Preferred Stock - Liquidating Value	240988	21.56	0.00
Property, Plant and Equipment - Total (Net)	238896	548.16	21.46
Common Stock Equivalents - Dollar Savings	240043	-0.08	0.00
Income Before Extraordinary Items	240936	78.41	3.24
Dividends - Preferred/Preference	240648	1.59	0.00
Extraordinary Items	233500	-1.37	0.00
Revenue - Total	240298	1392.95	110.31
Extraordinary Items and Discontinued Operations	240275	-0.54	0.00
Income Taxes - Total	240651	41.91	1.87
Dividends Common/Ordinary	240394	33.33	0.00
Nonoperating Income (Expense) - Other	220303	9.72	0.09
Long-Term Debt - Total	240024	601.95	11.5
Earnings Per Share (Basic)	239737	1.14	0.69
- Excluding Extraordinary Items			
Gross Profit (Loss)	238971	452.14	35.12
Assets - Other	238811	302.99	2.93
Inventories - Total	233805	240.40	7.78
Receivables - Total	235618	1027.35	20.05
Debt in Current Liabilities - Total	237736	384.62	2.1
Notes Payable - Short-Term Borrowings	237577	305.60	0.0
Receivables - Current - Other	213064	72.50	0.0
Common Equity - Tangible	237079	383.25	38.6
Book Value Per Share	236985	59.05	7.50
Current Liabilities - Total	208483	351.39	21.9
Working Capital (Balance Sheet)	206523	147.72	19.20
Special Items	227210	-15.87	0.0
Preferred/Preference Stock (Capital) - Total	235968	18.44	0.0
Preferred/Preference Stock - Nonredeemable	231536	14.55	0.0
Convertible Debt and Preferred Stock	214176	19.73	0.0
Investment Tax Credit (Balance Sheet)	212234	2.74	0.0
Deferred Taxes (Balance Sheet)	219484	89.87	0.15
Common/Ordinary Stock (Capital)	231995	50.76	0.8
Retained Earnings	231883	404.23	12.58
Capital Surplus/Share Premium Reserve	231442	397.69	25.2
Discontinued Operations	224571	0.82	0.00

Table 1: Descriptive Statistics of the Variables in the Deep Learning Model

Table 1 (Cont'd)

Variables	Obs.	Mean	Median
Assets - Total	241222	3411.01	144.01
Cash and Short-Term Investments	238598	383.88	11.38
Liabilities - Other - Total	238841	570.31	1.02
Current Assets - Other - Sundry	215323	43.90	0.80
Preferred Stock - Liquidating Value	240988	21.56	0.00
Property, Plant and Equipment - Total (Net)	238896	548.16	21.46
Common Stock Equivalents - Dollar Savings	240043	-0.08	0.00
Income Before Extraordinary Items	240936	78.41	3.24
Dividends - Preferred/Preference	240648	1.59	0.00
Extraordinary Items	233500	-1.37	0.00
Revenue - Total	240298	1392.95	110.31
Extraordinary Items and Discontinued Operations	240275	-0.54	0.00
Income Taxes - Total	240651	41.91	1.87
Dividends Common/Ordinary	240394	33.33	0.00
Nonoperating Income (Expense) - Other	220303	9.72	0.09
Long-Term Debt - Total	240024	601.95	11.55
Earnings Per Share (Basic)	239737	1.14	0.69
- Excluding Extraordinary Items			
Gross Profit (Loss)	238971	452.14	35.12
Assets - Other	238811	302.99	2.93
Inventories - Total	233805	240.40	7.78
Receivables - Total	235618	1027.35	20.02
Debt in Current Liabilities - Total	237736	384.62	2.11
Notes Payable - Short-Term Borrowings	237577	305.60	0.00
Receivables - Current - Other	213064	72.50	0.00
Common Equity - Tangible	237079	383.25	38.65
Book Value Per Share	236985	59.05	7.50
Current Liabilities - Total	208483	351.39	21.97
Working Capital (Balance Sheet)	206523	147.72	19.26
Special Items	227210	-15.87	0.00
Preferred/Preference Stock (Capital) - Total	235968	18.44	0.00
Preferred/Preference Stock - Nonredeemable	231536	14.55	0.00
Convertible Debt and Preferred Stock	214176	19.73	0.00
Investment Tax Credit (Balance Sheet)	212234	2.74	0.00
Deferred Taxes (Balance Sheet)	219484	89.87	0.12
Common/Ordinary Stock (Capital)	231995	50.76	0.81
Retained Earnings	231883	404.23	12.58
Capital Surplus/Share Premium Reserve	231442	397.69	25.25
Discontinued Operations	224571	0.82	0.00

#### 3.3. The Benchmark: Cross-sectional Linear Regression

Following the steps introduced in Hou et al. (2012), from 1968 to 2018, I use the past 10year data to do pooled cross-sectional linear regressions for each year *t*:

$$E_{i,t+\lambda} = \alpha_0 + \alpha_1 A_{i,t} + \alpha_2 D_{i,t} + \alpha_3 D D_{i,t} + \alpha_4 E_{i,t} + \alpha_5 Neg E_{i,t} + \alpha_6 A C_{i,t} + \epsilon_{i,t+\lambda}$$
(3)

where  $E_{i,t+\lambda}$  is the firm *i*'s earnings in year  $t + \lambda$ , for  $\lambda = 1, 2, 3, 4$ , or 5;  $A_{i,t}$  denotes the total assets of year *t* from Compustat;  $D_{i,t}$  is the dividend payment of year *t*;  $DD_{i,t}$  equals to 1 if firm *i* is a dividend payer for year *t*, and equals to 0 if not;  $E_{i,t}$  is the earnings of year *t*;  $NegE_{i,t}$  equals to 1 if firm *i* has negative earnings for year *t*, and 0 if not;  $AC_{i,t}$  represents accruals in year *t*.

Also, based on the consideration of accounting information availability, a three-month lag is applied in the calculation of earnings forecasts. For each year *t* and each firm *i*, I multiply the accounting numbers from firms whose fiscal years are located within the range of April of the previous year t - 1 and March of the current year *t* with the coefficients from the linear model above for a certain year *t* to calculate the forecasted earnings on year  $t + \lambda$ , for  $\lambda = 1$  to 5.

## 3.4. Computing the ICCs

The ICC is the discount rate equating the security share price to the sum of present values of future cash flows. In the past literature, there are mainly five methods of calculating ICC. Following Hou et al. (2012), in this study, I use the average of ICCs calculated by the five most commonly used methods: Gordon and Gordon (1997), Gebhardt et al. (2001), Claus and Thomas (2001), Easton (2004), and Ohlson and Juettner-Nauroth (2005).

I first calculate the five ICCs individually using the future five years forecasted earnings

from both deep learning and linear regression. The market equity and stock price at the end of each June are also used when doing calculation. After getting the five ICCs for both deep learning model and linear regression model, I compute the "composite" ICC, the equal-weighted average of these five Implied Cost of Capital values for each firm and each year, for both deep learning model and linear regression model. Since in some cases, the equation can not get a converged solution using the earnings forecasts and other accounting variables, I may not generate five available ICCs for each firm and each year. I will take an observation into account if there is at least one implied cost of capital value available and calculate the "composite" ICC only considering available ICCs for the firm on year *t*.

#### 4. RESULTS

4.1. Descriptive Statistics of the Deep-Learning-Based and Linear-Regression-Based Earnings Forecasts

Table 2 includes the averages across years of mean and median of one-year, two-year, and three- year ahead forecasted earnings estimated by the deep learning model (Panel A) and Hou et al. (2012) linear regression (Panel B) in each period listed, and it also includes the correlation between earnings forecasts from deep learning and linear regression (Panel C). When constructing the table, the market equity at the last day of June for each firm is used to scale the earnings forecasts from both models. Although extreme earnings forecasts for certain firms affect the mean of the estimated earnings, the results in median columns in Panel A show that there is a decreasing trend of earnings forecasts from late 1970, and the trend reverses for the period after 2008, both for deep learning based forecasts and for linear regression based forecasts. In general, the medians of earnings forecasts estimated by linear regression are larger than the ones estimated by deep learning model for all the three horizons in each period after late 1970, showing that the linear model provides more optimistic earnings forecasts comparing to the deep learning model. However, to check which model can provide relatively less biased or better estimates, further tests using actual earnings are needed, and more results will be shown in the following subsection.

Table 2 also indicates the number of firms with estimated earnings by deep learning model in Panel A and the number of firms with estimated earnings by the linear model in Panel B in each defined period. For the deep learning model, the number of firms has increased from around 2500 in the 1968-1973 period to almost 7000 at the end of 20 century, and decreased to a level of around 4000 in recent years. The linear model has a similar pattern, but it covers more distinct firms in every sub-period. The major reason behind is that for the deep learning model, all the 38 variables need to be non-missing to provide an estimate on earnings, while for the linear model, I only need the information from six variables to do the estimation. Obtaining earnings forecasts using the deep learning model needs more available data than doing so using the linear model.

Panel C shows the correlations between the deep learning based forecasted earnings and the linear regression based forecasted earnings. Since both models are relying on the quantitative information from financial statements, correlations between earnings forecasts are generally higher than 0.90, except the correlations between two-year horizon earnings forecasts from deep learning and other forecasts (from 0.74 to 0.94). To further check which method can generate better estimates of earnings and ICCs, I will discuss the analysis on earnings forecasts and ICCs in the following sections.

## Table 2: Descriptive Statistics of Estimated Earnings

This table includes descriptive statistics of the one-year, two-year, and three-year ahead estimates of earnings generated by deep learning (Panel A) and the linear forecast model (Panel B), and the correlations between the deep-learning-based forecasts and the linear regression-based forecasts (Panel C).  $E_{t+1}$ ,  $E_{t+2}$ , and  $E_{t+3}$  represent the one-year, two-year, and three-year ahead earnings forecasts. They are divided by market equity at the end of June of each year for scaling purposes. N is the number of distinct companies for which the forecasts are obtainable during each period. The sample period is 1968-2015.

Period	N	$E_{t+1}$ Mean	$E_{t+1}$ Median	$E_{t+2}$ Mean	$E_{t+2}$ Median	$E_{t+3}$ Mean	$E_{t+3}$ Median	Ν	$E_{t+1}$ Mean	$E_{t+1}$ Median	$E_{t+2}$ Mean	$E_{t+2}$ Median	$E_{t+3}$ Mean	$E_{t+3}$ Median
	Panel	A: Deep I	learning l	Model-bas	sed Foreca	asts			Panel B: I	Linear Mo	del-based	Forecast	s	
1968-1973	2546	0.0569	0.0582	0.1485	0.1005	0.0867	0.0736	3099	0.1131	0.0937	0.1599	0.1153	0.2059	0.1359
1974-1978	3681	0.1111	0.1197	0.2436	0.1634	0.2988	0.1750	4238	0.1837	0.1562	0.2757	0.1919	0.3560	0.2249
1979-1983	4599	0.0949	0.0929	0.0261	0.0790	0.0239	0.0763	5127	0.1538	0.1330	0.2454	0.1678	0.3230	0.1971
1984-1988	5884	0.7031	0.0401	0.5774	0.0767	0.8962	0.0391	6686	1.3380	0.0663	1.4176	0.0851	1.5160	0.1047
1989-1993	5581	-0.0592	0.0321	-0.0155	0.0386	0.0191	0.0457	6191	-0.1339	0.0483	0.0539	0.0591	0.1818	0.0722
1994-1998	7059	-0.0271	0.0265	-0.0293	0.0235	-0.0235	0.0338	7966	-0.0330	0.0350	0.0489	0.0467	0.1132	0.0616
1999-2003	6105	-0.1760	0.0123	0.0816	-0.0075	-0.0049	0.0226	6962	-0.5392	0.0200	-0.3786	0.0306	-0.2436	0.0464
2004-2008	4511	-0.1003	0.0126	-0.0050	0.0250	-0.0270	0.0264	5814	-0.0925	0.0253	-0.0523	0.0273	-0.0040	0.0343
2009-2015	4025	-0.0767	0.0249	-0.1371	0.0210	-0.0358	0.0323	5263	-0.1605	0.0309	-0.1037	0.0356	-0.0256	0.0458

		Deep Learning Model-based Forecasts			Line	ear Model-based Fore	casts
		$E_{t+1}$	$E_{t+2}$	$E_{t+3}$	$E_{t+1}$	$E_{t+2}$	$E_{t+3}$
Deep Learning	$E_{t+1}$	1.00					
Model-based Forecasts	$E_{t+2}$	0.75	1.00				
	$E_{t+3}$	0.93	0.94	1.00			
Linear Model	$E_{t+1}$	1.00	0.74	0.93	1.00		
-based Forecasts	$E_{t+2}$	1.00	0.75	0.93	1.00	1.00	
	$E_{t+3}$	1.00	0.76	0.94	1.00	1.00	1.00

4.2. Comparison of the Deep-Learning-Based and Linear-Model-Based Earnings Forecasts

Table 3 includes a comparison of the performance of deep-learning-based earnings forecasts to the performance of linear-model-based forecasts. I generate averages of the mean as well as median forecast bias, accuracy, and the annual earnings response coefficients across all years from 1968 to 2018, and the difference between the two sets of earnings forecasts. I use this set of sample because the generating process of earnings forecasts needs past ten years' sample observations for each year, and earnings forecasts as well as three years' actual earnings are included in this analysis. The comparison is based on a common firm-year range, meaning that all the observations included in the comparison have available actual values (income before extraordinary items) from Compustat and available earnings forecasts from both deep learning and cross-sectional linear regression. Each comparison includes one-year, two-year, and three-year periods as the horizons.

## Table 3: Comparison of Deep Learning's Forecasts and Model's Forecasts on Earnings

This table compares the bias, accuracy and annual earnings response coefficient (ERC) of deep learning's forecasts on earnings and linear regression model's earnings forecasts. Bias equals actual earnings minus earnings forecasts, divided by market equity at the end of June each year for scaling. Accuracy refers to the absolute value of forecast bias. Annual ERC is the coefficient by regressing buyand-hold returns on forecast bias. To make annual ERCs of deep learning model and linear model comparable, I also standardize the bias by one unit variance before the regression. All numbers are calculated in time-series averages. The sample is from 1968 to 2018. Panel A includes averages of mean as well as median of bias across years. Panel B includes averages of mean as well as median of accuracy across years. Panel C includes averages of annual ERC. The numbers under each average are the t-statistics.

