

Group Identity and Agency Frictions: Evidence using Big Data*

Jitendra Aswani [†]

Abstract

This paper examines whether similarity in social identities between a manager and the board affects executive compensation, firm value, and agency frictions. By using a novel dataset on surnames with multiple identities (native language, native place, and caste), developed by merging micro census data of 474 million Indians with data from Linguistic Survey of India (LSI), I provide evidence that the firms with a shared group identity between a manager and the board do well compare to other firms. Due to in-group favoritism, managers of such firms earn higher compensation. These results are more substantial for group identity based on native language and native place. I also find that the firm benefits from taking on the cost of in-group favoritism because, for such firms, the manager's reciprocation toward the firm reduces the agency frictions and increases firm value in the long run. These results are robust to the endogeneity test, managerial influence on firm, college ties, ties from past employment, and various other checks.

Keywords: Group Identity, In-group Favoritism, Managerial Compensation, Firm Value

JEL Codes: C92, M12, D21, D22

*Jitendra Aswani: jaswani@fordham.edu

I am deeply grateful to Kevin Donovan, Nadya Malenko, Thomas Philippon and Sanjai Bhagat for the discussions and detailed feedback. I also thank Mahazarin Banaji (Harvard Behavioral Lab), Mahesh Srinivasan (UC Berkeley Behavioral Lab), and Jay Van Bavel (NYU Behavioral Lab) for providing guidance on social psychology and social neuroscience literature. I also appreciate comments from Hersh Shefrin, Alex Butler, Ross Levin, Raymond Fisman, Ganet Gao, Cian Ruane, Vishan Gandhi Nigam (Discussant), Sitian Liu, Gaurav Sood, Tristan Fitzgerald, Yeejin Jang, Svenja Dube, Elizabeth Blankespoor, Darya Yuferova, Sris Chatterjee, Iftexhar Hasan, N.K. Chidambaran, Gayane Hovakimian, Hongyu Shen, Md Moshi Ul Alam, and seminar participants at Ph.D. Colloquium of Financial Management Association (FMA, 2020), Australasian Finance & Banking Conference (2020) at UNSW, Royal Economic Society Annual Meeting 2021, North American Summer Meeting of the Econometric Society (2021) at Montréal, Asian Meeting of the Econometric Society (2021) at Malaysia, and North East Universities Development Consortium (NEUDC) Meeting 2021. The author acknowledges the help from Fordham University for cloud computing.

[†]Author is a Ph.D. student at Gabelli School of Business, Fordham University and can be contacted at jaswani@fordham.edu.

1. Introduction

Recent social movements in the US related to racial identity (such as ‘Black Lives Matter’) and gender identity (such as ‘He for She,’ ‘Women’s March’ and others), demanding for equality in society, reignited the debate on diversity not only on news media outlets and public forums but in the boardrooms too. In Fortune 500 companies, white executives fill more than 80 % of board seats (Financial Times, 2020).¹ With only about 17% of board seats worldwide held by women, improvement in gender diversity at the corporate board level remains slow (WSJ, 2020).² The same pattern is prevalent in UK too (Guardian, 2019).³ The corporate leaders who have championed diversity and inclusion are under pressure to address the homogeneity of the teams they have assembled around them (Financial Times, 2021). This reflects two essential perspectives about identity dynamics in the firms. First, it matters in corporate decisions. Second, how does homogeneity in the boardroom worked for so long, and how does it affect the endogenous decision of executive compensation and firm value? This paper provides the answers to these questions.

Standard neoclassical models in finance posit that corporate decisions are made virtually in a vacuum and independently of agents’ social identities. However, since the seminal work of Akerlof and Kranton (2000), there has been increased interest in social identity research in economics and finance. Social identity models widely applied in understanding the discrimination at the workplace (Akerlof and Kranton, 2000; Bertrand and Mullainathan, 2004), the economics of poverty, social exclusion, and household division of labor (Akerlof

¹<https://www.ft.com/content/c47c9b63-26c0-4487-94a9-4db7da585fbd>

²<https://deloitte.wsj.com/cmo/2020/03/06/gender-disparity-persists-among-corporate-directors/>

³<https://www.theguardian.com/business/2019/dec/04/top-uk-firms-failing-to-increase-boardroom-diversity-study-shows>

and Kranton, 2000), contract theory (Akerlof and Kranton, 2005), economic development (Basu, 2006), public goods provision (Croson, Marks, and Snyder, 2008), team identity at the workplace (Eckel and Grossman, 2005), banking (Fisman, Paravisini, and Vig, 2017), and investment decisions (Morse and Shive, 2011; Grinblatt and Kelohraju, 2001). But there is limited, theoretical, and empirical, research on how similarity of social identity (or group identity) plays a role in corporate decisions, which are endogenous, inter-linked, and complex. This work attempts to fill that gap by providing theoretical perspectives and empirical analyses on how shared group identity between a manager and the board can affect the managerial compensation, firm value, and agency frictions. The current study is inspired by recent debate on diversity in the boardroom and massive literature on identity economics, social psychology, social neuroscience, and their applications in finance.

The paper begins by providing an intuition on how similarity of social identities between a manager and the board members can change the endogenous decisions of executive compensation and firm value. The agent’s utility function of Akerlof and Kranton (2005) in Holmstrom’s (1979) model shows that the manager gets higher compensation due to the in-group bias and that firm value increases as manager exert additional effort for the firm due to the same reason.

As theoretical predictions can differ from empirical findings, it is crucial to verify whether predictions sustain empirical analyses. Social psychology and neuroscience studies have shown that favoritism or discrimination based on natural identities exists and is strong than induced identities⁴ and it is detected in frequency magnetic resonance images (fMRIs) and through implicit association tests (IATs).⁵ But most of these findings are based on lab

⁴According to Huettel and Kranton (2012), humans’ identities can be categorized into natural and induced. The identities such as gender, race, caste, place of birth, native language, religion, and similar others are considered natural identities. The identities such as college alumni, club membership, experience-based, and others are categorized as induced identities.

⁵The IAT relies on test-takers’ speed of response to represent the strength of their unconscious mental associations. IATs are used to measure a wide range of implicit attitudes about social groups, products, or self-identity (Greenwald et al. 1998).

experiments and, therefore, hard to scale for cross-industry analysis. To work around this shortcoming, I followed Freshtman and Gneezy’s (2001) empirical design and chose a setting where somebody can discern identities easily due to well-established stereotypes based on them. In India, a surname (i.e., a family name or last name) can provide information such as family lineage, native language, place of origin, and caste ⁶, about a person. As a result, surnames are likely an important basis of social ties and is a source of favoritism or discrimination. The details for using surnames as a source of social identities is discussed in section 3.1, but the main reason is that people belong to a particular surname lives in tight kinship and provide effective cooperation to the group members, while people outside the group are distrusted (Alesina and Giuliano 2013; Moscona, Nunn, and Robinson 2017).

A novel dataset on surnames mapping to their native language, caste, and native place in India is developed to check how group identity based on social traits can affect executive compensation, agency frictions, and long-term firm value. I use these social identities because implicit stereotypes or biases are stronger for these social identities and one can easily detect these identities from a surname. As there is no readily available database to provide managers’ and directors’ identities, a surname dataset developed using the information of 474 million Indians from the Socio-Economic Caste Census (SECC) and the Linguistic Survey of India (LSI) data to know the identities of managers and directors. Section 3.2 provides the details of this dataset. These identities related to surname form the identities

⁶According to ancient literature, the caste is mainly divided into four - the Brahmins (priestly people), the Kshatriyas (also called Rajanyas, who were rulers, administrators, and warriors), the Vaishyas (artisans, merchants, tradesmen, and farmers), and Shudras (laboring classes). People who do not come in any of these castes are called ‘Harijans’ (untouchables). Brahmins, Kshatriyas, and Vaishyas are collectively called upper caste and, Shudras and Harijans are called lower caste. After establishing the Indian constitution in 1950, caste is redefined based on reservation provided to the under-privileged in government jobs, education, and other services. As per this new categorization, Brahmins, Kshatriyas, and Vaishyas are called ‘other’ caste, Shudras as ‘scheduled tribes’ (ST), and Harijans as ‘scheduled caste’ (SC). The caste in Socio-Economic Caste Census (SECC) data is according to new categorization. I also include caste as an identity because caste remains an enduring predictor of economic status in India. It is correlated with occupation and employment (Prakash, 2015; Ito, 2009; Thorat and Attewell, 2007), income and expenditure (Deshpande, 2000), and capital more generally (Kijima, 2006).

of the principal (board) and an agent (manager).⁷

By discerning the identities of managers and directors, homophily (or group identity) variables are computed based on native language, native place, and caste. To examine the role of homophily in executive compensation, firm value, and agency frictions, I use panel data of 2,324 non-financial firms listed on two leading Indian stock exchanges - Bombay Stock Exchange (BSE) and National Stock Exchange (NSE) from 2004-2018. The main empirical results show that the manager gets 6-8 % higher compensation if the manager speaks the same native language or belongs to the same native place as the majority of the board members. It is construed as a cost of in-group favoritism borne by the shareholders. I also find that the firm value (measured as Price to Book ratio) is 11-12% higher for such firms than their counterparts. Even if the cost of in-group favoritism is excluded, the residual firm value (a proxy for manager's reciprocation to in-group favor) is 8-9% higher for such firms. As managerial compensation and firm value are endogenously determined; therefore, there is a possibility that residual firm value do not capture the simultaneity issue completely. To mitigate such concerns, following Palia (2000), I also confirm the results using the two-stage least square (2SLS) method. The latter results imply that the manager exerts extra effort due to shared group identity, which leads to a reduction in agency frictions. These results are robust to firm fundamentals and corporate governance characteristics. The result that the individuals put the extra effort for in-group members compared to out-group members is in line with the literature on group identity in social psychology and social neuroscience (Cikara and Bavel, 2014; Allport, 1954).

To check the homophily measures' external validity, I reconstructed the identity variables based on an alternative dataset from "Indianchildnames.com". This website features the most extensive collection of names (roughly 125,000) and surnames (roughly 15,000) of Indian

⁷The surname dataset provides a composition of all the Indian surnames at the state level, language level, and caste level. As there is no significant intra-country movement of any community in India except Kashmiri pandits in 1990, this surname database would provide managers and directors' correct identities.

and international origins. The results are robust to these alternative measures of native place and native language. The total compensation and Price to Book ratio are replaced with compensation relative to the industry peers and sales growth to assess dependent variables' external validity. The results are consistent; the manager gets higher compensation due to in-group favoritism, and group identity reduces agency frictions and increases firm value.

As the board can select a manager who has the same identity to reduce the frictions in corporate decision making (Damaraju and Makhija, 2018), there can be an endogeneity issue. Therefore, following Fracassi and Tate (2012), I use the director's death or retirement as a shock to the board's identity. I still find results that are consistent with the main results. One can argue that a director's death or retirement may or may not change the board identity and hence, the homophily constructs. To mitigate such concerns, an additional test is conducted. Rather than using the death or retirement of a director as a shock, change in homophily due to it is considered a shock; otherwise, not. The additional test results suggest that the coefficient of language and the place homophily are positively associated with managerial compensation, firm value, and residual firm value; it confirms the paper's main findings.

I also test for parallel conjectures such as managerial influence on firm, college ties, or past employment ties between a manager and board members. Adams et al.(2005) show that firm has higher performance variability if the CEO has greater power to influence decisions. Therefore, I control for 'powerful CEO' measures of Adams et al.(2005) and show that homophily results still hold. Farcassi and Tate (2012) show that a powerful CEO is more likely to appoint known directors, and these CEO-director ties reduce firm value. From the network connections used in Farcassi and Tate (2012), the most common are past employment, and the least common are education and current employment. To mitigate the concerns that results are not driven by ties from past employment or college ties, I show results controlling the measures based on these ties. The homophily results are still robust.

The coefficient of ‘powerful CEO’ measures and past experience ties are in line with Adams et al. (2005) and Farcassi and Tate (2012).

In India, the business groups are collections of publicly traded firms with a presence in different industries and a significant amount of common ownership and control, usually by a family. Tarun and Palepu (2000) compare the performance of business group firms with the non-business group firms of India, and show that business group firms outperform non-business group firms. They provide the channel that the group-affiliated firms have disproportionately good access to international sources of capital due to group’s reputation. The current study proposes an alternative channel for better performance of group affiliated firms’ i.e., manager’s reciprocation towards a firm due to homophily. I find that although managerial compensation due to in-group favoritism is higher in non-business group firms, but the residual firm value is higher in business group firms. It signifies that managerial reciprocation toward a firm due to homophily is higher in group-affiliated firms compare to others.

Apart from the external validity tests, checking parallel conjectures, testing the role of group identity in business group firms versus non-business group firms, I also conduct battery of robustness checks. Like, following Bhagat and Bolton (2008), I use current profitability and future profitability as dependent variables. I replace individual homophily constructs with principal component based homophily index. I also examine the impact of geographical variation in homophily constructs on compensation, firm value, and residual firm value.

Lastly, I discuss the two possible channels - trust and reduced information asymmetry, for the decrease in agency frictions if a manager and the board share similar social identities. By analyzing comparative papers, I also explain why homophily based on some identities affect firm decisions (such as selecting the manager (or director), executive compensation, job assignments, etc.) and firm performance (or value) but not all.

This paper is close to Cai et al.(2013), which investigated how family ties with firm owners

affect managerial compensation and job assignment in Chinese private firms. Cai et al.(2013) show that family managers earn higher salaries and, receive more bonuses, hold higher positions, and are given more decision rights and job responsibilities than non-family managers in the same firm. However, unlike Cai et al. (2013), this paper shows that favoritism due to in-group bias (not familial bias) also leads to higher managerial compensation. The firms also benefited from the bias as it increases the firm value and reduces agency frictions. This paper can also be seen in accord with Fracassi and Tate (2012), which shows that CEO-director ties based on shared experiences reduce firm value in the absence of other governance mechanisms to substitute for board oversight. I show that shared identities between CEO and board increase firm value under strong governance, such as high promoters(founders)holding.

In sum, this paper shows that shared group identity between a manager and the board members can increase executive compensation due to in-group favoritism, and the firms with group identity have a higher value than other firms. Furthermore, the residual firm value (i.e., the firm value conditional on executive compensation of the last year) is higher for such firms, implies that the manager puts extra effort when there is a shared group identity between him and the board. These results indicate that although in-group favoritism is a cost, bearing this cost is beneficial to a firm as it reduces agency frictions and can increase long-term firm value through the manager’s reciprocation towards the firm.

Structure of the Paper: The rest of this paper proceeds as follows. Section 2 discusses the contribution to the related literature. Section 3 provides the institutional details. Section 4 provides a theoretical model on how the similarity of a manager’s identity with the board can change managerial compensation and firm value. Section 5 explains the datasets and the variables used in the analysis. Section 6 reviews the methodology used for surnames matching and the empirical specification used for the analysis. Section 7 provides the perspective on endogeneity concerns and the identification strategy used to mitigate those concerns. Section 8 discusses the results. Finally, Section 9 presents the concluding remarks.

2. Related Literature and Contribution

Psychologists have long posited a self or “ego” as a primary force of individual behavior. They have further associated an individual’s sense of self to social categories, and individuals identify with people in some categories and differentiate themselves from others (Akerlof and Kranton, 2010). Research in social psychology has shown that people categorize themselves and others based on trivial criteria (Tajfel et al., 1971). This categorization can be very fluid and often context-dependent (Turner et al., 1994). For instance, Steele and Aronson (1995) find that African-American students, who are stereotyped as weak students, underperformed relative to white students when they are told that the test is diagnostic of their abilities. In a similar kind of experiment, Levy (1996) find that older people perform worse on a memory task if they have previously been primed with a negative stereotype of aging. Using gender as an identity, Carlana (2019) shows that implicit stereotypes in teachers regarding gender induce girls to underperform in math and self-select into less demanding high-schools, following their teachers’ recommendation. This form of implicit biases due to identities is likely to be the primary driver of prejudice, stereotyping, discrimination, and in-group favoritism (Crisp and Hewstone, 2001). According to the social psychology literature, if identity remains in the sub-conscious mind due to strong priming effect and legitimized by groups of people in society, then identity gives a birth to a stereotype related to that identity.

The recent literature in economics and finance have shown that how in-group favoritism affects the decision making of loan officers (Fisman, Paravisini, and Vig, 2017), sell-side analysts (Jannati et al. 2019), the board of governors in M&A deals (Bhalla et al. 2018), employers (Becker, 1957; Bertrand and Mullainathan, 2004; Banerjee et al. 2009, Cornell and Ivo, 1996; Charles Sarsons, 2019), retail investors (Morse and Shive, 2011; Grinblatt and Kelohraju, 2001), mutual fund investors (Kumar et al. 2015), auditors (Cunnigham et

al. 2018; Du, 2017), students & instructors (Fairlie, Hoffmann, and Oreopoulos, 2014), and venture capitalist syndicates (Gompers et al. 2016). But the literature is silent on what role social identities play in the board room and the capability of the identities to change executive compensation, agency frictions, and long term firm value. This paper seeks those answers.

Contribution: This paper contributes to the literature on agency problem, executive compensation, corporate governance, and the role of identity economics in corporate finance.