		$E_{t+1}$			$E_{t+2}$	1		$E_{t+3}$	
	Deep Learning	Linear Model	Difference	Deep Learning	Linear Model	Difference	Deep Learning	Linear Model	Difference
Panel A:	Bias of Earn	nings Forecasts							
Mean	-0.0498 -4.20	-0.0740 -5.22	0.0242 1.77	-0.0274 -1.47	-0.0852 -5.74	0.0578 2.95	-0.0184 -0.88	-0.1153 -8.06	0.0968 4.49
Median	0.0024 0.78	-0.0057 -2.36	0.0081 2.53	-0.0028 -0.43	-0.0176 -4.28	0.0148 2.54	-0.0011 -0.12	-0.0308 - 6.54	0.0296 3.30
Panel B:	Accuracy of	Earnings Foreca	sts						
Mean	0.1601 14.22	0.1837 13.63	-0.0237 -4.07	0.1995 19.62	0.1984 18.77	0.0011 0.10	0.2138 15.98	0.2224 19.65	-0.0086 -0.75
Median	0.0464 20.86	0.0454 20.75	0.0009 0.67	$0.0731 \\ 20.40$	0.0650 20.44	0.0081 2.80	0.0865 13.83	0.0799 19.66	0.0066 1.40
Panel C:	Annual Ear	nings Response (	Coefficient (ER	.C)					
Mean	$0.0624 \\ 3.81$	$0.0501 \\ 2.95$	0.0123 1.09	$0.2034 \\ 8.19$	0.1522 6.65	0.0512 1.87	$0.3506 \\ 9.64$	$0.2531 \\ 9.40$	$0.0975 \\ 3.23$

According to definitions from Hou et al. (2012), forecast bias equals actual earnings minus forecasted earnings, divided by market equity value of the end of June for scaling purposes. In Table 3, the biases calculated by the earnings forecasts of deep learning model and linear model are all negative for one-year, two-year and three-year horizons, showing that the two models provide optimistic forecasts on average comparing to the realized earnings. Relying on the same sample base, the deep learning models provide less biased earnings forecasts (-0.0498, -0.0274, -0.0184 for one-year, two-year, and three-year ahead forecast), comparing to the earnings forecasts by the linear regression (-0.0740, -0.0852, -0.1153 for one-year, two-year, and three-year ahead forecast, based on the average of means for each year. The differences between the forecasting bias of the two approaches are 0.0242 for one-year ahead forecast, 0.0578 for two-year ahead forecast, and 0.0968 for three-year ahead forecast. The results indicate that the estimates of earnings by deep learning are less biased than those by the linear model for all the three horizons. When using the average of medians across years instead, I reach a consistent conclusion.

The forecast accuracy, discussed in Panel B, is the absolute value of forecast bias. For deep learning model, the time-series averages of the means of the forecast accuracy are 0.1601 for one-year horizon, 0.1995 for two-year horizon, and 0.2138 for three year horizon, while for the linear regression, the averages are 0.1837, 0.1984, and 0.2224 for the three horizons. The differences between the average of the means of forecast accuracy are -0.0237 for one-year ahead estimation, 0.0011 for two-year ahead estimation, and -0.0086 for three-year ahead estimation, showing that the deep learning model performs significantly better for one-year horizon (t-stat of the difference is -4.07). For the two-year and three-year horizon, the averages of earnings forecasts accuracy are very close for both the two models.

The earnings response coefficient (ERC), representing the response of the market on the

difference between the actual value and forecasted value, is a better way of measuring the quality of earnings forecasts. Following previous literature, I apply the "annual ERC" approach defined in Hou et al. (2012) in the article. It checks if earnings forecasts can meet the expectation of the market. I regress the buy-and-hold returns on forecast bias, and the coefficients are the annual ERCs. To make the annual ERCs from the two models comparable, I standardize forecast bias by one-unit variance before doing regression. Then, I take the average of the means of annual ERCs for each year. For the deep learning model, the averages are 0.0624 for one-year horizon forecast, 0.2034 for two-year horizon forecast, and 0.3506 for three-year horizon forecast, which are larger than those from the linear model (0.0501 for one-year horizon, 0.1522 for two-year horizon, and 0.2531 for three-year horizon). The differences between the averages are 0.0123, 0.0512 and 0.0975 for one-year, two-year and three-year ahead forecasts, which are all positive values, showing that the market reactions to the unexpected earnings are larger for the deep learning model. Because the difference between actual earnings and forecasted earnings causes a larger magnitude of reactions from the market, it shows that the estimates from the deep learning model match the expectation of the market better, and the earnings forecasts by the deep learning model proxy the expected earnings better.

### 4.3. Performance of Deep Learning Model and Linear Regression on Estimating Composite ICC

Following Guay, Kothari, and Shu (2011) and Hou et al. (2012), I measure the performance of deep learning model and linear regression on estimating composite ICC by evaluating the ability of the composite ICC on predicting future realized returns. Since implied cost of capital is a proxy of expected returns, a good ICC estimate should be able to predict the real returns in the future.

After estimating the earnings forecasts, I generate individual ICCs and composite ICC using the earnings forecasts from deep learning and linear regression model. In the following

analysis, I group firms into ten deciles according to the rank of their composite ICCs at the end of June of each year, in which group 1 includes firms having the lowest ICC while group 10 has firms with the highest ICC. Each group is treated as a portfolio. Next, I calculate the equal-weighted buy-and-hold returns of each group for each year, starting from July of the current year to June of the next year. Then, I compute the average of all the equal-weighted buy-and-hold returns across all years in the sample for each portfolio separately. Since it is equal-weighted average returns, the result can be easily affected by extreme return values from specific firms, such as small firms. To avoid this effect, I trim the annualized returns for each year at the 1st and 99th percentiles (excluding the extreme values beyond percentiles). As a proxy of expected returns, ICC should have the forecasting power for future returns. Thus, we are expected to see a higher average return for the decile with a higher rank of ICC, and the monotonically increasing trend should be more obvious for the model with better estimates of ICC.

In Table 4, comparing the results from deep learning and linear regression, the study shows that the difference between group 10 and group 1 for deep learning model is 0.0685 (0.1513-0.0828), while the difference for linear regression model is 0.0438 (0.1336-0.0899), for the annualized buy-and-hold return between year *t* and year *t*+1. The difference for the deep learning model (0.0685) is larger than the one for the linear regression model (0.0438). When discussing the annualized buy-and-hold returns from year *t* to year *t*+2, the difference for deep learning model (0.0873-0.0159=0.0714) is still larger than that for linear regression model (0.0894-0.0234=0.0659), though the magnitude is much smaller than the one for one-year horizon (0.0714-0.0659=0.0055 for two-year horizon). For three-year horizon, though the result shows that the performance of the linear regression is slightly better than the deep learning, the magnitude is very small (0.0719-0.0711=0.0008). Thus, from the spread of group 10 and group 1, the deep learning

model can generate earnings forecasts which will contribute to better estimated ICCs on average for a relatively short term (one-year horizon), and for a relatively long term (two-year and threeyear horizon), earnings forecasts from the two models will generate ICCs with similar predicting power for the future realized return. Besides, comparing to the average returns of portfolios in linear regression case, the average returns for deep learning model are in a relatively monotonic increasing pattern, especially for one-year horizon, showing that the deep-learning-based ICC has a better predictive ability than the one from linear regression on average.

The table compares the averages of annualized equal-weighted average buy-and-hold returns across years for each decile. Firms in each decile are ranked by the composite ICC from deep learning and linear model at the end of June each year, and are reclassified on each June. Time-horizons include one-year, two-year, and three-year ahead, and the range of the sample is from 1968 to 2018.

	Panel A:Comp	osite Deep-Learn	ning-Based ICC	Panel B:Composite Linear-Regression-Ba				
Decile	$r_{t,t+1}$	$r_{t,t+2}$	$r_{t,t+3}$	$r_{t,t+1}$	$r_{t,t+2}$	$r_{t,t+3}$		
1	0.0828	0.0159	0.0072	0.0899	0.0234	0.0108		
2	0.0907	0.0433	0.0345	0.0952	0.0524	0.0472		
3	0.0931	0.0556	0.0518	0.1049	0.0634	0.0567		
4	0.1017	0.0610	0.0550	0.0978	0.0618	0.0529		
5	0.1099	0.0663	0.0612	0.0980	0.0605	0.0515		
6	0.1271	0.0722	0.0610	0.1047	0.0599	0.0511		
7	0.1152	0.0672	0.0587	0.1085	0.0633	0.0532		
8	0.1294	0.0710	0.0553	0.1146	0.0641	0.0542		
9	0.1329	0.0747	0.0632	0.1100	0.0638	0.0599		
10	0.1513	0.0873	0.0784	0.1336	0.0894	0.0828		
10-1	0.0685	0.0714	0.0711	0.0438	0.0659	0.0719		

4.4. Comparison of the Individual Deep Learning Model-based ICC and the Individual Linear Model-based ICCs

I follow the same principle behind the results of Table 4 to conduct Table 5 using the five types of individual ICCs from both the deep learning and the linear regression. In Panel A, I calculate the correlations between individual ICCs from deep learning model and linear regression. The correlations between the individual ICCs from the linear model are higher than the associated

Table 4: The Predictability Power of Composite ICC on Future Returns, Based on Deep Learning and Linear Model
correlations from deep learning. The correlations from the linear model are ranging from 0.6656 (OG and CT) to 0.9048 (Gordon and OJ), while the correlations from the deep learning model are from 0.2367 (OG and GLS) to 0.6083 (MPEG and CT). Also, considering the correlations between the deep learning based ICCs and linear model based ICCs, I find that excepting OJ, for all the other four methods, the correlations between different methods under the same deep learning model are lower than the correlations between the same method under the two different models. More specifically, the correlation between GLS under deep learning and GLS under linear model is 0.5298, which is higher than the correlations between GLS under deep learning and other methods under deep learning (from 0.2367 to 0.4261). The rest three methods are in the same pattern (0.7035 vs. from 0.4124 to 0.6038 for CT, 0.6435 vs. from 0.3160 to 0.6083 for MPEG, 0.6427 vs. from 0.3160 to 0.5588 for Gordon). The results indicate that, at least under deep learning model, the choice of ICC estimation method matters, and taking average of ICCs is necessary for the analysis in this study. Besides, the correlations between composite ICC and individual ICCs are higher than the correlations between individual ICCs. For the deep learning model, the correlations are from 0.5857 (between composite and GLS) to 0.8715 (between composite and OJ), while for the linear model, the correlations are from 0.8558 (between composite and CT) to 0.9259 (between composite and Gordon). The potential reason behind the improvement of correlations may be the smaller number of available individual ICCs observations comparing to composite ICC. The results show that the composite ICC can better represent the implied cost of capital in this research comparing to the five individual ICCs solely.

Panel B shows the results of the predictability power tests applied in Table 4 applying individual deep learning model-based ICCs, while Panel C shows the results using individual linear model-based ICCs. Only results of 10-1 spreads are included in the two panels. Comparing

the results from Panel B and Panel C, I find that for one-year period, except that of OJ method (0.0439 vs. 0.0476), the 10-1 spreads of the average returns for all other four individual deep learning model-based ICCs are larger than those associated spreads for the individual linear modelbased ICCs accordingly, showing that not only the composite ICC, but also individual ICCs obtained by using the deep learning model have an equivalent or stronger predicting power on future real returns. This result further indicates that the deep learning model-based ICCs are better proxies of expected returns comparing to the linear model-based ICCs. For two-year and threeyear horizons, the results are relatively mixed considering different individual ICCs. It shows that when evaluating the predicting power of ICC on expected returns for relatively longer horizons, specific methods matter for evaluating the performance of the deep learning model and the linear model. For example, when using GLS as the method of generating ICCs, the deep learning model will get ICCs with stronger predicting power on future realized returns for all the three horizons, while the ICCs from the linear model can outperform those from the deep learning model when applying MPEG as the method of calculating ICCs. The potential reasons behind include the lack of available estimates as well as the various structures of different methods. Thus, the composite ICC, taking the average of ICCs from the five methods and bringing the largest number of observations into the analysis, is more reliable in the analysis of the predicting power than individual ICCs.

# Table 5: Individual ICCs and Realized Returns

The table indicates the correlations between the individual implied cost of capital from different models and composite implied cost of capital (Panel A) and the means of the 10-1 return spreads across years related to the individual deep-learning-based ICCs (Panel B) and the individual linear-regression-based ICCs (Panel C). The correlation between any pair of ICCs is calculated using the same set of observations having available values for both ICCs. The sample period is from 1968 to 2018.

	1		Deep	Learning M	Model-base	ICC	1		L	inear Mod	el-based IC	C		
ICC estimate	28	GLS	$\mathbf{CT}$	OJ	MPEG	Gordon	Composite	GLS	CT	OJ	MPEG	Gordon	Composite	
Panel A: Correlatio	ns between l	ndividual	ICCs (1968	8-2018)										
Deep Learning Model	GLS	1.0000												
	$\mathbf{CT}$	0.4261	1.0000											
	OJ	0.2367	0.4124	1.0000										
	MPEG	0.4053	0.6083	0.4770	1.0000									
	Gordon	0.4057	0.5588	0.4896	0.3160	1.0000								
	Composite	0.5857	0.6857	0.8715	0.8405	0.7436	1.0000							
Linear Model	GLS	0.5298	0.5386	0.2612	0.5233	0.5218	0.5372	1.0000						
	CT	0.4659	0.7035	0.3508	0.5307	0.5866	0.4924	0.8359	1.0000					
	OJ	0.4350	0.5460	0.3720	0.4948	0.5958	0.5028	0.6980	0.6656	1.0000				
	MPEG	0.3647	0.5871	0.3919	0.6435	0.4738	0.5628	0.7758	0.8182	0.7474	1.0000			
	Gordon	0.4773	0.5353	0.3303	0.4432	0.6427	0.4910	0.7460	0.7203	0.9048	0.7146	1.0000		
	Composite	0.4439	0.5925	0.4256	0.6362	0.5874	0.6024	0.8650	0.8558	0.9189	0.9224	0.9259	1.0000	
ICC estimate	25	$r_{t,t+1}$		$r_{t,t+2}$		$r_{t,t+3}$		$r_{t,t+1}$		$r_{t,t+2}$		$r_{t,t+3}$		
		Panel B: 10-1 Spread, Individual Deep Learning Model-based ICCs (1968-2018)							Panel C: Mod	10-1 Sprea lel-based IC	Spread, Individual Linear ed ICCs (1968-2018)			
GLS		0.0445		0.0778		0.0790		0.0360		0.0637		0.0694		
CT		0.0708		0.0587		0.0517		0.0334		0.0441		0.0543		
OJ		0.0439		0.0535		0.0503		0.0476		0.0730		0.0765		
MPEG		0.0550		0.0381		0.0482		0.0477		0.0540		0.0594		
Gordon		0.0630		0.0824		0.0772		0.0540		0.0787		0.0809		

### 4.5. Realized Returns, ICCs, and Firm Characteristics

In previous asset pricing literature, researchers mainly treat realized returns as the proxy of expected stock return to test if certain characteristics of firms can explain the variation of expected returns. Since in the previous sections, I reach to the conclusion that deep-learning-based implied cost of capital is a better proxy of expected returns, we should expect that the company's characteristics which can explain the variation of realized returns should also be able to explain the variation of the deep learning model-based composite ICCs, if the variation of realized returns explained by those characteristics indeed represents the variation in expected returns.

In this section, following Hou et al. (2012), I include 13 firm-level characteristics to do the tests. For each year, Beta is the market beta computed by the past sixty monthly returns (24 months availability is the minimum) for each stock in June. Size represents the natural logarithm of market equity at the last day of June. Leverage stands for the ratio of book debt to book equity. NOA is the ratio of net operating assets to lagged total assets. BE/ME is the natural logarithm of the quotient of book equity and market equity at the end of last fiscal year. CAPEX is generated from dividing capital expenditure by lagged total assets. Idiosyncratic volatility represents the standard deviation of the residuals calculated by the last sixty monthly returns (24 months availability is the minimum) applying the market model in each year's June. Asset growth represents total assets growth rate. Accruals is the ratio of accruals to lagged total assets. Analyst coverage represents the number of analysts following a certain company in June. Analyst dispersion is the quotient of the standard deviation of analysts' estimates and the stock price of the company at the last day of June. Earnings smoothness is the ratio of earnings volatility to operating cash flow volatility generated by last ten years' data (five years availability is the minimum). Accruals quality is calculated according to a modified Jones (1991) method introduced in Francis, LaFond, Olsson, and Schipper (2005a), utilizing a cross-sectional regression of accruals on one, the change in sales revenue, and gross property, plant, and equipment (PPE), after all dependent and independent variables being divided by lagged total assets for scaling purposes. I do the regression for each Fama and French (1997) industry which has twenty or more observations in a certain year. Then, I use the parameter estimates from the regression to multiply one, the difference between change in sales revenues and change in accounts receivable, and PPE respectively, and scale them by lagged total assets, to calculate the firm-level normal accruals. Accruals quality is the absolute value of abnormal accruals, which is the difference between actual accruals divided by lagged total assets and normal accruals. Following the steps of Hou et al. (2012), I put an additional negative sign in front of the original earnings smoothness and accruals quality values, making these variables better represent the smoothness of earnings and accruals quality of firms.

I apply Fama-MacBech regressions of realized returns as well as the deep learning modelbased ICCs on the 13 firm-level characteristics to check if these characteristics can explain the variation of realized returns and ICCs. The sample period is from 1968 to 2018 for both realized returns and ICCs. The composite ICCs are estimated on June each year, and realized returns are the returns from July of the current year to June of the following year. In Table 6, I list the averages of coefficients across years, as well as time-series Newey-West t-statistics, for different regressions. Panel A presents the results from regressing realized returns on the 13 characteristics separately, while Panel B includes those from regressing composite ICCs on those characteristics.

# Table 6: Realized Returns, Composite ICCs of Deep Learning, and Firm Characteristics

This table shows the means of the coefficients across years (as well as the time-series Newey-West t-stat) generated by Fama-MacBeth regressions of firm-level annual realized returns or deep learning model-based composite implied cost of capital on different firm characteristics annually. Panel A is for realized returns, while Panel B and C are for ICC. Firm characteristics are defined following Hou et al. (2012). The sample period is from 1968 to 2018. \*, \*\*, \*\*\* indicate significance at 10% level, 5% level, and 1% level separately.

Beta	Size	BE/ME	Leverage	Idiosyncra volatility	tic CAPEX	Asset growth	Accruals	NOA	Analyst disper- sion	Analyst coverage	Earnings smooth- ness	Accruals quality
Panel A:	: Univariate	Regression	ns of Realiz	ed Returns	s on Firm C	haracterist	tics					
-0.003 -0.28	-0.003 -0.42	0.044*** 5.24	0.001 1.54	0.160 0.69	-0.078*** -3.14	-0.036** -2.65	-0.079** -2.51	-0.016 -0.62	0.290 1.18	-0.002 -0.97	0.016 0.76	0.009 0.10
Panel B:	Univariate	Regression	ns of Comp	osite Deep	Learning N	fodel-based	l ICCs on l	Firm Chara	cteristics			
0.005 0.67	-0.043*** -9.79	0.048*** 13.32	0.001*** 4.18	0.741*** 6.03	-0.083*** -4.13	-0.029** -2.60	-0.058*** -3.73	-0.130*** -3.33	0.803*** 6.92	-0.003*** -8.21	0.001 0.15	-0.108*** -3.88
Panel C:	Multivaria	te Regressi	ions of Con	nposite Mo	del-based I	CC on Firn	n Characte	ristics				
0.006 1.35 0.003 0.83 0.000 0.02 0.006 1.38 0.005 1.39 0.005 1.25 0.002 0.53 $0.009^{**}$ 2.25 $0.010^{**}$	-0.038*** -9.00 -0.035*** -8.54 -0.034*** -9.33 -0.037*** -8.92 -0.037*** -8.98 -0.037*** -8.93 -0.033*** -8.78 -0.013*** -6.07 -0.030***	$0.029^{***}$ 13.04 $0.033^{***}$ 15.73 $0.032^{***}$ 14.59 $0.028^{***}$ 12.80 $0.026^{***}$ 11.35 $0.029^{***}$ 13.46 $0.030^{***}$ 10.31 $0.024^{***}$ 9.02 $0.028^{***}$	0.003*** 3.79	0.203*** 4.01	-0.041*** -3.59	-0.031*** -4.19	-0.106*** -9.10	-0.144*** -6.15	0.725*** 7.50	0.003***		
2.52 0.050 1.32 0.005 1.20 0.006*** 3.41	-8.20 -0.037*** -8.32 -0.037*** -8.84 -0.015*** -7.85	9.04 0.028*** 10.71 0.030*** 12.84 0.023*** 11.47	0.004*** 4.14	-0.014 -0.41	-0.004 -0.20	-0.008 -1.60	-0.033* -1.75	-0.050*** -3.66	0.617*** 6.45	7.22 0.001*** 6.89	0.020 1.52 0.051 1.30	-0.044*** -4.60 -0.010* -1.97

The results from Panel A and Panel B can be categorized into three groups. First, both the realized returns and deep learning based ICCs have positive relations with five characteristics: BE/ME (0.044 for realized returns vs. 0.048 for deep learning based ICCs), leverage (0.001 for both realized returns and ICCs), idiosyncratic volatility (0.160 for realized returns vs. 0.741 for ICCs), analyst dispersion (0.290 for realized returns vs. 0.803 for ICCs), and earnings smoothness (0.016 for realized returns vs. 0.001 for ICCs). Within the characteristics above, BE/ME has significant coefficients for both realized returns (t-statistics of 5.24) and ICCs (t-statistics of 13.32), while leverage (t-statistics of 4.18), idiosyncratic volatility (t-statistics of 6.03), and analyst dispersion (t-statistics of 6.92) have significant coefficients in the regressions of ICCs only.

Second, both realized returns and composite ICCs have negative relations with six characteristics, including size (-0.003 for realized returns and -0.043 for ICCs), CAPEX (-0.078 for realized returns vs. -0.083 for ICCs), asset growth (-0.036 for realized returns vs. -0.029 for ICCs), accruals (-0.079 for realized returns vs. -0.058 for ICCs), NOA (-0.016 for realized returns vs. -0.130 for ICCs), and analyst coverage (-0.002 for realized returns vs. -0.003 for ICCs). Especially, size has a significant coefficient in the regression of ICCs (t-statistics of -9.79) only; CAPEX has significantly negative coefficients in regressions of both realized returns (t-statistics of -3.14) and ICCs (t-statistics of -4.13); asset growth has significant coefficients in the regressions of both realized returns (t-statistics of -2.65) and ICCs (t-statistics of -2.60); accruals is a significant independent variable for both regressions (t-statistics of -2.51 for realized returns vs. t-statistics of -3.73 for ICCs); NOA is significant only in the regression of ICCs (t-statistics of -3.33); analyst coverage (t-statistics of -8.21) has significant effect in explaining ICCs.

Third, two characteristics have coefficients with different signs in the regressions of realized returns and composite ICCs. For market beta, the coefficients are -0.003 in the regression

of realized returns and 0.005 in the regression of ICCs, but both the relations are not significant (tstatistics of -0.28 for realized returns vs. t-statistics of 0.67 for ICCs). Another one is accruals quality, with positive coefficient (0.009) in the regression of realized returns but negative coefficient (-0.108) in the regression of composite ICCs. Only the coefficient in the regression of ICCs is significant (t-statistics of -3.88).

From the results in Panel A and Panel B, at least three parts of conclusions can be reached. First, the firm characteristics having explanation power on expost realized returns in the previous research can explain the cross-sectional variation of ICC, and the coefficients of 11 characteristics are in the same directions, though accruals quality is one exception out of the 13 characteristics. Market beta is another exception, but there is little evidence that market beta effect exists on both realized returns and ICCs. This result provides indirect evidence that deep learning based composite ICCs can be a good proxy of expected returns, since ICCs and future realized returns of firms can be explained by the same set of characteristics and have coefficients with the same sign for most cases. Second, the 11 characteristics other than market beta and earnings smoothness have significant effects on explaining the variation of ICCs, while for realized returns, only BE/ME, CAPEX, asset growth and accruals have significant effects. Third, the market participants require expected returns for owning the stocks of firms with small size, high book-market ratio, high leverage, high idiosyncratic volatility, high analyst dispersion, low accruals, low capital expenditure, low asset growth speed, low analyst coverage, low NOA, or low accruals quality. Among firms with these features, those with high book-to-market ratio, low CAPEX or asset growth, or low accruals have large realized returns, but not for firms with other features.

Panel C includes regressions of deep learning ICCs on multiple independent variables simultaneously. The first regression has three independent variables: beta, size, and BE/ME. This

regression's results strengthen the findings from the regressions with individual independent variables in Panel B: market beta has a positive but not significant coefficient; the coefficient of size is significantly negative, and BE/ME has a positive and significant relation with deep-learning-based implied cost of capital.

For the following regressions, I regress ICCs on each of the rest characteristics after controlling beta, size, and BE/ME in each regression. In these regressions, comparing to Panel B, the sign of the coefficients in front of the independent variables stays the same except analyst coverage, and all the independent variables with significant results in Panel B still have a significant relation with ICCs even adding the three control variables, providing a further confirmation of the relations shown in Panel B. It indicates that the firm characteristics discussed in this section can explain the cross-sectional variation of deep learning based ICCs in a relatively consistent way. In the new regression, analyst coverage has a positive and significant coefficient instead of a negative and significant coefficient in the regression without control variables. In the end, I regress ICCs on all the 13 firm characteristics. After taking all characteristics into consideration, I find that the coefficients in front of idiosyncratic volatility and analyst coverage have an opposite sign comparing to the ones in Panel B, and the coefficients of idiosyncratic volatility (t-statistics of -0.41), CAPEX (t-statistics of -0.20), and asset growth (t-statistics of -1.60) become insignificant, while market beta has a significant coefficient with a t-statistics of 3.41.

In summary, the results of Table 6 indicate that the firm characteristics discussed can describe the cross-sectional variation of the realized returns and ICCs in a similar way but not in the same significant levels for certain characteristics.

### 5. ADDITIONAL EVIDENCE

### 5.1. Extra Predictive Power of Deep Learning on Earnings

In the previous section, I have discussed how deep learning can outperform linear models on forecasting earnings as well as estimating ICCs. In this section, instead of comparing models, I will link deep learning to analysts' forecasts for the first step. Hou et al. (2012) show that the cross-sectional linear regression approach can make better forecasts on earnings than financial analysts. However, since financial analysts have a different information set with the model when forecasting earnings, and some of the information may come from the sources other than financial statements, we are not sure if a better model based on publicly available information only can still provide extra predictive power on earnings once controlling the forecasts of financial analysts. Although the comparison between models and financial analysts is "unfair" due to different information sets, we are still curious that, once machines can have the same set of information, what will happen. More specifically, do machines have a stronger ability of information analysis than human beings?

Although based on the current level of science development, it is very hard to input all the information available for financial analysts to machines and finish the comparison above, we can check if machines can provide extra predictive power even relying on publicly available information only, a subset of the information set of financial analysts. In other words, if machines can do that, it means that machines do a better job than financial analysts on analyzing available information, which will give the incentive for combining machines and human beings' works together in the future.

In this section, based on publicly available information, I first apply deep learning in the forecast of Street earnings per share, which is the target value for analysts' forecasts in I/B/E/S, to

see if machines can provide extra predictive power to earnings when controlling analysts' forecasts. To simplify the operation and reduce the information gap between machines and analysts, different from the sections above, I include the top 25 commonly appeared variables from quarterly financial statements to construct the deep learning model forecasting earnings per share. In Aubry, Kräussl, Manso, and Spaenjers (2019), the authors estimate prices of artworks by machine learning to check if machines can provide extra explanation power for actual prices. Following their principles, I apply the following regression model to analyze earnings issues:

$$EPS_{i,t} = \alpha + \beta_1 EPS \widehat{analyst}_{i,t} + \beta_2 EPS \widehat{machine}_{i,t} + \epsilon_{i,t}$$
(4)

where  $EPS_{i,t}$  represents the actual earnings per share (EPS) for firm *i* in year *t*,  $EPSanalyst_{i,t}$  represents the EPS forecast for firm *i* in year *t* from financial analysts, while  $EPSanachine_{i,t}$  represents the EPS forecast for firm *i* in year *t* from deep learning model.

When constructing the sample for this analysis, I only consider the firm-year observations with both analysts' forecasts and deep learning model forecasts. Actual Street earnings per share  $EPS_{i,t}$  and analysts' forecasts are available from I/B/E/S. For the estimates by analysts, I only rely on the latest forecasts by analysts before the end of each fiscal year, so that the estimates in this analysis are the most accurate estimates in the belief of analysts before the end of the fiscal period. Since there may be multiple analysts focusing on the same firms, I use the mean of all forecasts by analysts on firm *i* in year *t* as  $EPSanalyst_{i,t}$ . When constructing the deep learning model, I use the most recent quarterly financial data items before the fiscal period ending date of actual EPS (at least one quarter before the fiscal period ending date).