Modern finance theory argues that the objective of managers is to maximize the value of the firm, which in general means running the firm in the shareholders' interests. But managers are self-interested individuals. Because managers have effective control over the firm, they can make decisions and take actions that serve their interests (Jensen and Meckling (1976), Aggarwal and Samwick (1999)). The problem that surfaced due to the separation of owners and managers is known as 'agency problem'. Adam Smith forecasted in his work 'The Wealth of Nations' that if an organization is managed by a person or group of persons who are not the real owners, then there is a chance that they may not work for the owners' benefit. Berle and Means (1932) later supported this concern in their thesis, where they analyzed the ownership structure of the large firms of the U.S. and observed that agents appointed by the owners control large firms and carry the business operations. They argued that the agents might use the firm's property for their end, which will create the conflict between the principals and agents. Jensen and Meckling (1976) discusses the cost of this conflict (also called as 'agency cost') for the firm's securities owners'.⁸

There are different theories on agency conflict and its existence. Arrow (1971) and Wilson (1968) postulate that agency conflict occurs due to risk-sharing between the principal and an agent who have opposite risk preferences. The principal or the owners invest their

⁸In contrast, Fama (1980) suggested that this arrangement of separation of ownership and the control is efficient for the firm. The incentives for a manager and the monitoring effect would be provided by the market forces such as competition.

capital and expect to gain economic benefits from it, whereas the agents are risk-averse and concerned with maximizing their private benefits. Ross (1973) regarded the agency problem as the problem of incentives, while Mitnick (1975) considered the problem occurs due to the institutional structure, but the central idea behind their theories is similar. Grossman and Hart (1983) explained that the principal’s consumption gets affected by the agent’s output. The agent’s level of effort affects the firms’ output, where the principals desire for a higher level of effort from agents. Hence, the principal should trade-off the agent’s behavior with a proper payment structure. The incentive structure is affected by the agents’ attitude towards the principals’ risk and information quality, and no incentive problem arises if the agent is risk-neutral. The current study proposes that the similarity of social identities between the principal and an agent can reduce agency frictions. In that respect, this work contributes towards the literature on behavioral agency theory, which is propounded by Wiseman and Gomez-Mejia (1998), Sanders and Carpenter (2003), and Pepper and Gore (2012).

Due to the agency problem, there is considerable debate among both academics and practitioners on executive pay and how it affects the firm value. According to a literature review by Edmans, Gabaix, and Jenter (2017), there are three broad perspectives on executive compensation - “shareholder view”, “rent extraction view”, and “institutional view”. The “shareholder view” suggests that compensation contracts are chosen to maximize value for shareholders, taking into account the competitive market for executives and the need to provide adequate incentives. The “rent extraction view” argues that executives themselves set contracts to maximize their rents. Lastly, the “institutional view” explains the role of institutional forces, such as regulation, tax, and accounting policies in the executive compensation setting. The rent extraction argument starts with the observation that, in practice, executive pay is set by the board of directors and its compensation committee. This creates another agency problem, as directors on the compensation committee have their agenda and may have incentives to favor executives (Edmans, Gabaix, and Jenter (2017)). In theory,

market forces, including the market for corporate control, capital markets, product markets, and the managerial labor market, impose constraints on how much value destruction directors (and executives) can allow from rent extraction. In reality, constraints from market forces can be loose and permit large deviations from efficient contracting (Bebchuk, Fried, and Walker, 2002). There is limited theoretical work in this area, and empirical papers are mainly related to pay for non-performance (such as Pay-for-luck, severance pay, pay for acquisitions, option repricing, incentive rigging, or hedging), hidden pay (option compensation, backdating options, perks, pension, performance-based equity, or severance pay) and the association of certain practices with poor corporate governance. The current study connects the “rent extraction view” with the “shareholder view”. It suggests that rent extraction by a manager (hidden pay due to similar social identities as the board members) should not be considered as a cost to the firm if the manager increases the firm value more than the compensation paid.

Corporate governance and its effectiveness plays a big role in executive compensation (Core, Holthausen, and Larcker, 1999; Fahlenbrach, 2009; Armstrong, Ittner, and Larcker, 2012). Researchers have checked the impact of board size, number of independent directors, gray directors, and staggered board on executive compensation. But the identities of the board members are also essential. Delis et al.(2017) show that the impact of genetic diversity in the country of origin of the firms’ board members on corporate performance and find that adding a board of directors from countries with different levels of genetic diversity (either higher or lower) can increase firm performance. On the contrary, using a similar kind of data, Giannetti and Zhou (2019) show that greater ancestral diversity may lead to inefficiencies in the decision-making process and conflicts in the boardroom, and also increases the firm’s performance volatility. Using a sample of US firms and gender as identity, Adams and Ferreira (2009) find a negative impact of having females on the board on firm performance, despite better attendance records and more effective monitoring in firms with more gender-

balanced boards. On the contrary, Gregory-Smith et al. (2014) find no evidence that the gender composition of the board affects firm performance. These papers imply that diversity in the boardroom is still a mootable topic. Core, Holthausen, and Larcker (1999) showed that firms with weaker governance structures have greater agency problems and CEOs at such firms receive greater compensation, and firms with more significant agency problems perform worse. One part of this paper examines whether similarity in social identities between a manager and the board reduces agency problems or not.

Lastly, this paper also contributes to the literature on identity economics in corporate finance. Identity economics literature mainly discusses how identities can change outcomes due to implicit biases (Bertrand et al., 2005) or taste-based discrimination (Becker, 1971). Identities can be divided into two types - induced (educational institute, past employer, same club etc.) or by birth (gender, native place, native language, race, caste, etc.). The implicit biases and stereotypes are far stronger in the latter case. Like, Fisman et al. (2018) show how a candidate's hometown ties with fellow selection committee members can increase the selection probability in Chinese Academies of Sciences and Engineering by 39%. Fisman, Paravisni, and Vig (2017) show how group identity between the loan officer and the borrower based on the identities like religion and caste increases credit access and loan size dispersion, reduces collateral requirements, and induces better repayment even after the in-group officer leaves. Similarly, related to in-group favoritism and its impact, Gompers et al. (2016) show that venture capitalists who share the same ethnic, educational, or career background are more likely to syndicate with each other, and this homophily reduces the probability of investment success, and the detrimental effect is most prominent for early-stage investments. As the role of homophily in an organization, the theoretical work of Moisson and Tirole (2020) explores the trade-off between meritocracy and homophily in the selection and promotion process in the organization setup and checks the entrenchment and welfare properties of an organization. Unlike Moisson and Tirole (2020), I omitted the meritocracy of a manager

in the theoretical setting, however, in my empirics, I control for manager’s education as a proxy of meritocracy for the sub-sample. This paper extends the ongoing work on the role of identities in an organization and provides a theoretical perspective and empirical evidence on how identities change managerial compensation, firm value, and agency frictions.

3. Institutional Details

To examine the role of homophily based on native language, native place, and caste in executive compensation and firm value, India provides a perfect setting due to cultural diversity and persistent caste-based discrimination in the labor market.⁹ Because Indian last name provides these details about identities; therefore, this makes it easier for someone to discriminate or provide favorable behavior.

Although Indian cultural diversity strengthens this research’s prominence, there are some institutional differences between Indian firms compared to firms from developed markets such as the United States. Compared to developed markets, where institutional context is characterized by well-functioning capital, labor, and product markets, in emerging markets, such as India, China, or Brazil, various market failures are caused by information and agency problems (Khanna and Palepu (2000)).¹⁰ Perhaps this makes it more important to know whether a shared identity between manager and board of directors reduces the wedge between them and affects the agency frictions or not. Another significant institutional difference between firms in emerging markets from developed markets is that, in the former, firms have a substantial amount of common ownership and control, usually by founders (promoters). Concentrated ownership has its pro and cons. On one side, it aligns incentives and encourages monitoring, but on the other side, it weakens further corporate governance mechanisms and

⁹To review the literature on caste-based discrimination in the Indian labor market, refer to Vaid (2014).

¹⁰Appendix in Khanna and Palepu (2000) provides the detailed comparison of India’s institutional context with that in the United States and Japan.

can impose significant costs (Berglof and Classens (2006)). In the analyses, I control for additional corporate governance mechanisms discussed in Berglof and Classens (2006) to reduce common ownership's negative impact.

4. Theoretical Model

Holmstrom (1979) and Edmans, Gabaix, and Jenter (2017) provide the model's motivation.¹¹ According to this model, the board of directors hires a manager to run the firm. The firm value is indicated by $V(a, S, \epsilon)$, which is increasing in the manager's action a and firm size S . Suppressing the dependence on S and ϵ for simplicity, firm value is a function of only managerial effort.

$$V(a) = S + b(S)a + \epsilon$$

The manager is paid a compensation $c(V)$ contingent upon firm value. The limited liability on the board is assumed, which means $c(V) \leq V$; that is manager pay should not be more than firm value. Similarly, the limited liability of the manager is assumed which means $c(V) \geq 0$ and reservation wage, $w \geq 0$.

The model is modified by introducing group identity into the preference function by providing different weights to the managerial effort component. It defines the agent utility function with identity. This kind of agent utility function is suggested by Akerlof and Kranton (2005), and the similar way of incorporating into group identity in preferences is adopted in Chen and Li (2009), and Chen and Chen (2011) and this is in line with the literature on minimum-effort coordination games.¹²

¹¹In the first-order model of principal and agent by Holmstrom (1979), the agent's utility maximization problem is replaced with first order constraint and it is based on the assumption that an optimum exists and it is differentiable. To know more about first order model and other related models, see Mirrlees (1976) and Rogerson (1985).

¹²For more details on this topic, refer to Camerer (2003).

The model setup from Holmstrom (1979),¹³

$$\begin{aligned}
\max E[V-c] & \quad (\text{Principal's Problem}) \\
\text{s.t. } E[u(c-g(a))] & \geq E[u(w)] & \quad (\text{Participatory Constraint}) \\
c = \phi + \theta V & \quad (\text{Managerial Compensation}) \\
a^* \epsilon \arg\max E[c - g(a)] & \quad (\text{Agent's Problem}) \\
V = S + b(S)a + \epsilon & \quad (\text{Firm Value})
\end{aligned}$$

$g(a)$: cost function of manager's efforts

ϕ : Fixed component of the manager's compensation (Salary)

θ : Variable component of the manager's compensation

w : Reservation wage

ϵ : Random variable with normal distribution. $\epsilon \sim N(0, \sigma^2)$

Agent utility function from Akerlof and Kranton (2005)¹⁴

$$U(c, a, m) = c - a + I_m - t_m |\hat{a} - a|$$

$$\text{Agent's effort including identity}^{15} = a - t_m |\hat{a} - a| = \begin{cases} a - t_m (\hat{a} - a), & \text{if } \hat{a} > a \text{ [Complete Homophily]} \\ a + t_m (\hat{a} - a), & \text{if } \hat{a} < a \text{ [In-complete Homophily]} \end{cases}$$

\hat{a} : Ideal effort

I_m : Additional utility due to group identity (or homophily)

¹³For simplicity, I follow the same assumptions as in Holmstrom and Milgrom (1987) i.e. manager exhibits exponential utility ($u(x) = -e^{-\nu x}$) and cost of effort is pecuniary ($v(c) = c$). The terms are shown without the exponential part. For details, refer to Edmans, Gabaix, and Jenter (2017, Pp-43-45).

¹⁴ I_m is identity utility, which is considered as 0 for simplicity, and $t_m |\hat{a} - a|$ is disutility for diverging from the ideal effort level for in-group member, denoted by \hat{a} .

¹⁵In Akerlof and Kranton's (2005) model, agent's effort are bounded by high effort (i.e. high ideal effort and low ideal effort. If the employee identifies with the firm (insider), the deviation in effort is measured w.r.t high effort, and when the employee is outsider, the deviation in the effort is measured w.r.t. low effort (due to his tendency). The modulus part represents the deviation component.

Complete Homophily (or in-group): When a manager and board have the social identity

In-complete Homophily (or out-group): When a manager and board have different social identities.

m : Variable to categorize different social identities such as native language, native place, or caste.

Proposition. *Managerial efforts are higher in the case of complete homophily than from incomplete homophily ($0 \leq m < 1$).*

Proof: Firm value and manager's compensation for complete homophily case,

For complete homophily

$$V = S + b(S) [a - t_m(\hat{a} - a)] + \epsilon$$

$$c = \phi + \theta[S + b(S)[a - t_m(\hat{a} - a)]]$$

Following Holmstrom and Milgrom (1987), I use quadratic cost function of manager's efforts, i.e,

$$g(a) = \frac{1}{2}ka^2$$

$$E(u) = E[c - \frac{1}{2}k[a - t_m(\hat{a} - a)]^2]$$

First order condition (FOC) :

$$\frac{dE[u]}{da} = \theta b(s)(1 - t_m) - k[a - t_m(\hat{a} - a)](1 - t_m)$$

$$\frac{\theta b(s)}{k} = a^* - t_m(\hat{a} - a^*)$$

$$\frac{\theta b(s)}{k} = a - t_m(\hat{a} - a)$$

$$a_{in-group}^* = \frac{1}{1-t_m} \left[\frac{\theta b(s)}{k} - t_m \hat{a} \right] \quad (1)$$

Similarly, for incomplete homophily case,

$$a_{out-group}^* = \frac{1}{1+t_m} \left[\frac{\theta b(s)}{k} + t_m \hat{a} \right] \quad (2)$$

$a_{in-group}^*$ and $a_{out-group}^*$ depends on managerial effort depends on three factors and their dynamics sensitivity of effort to identity (t_m), sensitivity of firm value to effort ($b(S)$), and sensitivity of

managerial compensation (variable component) to firm value(θ).¹⁶

If, $t_m = \theta = b(S) = 0.5$

$a_{in-group}^* = 1$ and $a_{out-group}^* = 0.75$

$a_{in-group}^* > a_{out-group}^*$

This implies that in the case of complete homophily, the manager's effort at equilibrium is higher than for incomplete homophily.

Lemma 1. *Higher managerial efforts lead to higher firm value at equilibrium for complete homophily compare to incomplete homophily case.*

Proof. As,

$a_{in-group}^* > a_{out-group}^*$

Therefore, at equilibrium (for same $b(S)$),

$V_{in-group}^* > V_{out-group}^*$

This implies that a firm in which there is complete homophily has a higher value than a firm with incomplete homophily, which derives from the difference in managerial efforts in two cases.

Lemma 2. *Higher managerial efforts and firm value for complete homophily lead to higher managerial compensations at equilibrium compared to incomplete homophily.*

Proof. From participation constraint,

¹⁶Simulation is performed using the Desmos.com.Graph link:<https://www.desmos.com/calculator/eywticjr34>

$$E[c] = \begin{cases} E[c]_{in-group} = w + \frac{1}{2}ka_{in-group}^*{}^2 - \frac{\nu}{2}\theta_{in-group}^2\sigma^2, & \text{For complete homophily} \\ E[c]_{out-group} = w + \frac{1}{2}ka_{out-group}^*{}^2 - \frac{\nu}{2}\theta_{out-group}^2\sigma^2, & \text{For in-complete homophily} \end{cases} \quad (3)$$

$$\text{Principal's utility} = E[V - c] = \begin{cases} S + b(S)a_{in-group}^* - E[c]_{in-group} \\ S + b(S)a_{out-group}^* - E[c]_{out-group} \end{cases}$$

First Order Condition (FOC) with respect to θ ,

$$\theta^* = \begin{cases} \theta_{in-group}^* = \frac{1}{b(S)} \left[\frac{1+t_mk\hat{a}}{1-\frac{\nu k(1-t_m)}{(\frac{b}{\sigma})^2}} \right] \\ \theta_{out-group}^* = \frac{1}{b(S)} \left[\frac{1-t_mk\hat{a}}{1-\frac{\nu k(1+t_m)}{(\frac{b}{\sigma})^2}} \right] \end{cases} \quad (4)$$

if $t_m = \theta = b(S) = k = \sigma = \nu = 0.5$ and $\hat{a} \leq 0.667$,

$$\theta_{out-group}^* > \theta_{in-group}^*$$

As, $a_{in-group}^* > a_{out-group}^*$

Therefore, from eq(3), $E[c]_{in-group}^* > E[c]_{out-group}^*$

It implies that the manager's total compensation is higher for a firm with group identity. This is the cost of in-group favoritism to the firm.¹⁷

There is one limitation of this model i.e, the cost of managerial effort can be pecuniary and non-pecuniary, but following Holmstrom and Milgrom (1987), only pecuniary cost of effort is considered.

In sum, the firm value and a manager's total compensation (and salary) are higher if the manager and the board share the same identity (i.e., group identity or homophily). Although additional managerial compensation is a cost to the firm due to in-group favoritism, this cost is compensated by the manager's extra efforts to increase the firm value. In the next section,

¹⁷The simulation graph is available at following link: <https://www.desmos.com/calculator/m7sscqeyie>

I verify whether these theoretical predictions sustain empirical analysis.

5. Data

5.1. Surname Data

In current study, the surname is invoked as the link to know the identities of managers and the directors and it follows from sociology, anthropology, and social neuroscience literature. Sociological studies find that “surname remains central to the establishment of a person’s sense of belonging to a kin group” in contemporary times (Davies, 2011) and anthropologists have long noted that kinship systems differ in their tightness: the extent to which people are embedded in very large extended family networks. With tight kinship, effective cooperation is believed to take place within cohesive in-groups, while people outside the group are distrusted (Alesina and Giuliano, 2013; Moscona, Nunn, and Robinson, 2017). Conversely, in loose kinship societies, people are said to enter productive interactions with strangers but do not emphasize the in-group member. Surname is widely perceived as an important means of genealogy, potentially revealing one’s DNA information (Jobling, 2001). It may also indicate one’s ethnic minority background and engender race biases (Kumar et al., 2015). Notably, a surname connects the bearer not only to family members, but also potentially to many non-family members who share the same surname (Carr and Landa, 1983; Kiong and Kee, 1998). This creates a network and a bond embedded with rich social, cultural, or even biological connections. These connections can be invoked when people bearing the same surname meet or work together, making it easier for them to understand and cooperate. In addition to all these characteristics, a surname in India can provide information such as family lineage, native language, native place, and caste, which makes it a source of discrimination and

favoritism.¹⁸

Because identities of CEOs and directors of Indian firms were unavailable in any existing database, I needed to construct my own. I relied on two large databases- Socio-Economic Caste Census (SECC) and the Linguistic Survey of India (LSI), to develop my dataset of Indian surnames linked to native language, native place, and caste. The surname database is used to determine the identities of managers and the directors.