Table 7: Efficiency of Deep-Learning-Based Earnings Per Share (EPS) The table includes the regression results of the equation:

 $EPS_{i,t} = \alpha + \beta_1 EPSanalyst_{i,t} + \beta_2 EPSnachine_{i,t} + \epsilon_{i,t}$ 

The dependent variable is the actual earnings per share of firm i on year t; the independent variables are analysts' EPS forecasts and deep learning model's EPS forecasts. The deep learning model includes the top 25 commonly appeared variables from quarterly financial statements. Analysts' EPS forecasts are the averages of the latest forecasts on firm i by analysts before the end of each fiscal year.

Dependent Variable: Actual EPS						
Variable	Parameter Estimate	t Value				
Intercept	-0.12618	-1.17				
EPSanalyst	0.99979	190.53				
$EPS \hat{machine}$	0.06444	11.70				

I regress firms' actual earnings per share on analysts' forecasts and the predictive EPS by deep learning model. The results in Table 7 show that, the coefficients for analysts' forecasts and deep learning's forecasts are both positive and significant, with t-statistics of 190.53 for analysts and t-statistics of 11.70 for machines, stating that the estimates by deep learning model can provide extra explanation power to the actual value of EPS after controlling the forecasts provided by analysts. The coefficient of the variable  $EPSanalyst_{i,t}$  is 0.99979 while the coefficient of the variable  $EPSmachine_{i,t}$  is 0.06444, showing that though financial analysts can provide relatively close estimates on EPS (or earnings) on average, machines can still extract information which is useful for the prediction of actual EPS (or earnings).

### 5.2. Comparison of Deep Learning and Analysts on Earnings Forecasts

In Hou et al. (2012), the authors compare the linear regression's earnings forecasts to analysts' forecasts, showing that the cross-sectional linear regression can outperform analysts' forecasts in forecasting earnings. Since the coverage of analysts' forecasts, linear model forecasts, and deep learning model forecasts are various, to avoid the loss of observations in the comparison process, I compare the deep learning model to the linear model separately in Section 4.2. In this section, I will compare deep learning model to analysts on earnings forecasts. I focus on the common firm-year observations appeared in both analysts' earnings forecasts and deep learning earnings forecasts, meaning that all the observations in the test should have earnings forecasts as well as actual earnings in both deep learning related format and analysts related format. Since the number of firms covered by financial analysts is smaller than the number of firms covered by deep learning model, the sample of doing this comparison, having a smaller sample size, is different from the sample I used in Section 4.2. Since the number of the three-year ahead earnings forecasts in I/B/E/S is relatively small, in this section, I only discuss one-year and two-year ahead earnings forecasts.<sup>2</sup>

I generate actual earnings as well as analysts' forecasts from I/B/E/S. Since there may be several analysts who are focusing on a certain firm, and analysts may not announce and update forecasts in each month, I assign the most recent announced forecast on this firm by this analyst for a certain month as the forecasts of this month. Analysts' forecasts for firms equal to the mean of forecasts from all analysts providing forecasts on the firms at the end of June, scaled by the price of stocks of the firms. For deep learning model, the methodology is the same as the one applied in Section 4.2, relying on the data items from annual financial statements. To make the values from analysts forecasts and those from deep learning comparable, earnings forecasts and actual earnings for deep learning are scaled by market equity at the end of June in each year.

<sup>&</sup>lt;sup>2</sup>In the previous research, people may also use the estimation of long-term growth rates in I/B/E/S and the two-year ahead forecasted earnings to compute the three-year ahead forecasted earnings to increase the available observations in the analysis if the three-year ahead earnings forecasts made by analysts are not enough.

Table 8: Comparison of Analysts' Forecasts and Deep Learning's Forecasts on Earnings This table compares the bias, accuracy and annual earnings response coefficient (ERC) of deep learning's earnings forecasts and analysts' earnings forecasts. Bias equals actual earnings minus earnings forecasts, divided by market equity at the end of June each year for deep learning and by the stock price at the end of June each year for analysts for scaling purposes. Accuracy refers to the absolute value of forecast bias. Annual ERC stands for the coefficient by regressing buy-andhold returns on forecast bias. To make annual ERCs of deep learning model and analysts comparable, I also standardize the forecast bias by one unit variance before the regression. All numbers are calculated in time-series averages. The sample is from 1983 to 2018. Panel A includes averages of mean as well as median of bias across years. Panel B includes averages of the mean as well as median of accuracy across years. Panel C includes the averages of annual ERC. The numbers under each average are the t-statistics.

		$E_{t+1}$			$E_{t+2}$	
	Deep Learning	Analysts	Difference	Deep Learning	Analysts	Difference
Panel A	: Bias of Earnings	Forecasts				
Mean	-0.0372 -5.86	-0.0151 -0.49	-0.0220 -0.72	-0.0163 -2.00	-0.0404 -1.39	0.0241 0.84
Median	0.0035 2.72	-0.0066 -4.86	0.0101 5.91	0.0033 1.05	-0.0267 -6.90	0.0300 8.23
Panel B	Accuracy of Ear	nings Forecasts				
Mean	0.1013	0.1016	-0.0003	0.1093	0.1347	-0.0254
Median	14.76 0.0303 29.57	3.14 0.0153 9.21	-0.01 0.0150 7.15	23.76 0.0449 33.55	4.12 0.0381 10.35	-0.75 0.0068 1.65
Panel C	Annual Earning	Response Coe	fficient (ERC)			
Mean	$0.0688 \\ 7.45$	0.0488 4.31	$0.0200 \\ 1.33$	$0.2123 \\ 11.56$	0.1332 5.98	0.0791 3.30

The results in Table 8 show that deep learning can provide less biased earnings forecasts than financial analysts. Though from the mean of the bias, for one-year ahead forecasts, deep learning seems to provide estimates with a higher level of bias, the potential reason behind is the influence of extreme values in the estimating process. From the median, deep learning provides less biased one-year ahead forecasts (0.0035) than financial analysts (-0.0066). For the two years time horizon, deep learning's forecasts are still less biased than forecasts of financial analysts, for both median (0.0033 vs. -0.0267) and mean (-0.0163 vs. -0.0404). More importantly, when comparing the annual earnings response coefficient, I find that the mean annual ERCs of deep

learning model (0.0688 for one-year horizon and 0.2123 for two-year horizon) are larger than those of analysts (0.0488 for one-year horizon and 0.1332 for two-year horizon), indicating that deep learning model can make earnings forecasts which meet the market expectation more. The result is consistent with the conclusion in Section 4.2 and Hou et al. (2012), showing that deep learning model can generate earnings estimates which are less biased and better fit the market expectation, comparing to the linear model as well as financial analysts.

### 5.3. Deep Learning Model with the Common Variables in the Linear Model

In Section 4.2, I apply a deep learning model with 38 variables and find that this model can outperform the linear model in Hou et al. (2012) in forecasting earnings. But the reason behind this result may also be the different amount of information used in each model, since the deep learning model used in Section 4.2 has more variables than the linear model. Thus, I construct the deep learning model with the same 6 variables used in the linear model, repeat the test in Section 4.2, and see if the deep learning model can still outperform the linear model based on the same amount of information. In Table 9, the deep learning model, with only six variables, has more common observations with the linear model in the analysis, leading to a larger set of firm-year observations. For forecasting one-year horizon earnings, deep learning has a bias of -0.0631 on average, while the linear model's bias is -0.0705 on average. The deep learning model is also less biased than the linear model in forecasting earnings two-year (-0.0072 vs. -0.0792) and three-year ahead (-0.0075 vs. -0.1043) on average. Averages of medians further confirm that earnings forecasts by deep learning model are less biased than those by the linear model (-0.0033 vs. -0.0042 for one-year ahead, 0.0070 vs. -0.0147 for two-year ahead, and -0.0029 vs. -0.0261 for three-year ahead). More importantly, deep learning approach has a larger annual ERC than the linear regression approach in all the three horizons (the differences are 0.0210 for one-year ahead earnings forecasts, 0.0592 for two-year, and 0.1120 for three-year), showing that the forecasts by deep learning model better reflect the expectation of the market on future earnings. With the same amount of information and a larger range of firm-year pairs in the sample, the deep learning model still outperforms the linear model, showing that deep learning is a better approach to predict earnings.

# Table 9: Comparison of Deep Learning and Linear Model, with the Same 6 Independent Variables

This table compares the bias, accuracy and annual earnings response coefficient (ERC) of deep learning's earnings forecasts and linear model's earnings forecasts. The deep learning here uses the same variables from the linear regression approach of Hou et al. (2012). Bias equals actual earnings minus earnings forecasts, divided by market equity at the end of June each year for scaling. Accuracy refers to the absolute value of forecast bias. Annual ERC is the coefficient by regressing buy-and-hold returns on forecast bias. To make annual ERCs of deep learning model and linear model comparable, I also standardize the bias by one unit variance before the regression. All numbers are calculated in time-series averages. The sample is from 1968 to 2018. Panel A includes averages of mean as well as median of bias across years. Panel B includes averages of mean as well as median of accuracy across years. Panel C includes averages of annual ERC. The numbers under each average are the t-statistics.

Ĩ		$E_{t+1}$	1		$E_{t+2}$			$E_{t+3}$	
	Deep Learning	Linear Model	Difference	Deep Learning	Linear Model	Difference	Deep Learning	Linear Model	Difference
Panel A:	Bias of Earn	nings Forecasts							
Mean	-0.0631 -5.58	-0.0705 -5.75	0.0074 0.64	-0.0072 -0.50	-0.0792 -5.96	0.0720 4.18	-0.0075 -0.44	-0.1043 -8.30	0.0968 6.40
Median	-0.0033 -1.23	-0.0042 -2.00	0.0009 0.33	0.0070 1.73	-0.0147 -4.10	0.0217 4.92	-0.0029 -0.46	-0.0261 -6.29	0.0232 4.80
Panel B:	Accuracy of	Earnings Foreca	sts						
Mean	0.1742 12.57	0.1801 12.58	-0.0059 -0.85	0.1784 17.36	0.1906 18.94	-0.0123 -1.34	0.2036 18.00	0.2125 20.89	-0.0090 -0.85
Median	0.0455 21.26	0.0427 20.90	0.0028 1.56	0.0618 23.42	0.0610 21.87	0.0008 0.31	0.0773 21.30	0.0753 22.12	0.0019 0.71
Panel C:	Annual Ear	nings Response C	Coefficient (ER	C)			110		1.1
Mean	$0.0711 \\ 10.41$	$0.0501 \\ 3.59$	0.0210 1.65	0.2133 11.36	$0.1541 \\ 7.36$	0.0592 2.43	0.3648 13.57	0.2528 10.02	0.1120 3.91

### 6. CONCLUSION

In this article, I demonstrate that a specific deep learning model, a deep neural network, can be used productively in the contexts of earnings forecasting and uncovering a firm's implied cost of capital. I show that deep learning techniques can lead to earnings forecasts that offer marginal information content above and beyond human analyst forecasts. Thus, machines taught to think in some ways like humans can, at least on some dimensions, extract information from observable data that human analysts either miss or process incorrectly. Of particular note, I show that my deep learning model leads to much less bias than what is exhibited by analysts. However, the machine learning approach is imperfect, as there is clearly some valuable predictive information detected by analysts, perhaps from outside the system, which is not fully captured by my model. Turning to the implied cost of capital (ICC), I show that the ICC derived from deep-learning earnings forecasts are better predictors of future realized returns than corresponding ICCs derived from linear-regression earnings forecast models of the type advanced by Hou et al. (2012).

My evidence indicates that combining analyst predictions and deep-learning techniques may lead to substantially superior forecasts than either approach used in isolation. In some sense, my study shows that observable data can be used much more effectively than the linear regression approach by unleashing the creativity of artificial intelligence and allowing it to search in an unstructured way for nonlinear predictive patterns. On the human side, research such as Loudis (2019) shows that analyst forecasts can be decomposed in a novel way that extracts useful information from the brain of an analyst while eliminating certain biases in their incentives or thought patterns. Hopefully, a combination of improved machine models along the lines of what I present, coupled with improved adjustments to human analyst predictions, will lead to far more accurate predictive models of earnings and a firm's implied cost of capital. Artificial intelligence techniques are improving at a rapid rate, so the prospect of substantial research progress along these lines in the near future appears high. BIBLIOGRAPHY

#### BIBLIOGRAPHY

- Abdou, H. A., S. T. Alam, and J. Mulkeen. 2014. Would Credit Scoring Work for Islamic Finance? A Neural Network Approach. *International Journal of Islamic and Middle Eastern Finance and Management* 7:112 – 125.
- Altman, E. I., G. Marco, and F. Varetto. 1994. Corporate Distress Diagnosis: Comparisons Using Linear Discriminant Analysis and Neural Networks (the Italian Experience). *Journal of Banking & Finance* 18:505 – 529.
- Amel-Zadeh, A., J.-P. Calliess, D. Kaiser, and S. Roberts. 2020. Machine Learning-Based Financial Statement Analysis. *Working Paper*.
- Ashbaugh-Skaife, H., D. W. Collins, W. R. Kinney, and R. Lafond. 2009. The Effect of SOX Internal Control Deficiencies on Firm Risk and Cost of Equity. *Journal of Accounting Research* 47:1–43.
- Aubry, M., R. Kräussl, G. Manso, and C. Spaenjers. 2019. Machine Learning, Human Experts, and the Valuation of Real Assets. *HEC Paris Research Paper No. FIN-2019-1332*.
- Ball, R., and R. Watts. 1972. Some Time Series Properties of Accounting Income. *The Journal of Finance* 27:663–681.
- Bao, Y., B. Ke, B. Li, Y. J. Yu, and J. Zhang. 2019. Detecting Accounting Fraud in Publicly Traded U.S. Firms Using a Machine Learning Approach. *Working Paper*.
- Bertomeu, J., E. Cheynel, E. Floyd, and W. Pan. 2019. Using Machine Learning to Detect Misstatements. *Working Paper*.
- Bianchi, D., M. Büchner, and A. Tamoni. 2019. Bond Risk Premia with Machine Learning. WBS Finance Group Research Paper No. 252.
- Bloch, D. A. 2019. Option Pricing with Machine Learning. Working Paper.
- Bonelli, F., S. Figini, and E. Giovannini. 2017. Solvency Prediction for Small and Medium Enterprises in Banking. *Decision Support Systems* 102.
- Botosan, C. A. 1997. Disclosure Level and the Cost of Equity Capital. *The Accounting Review* 72:323–349.
- Botosan, C. A., and M. A. Plumlee. 2002. A Re-Examination of Disclosure Level and the Expected Cost of Equity Capital. *Journal of Accounting Research* 40:21–40.