5.2. Microcensus Data

To generate identity variables based on native place and the caste, the Socio-Economic and Caste Census (SECC) 2011 data of 474 million Indians is used.¹⁹ This census is unique because, in independent India’s history, only two times caste is measured - in 1950 and 2011. Registrar General of India conducted SECC 2011 and has approximately 4.6 million categories of caste, sub-caste, synonyms, different surnames, gotras²⁰ in the caste and clan names.²¹ The data contains an individual’s name, age, gender, address (state, district, and gram panchayat (similar to county)), caste, parents’ name, and other details-the preview of SECC data is provided in **Appendix C1**. To keep individuals’ information private, only surnames, caste, district, and state variables are extracted.

Extracting surnames from Indian names using SECC data is a complex process due to the following reasons: First, names in the SECC data are given in English as well as in native language separated by backslash (“\”). Second, occasionally, first name and surname are amalgamated, such as “DilipChauhan” (surname: Chauhan). When this occurs with surnames of the same character length as only first name or amalgamated first-last name,

¹⁸Indian media has covered this issue many times, such as “Ban Surnames to End Casteism” (Times of India, 8th Aug 2016), “Slash Surname to kill caste” (The Telegraph, 14th Oct 2005) and similar others.

¹⁹The exact number of observations in raw SECC data is 473,967,954.

²⁰Lineage segment within an Indian caste that prohibits intermarriage by virtue of the members’ descent from a common mythical ancestor, an essential factor in determining possible Hindu marriage alliances.

²¹<https://www.thehindu.com/opinion/lead/the-tools-for-counting/article24247791.ece>

it makes it difficult to extract surnames without including the noise. Third, names in north India and south India are differently written in SECC data, such as "Ram Singh" is a north Indian name with the surname 'Singh' and 'Chandra Rao P.K.' is a south Indian name with the surname 'Rao'. It exhibits a different position of surnames in north Indian and south Indian names in SECC data, which makes extraction further complicated. The long names such as 'Mohammad Kasim P.' (surname: Kasim) also have the same pattern as south Indian names. Details of generating a surname database from SECC 2011 using cloud computing are explained in **Appendix C2**. The final dataset is consolidated at the state level and contains variables such as surname, caste, and place of origin (state). The data is consolidated at the state level because stereotypes and implicit biases are stronger at the state level rather than at the district level.²²

5.2.1. Linguistic Survey of India (LSI) Dataset

The LSI is a comprehensive survey of Indian languages conducted for the first time in 1898 by George Abraham Grierson, a linguist, and obtained information on 364 languages and dialects. The second survey was conducted in 1984, and by 2010, it was 40% complete, but the project was abandoned after that. The third survey started as the People's Linguistic Survey of India (PLSI) in 2010. According India's 1991 census, there are about 1,576 "mother tongues" with separate grammatical structures and 1,796 languages classified as "other mother tongues."

For this paper, I use Lacina's (2017) data, which contains a linguistic composition of all the 640 districts of India. This data contains information on district id, state, district name, population, Hindu population in a district, and the seven most popular languages in a district with a percentage of the population speaking them.

For simplicity, I use only the most popular language in a district and further consolidated

²²Python code to scrape surnames with identities such as caste and state is available on request.

the data at the state level, which provides the data on the most popular language in 29 states and 7 union territories. The final surname dataset contains the native place, native language, and the caste associated with Indian surnames.

5.2.2. Alternative Surname Data

To externally validate the surname database developed using SECC and LSI, I develop an alternate surname dataset from the website “Indianchildnames.com”. This website features a large collection of about 125,000 names and 15,000 surnames of Indian and international origins. Data on Indian surnames, their native language, and their native place is collected using a web scraping technique. This dataset contains information on unique 2,225 Indian surnames with their native language and native place. Merging this data with managers and board members’ surnames from ProwessDx (CMIE) database using a text similarity algorithm gives information on language and place of origin of directors and managers. The details of the algorithm is provided in section 4.1.

5.3. Firm characteristics and Manager characteristics

ProwessDx (CMIE), Global Compustat, and the BoardEx databases are used for the manager and board member names, executive compensation data, firm fundamentals, board characteristics, and managerial demographics. Extracting surnames from manager and board members names are not straightforward because of various designations and their patterns such as Mr./MR/Mr/(MR)/(Mr.)/(Mr), Dr./Dr/(Dr), Prof., Sh./Shri, I.A.S., army designations (Maj. (Major), Gen. (General), Col. (Colonel), navy designations (Vice Admiral, Commodore) and many others. Therefore, an algorithm employed to extract surnames from a manager and the board members’ names are different from the way surnames extracted from SECC names.

To avoid data complexities or unavailability of firm fundamentals, the sample comprises

only non-financial firms listed on two leading Indian stock exchanges, Bombay Stock Exchange (BSE) and National Stock Exchange (NSE), from 2004-2018. There are 2,324 firms in the sample and 28,026 firm-year observations. **Appendix A1** shows the filtration of observations from the ProwessDx database to sample data. As the concept of CEO is new in India, therefore, CEO-equivalent positions like President, Managing Director, and Chairman are also part of the sample.²³ The distribution of managerial designation is provided in **Appendix B**. The yearly distribution of the observations, provided in (**Appendix A2**), shows that the sample is balanced.

The compensation data from ProwessDx includes information on total compensation, salary, perks, and retirement benefits. For the analysis, only total compensation and salary data are used as most of the observations for perks and retirement benefits data are missing. Log values of total compensation and salary are used due to skewness in the distribution.

Price to Book ratio is used as a proxy for firm value, and it is measured as the ratio of market price of the equity by book value of the equity. For the robustness check, *Sales Growth* is used as another proxy of firm value. *Sales Growth* is calculated as growth in sales compared to the previous year.

For the controls, *Size*, *Firm age*, *Institutional Ownership*, *Promoters Ownership*, *Leverage*, *Volatility*, *BIG5*, *Board Size*, *Number of Meetings*, and *ROA* are used. The data to construct these variables are from ProwessDx (CMIE). *Size* is measured as a log of market capitalization. *Firm Age* is the age of the firm from year of incorporation. *Promoters Ownership* is percentage of promoters equity holding to the total equity issued by the firm. *Institutional Ownership* is the percentage of institutional investors equity holding to the total equity issued by the firm. *Leverage* is the ratio of total debt by total assets. *ROA* is measured as the ratio of net income to total assets. *Board Size* is total number of directors

²³Bertrand and Hallock (2001) uses the CEO, chairman, Managing Director positions for the top corporate jobs.

on board. *Volatility* is return volatility. *BIG5* is a dummy variable which takes value 1 if the firm’s auditor is BIG 5 auditor; otherwise 0. *Number of Meetings* is the number of board meetings in a year. The institutional ownership, promoters ownership, board size, number of board meetings, and BIG 5 are internal and external governance characteristics.

Because managerial characteristics like age, gender, and education are not available in ProwessDx, this data is gathered from the BoardEx database. Due to constraints in matching these two databases, these characteristics are not available for all the managers. The number of firm-year observations is limited to 3,542 due to this constraint. The results controlling for managerial demographics are provided in **Appendix D**.

6. Methodology

6.1. Surname Matching using Unsupervised Machine Learning

Due to the lack of a common identifier between the surname dataset and ProwessDx dataset, a text similarity algorithm is used to discern a manager’s identities and the remaining board members. Text similarity algorithm is a simple form of unsupervised machine learning. The two most common text similarity algorithms are Jaccard and Cosine. These two algorithms give different results if the two documents or sentences in question are too different in length. In this case, both would give similar results as the comparison is between two surnames rather than between two documents.

For matching, surnames are first divided into bi-grams and by using Jaccard similarity algorithm, similarity scores are computed. For this case, the Jaccard similarity is the ratio of the number of common bi-grams in two surnames divided by the union of bi-grams in two surnames.

$$\text{Jaccard Similarity} = \frac{A \cap B}{A \cup B} \quad (6)$$

Here, A and B are two surnames.

It can be understood by computing the Jaccard similarity score of ‘smith’ and ‘thomas’. The bi-grams of ‘smith’ are {sm,mi,it,th} and bi-grams of ‘thomas’ are {th,ho,om,ma,as}. As there is only one bi-gram is matching, i.e. ‘th’, therefore, Jaccard similarity is 0.125. It is quite low because the surnames in this example are too discrete. For exact matching, the threshold of similarity score is kept at 0.9 and above, and to confirm the matching, the manual check was also conducted.

6.2. Empirical Specification

To test the role of in-group favoritism in setting executive pay, following empirical specification is used:

$$Y_{i,t} = \alpha_{i,t} + \beta_1 \text{Homophily}_{i,t} + \beta_2 X_{i,t} + \beta_3 L_{i,t} + \beta_4 Z_{i,t} + \gamma_{firm} + \gamma_{year} + \gamma_{Designation} + \epsilon_{i,t} \quad (7)$$

The main dependent variable, Y, is a log of total compensation. For external validation, in place of a log of total compensation, a log of salary and total compensation relative to industry peers are used. Homophily (a measurement of group identity) is a dummy variable that takes value 1 if the majority of board members have the same identity trait (same native language, same caste, or the same place of origin) as the executive/manager; otherwise, 0. This paper uses a simple baseline homophily measure as an estimate of in-group

favoritism.²⁴ The detailed literature surveys on executive compensation by Core, Guay, and Larcker (2002), Frydman and Jenter (2010), and Edmans, Gabaix, and Jenter (2017) suggest that main determinants of executive pay are mainly firm fundamentals, governance characteristics, and manager demographics. Following that suggestion, X is a vector of firm characteristics such as firm size, firm age, return on assets (ROA), leverage, and volatility. L is the vector of governance characteristics such as board size, number of board meetings, BIG 5 auditor, institutional ownership, and promoters ownership. Z is the vector of manager's demographics (such as education, age, gender, and others). To address the impact of unobservables, firm, and year fixed effects are used. The γ_{firm} and γ_{year} are firm and year fixed effects. As results can be affected due to managerial position; therefore, $\gamma_{Designation}$ fixed effects is also used. As managerial demographics are not available for the full sample, therefore, subsample results controlling for the manager's age, gender, and education are provided in **Appendix D**.

To check the impact of homophily on firm value, following empirical specification is implemented:

$$FirmValue_{i,t} = \alpha_i + \beta_1 Homophily_{i,t} + \beta_2 X_{i,t} + \beta_3 L_{i,t} + \gamma_{firm} + \gamma_{year} + \epsilon_{i,t} \quad (8)$$

The coefficient of interest is β_1 . As homophily is based on language, place, and caste, β_1 can be seen as $\beta_{Language}$, β_{Place} , and β_{Caste} . Firm Value is measured as the price to book ratio. X and L are the firm characteristics and governance characteristics listed earlier.

²⁴The extensive literature on homophily has demonstrated such patterns across many dimensions, including ethnicity, age, gender, religion, profession, as well as things like political opinions and other behaviors. Homophily can arise for many reasons, such as biases in the amount of contact that people have with each other (Blau (1977), Feld (1981), Rytina and Morgan (1982)), preferences for associating with individuals with similar traits, behaviors, or backgrounds (Cohen (1977), Kandel (1978), Knoke (1990), and Currarini, Jackson and Pin (2008) and Bramouille and Rogers (2009)), competition among groups (Giles and Evans (1986)), social norms and culture (Carley 1991)), and institutional and organizational pressures (Meeker and Weiler (1970), Khmelkov and Hallinan (1999), Kubitschek and Hallinan (1998), Stearns (2004)).

Based on the prior literature, all β s of firm fundamentals and governance are expected to be positive except for volatility and leverage (Hovakimian, Hovakimian, and Tehranian (2004)).

Next, to check the reciprocation of in-group favor by the manager, the impact of homophily on the residual firm value is examined. The residual firm value is firm value conditioned on managerial compensation of last year.

$$FirmValue_{i,t} = \xi_i + \phi TotalCompensation_{i,t-1} + \varepsilon_{i,t} \quad (9)$$

The Residual Firm Value is ε . It captures the firm value conditioned on managerial compensation; an increase in Residual Firm Value proxies higher managerial effort to increase the firm value. To test whether Residual Firm Value is positively associated with homophily variables or not, the following regression specification is used,

$$ResidualFirmValue_{i,t} = \alpha_i + \beta_1 Homophily_{i,t} + \beta_2 X_{i,t} + \beta_3 L_{i,t} + \gamma_{firm} + \gamma_{year} + \epsilon_{i,t} \quad (10)$$

The interpretation and the expectations for the β s are the same as for the firm value regression. If Residual Firm Value is only explained by firm characteristics and governance characteristics then we should expect β_1 to be statistically insignificant. However, if β_1 is positive and significant, that implies the manager exerts extra effort to increase firm value more than what he is paid and it happens due to shared group identity with the board.

6.3. Endogeneity Concerns and Identification Strategy

As the board can select the same identity manager to reduce frictions in corporate decision-making (Damaraju and Makhija, 2018), this can affect the causal inference of the results. Therefore, to mitigate this concern, I use a similar identification strategy as in Fracassi and Tate (2012). The director's death or retirement is used as an exogenous shock as these events

are not related to conditions inside the firm. Voluntary retirement is not included as it can be due to firm related issues.

As ProwessDx does not have information on director's date of death, therefore, this information is collected from BoardEx database.²⁵ The empirical specification for this test is,

$$Y_{i,t} = \alpha_i + \beta_1 Death + \beta_2 Homophily_{i,t} + \beta_3 Death \times Homophily_{i,t} + \beta_4 X_{i,t} + \beta_5 L_{i,t} + \gamma_{firm} + \epsilon_{i,t} \quad (11)$$

Here Y is log of total compensation (*Total Compensation*), firm value (*Firm Value*), or firm value conditioned on managerial compensation (*Residual Firm Value*). *Death* is a dummy variable that takes value 1 if any director died or retired in that firm-year; otherwise, 0. X is a vector of firm fundamentals, and L is a vector of governance characteristics. The coefficient of interest is β_3 . The firm (γ_{firm}) fixed effects is used to control for firm-level unobservables. As time fixed effects can affect *Death* variable's coefficient, it is not included in the specification.

One can argue that a director's death or retirement may or may not change the board identity and hence, the homophily constructs. To mitigate such concerns, an additional test is conducted. Rather than using the death or retirement of a director as a shock, change in homophily is considered as shock. Based on that, a *DR_C_HS* dummy variable is constructed, which takes value 1 if any director died or retired in that firm-year and homophily changed relative to next year; otherwise 0. S designates a form of homophily, i.e., due to language, place, or caste. The coefficient of *DR X Homo_S* determines whether homophily affected the managerial compensation, firm value, and residual firm value before or after the shock. I

²⁵Boardex dataset on India uses various sources of information like a specific database of Indian directors (Directors Database), media information, etc., to construct their dataset.

find that the coefficient of interest for the language and the place homophily is positively associated with managerial compensation, firm value, and residual firm value, confirming the paper’s main findings. Results are provided in table D4.

7. Results and Discussion

7.1. Univariate Analysis

7.1.1. Identities of the Managers and the Directors

Figure 1 shows maps showcasing the distribution of managers and directors as per their native place in India. These maps exhibit that although directors and managers are from all over India, most of them originally from southern India. In a parallel analysis, I find that most founders come from northern India. This implies that wealth creators come from the northern region in India while wealth managers come from the southern.²⁶

[Insert Figure 1 Here]

The pie charts in Figure 1 shows that most of the directors and the managers in India belong to the upper caste, and only 21-24% are from a lower caste. This distribution is more skewed at a managerial level than at the directors level.

Figure 2 shows the distribution of managers and directors as per their native language and conveys that most of the managers and directors speak Marathi, Tamil, or Western Hindi and this group comprises 14-18% of the sample. Other popular native languages spoken by managers and directors are Malayalam, Gujarati, and Kannada. The native place, native

²⁶Indian Media has also pointed out that half of the top ten Indian billionaires figuring in the list of Bloomberg Billionaires are Gujaratis (state in a northern India) and one in three CFOs of Nifty index companies is south Indian. For details check, <https://economictimes.indiatimes.com/news/politics-and-nation/stereotypes-dominate-indias-rich-list/articleshow/46857906.cms?from=mdr>

language, and caste forms the identities of the managers and the directors.²⁷

[Insert Figure 2 Here]

As there can be a concern that homophily constructs might be time-invariant, but graphs in Appendix A5 show that it is not true, and these constructs change with time. It happens because the change in manager or director's leaving or joining the board can affect these constructs.

7.1.2. Summary Statistics

Table 1 Panel A shows the summary statistics of the main variables used in the analysis. The average (maximum) total compensation of CEO or CEO equivalent position in India is 192,000 USD (2,461,000 USD).²⁸ The major part of the total compensation is salary, and the variable components of the pay comprise perquisites and retirement benefits. The average *Salary* is 132,000 USD, and maximum is 1,116,000 USD. Unlike in the US, the mean variable pay in India is 17% of total remuneration. Summary statistics and distribution of managerial compensation are compared with Hill and Thomas (2014). As total compensation and salary are skewed, a log of total remuneration and salary is used for the analysis. Figure 3 shows the kernel density curves of *Log Total Compensation* and *Log Salary* with normal density curves for reference. It shows that taking the logarithm of total remuneration and salary reduced the skewness in the variables.

[Insert Figure 3 Here]

²⁷The table A4 shows the industrial distribution of homophily variables. In certain industries such as metal and mining, consumer durables, media print, wines cables, wood wood products, fertilizers, and similar others have more homophily between a manager and the board as compared to other industries.

²⁸To provide the sense of these numbers compared to US executive compensation, for the period 1982 through 1988, the average salary and bonus for CEOs of large public companies were 843,000 USD. (Jensen and Murphy, 1990)

The mean (max) of the log of total compensation (*Total Compensation*) is 10.611 (13.634). *Log Salary* is a log of the salary component. The mean(max) of *Log Salary* is 10.4 (13.493). For the external validity test, I also use compensation relative to the industry average as a dependent variable. The mean(max) of industry relative compensation (*Industry Relative Compensation*) is 0.003 (3.752). The standard deviation of total compensation is higher than the standard deviation of the salary component, which is probably due to variation in perks. Log compensation and industry relative compensation variables show how executive compensation is affected by in-group favoritism.

[Insert Table 1 Here]

To show how the firm value is affected by group identity (homophily), *Price to Book ratio* and *Sales Growth* are used as proxies of firm value. *Price to Book ratio* is the ratio of the market price of equity to the book value of the equity. The mean (max) value of the *Price to Book ratio* is 2.238 (19.260). *Sales Growth* is change in sales compared to last year. The mean (max) of *Sales Growth* is 0.145(7.067). To show how the firm value is affected by the homophily even after paying the cost of in-group favoritism, *Residual Firm Value* is computed. The *Residual Firm Value* is the residual of regressing the firm value on last year's managerial compensation. The underlying theory for using it as a proxy of managerial reciprocation is that the determinants of residual firm value should be either from firm characteristics or from governance characteristics; if homophily explaining a part of it, that means it is due to the manager's extra effort for the firm. The mean (max) of *Residual Firm Value* is -0.014 (8.470).

The three main independent variables are *Homophily Language*, *Homophily Place*, and *Homophily Caste*. These variables are dummy variables and take value 1 if the board and the manager share the same identity, i.e., native language, native place, or caste. In the 40-42% of the sample's observations, the board and the manager speak the same native language or come from same native place. This indicates that the sample is balanced on these identities.

However, there is a skewness in the *Homophily Caste* variable. In 89% of the observations, the board and the manager belong to the same caste. As the board and managerial positions are dominated by upper-caste individuals, the skewness in *Homophily Caste* results from it. Table 1 Panel B shows the distribution of Homophily variables in detail.

The control variables used are *Size*, *ROA*, *Firm Age*, *Leverage*, *Big5*, *Promoters Ownership*, *Number of Meetings*, *Institutional Ownership*, *Volatility*, and *Board Size*. The *Size* is measured as a log of market capitalization. The mean *Size* is 16.982. The average firm age in the sample is 34 years, and the firm’s average leverage is 15.9%. As in India, the promoters own the majority of shares, and *Promoters Ownership* clearly reflects that. The mean *Promoters Ownership* and *Institutional Ownership* are 54% and 9.5% . On average, 27.5% of observations have BIG5 auditors, and on average, Indian boards conduct 8 meetings in a year. The definition of all the variables is provided at the end of the paper.

Regarding board characteristics, the average board size is 10 members. The minimum (average) age of a manager is 40 (64) yrs old, and the average qualification is a masters degree. The number of women on board (*Gender Diversity*) is only 9-10 % on average. Panel B shows that the sample is balanced for homophily language and homophily place. The homophily caste is skewed as majority of managers and directors in India belongs to upper caste. It implies that either lower caste individuals do not get opportunities to reach to such posts or those individuals are crowded out due to lobby by upper caste managers and directors. Testings those implications are beyond the scope of this paper but can be examined in future research. Panel C reports the summary statistics of different components of managerial compensation in million USD. Panel D shows the univariate analysis of total compensation, firm value, and residual firm value based on homophily constructs. Panel E reports the balance between control variables for three homophily constructs.

7.2. Main Results

7.2.1. Executive Compensation and In-Group Favoritism

Table 2 shows the baseline results of how group identity (or homophily) affects managerial compensation. The results show that when a manager and the board share the same native language or come from the same native place, it helps the manager to earn higher compensation. The impact of group identity is stronger for native language than the native place as $\beta_{Language} > \beta_{Place}$. The $\beta_{Language}$ is 0.066, and the β_{Place} is 0.071, but the difference is not statistically significant. But when the board and the manager have the same caste, homophily affects total compensation negatively. I attribute these contradictory results to the skewness in the caste variable, which requires further analysis. To control the impact of unobservables, managerial designation, firm, and year fixed effects are used.

[Insert Table 2 and Table 3 Here]

Table 3 shows the impact of group identity on executive compensation with controls. These results confirm that if the manager and the board speak the same native language or belong to the same native place then the manager earns higher compensation compared to his peers. The $\beta_{Language}$ is 0.078 and it is significant at 1% level. The β_{Place} is 0.077 and it is also significant at 1 % level. This implies that in a firm where the board is more homophilic to the manager due to native language or native place, the manager earns 7-8% higher compensation. The coefficient of the Homophily caste a variable is negative and significant, but it is hard to infer as variable is highly skewed. Among the controls, *Size*, *ROA*, *Institutional Ownership*, *Promoters Ownership*, *Number of Meetings*, *Firm Age*, and *BIG5* are significant and positively associated with managerial compensation. As expected, the coefficients of *Leverage* and *Volatility* are significant and negatively associated with compensation. The direction and magnitude of the controls' coefficients are in line with the

findings in the executive compensation literature.

7.2.2. Firm Value and Group Identity

Table 4 shows the impact of homophily on firm value. The results show that homophily in native language is positively associated with firm value, and the results are significant at the 5% level. Similar to the executive compensation results, homophily based on native language is stronger than homophily based on native place, i.e., $\beta_{Language} > \beta_{Place}$. The $\beta_{Language}$ is 0.115, and it is significant at 1% level. The β_{Place} is 0.058, but it is insignificant. This implies that firms in which there is more homophily between a manager and board have 11.5 % higher value. Among the controls, *Size*, *Institutional Ownership*, *Promoters Ownership*, *Number of Meetings*, and *BIG5* are significant and positively associated with firm value. The coefficient of *Leverage* and *Volatility* are significant at 5 % level and negatively associated with firm value. To avoid time trend and firm-level unobservables, firm and year fixed effects are used.

[Insert Table 4 Here]

7.2.3. Firm Value and Reciprocation of In-Group Favoritism

Table 5 shows the results of group identity on Residual firm value. The results indicate that homophily due to language is positively associated with residual firm value. Similar to the previous results, the impact of homophily based on native language is stronger than homophily based on native place, i.e., $\beta_{Language} > \beta_{Place}$. The $\beta_{Language}$ is 0.086, and it is significant at 5 % level. The β_{Place} is 0.056. This implies that firms in which there is more homophily between a manager and board have 5.6-8.6% higher residual value, i.e., higher firm value excluding the cost of in-group group favoritism.

[Insert Table 5 Here]

7.3. External Validity Test

7.3.1. Alternative Measure of Compensation and Firm Performance Variable

As Bizjak, Lemmon, and Nguyen (2011) exhibit that many firms target a specific range or percentile of their peers' pay levels when setting their CEO compensation. Therefore, to externally validate executive compensation, log total compensation is replaced with compensation relative to the industrial average. The results of these tests are provided in Table 6 Panel A and are in line with the previous results confirming that in-group favoritism based on native language helps a manager to earn higher relative compensation and the firms with group identity have higher sales growth compare to other firms (see Appendix D). This provides evidence of the external validity of my dependent variables and confirms the main findings.

7.3.2. Alternative Homophily Variable

To check the external validity of the homophily variables, the same constructs are computed using data from "Indianchildnames.com". The empirical specification for this test is,

$$Y_{i,t} = \alpha + \beta_1 Homophily_{i,t} + \beta_2 Size_{i,t} + \beta_3 ROA_{i,t} + \beta_4 Leverage_{i,t} + \beta_5 Board_Size_{i,t} + \gamma_{firm} + \gamma_{year} + \epsilon_{i,t} \quad (12)$$

Here Y is either log of total compensation or a log of salary. *Homophily* is a group identity dummy based on the native language. In Table 6 Panel B, The coefficient of *Homophily* is 0.138 in the regression of total compensation and 0.102 in the regression of the salary. In both cases, results are significant at the 5 % level.²⁹ To avoid the effect of firm-level

²⁹I also check the results using the group identity dummy based on native place and the impact of group identity on firm performance. Results support the main hypothesis of the paper.

unobservables or any time trend, firm fixed effects and year fixed effects are used.

[Insert Table 6 Here]

These results confirm that if a manager and the board share the same identity, then manager and the firm are better off as it increases managerial compensation, increases the firm value, and reduces agency frictions.

7.4. Endogeneity Issue and the Identification Test

The endogenous relationship between executive compensation and corporate governance is always an issue (Hermalin and Weisbach, 1998; Wintoki et al., 2012) for this kind of research question. Recall, there is a possibility that the board chooses a manager that matches their own identity just to avoid conflict in corporate decision making (Damaraju and Makhija, 2018). Thus to mitigate self-selection bias, following Fracassi and Tate (2012) , the director’s death or retirement is used as a shock to the board identity.

[Insert Table 7 Here]

I find that β_3 is positive in all the three regression specification. However, it is statistically significant only for compensation and firm value regressions. This confirms paper’s main results that group identity between a manager and the board based on native language leads to higher managerial compensation. Although the increment in managerial compensation is a cost to the firm, group identity also incentivizes the manager to put more effort toward the firm’s growth and increase its value.

7.5. Other Robustness Tests

7.5.1. Community Migration and Group Identity

The underlying assumption of using the surname data to discern managers and directors' identities is that there is no migration of whole communities (people belong to a particular surname) from one part of India to another. For example, if the majority of 'Gupta' community decided to live at Bengal after Independence, this assumption implies most 'Gupta' community still reside there. This premise is quintessential because change in place would also affect the linguistic profile and the caste profile related to that surname and hence, manager and directors' identities. Although this premise is valid for most of the communities (or surnames) but not valid for all. The community such as 'Kashmiri Pandits' migrated from India's northernmost part known as Kashmir region to other parts of India due to the riots in the 1990s in that region.³⁰ As it is hard to track each 'Kashmiri Pandit' surname and to exclude them from the surname database, therefore, I exclude all the surnames that are associated with the Kashmir region and do a robust check with remaining surnames.³¹

[Insert Table 8 Here]

This test results are provided in Table 8 Panel A. Like, for compensation regression, the coefficient of *Homophily Place* is 0.077, and it is statistically significant at 10%. Similarly, with the same specification, the coefficient of *Homophily Language* is 0.079, and it is statistically significant at 1%. For the firm value regression, the coefficients are 0.112 (significant at 1 %) and 0.058 (insignificant). *Homophily Place* and *Homophily Language*'s coefficient for Residual Firm Value are 0.060 and 0.090, and both are statistically significant. It implies that the main results are robust to community migration. The detailed table for this test is provided in the Appendix.

³⁰<https://www.bbc.com/news/world-asia-india-35923237>

³¹ 318 observations with surnames from Kashmir region excluded from the sample to conduct this test.

7.5.2. Geographical Differences

As maps in Figure 3 shows there is a dominance of people from the southern part of India at CEO or CEO equivalent positions or at directors' posts, it is important to check the difference between the homophily arising between a manager and the board members due to identities from the southern part of India w.r.t to the northern part of India. Also, as the southern part of India has a different culture, language, food habits, and lifestyle from the northern part of India, there is a possibility that the intensity of homophily between a manager and the board members could be different for these two geographical locations. To verify that, I compute dummy variable (*'South'*), which takes value 1 if the native place is from southern states - Karnataka, Tamil Nadu, Andhra Pradesh, or Kerala; otherwise 0. I interacted this dummy variable with homophily variables and modified the regression specification.

The results for these regressions are provided in Table 8 Panel B. The coefficient of *Homophily Place* and *Homophily Language* are positive and significant for the compensation specification. These coefficients are also positive for residual firm value but insignificant. It supports the previous results and the main argument of the paper that there is in-group favoritism in deciding the managerial compensation based on native language and a native place, and homophily based on these identities also increases the firm value in the long run, it is beneficial for the firm to bear this cost. Thus, the net benefit from homophily between a manager and the board members shows that the homophily reduces agency frictions.

On the other side, the coefficient of *HomophilyXSouth* is negative but insignificant in all the regressions. It implies that, although managerial positions and directors' posts are dominated by people from the southern part of India, the association of homophily with total compensation, long term firm value, and residual firm value does not differ due to homophily based on identities from the southern region of India or the northern region of India.

7.5.3. Powerful CEO

As powerful managers have a higher influence on the decision making process in an organization, there is a possibility that s/he can also influence the executive compensation decision and can affect firm performance differently. Adams et al.(2005) demonstrate that firms in which the CEO has greater power to influence decisions such firms also have higher performance variability. To mitigate such concerns , I also control for manager’s influence in the organization. Following Adams et al. (2005), I measure powerful manager (*Powerful_CEO*) as a dummy variable that takes value 1 if either manager is also a founder (promoter) of the firm or manager’s concentration of job titles, i.e., chairman, managing director or CEO of the firm. Results are provided in Table 9 and show that even controlling for manager’s influence in the organization, homophily due to language and place are still positively associated with executive compensation and residual firm value.

[Insert Table 9 Here]

These findings support the main results of the paper. I also find that there is a positive and statistically significant relation between powerful manager and executive compensation but there is no relation between powerful manager and firm value. This implies rent extraction due to the manager’s influence on the organization’s decision making process. As this is not the primary focus of the paper, it needs further research.

7.5.4. College Ties and Past Experience

Farcassi and Tate (2012) show that powerful CEOs are more likely to appoint directors with ties to the CEO and that CEO-director ties reduce firm value. CEO-director ties measured in the form of Social Network Index (SNI) and defined as the sum of Current Employment Connection, Prior Employment Connection, Education Connection, and Other Activity Connection. From these network connections, the most common are past employment and other

activities, and the least common are education and current employment. As there is a possibility that an increase in compensation, firm value, and residual firm value is not due to the homophily but due to the college ties or past employment network between a manager and the directors, and homophily is just a proxy for these ties. To mitigate such concern, I control for college ties and the past employment connections between a manager and the directors. Using BoardEx’s education database and the network database, I created the variables- *College Ties* and *Experience Ties*. *College Ties* is a dummy variable that takes value 1 if there is at least one director in the board who is from the same college as manager; otherwise 0. Similarly, *Experience Ties* is a dummy variable that takes value 1 if the manager has an experience of working with at least one director of the board; otherwise, 0.

The results for this test are provided in Table 10. Panel A shows the results controlling for college ties. The $\beta_{Language}$ is 0.131, and β_{Place} is 0.126 for compensation, and both are statistically significant. The $\beta_{Language}$ is also positive and significant for residual firm value. It implies that homophily based on the native language between a manager and the board increases executive compensation and residual firm value, even controlling for college ties. I also find that college ties between a manager and the board do not impact compensation, firm value, and residual firm value.

[Insert Table 10 Here]

Panel B shows the results controlling for ties from past employment. $\beta_{language}$ and β_{place} are positive and statistically significant for executive compensation and residual firm value. This confirms the paper’s main results that homophily based on native language and native place increases executive compensation and reduces agency frictions by increasing the residual firm value. I also find that ties due to past employment reduce total compensation, firm value, and residual firm value. These results are in line with Farcassi and Tate (2012).

7.5.5. Promoter as a Manager

In India, the promoters (or founders) have higher influence on firms as they are the major shareholders. Therefore, it is crucial to check whether results are different for firms where one of the promoters is the manager. As this information is not available for the full sample, therefore, sub-sample test is conducted. Results are provided in Table 11.

[Insert Table 11 Here]

column (1) and column (2) of Table 11 show that there is no impact of homophily on compensation. These results are intuitive as promoters have no incentive to earn extra compensation based on identity as they hold majority of shares in that firm. Column (3) shows that homophily based on caste reduces compensation and this result is statistically significant and is inline with result from Table 3. However, as articulated before, caste is highly skewed towards upper caste in Indian boards, therefore, the latter result can be biased.

Column (4)-(5) show that homophily based on native language and native place increases firm value by 18.6% and 15.6 %. This is higher compared to full sample results, implying that promoter being a manager, governs the firm better and takes constructive investment decisions. Column (7)-(8) show the residual firm value, homophily due to language and place increases residual firm value by 14.6 % and 11.2 %. It implies that even after paying the cost of in-group favoritism, the firm value for these homophily variable increases. It is higher in case manager is also one of the promoters.

7.5.6. Simultaneity Issue and Two-Stage Least Square (2SLS) Regression

As managerial compensation and firm value are endogenously determined; therefore, there is a possibility that residual firm value (*Residual Firm Value*) do not capture the simultaneity issue completely. To mitigate such concerns, following Palia (2000), I also show results using the two-stage least square (2SLS) method. To implement 2SLS regression, I use the fitted

value of total compensation from the first stage as an independent variable in firm value regression. For first-stage regression, I run the eq(7) by excluding the homophily variable, adding the promoter dummy (*promoter*) and manager's age (*Age*) as instruments, and by replacing ROA with price to book ratio. The fitted value from this step is used as an additional variable in eq(8).