- Boubakri, N., S. El Ghoul, and W. Saffar. 2014. Political Rights and Equity Pricing. *Journal of Corporate Finance* 27:326 344.
- Boubakri, N., O. Guedhami, D. Mishra, and W. Saffar. 2012. Political Connections and the Cost of Equity Capital. *Journal of Corporate Finance* 18:541 559.
- Brogaard, J., and A. Zareei. 2018. Machine Learning and the Stock Market. Working Paper.
- Brooks, L. D., and D. A. Buckmaster. 1976. Further Evidence of the Time Series Properties of Accounting Income. *The Journal of Finance* 31:1359-1373.
- Brown, L. D., and M. S. Rozeff. 1978. The Superiority of Analyst Forecasts as Measures of Expectations: Evidence from Earnings. *The Journal of Finance* 33:1–16.
- Campbell, J. L., D. S. Dhaliwal, and W. C. Schwartz. 2012. Financing Constraints and the Cost of Capital: Evidence from the Funding of Corporate Pension Plans. *The Review of Financial Studies* 25:868–912.
- Chava, S., and A. Purnanandam. 2010. Is Default Risk Negatively Related to Stock Returns? *The Review of Financial Studies* 23:2523–2559.
- Chen, H., J. Z. Chen, G. J. Lobo, and Y. Wang. 2011. Effects of Audit Quality on Earnings Management and Cost of Equity Capital: Evidence from China\*. *Contemporary Accounting Research* 28:892–925.
- Chen, L., M. Pelger, and J. Zhu. 2019. Deep Learning in Asset Pricing. Working Paper.
- Claus, J., and J. Thomas. 2001. Equity Premia as Low as Three Percent? Evidence from Analysts' Earnings Forecasts for Domestic and International Stock Markets. *The Journal of Finance* 56:1629–1666.
- Dhaliwal, D., S. Heitzman, and O. Z. Li. 2006. Taxes, Leverage, and the Cost of Equity Capital. *Journal of Accounting Research* 44:691–723.
- Dhaliwal, D., J. S. Judd, M. Serfling, and S. Shaikh. 2016. Customer Concentration Risk and the Cost of Equity Capital. *Journal of Accounting and Economics* 61:23 48.
- Dhaliwal, D., L. Krull, and O. Z. Li. 2007. Did the 2003 Tax Act Reduce the Cost of Equity Capital? *Journal of Accounting and Economics* 43:121 150.
- Dhaliwal, D., L. Krull, O. Z. Li, and W. Moser. 2005. Dividend Taxes and Implied Cost of Equity Capital. *Journal of Accounting Research* 43:675–708.
- Ding, K., B. Lev, X. Peng, T. Sun, and M. A. Vasarhelyi. 2019. Machine Learning Improves Accounting Estimates. *Working Paper*.

- Donangelo, A. 2014. Labor Mobility: Implications for Asset Pricing. *The Journal of Finance* 69:1321–1346.
- Easton, P. D. 2004. PE Ratios, PEG Ratios, and Estimating the Implied Expected Rate of Return on Equity Capital. *The Accounting Review* 79:73–95.
- El Ghoul, S., O. Guedhami, Y. Ni, J. Pittman, and S. Saadi. 2013. Does Information Asymmetry Matter to Equity Pricing? Evidence from Firms' Geographic Location\*. *Contemporary Accounting Research* 30:140–181.
- Elton, E. J. 1999. Expected Return, Realized Return, and Asset Pricing Tests. *The Journal of Finance* 54:1199–1220.
- Fama, E. F., and K. R. French. 1997. Industry Costs of Equity. *Journal of Financial Economics* 43:153 193.
- Francis, J., R. LaFond, P. Olsson, and K. Schipper. 2005a. The Market Pricing of Accruals Quality. Journal of Accounting and Economics 39:295 – 327.
- Francis, J., D. Nanda, and P. Olsson. 2008. Voluntary Disclosure, Earnings Quality, and Cost of Capital. *Journal of Accounting Research* 46:53–99.
- Francis, J. R., I. K. Khurana, and R. Pereira. 2005b. Disclosure Incentives and Effects on Cost of Capital around the World. *The Accounting Review* 80:1125–1162.
- Froot, K. A., and J. A. Frankel. 1989. Forward Discount Bias: Is it an Exchange Risk Premium? *The Quarterly Journal of Economics* 104:139–161.
- García Lara, J. M., B. García Osma, and F. Penalva. 2011. Conditional Conservatism and Cost of Capital. *Review of Accounting Studies* 16:247–271.
- Gebhardt, W. R., C. M. C. Lee, and B. Swaminathan. 2001. Toward an Implied Cost of Capital. *Journal of Accounting Research* 39:135–176.
- Goh, B. W., J. Lee, C. Y. Lim, and T. Shevlin. 2016. The Effect of Corporate Tax Avoidance on the Cost of Equity. *The Accounting Review* 91:1647–1670.
- Gordon, J. R., and M. J. Gordon. 1997. The Finite Horizon Expected Return Model. *Financial Analysts Journal* 53:52–61.
- Guay, W., S. Kothari, and S. Shu. 2011. Properties of Implied Cost of Capital Using Analysts' Forecasts. *Australian Journal of Management* 36:125–149.
- Heaton, J. B., N. G. Polson, and J. H. Witte. 2017. Deep Learning for Finance: Deep Portfolios. *Applied Stochastic Models in Business and Industry* 33:3–12.

- Hou, K., M. A. van Dijk, and Y. Zhang. 2012. The Implied Cost of Capital: A New Approach. *Journal of Accounting and Economics* 53:504 – 526.
- Hwang, L.-S., W.-J. Lee, S.-Y. Lim, and K.-H. Park. 2013. Does Information Risk Affect the Implied Cost of Equity Capital? An Analysis of PIN and Adjusted PIN. *Journal of Accounting and Economics* 55:148 – 167.
- Jones, J. J. 1991. Earnings Management During Import Relief Investigations. *Journal of Accounting Research* 29:193–228.
- Kalev, P. S., K. Saxena, and L. Zolotoy. 2019. Coskewness Risk Decomposition, Covariation Risk, and Intertemporal Asset Pricing. *Journal of Financial and Quantitative Analysis* 54:335– 368.
- Khandani, A. E., A. J. Kim, and A. W. Lo. 2010. Consumer Credit-Risk Models via Machine-Learning Algorithms. *Journal of Banking & Finance* 34:2767 – 2787.
- Khurana, I. K., and K. K. Raman. 2006. Do Investors Care about the Auditor's Economic Dependence on the Client?\*. *Contemporary Accounting Research* 23:977–1016.
- Kim, Y. S., and S. Y. Sohn. 2004. Managing Loan Customers Using Misclassification Patterns of Credit Scoring Model. *Expert Systems with Applications* 26:567 573.
- Krishnan, J., C. Li, and Q. Wang. 2013. Auditor Industry Expertise and Cost of Equity. *Accounting Horizons* 27:667–691.
- Lee, C., D. Ng, and B. Swaminathan. 2009. Testing International Asset Pricing Models Using Implied Costs of Capital. *The Journal of Financial and Quantitative Analysis* 44:307–335.
- Lee, C. M. C., E. C. So, and C. C. Y. Wang. 2020. Evaluating Firm-Level Expected-Return Proxies: Implications for Estimating Treatment Effects. *The Review of Financial Studies*.
- Li, Y., D. T. Ng, and B. Swaminathan. 2013. Predicting Market Returns Using Aggregate Implied Cost of Capital. *Journal of Financial Economics* 110:419 – 436.
- Loudis, J. 2019. Expectations in the Cross Section: Stock Price Reactions to the Information and Bias in Analyst-Expected Returns. *Working Paper*.
- Messmer, M. 2017. Deep Learning and the Cross-Section of Expected Returns. Working Paper.
- Myers, L. A., M. S. Drake, M. T. Bradshaw, and J. N. Myers. 2012. A Re-examination of Analysts' Superiority over Time-series Forecasts of Annual Earnings. *Review of Accounting Studies* 17:944–968.
- Naiker, V., F. Navissi, and C. Truong. 2013. Options Trading and the Cost of Equity Capital. *The Accounting Review* 88:261–295.

- Ogneva, M., K. R. Subramanyam, and K. Raghunandan. 2007. Internal Control Weakness and Cost of Equity: Evidence from SOX Section 404 Disclosures. *The Accounting Review* 82:1255–1297.
- Ohlson, J. A., and B. E. Juettner-Nauroth. 2005. Expected EPS and EPS Growth as Determinants of Value. *Review of Accounting Studies* 10:349–365.
- Peng, Y., and P. H. M. Albuquerque. 2019. Non-Linear Interactions and Exchange Rate Prediction: Empirical Evidence Using Support Vector Regression. *Applied Mathematical Finance* 26:69–100.
- Pástor, u., M. Sinha, and B. Swaminathan. 2008. Estimating the Intertemporal Risk–Return Tradeoff Using the Implied Cost of Capital. *The Journal of Finance* 63:2859–2897.
- Sirignano, J., A. Sadhwani, and K. Giesecke. 2018. Deep Learning for Mortgage Risk. *Working Paper*.
- Wang, C. C. Y. 2017. Commentary on: Implied Cost of Equity Capital Estimates as Predictors of Accounting Returns and Stock Returns. *Journal of Financial Reporting* 2:95–106.
- Ye, T., and L. Zhang. 2019. Derivatives Pricing via Machine Learning. Boston University Questrom School of Business Research Paper No. 3352688.
- Zheng, C., X. Deng, Q. Fu, Q. Zhou, J. Feng, H. Ma, W. Liu, and X. Wang. 2020. Deep Learningbased Detection for COVID-19 from Chest CT using Weak Label. *medRxiv*.
- Zhou, B. 2019. Deep Learning and the Cross-Section of Stock Returns: Neural Networks Combining Price and Fundamental Information. *Working Paper*.
- Zhou, Z.-H., and J. Feng. 2017. Deep Forest: Towards an Alternative to Deep Neural Networks. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, pp. 3553–3559.

CHAPTER 2. Reputational Spillovers and Information Production by Analysts

### 1. INTRODUCTION

Psychological biases exist in individuals' daily activities, and people sometimes make judgments based on characteristics of others, including religion, age, race, etc. Among those characteristics, gender is a social topic with a long history and large influence. Different treatments due to gender exist not only in daily social life, but also in the business world. In the previous literature, gender bias is widely discussed, especially in the labor market. There is ample evidence that female employees face unequal treatments on many dimensions including compensation (Altonji and Blank, 1999) and job opportunities (Goldin and Rouse, 2000). In financial markets, specific areas with evidence of gender bias include mutual funds (Atkinson, Baird, and Frye, 2003; Niessen-Ruenzi and Ruenzi, 2019), hedge funds (Aggarwal and Boyson, 2016), venture capital (Gompers, Mukharlyamov, Weisburst, and Xuan, 2021), etc. In this paper, I investigate whether investors in financial markets treat financial analysts differently based on gender, especially when financial misconduct is revealed.

In the US, investors rely heavily on financial advisory services. In an effort to increase levels of information transparency, FINRA provides a search engine and databases of the historical misconduct records of financial advisers, or buy-side analysts. In this research, buy-side analysts consist of investment advisers (fund managers) and brokers, while sell-side analysts refer to financial analysts providing recommendations to investors or clients of a brokerage firm. Egan, Matvos, and Seru (2019), exploiting data from BrokerCheck, provide a full picture of financial misconduct by buy-side analysts, and Egan, Matvos, and Seru (2017) further discuss the "gender punishment gap" existing in the financial advisory industry, especially for buy-side analysts. They conclude that managers experience gender discrimination on punishment (taste-based or miscalibrated beliefs), which they characterize as "The Gender Punishment Gap." Female financial

advisers are more likely to receive harsher punishment from brokerages comparing to their male colleagues, and managers experience different treatments based on gender.

Following the discussion of gender discrimination in Egan et al. (2017), I consider possible explanations for "The Gender Punishment Gap." Clearly, financial market participants, investors and clients of financial advisory firms are very important for both buy-side analysts and brokerage firms. When making employment related decisions, brokerages may consider the "taste" of existing clients. Thus, a potential reason why brokerage firms treat male analysts and female analysts differently when misconducts happen is that investors or clients in the financial market have different opinions on misconducts by buy-side analysts with different genders, and brokerage firms follow the preference of investors to make the punishment decision. Thus, the question is, do investors/clients of brokerage have gender discrimination in face of misconducts? If so, does gender discrimination of clients/investors affect the business of brokerage firms, such as sell-side recommendations?

To answer of the questions, this paper conducts analysis to check if there is reputational spillover effect from the buy-side to sell-side within a brokerage, and more importantly, if investors have gender bias when facing the financial misconducts of buy-side analysts. I primarily focus on three main categories of misconduct: regulatory, criminal, and civil. I use Cumulative Abnormal Return (CAR) value around a sell-side analyst recommendation as a proxy of investors' trust on the sell-side analyst, and compare the average CAR value before the financial misconduct with that after the financial misconduct to see if there is any significant change on investors' reaction on recommendations. I find that after a buy-side analyst is associated with a financial misconduct event, there is a significant decrease on the absolute value of average CAR, indicating that the market reacts less to the recommendations from sell-side analysts who are in the same

brokerage, and financial misconducts of buy-side analysts negatively affect the reputation of their colleagues on the sell-side.

After identifying the reputational spillover effect, I use a popular gender-name database to classify the gender of all buy-side analysts with misconduct. I then repeat the reputation spillover tests for the two gender groups. Interestingly, when female buy-side analysts are associated with financial misconduct, investors appear to have a larger magnitude negative reaction, reflecting on the attitudes to the recommendations from the sell-side analysts in the same brokerage. The negative reputational impact on sell-side analysts is larger when a female buy-side analyst is related to the financial misconduct case. Compared with male analyst misconduct events, female analyst misconduct appears to destroy investors' trust of sell-side analysts from the same brokerage firm to a much greater extent.