[Insert Table 12 Here]

The results for the 2SLS regression are provided in Table 12. Column(1) shows the regression result of the first stage. Firm value and manager's age are positively associated with total compensation. However, both variables are statistically insignificant. The promoter dummy is positively associated with total compensation and it is statistically significant at 1 %. ³². Column(2)-(4) show the results for the second stage. $\beta_{Language}$ is 0.427 and it is significant at 1%. β_{Place} is also positively associated with firm value, however, it is insignificant. The coefficient of the fitted value of total compensation ($\overline{TotalCompensation}$) is 8.987 and it is significant at 1 %. It confirms that there is a positive relation between managerial compensation and firm value. These results also support the main hypothesis of the paper that homophily increases the firm value even if the cost of managerial compensation is considered.

7.5.7. Business Group vs. Non-Business Group Firms

India's business groups are collections of publicly traded firms with a presence in different industries and a significant amount of common ownership and control, usually by a family. Khanna and Palepu (2000) compare the performance of business group firms with the non-business group firms of India, and show that business group firms outperform non-business

³²These results cannot be directly compared to table 11 results as these are based on sub-sample of the data used in table 11

group firms. They argue that the group-affiliated firms have disproportionately good access to international sources of capital due to business group reputation.

[Insert Table 13 Here]

The current study proposes an alternative channel for better performance of group affiliated firms' i.e., manager's reciprocation towards a firm due to homophily. In Table 13, the regression results of residual firm value show that $\beta_{language}$ is 0.144 (14.4 %) for business group firms and it is significant at 1 % level. The same coefficient is 0.020 (2 %) for non-business group firms, but it is insignificant. In business group firms, managers mainly hire from the same family or someone known to the promoter (founder); therefore, the manager's reciprocation towards the firm increases its value due to homophily.

7.5.8. Other Robustness Checks

The principal and an agent have multiple identities and bias can exist due to any of it. Therefore, rather than using a particular homophily construct, I develop a homophily index using principal component of three homophily constructs and run the same regressions on this index. Results are reported in Appendix D5 and it suggests that managerial compensation increases by 2.9 % due to homophily index which implies the cost of in-group favoritism. However, net firm value also increases by 2.7 %. Although the firm value increases due to homophily even conditioned on additional compensation to a manager, however, this incremental firm value provides higher return to equity holders or not is to be examined. Therefore, to check that, I run the regression of homophily constructs on the return on equity (ROE). The results are provided in Appendix D6. The $\beta_{Language}$ and β_{Place} are positive and significant for ROE and Residual ROE. Lastly, following Bhagat and Bolton (2008), I use current and future profitability (measured as return on assets (ROA)) as dependent variables. Results suggest that *HomophilyLanguage* and *HomophilyPlace* are positively

associated with current and future ROA. However, only β_{Place} is statistically significant. These findings are in contrast with Cai et al. (2013) which shows that profitability reduces if a manager and owner are from the same family. I presume these difference in results arise due to difference in governance system. The high promoters ownership maintains higher scrutiny over manager which leads to higher profitability in case of homophily between manager and board members.

8. Homophily Constructs and Firm Value: A Channel

Managers and directors have multiple identities, such as age, sex, religion, race (or caste), affiliation with a native place, native language, and many others. Then, why homophily based on some identities affect firm decisions (such as selecting the manager (or director), executive compensation, job assignments, etc.) and firm performance (or value) and not all? The reason for this can be ‘trust’ and supervision (governance). As the underlying element of homophily is trust, homophily based on different identities reflects different levels of trust. Considering that, if a firm hires a trustworthy manager under proper supervision (or governance), the manager exerts additional effort to increase firm value. The firm would recruit that manager even if the firm has to pay extra compensation to the manager.

An alternative view of trust is the reduction in information asymmetry. The individuals similar in their personal or social characteristics and sharing a common identity (e.g., race, gender, culture) tend to have common interests and worldviews, which explains the formation of ties based on interpersonal attraction (McPherson, Smith-Lovin, and Cook, 2001). Due to this reason, board members understand the manager well that leads to better monitoring and a higher likelihood of the manager’s refraining from shirking because of possible sanctions from the board. In other words, trust and reduced information asymmetry imply lower agency and monitoring costs.

Using a survey of 600 Chinese private family firms, Cai et al. (2013) show that family managers earn 18-22 % higher salaries and receive more bonuses but face weaker incentives. Using demographic similarity between CEO and directors of US firms, Westphal and Zajac (1995) provides evidence that similarity results in more generous CEO compensation contracts. Fracassi and Tate (2012) show that the CEO-director ties based on shared experiences (such as past employment) reduce the firm value, particularly in the absence of other governance mechanisms to substitute for board oversight. The main results of this paper are that manager gets 6-8% higher compensation based on homophily such as native language and a native place, and even after paying higher compensation, such firms have 8-9% more increased value. Managerial efforts can drive the latter result due to promoters' intense supervision as they are major equity holders in Indian firms (Khanna and Palepu (2000)).

In sum, I argue that as homophily based on different identities reflects different levels of trust relationships, the level of association with executive compensation can be different, but mostly all lead to high executive compensation due to favoritism. However, different shared identities are differently associated with firm outcomes due to weaker or stronger governance.

9. Concluding Remarks

This paper examines how identities affect executive compensation, agency frictions, and firm value. I begin with providing an intuition on how identities can affect the agency frictions between a manager and the board by using the Akerlof and Kranton (2005) suggested identity-based agent's utility function in first-order principal-agent model of Holmstrom (1979). This change in the Holmstrom (1979) framework shows that the manager gets higher compensation due to in-group bias, and the same motivates the manager to exert more effort for the firm and this is reflected in the higher future firm value compared to its counterparts.

To test this empirically, a novel dataset on Indian surnames using the data on 474 million Indians from the Social Economic Caste Census 2011 and the data from the Linguistic Survey of India is developed. Indian surnames provide information on native language, native place, caste, and other identities about an individual, and therefore, it is a single source of multiple identities. Mapping this data with directors' and managers surnames provide their identities.

I test whether the group identity (or homophily) between a manager and the board based on the above-mentioned social traits results in any kind of in-group favoritism in managerial compensation, and if yes, then does the manager reciprocate that favor ? I find that manager gets 6-8% higher compensation if he speaks the same native language or belongs to the same native place as the majority of the board members. It is the cost of in-group favoritism. Furthermore, I also show that such firms have 11-12% higher value than their counterparts. Even if the cost of in-group favoritism is excluded, such firms have 5-8% higher value (measured as residual firm value), implying that the manager exerts additional effort to increase the firm value due to the shared group identity. As managerial compensation and firm value are endogenously determined; therefore, there is a possibility that residual firm value do not capture the simultaneity issue completely. To mitigate such concerns, following Palia (2000), I also confirm the results using the two-stage least square (2SLS) method. This result that the individual puts extra effort for their group members compared to outsiders is in line with literature on group identity in social psychology and social neuroscience (Cikara and Bavel, 2014; Allport, 1954).

To check the dependent variables' external validity, total compensation is replaced with salary and compensation with respect to industry peers, and the Price to Book ratio is replaced with sales growth. Results are robust to these changes. Furthermore, to check the external validity of the homophily constructs, an alternate dataset on surnames with the same identities from 'Indianchildnames.com' is used. The results are also robust to this alternate dataset.

As there is a possibility of self-selection that the board can select the manager of the same identity to reduce frictions in corporate decision making, I use the director's death and retirement as the exogenous shock to the board's identity. Results are robust even after tackling the endogeneity concerns. One can argue that a director's death or retirement may or may not change the board identity and hence, the homophily constructs. To mitigate such concerns, an additional test is conducted. Rather than using death or retirement of a director as a shock, change in homophily due to it is considered as a shock. The results from the additional test suggests that the coefficient of interest for the language and the place homophily are positively associated with managerial compensation, firm value, and residual firm value. It confirms the main findings of the paper.

Lastly, I check alternative conjectures such as the influence of powerful CEO on the firm and the role of college ties or ties from past employment between CEO and board members. The main results are robust to the influence of the CEO on the firm. Regarding college and ties from past employment, I find that the firm value and the residual value are still positively associated with homophily based on native language and native place, even controlling for college ties and ties due to past employment. The coefficient of ties due to past employment suggests that the total managerial compensation, firm value, and residual firm value reduces with manager-board members ties from past employment. The latter results are in line with Fracassi and Tate (2012).

In sum, using the CEO-board setting, this article provides evidence of how in-group favoritism can result in higher managerial compensation and how the manager returns this favor by increasing firm value. Latter results support the notion that group identity is not just a hidden cost but can also reduce agency frictions. These results have firm-level implications as it shows how homophily based on native language and native place between a manager and the board affects corporate governance, managerial compensation, and firm value.

References

- Abrams, Dominic, and Michael A. Hogg. "Prospects for research in group processes and intergroup relations." *Group Processes Intergroup Relations* 1.1 (1998): 7-20.
- Adams, Renée B., and Daniel Ferreira. "Women in the boardroom and their impact on governance and performance." *Journal of Financial Economics* 94.2 (2009): 291-309.
- Adams, Renée B., Heitor Almeida, and Daniel Ferreira. "Powerful CEOs and their impact on corporate performance." *The Review of Financial Studies* 18.4 (2005): 1403-1432.
- Akbari, Mahsa, Duman Bahrami-Rad, and Erik O. Kimbrough. "Kinship, fractionalization and corruption." Working paper, (2018).
- Akerlof, George A., and Rachel E. Kranton. "Economics and identity." *The Quarterly Journal of Economics* 115.3 (2000): 715-753.
- Akerlof, George A., and Rachel E. Kranton. "Identity and the Economics of Organizations." *Journal of Economic Perspectives* 19.1 (2005): 9-32.
- Akerlof, George A., and Rachel Kranton. "Identity economics." *The Economists' Voice* 7.2 (2010).
- Alesina, Alberto, and Paola Giuliano. "Culture and institutions." *Journal of Economic Literature* 53.4 (2015): 898-944.
- Allport. "The nature of prejudice." Addison-Wesley Press, (1954).
- Amodio, David M. "The social neuroscience of intergroup relations." *European Review of Social Psychology* 19.1 (2008): 1-54.
- Arrow, K. *Essays in the theory of risk bearing*. Markham, Chicago, IL (1971).
- Basu, Kaushik. "Gender and say: A model of household behaviour with endogenously determined balance of power." *The Economic Journal* 116.511 (2006): 558-580.
- Bhagat, Sanjai, and Brian Bolton. "Corporate governance and firm performance." *Journal of corporate finance* 14, no. 3 (2008): 257-273.
- Bebchuk, Lucian Arye, and Jesse M. Fried. "Executive compensation as an agency problem." *Journal of Economic Perspectives* 17.3 (2003): 71-92.
- Berglöf, Erik, and Stijn Claessens. "Enforcement and good corporate governance in developing countries and transition economies." *The World Bank Research Observer* 21.1 (2006): 123-150.
- Becker, Gary S. "The economics of discrimination." The University of Chicago Press, (1957).
- Bertrand, Marianne, and Sendhil Mullainathan. "Are Emily and Greg more employable

- than Lakisha and Jamal? A field experiment on labor market discrimination.” *American Economic Review* 94.4 (2004): 991-1013.
- Bertrand, Marianne, Dolly Chugh, and Sendhil Mullainathan. “Implicit discrimination.” *American Economic Review* 95.2 (2005): 94-98.
- Bertrand, Marianne, and Kevin F. Hallock. “The gender gap in top corporate jobs.” *ILR Review* 55.1 (2001): 3-21.
- Bhagavatula, Suresh, et al. “The Business of Religion and Caste in India.” Working Paper (2019).
- Bizjak, John, Michael Lemmon, and Thanh Nguyen. ”Are all CEOs above average? An empirical analysis of compensation peer groups and pay design.” *Journal of Financial Economics* 100.3 (2011): 538-555.
- Brown, Rupert. *Group processes: Dynamics within and between groups*. Basil Blackwell, (1988).
- Cai, Hongbin, et al. ”Family ties and organizational design: Evidence from Chinese private firms.” *Review of Economics and Statistics* 95.3 (2013): 850-867.
- Carlana, Michela. “Implicit stereotypes: Evidence from teachers’ gender bias.” *The Quarterly Journal of Economics* 134.3 (2019): 1163-1224.
- Camerer, Colin F. *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press, 2011.
- Carr, Jack L., and Janet T. Landa. “The economics of symbols, clan names, and religion.” *The Journal of Legal Studies* 12.1 (1983): 135-156.
- Cikara, Mina, and Jay J. Van Bavel. ”The neuroscience of intergroup relations: An integrative review.” *Perspectives on Psychological Science* 9.3 (2014): 245-274.
- Chen, Yan, and Sherry Xin Li. “Group identity and social preferences.” *American Economic Review* 99.1 (2009): 431-57.
- Chen, Roy, and Yan Chen. “The potential of social identity for equilibrium selection.” *American Economic Review* 101.6 (2011): 2562-89.
- Crisp, Richard J., Miles Hewstone, and Mark Rubin. “Does multiple categorization reduce intergroup bias?.” *Personality and Social Psychology Bulletin* 27.1 (2001): 76-89.
- Croson, Rachel, Melanie Marks, and Jessica Snyder. “Groups work for women: Gender and group identity in social dilemmas.” *Negotiation Journal* 24.4 (2008): 411-427.
- Cunningham, Lauren M., et al. “What’s in a name? Initial evidence of US audit partner identification using difference-in-differences analyses.” *The Accounting Review* (2018).
- Dahiya, Sandeep, and David Yermack. “You can’t take it with you: Sunset provisions for

equity compensation when managers retire, resign, or die.” *Journal of Corporate Finance* 14.5 (2008): 499-511.

Delis, Manthos D., et al. “The effect of board directors from countries with different genetic diversity levels on corporate performance.” *Management Science* 63.1 (2017): 231-249.

Devine, Patricia G. “Stereotypes and prejudice: Their automatic and controlled components.” *Journal of Personality and Social Psychology* 56.1 (1989): 5.

Eckel, Catherine C., and Philip J. Grossman. “Managing diversity by creating team identity.” *Journal of Economic Behavior & Organization* 58.3 (2005): 371-392.

Edmans, Alex, Xavier Gabaix, and Dirk Jenter. “Executive compensation: A survey of theory and evidence.” *The Handbook of the Economics of Corporate Governance*. Vol. 1. North-Holland, 2017. 383-539.

Enke, Benjamin. “Kinship, cooperation, and the evolution of moral systems.” *The Quarterly Journal of Economics* 134.2 (2019): 953-1019.

Eres, Robert, and Pascal Molenberghs. “The influence of group membership on the neural correlates involved in empathy.” *Frontiers in Human Neuroscience* 7 (2013): 176.

Fairlie, Robert W., Florian Hoffmann, and Philip Oreopoulos. “A community college instructor like me: Race and ethnicity interactions in the classroom.” *American Economic Review* 104.8 (2014): 2567-91.

Fisman, Raymond, et al. “Social ties and favoritism in Chinese science.” *Journal of Political Economy* 126.3 (2018): 1134-1171.

Fisman, Raymond, Daniel Paravisini, and Vikrant Vig. “Cultural proximity and loan outcomes.” *American Economic Review* 107.2 (2017): 457-92.

Fracassi, Cesare, and Geoffrey Tate. “External networking and internal firm governance.” *The Journal of finance* 67, no. 1 (2012): 153-194.

Frydman, Carola, and Raven E. Saks. “Executive compensation: A new view from a long-term perspective, 1936–2005.” *The Review of Financial Studies* 23.5 (2010): 2099-2138.

Frydman, Carola, and Dirk Jenter. “CEO compensation.” *Annual Review of Financial Economics* 2.1 (2010): 75-102.

Giannetti, Mariassunta, and Mengxin Zhao. “Board ancestral diversity and firm-performance volatility.” *Journal of Financial and Quantitative Analysis* 54, no. 3 (2019): 1117-1155.

Gompers, Paul A., Vladimir Mukharlyamov, and Yuhai Xuan. “The cost of friendship.” *Journal of Financial Economics* 119.3 (2016): 626-644.

Gregory-Smith, Ian, Brian GM Main, and Charles A. O’Reilly III. “Appointments, pay and performance in UK boardrooms by gender.” *The Economic Journal* 124.574 (2014): F109-

F128.

Grinblatt, Mark, and Matti Keloharju. "How distance, language, and culture influence stockholdings and trades." *The Journal of Finance* 56.3 (2001): 1053-1073.

Gorodnichenko, Yuriy, and Gerard Roland. "Culture, institutions, and the wealth of nations." *Review of Economics and Statistics* 99.3 (2017): 402-416.

Guay, Wayne R., John E. Core, and David F. Larcker. "Executive equity compensation and incentives: A survey." Working Paper (2002).

Hill, Jennifer G., and Randall S. Thomas, eds. *Research Handbook on Executive Pay*. Edward Elgar Publishing, 2012.

Holmstrom, Bengt. "Moral Hazard and Observability." *The Bell Journal of Economics*. (1979): 74.

Hovakimian, Armen, Gayane Hovakimian, and Hassan Tehranian. "Determinants of target capital structure: The case of dual debt and equity issues." *Journal of Financial Economics* 71, no. 3 (2004): 517-540.

Ito, Tiffany A., and Bruce D. Bartholow. "The neural correlates of race." *Trends in Cognitive Sciences* 13.12 (2009): 524-531.

Jannati, Sima, et al. "In-group bias in financial markets." Working Paper (2019).

Jobling, Mark A. "In the name of the father: surnames and genetics." *Trends in Genetics* 17.6 (2001): 353-357.

Khanna, Tarun, and Krishna Palepu. "Is group affiliation profitable in emerging markets? An analysis of diversified Indian business groups." *The Journal of Finance* 55.2 (2000): 867-891.

Kumar, Alok, Alexandra Niessen-Ruenzi, and Oliver G. Spalt. "What's in a name? Mutual fund flows when managers have foreign-sounding names." *The Review of Financial Studies* 28.8 (2015): 2281-2321.

Lacina, Bethany. *Rival claims: Ethnic violence and territorial autonomy under Indian federalism*. University of Michigan Press, 2017.

Lowes, Sara. "Matrilineal kinship and spousal cooperation: evidence from the Matrilineal belt." Working paper (2018).

McPherson, Miller, Lynn Smith-Lovin, and James M. Cook. "Birds of a feather: Homophily in social networks." *Annual Review of Sociology* 27, no. 1 (2001): 415-444.

Mehran, Hamid. "Executive compensation structure, ownership, and firm performance." *Journal of Financial Economics* 38.2 (1995): 163-184.

Mitnick, Barry M. "The theory of agency." *Public Choice* 24, no. 1 (1975): 27-42.

- Moisson, Paul-Henri, and Jean Tirole. "Cooptation: Meritocracy vs. Homophily in Organizations." Working paper, 2020.
- Morck, Randall, Andrei Shleifer, and Robert W. Vishny. "Management ownership and market valuation: An empirical analysis." *Journal of Financial Economics* 20 (1988): 293-315.
- Morse, Adair, and Sophie Shive. "Patriotism in your portfolio." *Journal of Financial Markets* 14.2 (2011): 411-440.
- Moscona, Jacob, Nathan Nunn, and James A. Robinson. "Keeping it in the family: Lineage organization and the scope of trust in sub-Saharan Africa." *American Economic Review* 107.5 (2017): 565-71.
- Palia, Darius. "The endogeneity of managerial compensation in firm valuation: A solution." *The Review of Financial Studies* 14, no. 3 (2001): 735-764.
- Ross, Stephen A. "The economic theory of agency: The principal's problem." *The American economic review* 63, no. 2 (1973): 134-139.
- Sanders, Wm Gerard, and Mason A. Carpenter. "Strategic satisficing? A behavioral-agency theory perspective on stock repurchase program announcements." *Academy of Management Journal* 46, no. 2 (2003): 160-178.
- Shih, Margaret, Todd L. Pittinsky, and Nalini Ambady. "Stereotype susceptibility: Identity salience and shifts in quantitative performance." *Psychological Science* 10.1 (1999): 80-83.
- Steele, Claude M., and Joshua Aronson. "Stereotype threat and the intellectual test performance of African Americans." *Journal of Personality and Social Psychology* 69.5 (1995): 797.
- Turner, John C., Rupert J. Brown, and Henri Tajfel. "Social comparison and group interest in ingroup favouritism." *European Journal of Social Psychology* 9.2 (1979): 187-204.
- Turner, John C., et al. *Rediscovering the social group: A self-categorization theory*. Basil Blackwell, 1987.
- Westphal, James D., and Edward J. Zajac. "Who shall govern? CEO/board power, demographic similarity, and new director selection." *Administrative Science Quarterly* (1995): 60-83.
- Wilson, R. On the theory of syndicates. *Econometrica* (1968), 36(1), 119-132
- Wiseman, Robert M., and Luis R. Gomez-Mejia. "A behavioral agency model of managerial risk taking." *Academy of management Review* 23, no. 1 (1998): 133-153.
- Vaid, Divya. "Caste in contemporary India: Flexibility and persistence." *Annual Review of Sociology* 40 (2014): 391-410.
- Van Vugt, Mark, and Claire M. Hart. "Social identity as social glue: The origins of group

loyalty.” *Journal of Personality and Social Psychology* 86.4 (2004): 585.

Volz, Kirsten G., Thomas Kessler, and D. Yves von Cramon. “In-group as part of the self: in-group favoritism is mediated by medial prefrontal cortex activation.” *Social Neuroscience* 4.3 (2009): 244-260.

Variable Definition

Variable	Description	Source
TotalCompensation	Log of manager's total compensation. Total compensation is in USD.	ProwessDx
Salary	Log of manager's salary. Salary is in USD.	ProwessDx
Firm Value	Ratio of market value of equity to book value of equity.	ProwessDx
SalesGrowth	Growth in sales compare to last year	ProwessDx
Residual Q	Residual from regression of Price to Book ratio on last year managerial compensation	ProwessDx
Homophily_Language	Dummy variable which takes value 1 if a manager and majority of board members have the same native language; otherwise 0.	Linguistic Survey of India (LSI)
Homophily_Place	Dummy variable which takes value 1 if a manager and majority of board members have the same native place; otherwise 0.	Socio-Economic Caste Census (SECC)
Homophily_Caste	Dummy variable which takes value 1 if a manager and majority of board members have the same caste; otherwise 0.	Socio-Economic Caste Census (SECC)
Size	Log of market capitalization	ProwessDx
ROA	Ratio of net income to total assets.	ProwessDx
PromotersOwnership	Perentage of equity ownership by promoters.	ProwessDx
InstitutionalOwnership	Perentage of equity ownership by institutional investors.	ProwessDx
Leverage	Ratio of total debt to total assets.	ProwessDx
South	Dummy variable which takes value 1 if the state is in south part of India; otherwise 0.	Socio-Economic Caste Census (SECC)
Firm Age	Age of the firm from incorporation	ProwessDx
Board Size	Number of board members	BoardEx
Board Independence	Percentage of independent board members	BoardEx
Board Diversity	Percentage of women in board	BoardEx
ID Rotation	Dummy variable which takes value 1 if there is change in independent director in the same gender	BoardEx

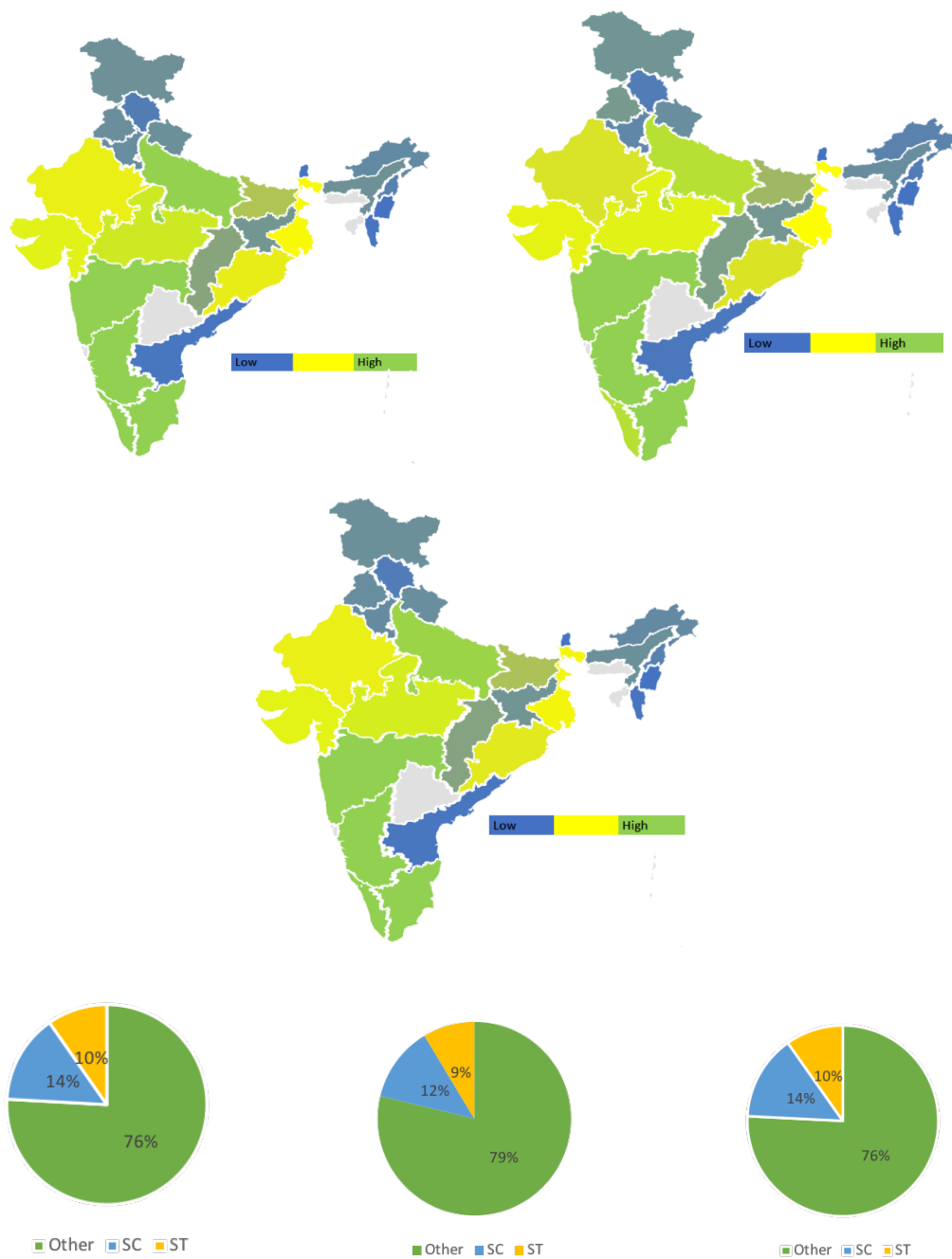


Figure 1: Maps and pie charts show the distribution of native place and caste of managers & directors, managers, and directors.

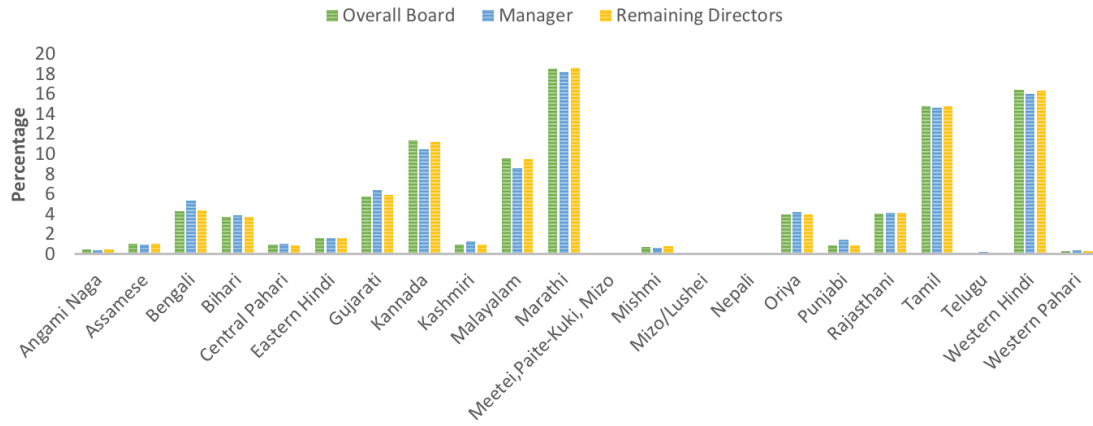


Figure 2: This histogram shows the distribution of language of managers & directors, managers, and directors.

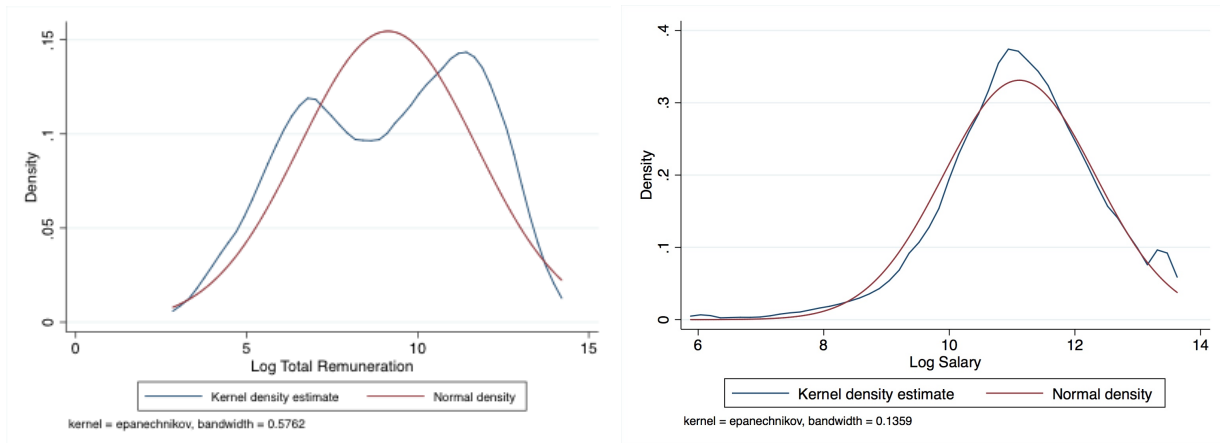


Figure 3: Kernel Density Curves : Log Total Compensation and Log Salary

Table 1
Panel A: Summary Statistics

This table provides the summary statistics of the main variables. *Total Compensation* is log of total compensation. *Industry Relative Compensation* is compensation relative industry average. *Firm value* is ratio of market price of equity to the book value of equity. Residual Firm Value is residual of regression of firm value on manager's total compensation of last year. *Homophily Language* is a dummy variable whose value equal to 1 if the board (on average) and the manager speaks same Indian language otherwise 0. *Homophily Place* is dummy variable whose value equal to 1 if the board and the manager belong to same Indian state otherwise 0. *Homophily Caste* is dummy variable whose value equal to 1 if the board (on average) and the manager belong to same caste otherwise 0. *Size* is measured as log of market capitalization. *Firm Age* is the age of the firm from year of incorporation. *Promoters Ownership* is percentage of promoters equity holding to the total equity issued by the firm. *Institutional Ownership* is percentage of institutional investors equity holding to the total equity issued by the firm. *Leverage* is the ratio of total debt by total assets. *ROA* is measured as ratio of net income to total assets. *Board Size* is total number of directors on board. *Volatility* is return volatility. *BIG5* is a dummy variables which takes value 1 if the firm's auditor is BIG 5 auditor; otherwise 0. *Number of Meetings* is number of board meetings in a year. *Death* is a dummy variable which takes value 1 if any director died or retired in a firm in that year; otherwise 0. This variable is based on Fracassi and Tate (2012). *Powerful CEO* is a dummy variable which takes value 1 if manager is also the founder of the firm or manager holds all the three key positions (CEO, Chairman, and Managing Director) in a firm; otherwise 0. This variable is based on Adams et al.(2005). *Age*, *Gender*, and *Education* are measures of age, gender, and education of a manager. *Experience Ties* is a dummy variable which takes value 1 if manager knew any director from past employment; otherwise 0. *College Ties* is a dummy variable which takes value 1 if manager and any of the director attend the same college; otherwise 0.

Variable	Obs	Mean	Std. Dev.	Min	Max
Firm Value	28,026	2.238	3.073	-3.07	19.26
Homophily Caste	28,026	0.892	0.310	0	1
Homophily Place	28,026	0.397	0.489	0	1
Homophily Language	28,026	0.421	0.494	0	1
Total Compensation	28,026	10.611	2.158	3.412	13.634
Industry Relative Compensation	28,026	0.003	2.070	-5.699	3.752
Residual Firm Value	27,908	-0.014	1.754	-5.143	8.470
Size	28,026	17.326	1.521	14.103	21.463
Firm Age	28,026	33.872	20.413	4	102
Institutional Ownership	28,026	9.540	12.006	0	48.365
Promoters Ownership	28,026	54.471	16.282	5.960	88.580
ROA	28,026	0.038	0.099	-1.435	0.416
Leverage	28,026	0.159	0.179	0	2.372
BIG 5	28,026	0.273	0.445	0	1
Number of Board Meetings	28,026	7.855	3.155	1	29
Board Size	28,026	10.476	3.252	3	23
Age	3,542	64.566	11.282	33	98
Gender	3,598	1.982	0.132	1	2
Qualification	3,430	2.131	1.103	1	9
Death	28,026	0.307	0.461	0	1
Experience Ties	2,921	0.755	0.430	0	1
College Ties	3,299	0.267	0.443	0	1
CEO-Chair Dummy	28,026	0.256	0.437	0	1
Powerful CEO	28,026	0.363	0.481	0	1

Panel B: Homophily Distribution (For Managers)

Homophily Language is dummy variable whose value equal to 1 if the board (on average) and the manager speaks same Indian language, otherwise 0. Homophily Place is dummy variable whose value equal to 1 if the board and the manager belong to same Indian state otherwise 0. Homophily Caste is dummy variable whose value equal to 1 if the board (on average) and the manager belong to same caste, otherwise 0.

	Language		Place		Caste	
Homophily	Freq.	Percent	Freq.	Percent	Freq.	Percent
0	16,240	57.95	16,897	60.29	3,018	10.77
1	11,786	42.05	11,129	39.71	25,008	89.23
Total	28,026	100	28,026	100	28,026	100

Panel C: Managerial Compensation - Distribution and Summary Statistics

This table shows the components of managerial compensation (not in log form) and their summary statistics. Total Remuneration is total managerial compensation in million USD. Salary is major component of total managerial compensation and it is in million USD. Variable pay comprises Retirement Benefits and Perks. Both are in million USD.