This paper contributes to the literature in at least three ways. First, this is the first study that identifies the reputational connection between buy-side analyst and sell-side analyst. Prior studies primarily focus on active connections between buy-side analysts and sell-side analysts with benefit considerations, including information flows in both directions (Cici, Shane, and Yang, 2019; Irvine, Lipson, and Puckett, 2007). This paper demonstrates the existence of reputational spillover effect between the buy-side and sell-side, showing that misconduct events or illegal activities affect a firm's overall reputation, including employees with different job functions.

Second, this study provides empirical evidence that investors/clients of brokerage have gender bias in face of analyst misconduct. Misconducts of female analysts have a larger impact on the business of brokerage firms, especially recommendations. Different treatments exist not only within the brokerage firm, but also widely in the financial market. Under certain situations, different treatments by gender may be client-oriented or market preference oriented. A third contribution of this study is to provide evidence of alternative explanations on the discriminatory behavior by gender in labor markets. Gender bias in the labor market may not fully originate from the labor market alone, and the underlying reasons for such bias may come from the influence of other related areas. Other than gender bias of managers, gender bias of investors/clients can be a reason why brokerage treats analysts differently by gender. According to Gurun, Stoffman, and Yonker (2021), even with misconducts, buy-side analysts with solid connections with clients receive less severe punishment from the brokerage firms. To keep the assets of clients, brokerage firms may hesitate to fire buy-side analysts with misconduct activities. Thus, a connection between gender discrimination in labor markets and that in financial markets may exist.

The rest of the article is organized in the following structure. Section 2 is the literature review, while Section 3 shows the data and empirical methodology. In Section 4, I present the main results regarding the reputational spillover effect and gender difference. Section 5 concludes and discusses future potential research directions.

### 2. LITERATURE REVIEW

### 2.1. Financial Misconduct

Misconduct and fraud are major topics in finance and accounting research. Since misconduct is a broad research issue that may be connected to various areas in business, when discussing misconducts or frauds, researchers are often interested in the specific identities of individuals associated with misconduct. For example, targeting the executives of firms, Davidson, Dey, and Smith (2015) show that CEOs and CFOs with past legal troubles have a higher possibility of committing subsequent fraud. McNichols and Stubben (2008) investigate the investment behavior of firms manipulating earnings and show that, compared to others, firms with misconducts overinvest in the period of financial reporting manipulation. Since they are studying firms' real activities, those researchers consider firm-level misconduct events.<sup>3</sup> Researchers rely on different fraud or misconduct data to exploit the relation between financial misconduct and investors' behavior (Gurun, Stoffman, and Yonker, 2017; Law and Mills, 2019), investors' benefits (Piskorski, Seru, and Witkin, 2015), disclosures (Dimmock and Gerken, 2012), competitors (Bertsch, Hull, Qi, and Zhang, 2020), etc. This paper is targeting the financial misconduct of buy-side analysts, or financial advisers, contributing to the literature on misconducts, frauds, or crimes of financial industry employees. There are also other papers discussing the misconducts of buy-side analysts. Egan et al. (2019) draw a full picture of the misconduct behavior of financial advisers or buy-side analysts. They report that a considerable portion of buy-side analysts have a misconduct record, and some of them have offended the law more than once and have had to pay compensation to clients for their illicit behavior. Dimmock, Gerken, and Graham (2018) find that buy-side analysts working in the same location have an effect

<sup>&</sup>lt;sup>3</sup> Karpoff, Koester, Lee, and Martin (2017) summarize classic data sources and proxies of financial misconducts in the past literature.

on each other when it comes to misconduct. Using mergers of brokerages, researchers have found that financial advisers will have a higher possibility to commit financial misconducts if their new coworkers have a record of misconduct previously. Honigsberg and Jacob (2021) find that more than 10% of misconducts have been deleted from the records by buy-side analysts, and also report that analysts who attempted to eliminate misconduct records have a significantly higher possibility to be associated with new misconducts.

#### 2.2. Gender Difference

Diversity is an important topic in much current academic research. The role of gender in labor markets and business decisions has been a particular issue of much recent discussion given the historical differences in how the different genders have been treated in the workplace. This paper contributes to this nascent but important literature.

Unfair treatments arising from gender are formally discussed in work on discrimination theory. When discussing discrimination, researchers classify cases into two major categories: statistical (Phelps, 1972; Arrow, 1973), and taste-based or implicit (Becker, 1957; Bertrand, Chugh, and Mullainathan, 2005). In the previous literature, researchers find that different treatments by gender exist in financial markets, especially with regard to investment activities. Adams, Kräussl, Navone, and Verwijmeren (2021) provide strong evidence of a significant gender discount in the art market. Especially in the countries with higher level of gender inequality, art works by female artists sell for a significant lower price. Gafni, Marom, Robb, and Sade (2020) show that though female entrepreneurs have a higher level of success rate on crowdfunding platforms, people prefer to invest in projects of same-gender entrepreneurs, strongly suggesting the presence of taste-based discrimination. Using data from AngelList, Ewens and Townsend (2020) find that male investors are less interested in startups led by females, though similar situation male-led startups chosen by male investors are ultimately less successful.

Studies discussing in-firm gender bias or treatments on analysts are more relevant to this paper. Fang and Huang (2017) show that, though connections with corporate board members will benefit financial analysts, female analysts receive less positive effect from such connections compared to male analysts. The gender bias exists even people know each other very well. According to Duchin, Simutin, and Sosyura (2020), even within a firm, male CEO is distributing more resources or budgets to divisions with a male manager, and this gap of treatments is also related to the growing background and family structure of male CEOs. Distinct from the previous literature, this paper will discuss gender difference on the spillover effect between different job functions within a firm.

### 3. DATA AND EMPIRICAL METHODOLOGY

3.1. Data

The data I use related to buy-side analysts comes from IAPD (Investment Adviser Public Disclosure database) and FINRA BrokerCheck. I use similar variables to those selected by Egan et al. (2017).<sup>4</sup> The data includes misconduct years, locations, names, etc. The observations are at the year-analyst level, from 2007 to 2015. In this paper, I focus on three major types of misconducts: regulatory misconduct, criminal misconduct, and civil misconduct. Thus, I select all the misconduct observations within the three categories. Next, I manually search each buy-side analyst on IAPD and BrokerCheck, one-by-one, and find out the exact misconduct date from the reports for each misconduct observation. Figure 3 exhibits an example of the financial misconduct report. For gender information, I exploit the Genderchecker database. It is a database containing 102240 authenticated names worldwide, and each name is categorized as male, female or unisex.<sup>5</sup>

<sup>&</sup>lt;sup>4</sup>I am very grateful that Dr.Mark Egan and his co-authors kindly share the investment adviser data used in their research with me.

<sup>&</sup>lt;sup>5</sup>According to the introduction on the official website of Genderchecker, 7% of names in the database are classified as "unisex," indicating that these names are used for both genders. I exclude these cases from the analysis in this paper.

# Figure 3: An Example of the Financial Misconduct Report

Criminal	
This disclosure event involves dismissal, plea, acquittal or cor (e.g., bribery, perjury, forgery, or Disclosure 1 of 1	a criminal charge against the Investment Adviser Representative that has resulted in a nviction. The criminal matter may relate to any felony or certain misdemeanor offenses counterfeiting, extortion, fraud, wrongful taking of property).
Reporting Source:	Regulator
Formal Charges were brought in:	Federal Court
Name of Court:	UNITED STATES DISTRICT COURT
Location of Court:	SOUTHERN DISTRICT OF OHIO
Docket/Case #:	2:09-CR-00103-GLF
Charge Date:	04/23/2009
Charge(s) 1 of 2	
Formal Charge(s)/Description:	CIVIL FORFEITURE
No of Counts:	1
Felony or Misdemeanor:	Felony
Plea for each charge:	GUILTY
Disposition of charge: Charge(s) 2 of 2	
Formal Charge(s)/Description:	MAIL FRAUD
No of Counts:	1
Felony or Misdemeanor:	Felony
Plea for each charge:	GUILTY
Disposition of charge:	
Current Status:	Final
Status Date:	10/15/2009
Disposition Date:	10/15/2009
Sentence/Penalty:	66 MONTHS INCARCERATION, 3 YEARS SUSPENDED RELEASES, \$100 SPECIAL ASSESSMENT, \$2,663,681.44 RESTITUTION, NO FINE.

The sell-side analyst data is from I/B/E/S. Since there is no common indicator (such as cik, ticker) between the buy-side analyst database, which is from IAPD and BrokerCheck, and the sell-side analyst database, which is from I/B/E/S, the only way of combining the two databases is to manually match the names of the brokerage firms. On I/B/E/S, there is only a masked code reported

for each brokerage firm. Following the previous literature, I rely on two data sources to find out the true name of the brokerages in I/B/E/S. First, I use the translation table from I/B/E/S. This data contains a linkage between the marked code and the true name information before 2010.<sup>6</sup> Second, following the new method introduced in Gibbons, Hiev, and Kalodimos (2021), I manually collect the actual names of brokerage firms and sell-side analysts from a Bloomberg terminal. By matching the recommendations on Bloomberg and I/B/E/S, using the initials of sell-side analysts on I/B/E/S, I can find a link between I/B/E/S data and Bloomberg terminal data.

Once I obtain the true name information, I manually match the brokerage firm names by searching the common words in the brokerage names of both the buy-side database and the sell-side database. After creating a link table between buy-side names and sell-side names, I also search each name from both sides using google to confirm that the two names with common words refer to the same firm or the same company group. After creating the brokerage names linktable, I merge the buy-side analyst data to this linktable, and further expand the data to the misconduct level. I then match this to I/B/E/S recommendations from sell-side analysts under the same brokerage firm as each buy-side analyst. In the end, after cleaning the data, I obtain a set of 882 misconduct cases, 756 for male buy-side analysts and 126 for female buy-side analysts.

Other than analysts related data, I also use daily stock price data from CRSP for the calculation of CAR values.

### 3.2. Empirical Methodology

The sell-side analysts generally provide various recommendations. In this research, I classify the recommendations with IRECCD=1 or 2 as "buy recommendation" and categorize recommendations with IRECCD=4 or 5 as "sell recommendation." The subsequent discussion is

<sup>&</sup>lt;sup>6</sup>I am very grateful that Dr.Alexander Ljungqvist, Dr.Huihao Yan and their co-authors for kindly sharing the linkage data with me.
separated into the buy recommendation group and the sell recommendation group. When investors are disappointed, if reputational spillover effect exists, they may question the information quality of sell-side analysts, causing a smaller market reaction to these recommendations of analysts. To measure the market reaction to the recommendations of sell-side analysts, I calculate the Cumulative Abnormal Return (CAR) values on the sell-side analyst recommendation dates, setting 150 days as the length of the estimation period in trading days for the market model, 15 days as the gap between the estimation period and the beginning of the event window, and an event window of [-1 day,1 day] around each recommendation announcement date.

After obtaining CAR values on the recommendation dates, I set a three months window before and after the misconduct date and take the average of CAR values in the before and after windows of all recommendations provided by sell-side analysts who are from the same brokerage firm as the buy-side analyst accused of misconduct. I then compare the average CAR value before the misconduct with the average CAR value after the misconduct, and test if there is a significant decrease in the absolute value of average CAR. Since I use CAR values as a measure of investors' reaction to sell-side analysts' recommendations, a decrease in the CAR value represents a lower level of attention or agreement on the recommendation.

I use full sample with both male and female analysts first and divide the full sample into three categories based on the type of financial misconduct. For each category, I repeat the calculation of CAR values, average CAR values, and differences before and after the misconduct. I also do extra analysis to further explore the main evidence.

When comparing CAR averages, I do not pair brokerage firms or sell-side analysts between the two sides of the window and do not require that they exist on both sides of the window. The only commonality is that they are all related to a certain buy-side analyst who engaged in misconduct. Also, when calculating the average CAR value on each side, I only delete repeated CAR values on the same recommendation for a certain misconduct case, but not across all cases, since for different misconduct cases they may have common related sell-side recommendations. If I delete the repeated sell-side recommendations across all the misconduct cases, the weight of the CAR values on certain recommendations will be reduced.

### 4. RESULTS

### 4.1. Comparison of Average CAR Value Before and After Buy-Side Analysts' Misconducts

After obtaining CAR values within the [-3 months, 3 months] window around the buy-side misconduct date, I calculate the average of the CAR values on sell-side recommendations 3 months before and after the misconduct date separately. All the sell-side analysts are in the same firm or under the same group as the buy-side analyst who engaged in misconduct. Then, I use the average of all CAR values in the 3 months before the misconduct date minus the average of all CAR values in the 3 months after the misconduct date. By making this calculation, I obtain the reduction of magnitude of the reaction from investors, and check if such reduction/difference is significant or not.

Type		3 M	onths Before	3 N	Ionths After		
Regulatory	Buy	N 41565	Average CAR 0.018	N 43360	Average CAR 0.0175	Diff 0.000527	
	Sell	13512	-0.0413	15153	-0.0363	-0.00497*** (.000)	
Criminal	Buy	17372	0.0167	17101	0.0173	-0.00058	
	Sell	6802	-0.0364	6830	-0.0357	(.4017) -0.00068	
Civil	Buy	1256	0.0175	1216	0.0207	(.0453) -0.00317 (.2216)	
	Sell	516	-0.0366	435	-0.0281	-0.00855*	
All Categories	Buy	59614	0.0176	61034	0.0175	(.0930) 0.000129 (.7482)	
	Sell	20600	-0.0396	22225	-0.036	-0.00359*** (.000)	

Table 10: Comparison of Average CAR Value Before and After Buy-Side Misconducts

In Table 10, I calculate the average of difference between CAR values before the misconduct date and CAR values after the misconduct date to see if there is any significant change on the magnitude of market reaction to the sell-side analysts' recommendations. First, I divide all sample observations into three categories by the type of misconducts, and within each category. I will discuss buy recommendation and sell recommendation separately, because generally, buy and sell recommendations will cause market reaction in opposite directions. Within the three types of misconducts I focus on, regulatory misconduct category has the largest number of observations in the sample. In the three months period before the regulatory misconduct dates, there are 41,565 observations for buy recommendations, and 13,512 observations for sell recommendations, while within three months after the regulatory misconduct dates, there are 43,360 observations in the buy recommendation group, and 15,153 observations in the sell recommendation group. I find that for the buy recommendation group, there is no significant change in average CAR value, but for the sell recommendation group, the average CAR value changes from -0.0413 to -0.0363, with a -0.00497 difference. The change is significant, showing that there is a significant decrease in the market reaction to sell recommendations of analysts after the misconduct date. In the criminal misconduct and civil misconduct categories, while there is a decrease in market reactions to sell recommendations, they are not very significant. When taking the full sample into consideration, for the sell recommendation group, the average CAR value changes from -0.0396 to -0.036, with a significant change of -0.00359. These results indicate that investors react to the recommendations of related sell-side analysts less after the misconduct dates of buy-side analysts from the same brokerage, especially for sell recommendations.

The results from Table 10 provide evidence of the existence of reputational spillover effects between buy-side and sell-side analysts within the same brokerage firms. With these initial results

in hand, I move to the analysis of differences by gender. Based on the structure of the test in Table 10, I further divide the sample based on buy-side analysts' gender and repeat the analysis steps for males and females separately. If there are differences on the reduction of the market reaction by gender, it will show that investors or clients of brokerage firms exhibit gender-dependent differences in judgment in the face of financial misconduct.

			Miscondu	ıc <mark>ts</mark> by	Male Analyst	Misconducts by Female Analysts					
Type		3 M	onths Before	3 Months After			3 Months Before		3 Months After		
10.000		Ν	Average CAR	Ν	Average CAR	Diff	Ν	Average CAR	Ν	Average CAR	Diff
Regulatory	Buy	35173	0.018	35736	0.0178	(.7588)	6623	0.0186	7677	0.0166	0.00199* (.0626)
	Sell	11601	-0.0416	12440	-0.0374	-0.00413*** (.0012)	1989	-0.0384	2760	-0.0313	-0.00702*** (.0057)
Criminal	Buy	15053	0.0169	14377	0.0176	-0.00075 (.3229)	2157	0.0149	2560	0.0147	0.000175 (.9228)
	Sell	6034	-0.035	5602	-0.037	0.00202 (.1973)	736	-0.0481	1200	-0.0302	-0.0179*** (.0002)
Civil	Buy	1256	0.0175	1216	0.0207	-0.00317 (.2216)					Address of the
	Sell	516	-0.0 <mark>3</mark> 66	435	-0.0281	-0.00855* (.0930)					
All Categories	Buy	50903	0.0177	50686	0.0179	-0.00019 (.6756)	8780	0.0176	10237	0.0161	0.00155* (.0918)
	Sell	17921	-0.0392	18284	-0.0371	-0.00212** (.0313)	2725	-0.041	3960	-0.031	-0.00999*** (.000)

## Table 11: Comparison of Average CAR Value Before and After Buy-Side Misconducts, by Gender

Table 11 includes analysis by gender of buy-side analysts who are associated with financial misconducts. I follow the same principles applied in the analysis of Table 10, but divide the sample by gender. From Table 11, we know that the number of observations related to female buy-side analysts (8,780 for all categories) is around 17% of that related to male buy-side analysts (50,903 for all categories). There are several possible reasons for this difference. First, in the financial advisory industry, female analysts are fewer than male analysts. Also, according to Egan et al. (2017), female analysts are less likely to have financial misconducts in general comparing to male buy-side analysts. In addition, the diversity of different brokerage firms and the availability of data may impact the portion of female related observations in the sample. When checking the results by gender and the type of financial misconduct, I find that for both the male and female subcategories, when regulatory misconducts happen, there is a significant decrease in the market reaction to sell recommendations made by related sell-side analysts. However, for the male analysts sub-category, the average CAR value moves from -0.0416 before the misconduct dates to -0.0374 after the misconduct dates, with a difference of -0.00413, while for the female analysts sub-category, the average CAR value moves from - 0.0384 to -0.0313, with a -0.00702 difference. This indicates that, when regulatory misconducts happen, the negative reputational spillover effect from buy-side to sell-side is larger if the buy-side analysts involved are women. For criminal misconducts, there is no significant decrease for the male sub-category, but a significant decrease is evident for the female sub-category (from -0.0481 to -0.0302, and P-value is 0.0002). There is no female analyst observation related to civil misconduct observations. For the full sample, though decreases in market reaction appear when male buy-side analysts have financial misconducts, the magnitude is much smaller than that of female analysts when considering sell recommendations (-0.00212 vs - 0.00999). These results establish that, although decreases in market reaction exist for

both male analysts related cases and female analysts related cases, the magnitude of the decrease is larger for female analysts related cases, indicating that the reputational spillover effect discussed earlier is stronger when the financial misconducts are related to female buy-side analysts.

One potential explanation behind the identified gender difference is that investors have a higher expectation of the moral standards of female analysts, since in general, female employees are less likely to be involved in fraud or illegal cases. Some investors may choose female analysts for an expectation of reliable services and low level of fraud risk. Once a fraud related to a female analyst is detected, investors and clients sharply revise downward their expectations in the face of such unexpected bad news. Due to this disappointment, investors treat the misconduct more seriously, and more negative feedback applies to the brokerage and its services, causing a stronger negative reputational spillover effect.

Because female analysts are less likely to be involved in financial misconduct, once female buy-side analysts are associated with financial misconduct, based on the analysis, investors and clients of brokerage firms may lose their trust in the general environment of the brokerage firm and suspect that there may be additional undiscovered illegal activities occurring at the firm. In other words, investors may treat the financial misconducts by female buy-side analysts as a signal of dysfunctional moral environment, reducing overall trust in the brokerage firms. Moreover, financial misconducts of female buy-side analysts may also signal poor performance or unstable operations at brokerage firms, because bad performance may induce financial analysts to take on the risk of acting illegally. If female analysts are willing to engage in illegal activities, it may indicate that the situation of the brokerage firm is quite precarious as employees with highexpected levels of moral standards are willing to engage in these activities to survive.

### 4.2. Comparison in One Year Before and After the Misconduct Dates

In the previous section, when comparing genders, I find that a reputational spillover effect exists and that the gender of buy-side analysts affects the level of this effect. To buttress this evidence, I turn next to ruling out some alternative explanations for these results. One alternative is that the phenomenon I detect is periodic, which not only happens on misconduct dates, but also the same days in other years. Thus, I move all misconduct dates to one year ahead and one year after the true date, and redo the analysis in the previous section using the moved misconduct dates. Since the misconduct dates have been moved to new dates, the observations, or sell-side analysts' recommendations, within the [-3 months, 3 months] window around the new dates form a new sample. The principles behind the new analysis are still the same as those in Section 4.1.

Type		3 M	onths Before	3 N	Ionths After		
		N	Average CAR	N	Average CAR	Diff	
Regulatory	Buy	42662	0.0181	45233	0.0174	0.00075	
						(.1077)	
	Sell	13393	-0.0371	14734	-0.0356	-0.0015	
						(.1812)	
Criminal	Buy	17568	0.0181	17091	0.018	0.000093	
	a 11					(.8919)	
	Sell	6349	-0.0343	6077	-0.0389	0.00466***	
0::1	D	1000	0.0007	1570	0.0100	(.0052)	
Civil	Buy	1682	0.0207	1578	0.0192	0.00149	
	Call	577	0.0221	591	0.0406	(.5832)	
	Sell	511	-0.0551	021	-0.0400	(1085)	
All Categories	Buy	61189	0.0181	63214	0.0176	0.000543	
rin categories	Duy	01100	0.0101	00211	0.0110	(1572)	
	Sell	20120	-0.036	21091	-0.0366	0.000593	
		11.0				(.5211)	

Table 12: Moving All Misconducts Dates 1 Year Earlier

First, I test if the decreases on the market reaction, or the reputational spillover effect as reported in Table 10, still exist. Table 12 repeats the analysis in Table 10 using the new sample constructed by applying the new dates, which are generated by subtracting one year from all misconduct dates. For example, if the original misconduct date is 01/01/2013, then the new date after the move is 01/01/2012, and the three months window is around the new date for the analysis. There is no significant decrease on the absolute value of average CAR after the new dates for all categories. The only significant result is for the sell recommendation group in the criminal misconduct category, but different from the decreased reaction we are looking for, there is a positively significant change on the market reaction to sell recommendations. Thus, when using alternative (incorrect) dates I do not find similar results as those using the original (true) misconduct dates in Table 10.

Type		3 M	onths Before	3 N	Ionths After		
	100	N	Average CAR	N	Average CAR	Diff	
Regulatory	Buy	39873	0.0177	38458	0.0176	0.000043	
	Sell	13006	-0.0402	12992	-0.0392	(.9334) -0.00102 (.4033)	
Criminal	Buy	17246	0.0167	15834	0.0182	-0.00147*	
						(.0688)	
	Sell	7860	-0.0343	7100	-0.042	0.00775***	
0: 1	D	1004	0.0107	1010	0.0155	(.000)	
Civil	Buy	1284	0.0197	1312	0.0157	$(0.00401^{*})$	
	Sell	516	-0.0298	474	-0.037	(.0723) (.00712) (.1599)	
All Categories	Buy	57738	0.0175	54876	0.0178	-0.00028	
	Sell	21171	-0.0378	20322	-0.0402	(.5105) $0.00236^{**}$ (.0154)	

Table 13: Moving All Misconducts Dates 1 Year Later

\*, \*\*, \*\*\* indicate significance at 10% level, 5% level, and 1% level separately.

Next, I use new dates one year after the misconduct dates to construct another new sample and redo the analysis. According to the results reported in Table 13, there is no significant decrease in market reaction. On the contrary, the absolute values of average CAR significantly increase after the new dates. When considering criminal misconduct only, the average CAR value around sell recommendations changes from -0.0343 to -0.042, while for the full sample, the average CAR value decreases from -0.0378 to -0.0402, and both changes are significant. This indicates that, one year after the original financial misconduct dates, there is no significant decreasing trend in the absolute value of average CAR around sell-side analysts' sell recommendations. It also shows that the reputational spillover effect does not last for a relatively long term, because after one year, the decreasing trend of market reaction is not evident. Based on the results from Table 12 and Table 13, we can conclude that the results in Table 10 are not appearing on pseudo-event dates and that the decrease in market reaction only occurs immediately after the actual financial misconduct date.

After confirming the general spillover effect is related to the actual misconduct using the preceding placebo-date analysis, I next repeat the gender analysis in Table 11 using the alternative dates by moving the original misconduct dates one year forward and afterward. Table 14 tabulates evidence for the one year earlier dates. Dividing the sample into gender groups, I find that for male analysts analysis, there is only a weak (10% significance level) result for sell recommendations in the criminal misconduct category, following the similar pattern without considering gender in Table 12. However, for misconducts by female analysts, I find significant results in multiple groups. For regulatory misconduct, there is no significant result for buy recommendations, while there is a 5% significance level decrease in the absolute value of average CAR after the new dates. Also, for criminal misconducts of female analysts, a significant decrease in average CAR value exists for buy recommendations. Both sets of results indicate decreases in the market reaction. Moreover, for sell recommendations in the criminal misconduct category for female analysts', a significant increase in market reaction is evident. This indicates that, one year before the actual financial misconduct dates, investors start paying more attention to the sell recommendations by sell-side analysts who are in the same brokerage as female buy-side analysts with future criminal misconduct.

			Misconduc	ts by N	Iale Analysts		Misconducts by Female Analysts					
Type		3 Months Before			3 Months After		3 Months Before		3 Months After			
Regulatory	Buy	N 35559	Average CAR 0.0181	N 38198	Average CAR 0.0174	Diff 0.000759 (.1322)	N 7145	Average CAR 0.0181	N 7086	Average CAR 0.018	Diff 0.000026 (.9830)	
	Sell	11322	-0.0367	12248	-0.0359	-0.00083 (.4946)	2129	-0.0405	2498	-0.0345	-0.00606** (.0392)	
Criminal	Buy	15365	0.0176	14848	0.0184	-0.00081 (.2663)	2059	0.0204	2128	0.0145	0.00588*** (.0013)	
	Sell	5691	-0.0346	5424	-0.0379	0.00334* (.0604)	607	-0.0298	600	-0.046	0.0162*** (.0007)	
Civil	Buy	1682	0.0207	1578	0.0192	0.00149 (.5832)						
	Sell	577	-0.0331	521	-0.0406	0.00747 (.1985)						
All Categories	Buy	51883	0.018	539 <mark>36</mark>	0.0177	0.000282 (.4954)	9204	0.0186	9214	0.0172	0.00136 (.1809)	
	Sell	17391	-0.0358	17952	-0.0365	0.000753 (.4502)	2736	-0.0381	3098	-0.0367	-0.00146 (.5619)	

## Table 14: Moving All Misconducts Dates 1 Year Earlier, by Gender

			Miscondu	cts by I	Male Analysts		Misconducts by Female Analysts					
Туре		3 Months Before		3 Months After			3 Months Before		3 Months After			
Regulatory	Buy	N 33259	Average CAR 0.0177	N 32306	Average CAR 0.0177	Diff -0.00004 (.9427)	N 6744	Average CAR 0.0174	N 6333	Average CAR 0.0171	Diff 0.000272 (.8356)	
	Sell	10495	-0.0412	10788	-0.0392	-0.00194 (.1522)	2604	-0.037	2297	-0.0398	0.00273 (.3292)	
Criminal	Buy	15200	0.0166	13718	0.0178	-0.000117 (.1644)	1930	0.018	<b>1969</b>	0.0221	-0.00407 (.1447)	
	Sell	6864	-0.0334	6044	-0.0413	0.0079*** (.000)	977	-0.0398	1034	-0.0464	0.00666 (.1933)	
Civil	Buy	1284	0.0197	1312	0.0157	0.00401* (.0725)						
	Sell	516	-0.0298	474	-0.037	0.00712 (.1620)						
All Categories	Buy	49078	0.0175	46608	0.0177	-0.00024 (.5943)	8674	0.0176	8302	0.0183	-0.00077 (.5219)	
	Sell	17664	-0.0379	17062	-0.04	0.00206* (.0513)	3581	-0.0378	3331	-0.0418	0.00405 (.1032)	

# Table 15: Moving All Misconducts Dates 1 Year Later, by Gender

In Table 15, I repeat the analysis by gender using new dates generated by adding one year to the original financial misconduct dates. I find no evidence of a significant (5% level or better) difference, except that for the male analysts' criminal misconduct category, there is an increase on the market reaction to sell recommendations, which is the same directional finding as reported in Table 13.