Variable	Obs	Mean	Std. Dev.	Min	Max
Total Compensation	28,026	0.192	0.373	0.000	2.461
Salary	21,328	0.132	0.183	0.002	1.116
Perks	9,471	0.042	0.079	0.000	0.530
Retirement Benefits	1,898	0.033	0.057	0.000	0.392

Panel D: Univariate analysis

This table shows univariate analysis of managerial compensation, firm value and residual firm value with respect to homophily variables. Total Compensation is total managerial compensation in million USD. Firm value is price to book ratio. Residual Firm Value is the residual of regressing price to book ratio on last year's managerial total compensation.

		TotalCompensation	Firm Value	Residual Firm Value
Homophily Language	1	10.632	2.375	0.001
	0	10.595	2.049	-0.024
	Diff	0.037	0.325	0.025
	t-value	1.435	8.767	1.178
Homophily Place	1	10.645	2.376	-0.007
	0	10.587	2.028	-0.019
	Diff	0.058	0.349	-0.012
	t-value	2.203	9.313	-0.564
Homophily Caste	1	10.607	2.315	0.003
	0	10.635	2.229	-0.016
	Diff	-0.027	0.085	0.019
	t-value	-0.67	1.452	0.584

Panel E: Covariate Balance

This table shows co-variate balance between treated group (*Homophily* equals to 1) and control group (*Homophily* equals to 0) for homophily constructs based on language, place, and caste.

	Homophily Language			Homophily Place			Homophily Caste		
	Treated	Control	P-value	Treated	Control	P-value	Treated	Control	P-value
Size	19.057	19.005	0.335	19.042	19.006	0.531	19.218	19.162	0.084
Firm Age	37.918	39.730	0.069	36.874	35.922	0.339	38.621	35.979	0.000
ROA	0.063	0.061	0.631	0.066	0.069	0.465	0.067	0.069	0.260
Institutional Ownership	23.502	23.937	0.428	24.002	23.743	0.643	23.877	25.812	0.000
Promoters Ownership	52.675	52.290	0.586	52.473	53.873	0.053	53.958	51.055	0.000
Leverage	0.176	0.182	0.443	0.160	0.158	0.820	0.165	0.167	0.584
Volatility	0.023	0.023	0.386	0.023	0.023	0.755	0.022	0.022	0.282
Big 5	0.540	0.564	0.242	0.541	0.544	0.895	0.587	0.621	0.006
Number of Meetings	9.385	9.451	0.477	9.342	9.379	0.697	9.252	9.377	0.015
Board Size	12.608	12.551	0.697	12.515	12.294	0.131	12.689	12.489	0.026
Age	64.737	64.417	0.499	64.888	64.881	0.989	64.247	64.227	0.944
Gender	1.985	1.988	0.477	1.979	1.982	0.636	1.981	1.984	0.336
Education	1.982	2.016	0.303	2.048	2.077	0.410	2.047	1.951	0.000

Table 2
Group Identity and Managerial Compensation (Baseline Scenario)

Total Compensation is log of total compensation paid to the manager. Here, manager is equal to CEO or CEO equivalent positions. CEO equivalent positions are chairman, president, and managing director. *Homophily Language* is a dummy variable whose value equals to 1 if the board (on average) and the manager speaks same Indian language otherwise 0. *Homophily Place* is a dummy variable whose value equals to 1 if the board and the manager belong to same Indian state otherwise 0. *Homophily Caste* is dummy variable whose value equals to 1 if the board (on average) and the manager belong to same caste otherwise 0. To control for unobservables, firm, year, and managerial designation fixed effects are used. All regressions include a constant. Standard errors are in parentheses. * indicates significance at 10% level; ** indicates significance at 5% level; *** indicates significance at 1% level.

VARIABLES	(1) Total Compensation	(2) Total Compensation	(3) Total Compensation
Homophily_language	0.066*** (0.021)		
Homophily_place		0.071*** (0.022)	
Homophily_caste			-0.141*** (0.041)
Constant	10.583*** (0.011)	10.583*** (0.011)	10.737*** (0.037)
Observations	28,026	28,026	28,026
R-squared	0.758	0.758	0.758
Manager Designation	Yes	Yes	Yes
Firm	Yes	Yes	Yes
Year	Yes	Yes	Yes

Table 3
Group Identity and Managerial Compensation (With Controls)

This table shows the impact of group identity (or homophily) on managerial compensation controlling for firm fundamentals and governance characteristics. *Log compensation* is log of total compensation paid to the manager. Here manager is equal to CEO or CEO equivalent positions (i.e. chairman, president, and managing director). *Homophily Language* is dummy variable whose value equal to 1 if the board (on average) and the manager speaks same Indian language otherwise 0. *Homophily Place* is a dummy variable whose value equal to 1 if the board and the manager belong to same Indian state, otherwise 0. *Homophily Caste* is a dummy variable whose value equal to 1 if the board (on average) and the manager belong to same caste, otherwise 0. *Size* is measured as log of market capitalization. *Firm Age* is the age of the firm from year of incorporation. *Promoters Ownership* is percentage of promoters equity holding to the total equity issued by the firm. *Institutional Ownership* is percentage of institutional investors equity holding to the total equity issued by the firm. *Leverage* is the ratio of total debt by total assets. *ROA* is measured as ratio of net income to total assets. *Board Size* is total number of directors on board. *Volatility* is return volatility. *BIG5* is a dummy variables which takes value 1 if the firm's auditor is BIG 5 auditor; otherwise 0. *Number of Meetings* is number of board meetings in a year. To control for unobservables, firm, year, and managerial designation fixed effects are used. All regressions include a constant. Standard errors are in parentheses. * indicates significance at 10% level; ** indicates significance at 5% level; *** indicates significance at 1% level.

VARIABLES	(1) Total Compensation	(2) Total Compensation	(3) Total Compensation
Homophily Language	0.078*** (0.021)		
Homophily Place		0.077*** (0.022)	
Homophily Caste			-0.138*** (0.041)
Size	0.094*** (0.017)	0.095*** (0.017)	0.093*** (0.017)
Firm Age	0.032* (0.017)	0.032* (0.017)	0.032* (0.017)
Institutional Ownership	0.015*** (0.001)	0.015*** (0.001)	0.015*** (0.001)
Promoters Ownership	0.007*** (0.001)	0.007*** (0.001)	0.007*** (0.001)
ROA	1.133*** (0.099)	1.130*** (0.099)	1.126*** (0.099)
Leverage	-0.148** (0.068)	-0.146** (0.068)	-0.149** (0.068)
Volatility	-5.924*** (0.961)	-5.915*** (0.961)	-5.883*** (0.961)
BIG5	0.096** (0.038)	0.094** (0.038)	0.096** (0.038)
Number of Meetings	0.015*** (0.004)	0.015*** (0.004)	0.015*** (0.004)
Board Size	-0.005 (0.004)	-0.005 (0.004)	-0.005 (0.004)
Constant	7.431*** (0.668)	7.414*** (0.668)	7.593*** (0.668)
Observations	28,026	28,026	28,026
R-squared	0.762	0.762	0.762
Manager Designation	Yes	Yes	Yes
Firm	Yes	Yes	Yes
Year	Yes	Yes	Yes

Table 4
Group Identity and Firm Value

This table shows the regression results of firm value (measured as price to book ratio) on group identity(measured as Homophily). *Firm Value* is the ratio of market price of a equity to the book value of a equity. *Homophily Language* is a dummy variable whose value equal to 1 if the board (on average) and the manager speaks same Indian language otherwise 0. *Homophily Place* is dummy variable whose value equal to 1 if the board and the manager belong to the same Indian state, otherwise 0. *Homophily Caste* is a dummy variable whose value equal to 1 if the board (on average) and the manager belong to same caste, otherwise 0. *Size* is measured as log of market capitalization. *Firm Age* is the age of the firm from year of incorporation. *Promoters Ownership* is percentage of promoters equity holding to the total equity issued by the firm. *Institutional Ownership* is percentage of institutional investors equity holding to the total equity issued by the firm. *Leverage* is the ratio of total debt by total assets. *Board Size* is total number of directors on board. *Volatility* is return volatility. *BIG5* is a dummy variables which takes value 1 if the firm's auditor is BIG 5 auditor; otherwise 0. *Number of Meetings* is number of board meetings in a year. To control for unobservables, firm and year fixed effects are used. All regressions include a constant. Standard errors are in parentheses. * indicates significance at 10% level; ** indicates significance at 5% level; *** indicates significance at 1% level.

VARIABLES	(1) Firm Value	(2) Firm Value	(3) Firm Value
Homophily Language	0.115*** (0.040)		
Homophily Place		0.058 (0.041)	
Homophily Caste			-0.142* (0.077)
Size	0.244*** (0.033)	0.245*** (0.033)	0.243*** (0.033)
Firm Age	0.049 (0.033)	0.049 (0.033)	0.049 (0.033)
Institutional Ownership	0.053*** (0.003)	0.053*** (0.003)	0.053*** (0.003)
Promoters Ownership	0.047*** (0.002)	0.047*** (0.002)	0.047*** (0.002)
Leverage	-0.608*** (0.125)	-0.606*** (0.125)	-0.608*** (0.125)
Volatility	-11.853*** (1.818)	-11.820*** (1.818)	-11.789*** (1.818)
Big5	0.322*** (0.072)	0.321*** (0.072)	0.322*** (0.072)
Number of Meetings	0.050*** (0.007)	0.050*** (0.007)	0.050*** (0.007)
Board Size	0.009 (0.007)	0.009 (0.007)	0.009 (0.007)
Constant	-6.913*** (1.266)	-6.884*** (1.266)	-6.719*** (1.267)
Observations	28,026	28,026	28,026
R-squared	0.577	0.577	0.577
Firm	Yes	Yes	Yes
Year	Yes	Yes	Yes

Table 5
Group Identity and Residual Firm Value

This table shows the regression results of residual firm value (measured as Residual Firm Value) on group identity(measured as Homophily). *Residual Firm Value* is the residual of regressing price to book ratio on last year's managerial total compensation. *Homophily Language* is a dummy variable whose value equal to 1 if the board (on average) and the manager speaks same Indian language, otherwise 0. *Homophily Place* is a dummy variable whose value equal to 1 if the board and the manager belong to same Indian state, otherwise 0. *Homophily Caste* is a dummy variable whose value equal to 1 if the board (on average) and the manager belong to same caste, otherwise 0. *Size* is measured as log of market capitalization. *Firm Age* is the age of the firm from year of incorporation. *Promoters Ownership* is percentage of promoters equity holding to the total equity issued by the firm. *Institutional Ownership* is percentage of institutional investors equity holding to the total equity issued by the firm. *Leverage* is the ratio of total debt by total assets. *Board Size* is total number of directors on board. *Volatility* is return volatility. *BIG5* is a dummy variables which takes value 1 if the firm's auditor is BIG 5 auditor; otherwise 0. *Number of Meetings* is number of board meetings in a year. To control for unobservables, firm and year fixed effects are used. All regressions include a constant. Standard errors are in parentheses. * indicates significance at 10% level; ** indicates significance at 5% level; *** indicates significance at 1% level.

VARIABLES	(1) Residual Firm Value	(2) Residual Firm Value	(3) Residual Firm Value
Homophily Language	0.086** (0.034)		
Homophily Place		0.056 (0.035)	
Homophily Caste			-0.114* (0.065)
Size	0.229*** (0.028)	0.230*** (0.028)	0.228*** (0.028)
Firm Age	0.040 (0.028)	0.040 (0.028)	0.040 (0.028)
Institutional Ownership	0.050*** (0.002)	0.050*** (0.002)	0.050*** (0.002)
Promoters Ownership	0.042*** (0.002)	0.042*** (0.002)	0.042*** (0.002)
Leverage	-0.541*** (0.106)	-0.539*** (0.106)	-0.541*** (0.106)
Volatility	-11.677*** (1.542)	-11.655*** (1.542)	-11.628*** (1.542)
Big5	0.268*** (0.061)	0.266*** (0.061)	0.268*** (0.061)
Number of Meetings	0.049*** (0.006)	0.049*** (0.006)	0.049*** (0.006)
Board Size	0.007 (0.006)	0.007 (0.006)	0.007 (0.006)
Constant	-8.251*** (1.074)	-8.242*** (1.074)	-8.100*** (1.074)
Observations	28,026	28,026	28,026
R-squared	0.062	0.061	0.061
Firm	Yes	Yes	Yes
Year	Yes	Yes	Yes

Table 6
Panel A: External Validity of Dependent Variable

This table shows the regression results using the alternative dependent variable. *Ind_Average Relative* is log compensation relative to industry average. *is log of salary*. *Homophily Language* is a dummy variable whose value equals to 1 if the board (on average) and the manager speaks same Indian language, otherwise 0. *Homophily Place* is a dummy variable whose value equal to 1 if the board and the manager belong to same Indian state, otherwise 0. *Homophily Caste* is a dummy variable whose value equal to 1 if the board (on average) and the manager belong to same caste, otherwise 0. *Size* is measured as log of market capitalization. *Firm Age* is the age of the firm from year of incorporation. *Promoters Ownership* is percentage of promoters equity holding to the total equity issued by the firm. *Institutional Ownership* is percentage of institutional investors equity holding to the total equity issued by the firm. *Leverage* is the ratio of total debt by total assets. *Board Size* is total number of directors on board. *Volatility* is return volatility. *BIG5* is a dummy variables which takes value 1 if the firm's auditor is BIG 5 auditor; otherwise 0. *Number of Meetings* is number of board meetings in a year. To control for unobservables, firm and year fixed effects are used. All regressions include a constant. Standard errors are in parentheses. * indicates significance at 10% level; ** indicates significance at 5% level; *** indicates significance at 1% level.

VARIABLES	(1) Ind_Average Relative	(2) Ind_Average Relative	(3) Ind_Average Relative
Homophily Language	0.046** (0.021)		
Homophily Place		0.024 (0.022)	
Homophily Caste			-0.117*** (0.041)
Size	0.083*** (0.018)	0.083*** (0.018)	0.082*** (0.018)
Firm Age	0.003 (0.017)	0.003 (0.017)	0.003 (0.017)
Institutional Ownership	0.012*** (0.002)	0.012*** (0.002)	0.012*** (0.002)
Promoters Ownership	0.006*** (0.001)	0.006*** (0.001)	0.006*** (0.001)
ROA	0.857*** (0.099)	0.855*** (0.099)	0.852*** (0.099)
Leverage	-0.115* (0.068)	-0.114* (0.068)	-0.116* (0.068)
Volatility	-3.661*** (0.968)	-3.649*** (0.968)	-3.634*** (0.968)
BIG5	0.093** (0.038)	0.092** (0.038)	0.093** (0.038)
Number of Meetings	0.012*** (0.004)	0.012*** (0.004)	0.012*** (0.004)
Board Size	-0.004 (0.004)	-0.005 (0.004)	-0.004 (0.004)
Constant	-1.957*** (0.673)	-1.946*** (0.673)	-1.833*** (0.673)
Observations	28,026	28,026	28,026
R-squared	0.737	0.737	0.737
Manager Designation	Yes	Yes	Yes
Firm	Yes	Yes	Yes
Year	Yes	Yes	Yes

Panel B : External Validity Test of Homophily Construct

This table shows the regression results of using alternative homophily variable. To check the external validity of the homophily variables, rather than using the SECC 2011 and LSI data, data from Indianchildnames.com is used. *Homophily* is a dummy variable whose value equals to 1 if the board (on average) and the manager speaks same Indian language, otherwise 0. *Size* is measured as log of market capitalization. *Leverage* is the ratio of total debt by total assets. *Board Size* is total number of directors on board. *ROA* is measured as ratio of net income to total assets. To control for unobservables, firm and year fixed effects are used. All regressions include a constant. Standard errors are in parentheses. * indicates significance at 10% level; ** indicates significance at 5% level; *** indicates significance at 1% level.

VARIABLES	(1) Total Compensation	(2) Salary
Homophily	0.138*** (0.032)	0.102*** (0.032)
Size	0.368*** (0.029)	0.337*** (0.024)
ROA	1.364*** (0.136)	0.854*** (0.145)
Leverage	-0.308*** (0.091)	-0.228** (0.089)
Board Size	0.004 (0.007)	-0.002 (0.005)
Constant	7.935*** (0.226)	8.016*** (0.189)
Observations	14,523	14,229
R-squared	0.805	0.804
Firm Fixed Effects	Yes	Yes
Year Fixed Effects	Yes	Yes

Table 7
Endogeneity Issue and the Identification Test

This table shows the regression results for identification test. *Total Compensation* is log of total compensation. *Firm Value* is the ratio of market price of an equity to the book value of an equity. Residual Firm Value is the firm value condition on managerial compensation of last year. *Death* is a dummy variable which takes value 1 if any director died or retired in a firm in that year; otherwise 0. This variable is based on Fracassi and Tate (2012). *Homophily Language* is a dummy variable whose value equal to 1 if the board (on average) and the manager speaks same Indian language, otherwise 0. *Homophily Place* is a dummy variable whose value equal to 1 if the board and the manager belong to same Indian state, otherwise 0. *Homophily Caste* is a dummy variable whose value equal to 1 if the board (on average) and the manager belong to same caste, otherwise 0. *Size* is measured as log of market capitalization. *Firm Age* is the age of the firm from year of incorporation. *Promoters Ownership* is percentage of promoters equity holding to the total equity issued by the firm. *Institutional Ownership* is percentage of institutional investors equity holding to the total equity issued by the firm. *Leverage* is the ratio of total debt by total assets. *ROA* is measured as ratio of net income to total assets. *Board Size* is total number of directors on board. *Volatility* is return volatility. *BIG5* is a dummy variables which takes value 1 if the firm's auditor is BIG 5 auditor; otherwise 0. *Number of Meetings* is number of board meetings in a year. To control for unobservables, firm and managerial designation fixed effects are used. All regressions include a constant. Standard errors are in parentheses. * indicates significance at 10% level; ** indicates significance at 5% level; *** indicates significance at 1% level.