By combining the results from Table 14 and Table 15, we can conclude that the stronger reputational spillover effect for misconducts by female analysts is not a seasonal phenomenon with a regular timeline. Although in the one year earlier period, there is significant reduction of market reaction to sell recommendations in the female analysts' regulatory misconduct category and to buy recommendations. For the results in regulatory misconduct category, we still reach the same overall conclusion. For the results in regulatory misconduct category, the magnitude of the decrease is larger and more significant for the real financial misconduct dates (-0.00702 with a 0.0057 P-value, vs. -0.00606 with a 0.0392 P-value), and the absolute value of average CAR after the misconduct dates reach to a bottom level (0.0313), smaller than for the one year earlier test (0.0345), showing that though the trends are in the same direction, the original analysis in Table 11 generates more powerful results on the spillover phenomenon. For the criminal misconduct dates on this specific group, the results for this group cannot challenge the evidence for a spillover effect in the other categories and groups.

### 4.3. Comparison of Non-Misconduct Related Recommendations

In Section 4.1 and 4.2, for each misconduct date and each buy-side analyst with misconduct, within the three months window, I focus on recommendations provided by sell-side analysts from the same brokerage. In this section, for each buy-side misconduct case, I select all

the recommendations which are targeting the same firms within the same windows around the misconduct dates in Section 4.1, but from sell-side analysts who are not from the same brokerage as the misconduct buy-side analyst. I repeat the analysis on all other unrelated recommendations, and compare the results with the ones from Table 10 and Table 11 to see if the results on the reputational spillover effect and the gender difference only exist for misconduct related analysts and recommendations rather than more generally.

I start with the sample of financial misconduct and collect all sell-side analysts' recommendations within the [-3 months, 3 months] window around the buy-side misconduct dates. Then, for each misconduct case, I identify all recommendations within the window which are not related to this misconduct buy-side analyst and his/her brokerage, but are targeting the same set of firms as the misconduct related recommendations. After constructing the sample, I calculate the average CAR values and the difference, repeating the analysis process in the previous sections. When calculating the average CAR value for the recommendations of non-related sell-side analysts, I focus on each specific misconduct case, but I do not delete his/her recommendations can be from non-related analysts.

Type		3 Mo	nths Before	3 Me	3 Months After				
Regulatory	Buy	N 1124070	Average CAR 0.0171	N 1112159	Average CAR 0.0172	Diff -0.00011 (.3260)			
	Sell	203430	-0.0308	196170	-0.0297	-0.00115*** (.0035)			
Criminal	Buy	403742	0.017	390257	0.0177	-0.00063*** (.0005)			
	Sell	72664	-0.0308	69535	-0.0321	0.00136**			
Civil	Buy	22678	0.0163	19426	0.0156	0.000679 (.3559)			
	Sell	4244	-0.0302	3482	-0.0237	-0.00648** (.0376)			
All Categories	Buy	1534172	0.0171	1504038	0.0173	-0.00023** (.0120)			
	Sell	277098	-0.0308	266020	-0.0303	-0.00052** (.01212)			

Table 16: Comparison of Average CAR Value of Brokerage Without Misconducts

\*, \*\*, \*\*\* indicate significance at 10% level, 5% level, and 1% level separately.

Table 16 contains results for the non-misconduct related recommendation sample, without considering gender. There are significant market reaction decreases in certain categories. For regulatory misconducts, the average CAR value around sell recommendations increases from - 0.0308 to -0.0297, with a significant reduction in the absolute value. Compared to the results for the misconduct related sample in Table 10, the magnitude of reduction is only 1/4 to 1/5 of the reduction for misconduct related recommendations (-0.00115 vs -0.00497). For civil misconducts, the decrease on the absolute value of average CAR around sell recommendations is 0.00648, which is smaller than the one in original misconduct related sample (0.00855). When considering all categories, the reduction in the absolute value of average CAR is only 0.00052 for non-misconduct related sell recommendations, while for misconduct related sell recommendations, the reduction is 0.00359 and has a higher level of significance. Thus, although certain categories of the non-

misconduct related recommendations sample share a similar reduction in market reaction as those in the original misconduct related sample, the magnitude is much smaller, thus not fully explaining the reduction of the absolute value of average CAR for misconduct related observations.

In addition to the results concerning same trend directions, Table 16 reveals evidence concerning different directions that may strengthen the case for the existence of reputational spillover effects. For example, while in the original analysis in Table 10, there is no significant change for the criminal misconduct category (both buy recommendations and sell recommendations), in Table 16, a significant increase in the absolute value of average CAR is evident. When market reactions to non-misconduct related recommendations rises, investors appear to be keeping the original level of attention on misconduct related recommendations. This indicates that the market reaction on misconduct related recommendations decreases in a relative sense, providing indirect evidence that buy-side misconducts negatively impact the reputation of sell-side analysts from the same brokerage.

			Miscondu	cts by N	Iale Analysts	Misconducts by Female Analysts					
Type		3 Months Before			3 Months After		3 Months Before		3 Months After		2
Regulatory	Buy	N 931350	Average CAR 0.0171	N 920734	Average CAR 0.0172	Diff -0.0009	N 193160	Average CAR 0.017	N 192825	Average CAR 0.0171	Diff -0.0001
	Sell	169344	-0.0308	165089	-0.0298	(.4233) -0.00097** (.0221)	32819	-0.03	32278	-0.0284	(.6830) -0.0016 (.1022)
Criminal	Buy	3558 <mark>3</mark> 8	0.0172	343148	0.0177	-0.00059*** (.0022)	41328	0.0149	41259	0.0158	-0.00089* (.0775)
	Sell	65646	-0.0311	62589	-0.032	0.000944 (.1741)	6074	-0.0273	5616	-0.0321	0.00478** (.0142)
Civil	Buy	22678	0.0163	19426	0.0156	0.000679 (.3559)					
	Sell	4244	-0.0302	3482	-0.0237	-0.00648** (.0376)					
All Categories	Buy	1294339	0.0171	1E+06	0.0173	-0.00022** (.0282)	234385	0.0167	234028	0.0169	-0.00024 (.2958)
	Sell	236164	-0.0309	228139	-0.0304	-0.00048 (.1861)	38893	-0.0296	37894	-0.029	-0.00064 (.4725)

## Table 17: Comparison of Average CAR Value of Brokerage Without Misconducts, by Gender

Table 17 compares average CAR value for non-misconduct related recommendations by gender separately. When adding gender into analysis, the results can directly or indirectly prove the existence of a spillover effect and a gender difference when misconducts occur. First, for the subsample related to female buy-side analysts' misconducts, criminal misconduct is the only category having significant results, showing that there is no evidence of changes in market reaction to recommendations not related to female misconduct analysts for most groups, emphasizing the evidence of reputational spillover effect on female side in Table 11. Second, there is only one 5%-or-higher-level significant result, namely a 0.00478 increase in the absolute value of average CAR around sell recommendations, indicated for females in Table 17. Comparing with the same category in Table 11, which has a significant 0.0179 decrease on the absolute value of average CAR after misconduct dates, the result for the non-misconduct related sample analysis goes in the opposite direction, further strengthening the reputational spillover interpretation related to female analysts reported in Table 11.

When non-misconduct related recommendations exhibit an increased market reaction, the decrease in the market reaction in the misconduct related recommendations strengthens the conclusions from Table 11. While the significance levels are different, the differences of average CAR for male and female analysts in Table 17 are small and similar in value, especially for the full sample comparison. For the male case, the differences of average CAR are -0.00022 for buy recommendations and -0.00048 for sell recommendations, while for the female case, the differences of average CAR are -0.00064 for sell recommendations. The male and female cases in Table 17 have very similar results, and, compared to Table 11, which are the results for misconduct related recommendations, the results for all other recommendations are very small. This strongly suggests that no gender difference exists for the

market reaction to non-misconduct related recommendations.

### 5. CONCLUSION AND FUTURE POTENTIAL DIRECTIONS

In this article, I investigate the reputational spillover effect between buy-side analysts and sell-side analysts, and examine whether gender difference affects the magnitude of the spillover effect within the same brokerage firm. After the revelation of misconduct by buy-side analysts, investors react less to recommendations of sell-side analysts in the same brokerage, especially to "sell recommendations." These results indicate the existence of a substantive reputational spillover effect between the two types of analysts. More surprisingly, female buy-side analyst misconducts have a larger negative impact on sell-side analysts from the same brokerage, indicating that investors/clients have different reactions to different genders in the face of misconducts. By comparing these findings with a non-misconduct related sample and samples based on placebo dates, I rule out the potential that my results are not specifically reflective of actual misconduct and the gender of the underlying analysts. The evidence I detect on gender appears to reflect differing expectations of the genders when it comes to following various rules and regulations, and it may have implications for understanding various puzzles regarding the organization of the market for analysts.

My study also builds on Gibbons et al. (2021) by collecting rich data from Bloomberg to illustrate that researchers can obtain detailed demographic characteristics of sell-side analysts including name, position, education, location of firm, employment history, etc. By adding these details into the analysis, researchers may be able to examine the role of salient characteristics that have not been previously studied and/or controlled for. For example, researchers can further group the sample by the gender of sell-side analysts, to see if reputational spillover effect between buyside analysts and sell-side analysts has an "in group" effect. When buy-side analysts and sell-side analysts have the same gender, will the spillover effect be of a different magnitude? After

considering the education level and employment history, will investors have a lower level of gender bias on these financial advisory employees, especially when they get involved in illegal activities?

BIBLIOGRAPHY

#### BIBLIOGRAPHY

- Adams, R. B., R. Kräussl, M. Navone, and P. Verwijmeren. 2021. Gendered Prices. *The Review* of *Financial Studies* URL https://doi.org/10.1093/rfs/hhab046. Hhab046.
- Aggarwal, R., and N. M. Boyson. 2016. The Performance of Female Hedge Fund Managers. *Review of Financial Economics* 29:23–36.
- Altonji, J., and R. Blank. 1999. Race and Gender in the Labor Market. In O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics*, vol. 3, Part C, 1st ed., chap. 48, pp. 3143–3259. Elsevier. URL https://EconPapers.repec.org/RePEc:eee:labchp:3-48.
- Arrow, K. 1973. The Theory of Discrimination. In O. Ashenfelter and A. Rees (eds.), *Discrimination in Labor Markets*, pp. 3–33. Princeton University Press.
- Atkinson, S. M., S. B. Baird, and M. B. Frye. 2003. Do Female Mutual Fund Managers Manage Differently? *Journal of Financial Research* 26:1–18.
- Becker, G. 1957. *The Economics of Discrimination*. Chicago & London: University of Chicago Press.
- Bertrand, M., D. Chugh, and S. Mullainathan. 2005. Implicit Discrimination. *The American Economic Review* 95:94–98.
- Bertsch, C., I. Hull, Y. Qi, and X. Zhang. 2020. Bank Misconduct and Online Lending. *Journal of Banking and Finance* 116:105822.
- Cici, G., P. B. Shane, and Y. S. S. Yang. 2019. Do Connections with Buy-Side Analysts Inform Sell-Side Analyst Research? *Working Paper*.
- Davidson, R., A. Dey, and A. Smith. 2015. Executives' "off-the-job" Behavior, Corporate Culture, and Financial Reporting Risk. *Journal of Financial Economics* 117:5–28. NBER Conference on the Causes and Consequences of Corporate Culture.
- Dimmock, S. G., and W. C. Gerken. 2012. Predicting Fraud by Investment Managers. *Journal of Financial Economics* 105:153–173.
- Dimmock, S. G., W. C. Gerken, and N. P. Graham. 2018. Is Fraud Contagious? Coworker Influence on Misconduct by Financial Advisors. *The Journal of Finance* 73:1417–1450.
- Duchin, R., M. Simutin, and D. Sosyura. 2020. The Origins and Real Effects of the Gender Gap: Evidence from CEOs' Formative Years. *The Review of Financial Studies* 34:700–762.

- Egan, M., G. Matvos, and A. Seru. 2019. The Market for Financial Adviser Misconduct. *Journal* of Political Economy 127:233–295.
- Egan, M. L., G. Matvos, and A. Seru. 2017. When Harry Fired Sally: The Double Standard in Punishing Misconduct. *Working Paper* 23242, National Bureau of Economic Research.
- Ewens, M., and R. R. Townsend. 2020. Are Early Stage Investors Biased Against Women? *Journal* of Financial Economics 135:653–677.
- Fang, L. H., and S. Huang. 2017. Gender and Connections among Wall Street Analysts. *The Review of Financial Studies* 30:3305–3335.
- Gafni, H., D. Marom, A. Robb, and O. Sade. 2020. Gender Dynamics in Crowdfunding (Kickstarter): Evidence on Entrepreneurs, Backers, and Taste-Based Discrimination\*. *Review of Finance* 25:235–274.
- Gibbons, B., P. Hiev, and J. Kalodimos. 2021. Analyst Information Acquisition via EDGAR. *Management Science* 67.
- Goldin, C., and C. Rouse. 2000. Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians. *American Economic Review* 90:715–741.
- Gompers, P. A., V. Mukharlyamov, E. Weisburst, and Y. Xuan. 2021. Gender Gaps in Venture Capital Performance. *Journal of Financial and Quantitative Analysis* p. 1–29.
- Gurun, U. G., N. Stoffman, and S. E. Yonker. 2017. Trust Busting: The Effect of Fraud on Investor Behavior. *The Review of Financial Studies* 31:1341–1376.
- Gurun, U. G., N. Stoffman, and S. E. Yonker. 2021. Unlocking Clients: The Importance of Relationships in the Financial Advisory Industry. *Journal of Financial Economics*.
- Honigsberg, C., and M. Jacob. 2021. Deleting Misconduct: The Expungement of BrokerCheck records. *Journal of Financial Economics* 139:800 831.
- Irvine, P., M. Lipson, and A. Puckett. 2007. Tipping. *The Review of Financial Studies* 20:741–768.
- Karpoff, J. M., A. Koester, D. S. Lee, and G. S. Martin. 2017. Proxies and Databases in Financial Misconduct Research. Accounting Review 92:129 – 163.
- Law, K. K. F., and L. F. Mills. 2019. Financial Gatekeepers and Investor Protection: Evidence from Criminal Background Checks. *Journal of Accounting Research* 57:491–543.
- McNichols, M., and S. R. Stubben. 2008. Does Earnings Management Affect Firms' Investment Decisions? *The Accounting Review* 83:1571–1603.

- Niessen-Ruenzi, A., and S. Ruenzi. 2019. Sex Matters: Gender Bias in the Mutual Fund Industry. *Management Science* 65:3001–3025.
- Phelps, E. S. 1972. The Statistical Theory of Racism and Sexism. *The American Economic Review* 62:659–661.
- Piskorski, T., A. Seru, and J. Witkin. 2015. Asset Quality Misrepresentation by Financial Intermediaries: Evidence from the RMBS Market. *The Journal of Finance* 70:2635–2678.