[illegible]

Table 8

Panel A: Community Migration and Group Identity

This table shows the regression results excluding the surnames for those communities which migrated from part of India to another after independence. *Total Compensation* is log of total compensation. *Firm Value* is the ratio of market price of an equity to the book value of an equity. *Residual Firm Value* is the firm value condition on managerial compensation of last year. *Homophily Language* is a dummy variable whose value equals to 1 if the board (on average) and the manager speaks same Indian language, otherwise 0. *Homophily Place* is a dummy variable whose value equal to 1 if the board and the manager belong to same Indian state, otherwise 0. *Homophily Caste* is a dummy variable whose value equal to 1 if the board (on average) and the manager belong to same caste, otherwise 0. *Size* is measured as log of market capitalization. *Firm Age* is the age of the firm from year of incorporation. *Promoters Ownership* is percentage of promoters equity holding to the total equity issued by the firm. *Institutional Ownership* is percentage of institutional investors equity holding to the total equity issued by the firm. *Leverage* is the ratio of total debt by total assets. *ROA* is measured as ratio of net income to total assets. *Board Size* is total number of directors on board. *Volatility* is return volatility. *BIG5* is a dummy variables which takes value 1 if the firm's auditor is BIG 5 auditor; otherwise 0. *Number of Meetings* is number of board meetings in a year. To control for unobservables, firm, year, and managerial designation fixed effects are used. All regressions include a constant. Standard errors are in parentheses. * indicates significance at 10% level; ** indicates significance at 5% level; *** indicates significance at 1% level.

[illegible]

Panel B: Geographical Differences

This table shows the regression results of homophily difference in southern Indian states compared to northern Indian states on executive compensation, firm value, and residual firm value. *South* is dummy variable which takes value if a manager comes from southern Indian states; otherwise 0. *Total Compensation* is log of total compensation. *Firm Value* is the ratio of market price of an equity to the book value of an equity. *Residual Firm Value* is the firm value condition on managerial compensation of last year. *Homophily Language* is a dummy variable whose value equals to 1 if the board (on average) and the manager speaks same Indian language, otherwise 0. *Homophily Place* is a dummy variable whose value equal to 1 if the board and the manager belong to same Indian state, otherwise 0. *Homophily Caste* is a dummy variable whose value equal to 1 if the board (on average) and the manager belong to same caste, otherwise 0. *Size* is measured as log of market capitalization. *Firm Age* is the age of the firm from year of incorporation. *Promoters Ownership* is percentage of promoters equity holding to the total equity issued by the firm. *Institutional Ownership* is percentage of institutional investors equity holding to the total equity issued by the firm. *Leverage* is the ratio of total debt by total assets. *ROA* is measured as ratio of net income to total assets. *Board Size* is total number of directors on board. *Volatility* is return volatility. *BIG5* is a dummy variables which takes value 1 if the firm's auditor is BIG 5 auditor; otherwise 0. *Number of Meetings* is number of board meetings in a year. To control for unobservables, firm, year, and managerial designation fixed effects are used. All regressions include a constant. Standard errors are in parentheses. * indicates significance at 10% level; ** indicates significance at 5% level; *** indicates significance at 1% level.

VARIABLES	(1) Total Compensation	(2) Total Compensation	(3) Total Compensation	(4) Firm Value	(5) Firm Value	(6) Firm Value	(7) Residual Firm Value	(8) Residual Firm Value	(9) Residual Firm Value
Homophily Place	0.037* (0.020)				-0.019 (0.039)			0.027 (0.024)	
South	-0.116*** (0.028)	-0.122*** (0.028)	-0.159* (0.085)	0.226*** (0.053)	0.245*** (0.054)	0.169 (0.164)	-0.048 (0.033)	-0.034 (0.033)	0.057 (0.102)
Homophily Place x South	0.131*** (0.044)				-0.105 (0.085)			0.032 (0.053)	
Homophily Language		0.020 (0.020)		0.014 (0.038)			0.034 (0.024)		
Homophily Language x South		0.156*** (0.045)		-0.056 (0.085)			0.073 (0.053)		
Homophily Caste			-0.045 (0.030)			-0.051 (0.058)			0.000 (0.036)
Homophily Caste x South			0.103 (0.088)			0.041 (0.169)			-0.084 (0.105)
Constant	7.730*** (0.128)	7.737*** (0.129)	7.799*** (0.130)	-1.248*** (0.245)	-1.221*** (0.245)	-1.201*** (0.248)	0.113 (0.152)	0.124 (0.152)	0.146 (0.154)
Observations	28,026	28,026	28,026	28,026	28,026	28,026	28,026	28,026	28,026
R-squared	0.542	0.542	0.542	0.165	0.165	0.165	0.007	0.007	0.007
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Manager Designation	Yes	Yes	Yes	No	No	No	No	No	No
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 9
Powerful CEO

This table shows the regression results including the variable on CEO's influence in firm i.e. CEO power. *Total Compensation* is log of total compensation paid to the manager. Here, manager is equal to CEO or CEO equivalent positions. CEO equivalent positions are chairman, president, and managing director.. *Firm Value* is the ratio of market price of an equity to the book value of an equity. *Residual Firm Value* is the firm value condition on managerial compensation of last year. *Homophily Language* is a dummy variable whose value equals to 1 if the board (on average) and the manager speaks same Indian language, otherwise 0. *Homophily Place* is a dummy variable whose value equal to 1 if the board and the manager belong to same Indian state, otherwise 0. *Homophily Caste* is a dummy variable whose value equal to 1 if the board (on average) and the manager belong to same caste, otherwise 0. *Powerful CEO* is a dummy variable which takes value 1 if manager is also the founder of the firm or manager holds all the three key positions (CEO, Chairman, and Managing Director) in a firm; otherwise 0. This variable is based on Adams et al.(2005). *Size* is measured as log of market capitalization. *Firm Age* is the age of the firm from year of incorporation. *Promoters Ownership* is percentage of promoters equity holding to the total equity issued by the firm. *Institutional Ownership* is percentage of institutional investors equity holding to the total equity issued by the firm. *Leverage* is the ratio of total debt by total assets. *ROA* is measured as ratio of net income to total assets. *Board Size* is total number of directors on board. *Volatility* is return volatility. *BIG5* is a dummy variables which takes value 1 if the firm's auditor is BIG 5 auditor; otherwise 0. *Number of Meetings* is number of board meetings in a year. To control for unobservables, firm, year, and managerial designation fixed effects are used. All regressions include a constant. Standard errors are in parentheses. * indicates significance at 10% level; ** indicates significance at 5% level; *** indicates significance at 1% level.

VARIABLES	(1) Total Compensation	(2) Total Compensation	(3) Total Compensation	(4) Firm Value	(5) Firm Value	(6) Firm Value	(7) Residual Firm Value	(8) Residual Firm Value	(9) Residual Firm Value
Homophily Language	0.043** (0.018)			0.008 (0.035)			0.049** (0.021)		
Homophily Place		0.054*** (0.018)			-0.029 (0.035)			0.033 (0.022)	
Homphily Caste			-0.039 (0.028)			-0.025 (0.054)			-0.012 (0.034)
Powerful CEO	0.152*** (0.019)	0.151*** (0.019)	0.155*** (0.019)	-0.169*** (0.035)	-0.166*** (0.035)	-0.168*** (0.035)	-0.002 (0.022)	-0.001 (0.022)	0.001 (0.022)
Constant	7.646*** (0.128)	7.642*** (0.128)	7.710*** (0.129)	-1.060*** (0.244)	-1.036*** (0.244)	-1.032*** (0.247)	0.094 (0.151)	0.108 (0.151)	0.139 (0.154)
Observations	28,026	28,026	28,026	28,026	28,026	28,026	28,026	28,026	28,026
R-squared	0.543	0.543	0.543	0.165	0.165	0.165	0.007	0.007	0.006
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Manager Designation	Yes	Yes	Yes	No	No	No	No	No	No
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 10
Panel A : College Ties

This table shows the regression results including the variable on college ties. *Total Compensation* is log of total compensation paid to the manager. Here, manager is equal to CEO or CEO equivalent positions. CEO equivalent positions are chairman, president, and managing director. *Firm Value* is the ratio of market price of an equity to the book value of an equity. *Residual Firm Value* is the firm value condition on managerial compensation of last year. *Homophily Language* is a dummy variable whose value equals to 1 if the board (on average) and the manager speaks same Indian language, otherwise 0. *Homophily Place* is a dummy variable whose value equal to 1 if the board and the manager belong to same Indian state, otherwise 0. *Homophily Caste* is a dummy variable whose value equal to 1 if the board (on average) and the manager belong to same caste, otherwise 0. *College Ties* is a dummy variable which takes value 1 if manager and any of the director attend the same college; otherwise 0. *Size* is measured as log of market capitalization. *Firm Age* is the age of the firm from year of incorporation. *Promoters Ownership* is percentage of promoters equity holding to the total equity issued by the firm. *Institutional Ownership* is percentage of institutional investors equity holding to the total equity issued by the firm. *Leverage* is the ratio of total debt by total assets. *ROA* is measured as ratio of net income to total assets. *Board Size* is total number of directors on board. *Volatility* is return volatility. *BIG5* is a dummy variables which takes value 1 if the firm's auditor is BIG 5 auditor; otherwise 0. *Number of Meetings* is number of board meetings in a year. To control for unobservables, firm, year, and managerial designation fixed effects are used. All regressions include a constant. Standard errors are in parentheses. * indicates significance at 10% level; ** indicates significance at 5% level; *** indicates significance at 1% level.

VARIABLES	(1) Total Compensation	(2) Total Compensation	(3) Total Compensation	(4) Firm Value	(5) Firm Value	(6) Firm Value	(7) Residual Firm Value	(8) Residual Firm Value	(9) Residual Firm Value
Homophily Language	0.131** (0.058)			0.182 (0.142)			0.154* (0.082)		
Homophily Place		0.126** (0.060)			-0.326** (0.146)			0.023 (0.085)	
Homophily Caste			-0.125 (0.087)			-0.221 (0.214)			-0.039 (0.124)
College Ties	0.071 (0.061)	0.068 (0.061)	0.080 (0.061)	-0.070 (0.149)	-0.034 (0.149)	-0.056 (0.149)	-0.014 (0.086)	-0.007 (0.087)	-0.005 (0.086)
Constant	10.976*** (0.435)	10.981*** (0.436)	11.190*** (0.435)	6.430*** (1.066)	6.937*** (1.066)	6.764*** (1.067)	2.449*** (0.618)	2.583*** (0.618)	2.633*** (0.618)
Observations	3,299	3,299	3,299	3,299	3,299	3,299	3,299	3,299	3,299
R-squared	0.412	0.412	0.411	0.237	0.238	0.237	0.075	0.074	0.074
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Manager Designation	Yes	Yes	Yes	No	No	No	No	No	No
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 10
Panel B : Past Experiences

This table shows the regression results including the ties from past employment. *Total Compensation* is log of total compensation paid to the manager. Here, manager is equal to CEO or CEO equivalent positions. CEO equivalent positions are chairman, president, and managing director. *Firm Value* is the ratio of market price of an equity to the book value of an equity. *Residual Firm Value* is the firm value condition on managerial compensation of last year. *Homophily Language* is a dummy variable whose value equals to 1 if the board (on average) and the manager speaks same Indian language, otherwise 0. *Homophily Place* is a dummy variable whose value equal to 1 if the board and the manager belong to same Indian state, otherwise 0. *Homophily Caste* is a dummy variable whose value equal to 1 if the board (on average) and the manager belong to same caste, otherwise 0. *Experience Ties* is a dummy variable which takes value 1 if manager knew any director from past employment; otherwise 0. *Size* is measured as log of market capitalization. *Firm Age* is the age of the firm from year of incorporation. *Promoters Ownership* is percentage of promoters equity holding to the total equity issued by the firm. *Institutional Ownership* is percentage of institutional investors equity holding to the total equity issued by the firm. *Leverage* is the ratio of total debt by total assets. *ROA* is measured as ratio of net income to total assets. *Board Size* is total number of directors on board. *Volatility* is return volatility. *BIG5* is a dummy variables which takes value 1 if the firm's auditor is BIG 5 auditor; otherwise 0. *Number of Meetings* is number of board meetings in a year. To control for unobservables, firm, year, and managerial designation fixed effects are used. All regressions include a constant. Standard errors are in parentheses. * indicates significance at 10% level; ** indicates significance at 5% level; *** indicates significance at 1% level.

VARIABLES	(1) Total Compensation	(2) Total Compensation	(3) Total Compensation	(4) Firm Value	(5) Firm Value	(6) Firm Value	(7) Residual Firm Value	(8) Residual Firm Value	(9) Residual Firm Value
Homophily Language	0.145** (0.064)			0.280* (0.159)			0.283*** (0.096)		
Homophily Place		0.183*** (0.067)			-0.017 (0.165)			0.252** (0.099)	
Homophily Caste			-0.229** (0.096)			-0.009 (0.239)			0.181 (0.144)
Experience Ties	-0.065 (0.068)	-0.066 (0.068)	-0.070 (0.068)	-0.331* (0.170)	-0.333* (0.170)	-0.333* (0.170)	-0.080 (0.102)	-0.082 (0.102)	-0.078 (0.102)
Constant	10.663*** (0.468)	10.604*** (0.469)	10.938*** (0.467)	5.951*** (1.153)	6.243*** (1.157)	6.231*** (1.152)	2.334*** (0.693)	2.336*** (0.696)	2.503*** (0.693)
Observations	2,921	2,921	2,921	2,921	2,921	2,921	2,921	2,921	2,921
R-squared	0.425	0.425	0.425	0.222	0.221	0.221	0.081	0.080	0.079
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Manager Designation	Yes	Yes	Yes	No	No	No	No	No	No
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 11
Promoter as a Manager

This table shows the regression results of managerial compensation and firm value (measured as price to book ratio) on group identity (measured as Homophily) when manager is one of the promoters. *Total Compensation* is log of total compensation paid to the manager. Here, manager is equal to CEO or CEO equivalent positions. CEO equivalent positions are chairman, president, and managing director. Firm Value is the ratio of market price of a equity to the book value of a equity. *Firm Value* is the ratio of market price of an equity to the book value of an equity. *Residual Firm Value* is the firm value condition on managerial compensation of last year. *Homophily Language* is a dummy variable whose value equals to 1 if the board (on average) and the manager speaks same Indian language, otherwise 0. *Homophily Place* is a dummy variable whose value equal to 1 if the board and the manager belong to same Indian state, otherwise 0. *Homophily Caste* is a dummy variable whose value equal to 1 if the board (on average) and the manager belong to same caste, otherwise 0. *Size* is measured as log of market capitalization. *Firm Age* is the age of the firm from year of incorporation. *Promoters Ownership* is percentage of promoters equity holding to the total equity issued by the firm. *Institutional Ownership* is percentage of institutional investors equity holding to the total equity issued by the firm. *Leverage* is the ratio of total debt by total assets. *Board Size* is total number of directors on board. *Volatility* is return volatility. *BIG5* is a dummy variables which takes value 1 if the firm's auditor is BIG 5 auditor; otherwise 0. *Number of Meetings* is number of board meetings in a year. To control for unobservables, firm and year fixed effects are used. All regressions include a constant. Standard errors are in parentheses. * indicates significance at 10% level; ** indicates significance at 5% level; *** indicates significance at 1% level.

VARIABLES	(1) Total Compensation	(2) Total Compensation	(3) Total Compensation	(4) Firm Value	(5) Firm Value	(6) Firm Value	(7) Residual Firm Value	(8) Residual Firm Value	(9) Residual Firm Value
Homophily Language	0.056 (0.038)			0.186*** (0.068)			0.146** (0.057)		
Homophily Place		0.017 (0.039)			0.156** (0.070)			0.112* (0.059)	
Homophily Caste			-0.189** (0.082)			-0.236 (0.147)			-0.207* (0.123)
Constant	6.829*** (1.031)	6.852*** (1.032)	7.036*** (1.033)	-10.207*** (1.849)	-10.219*** (1.847)	-9.869*** (1.857)	-10.961*** (1.558)	-10.980*** (1.559)	-10.670*** (1.561)
Observations	10,083	10,083	10,083	10,083	10,083	10,083	10,083	10,083	10,083
R-squared	0.773	0.773	0.773	0.568	0.570	0.568	0.172	0.172	0.172
Manager Designation	Yes	Yes	Yes	No	No	No	No	No	No
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 12
Two-Stage Least Square (2SLS) Regression

This table shows the regression results of two-stage least square. *Total Compensation* is log of total compensation paid to the manager. Here, manager is equal to CEO or CEO equivalent positions. CEO equivalent positions are chairman, president, and managing director. Firm Value is the ratio of market price of a equity to the book value of a equity. *Firm Value* is the ratio of market price of an equity to the book value of an equity. *TotalCompensation* is the fitted value of managerial compensation from first stage. *Homophily Language* is a dummy variable whose value equals to 1 if the board (on average) and the manager speaks same Indian language, otherwise 0. *Homophily Place* is a dummy variable whose value equal to 1 if the board and the manager belong to same Indian state, otherwise 0. *Homophily Caste* is a dummy variable whose value equal to 1 if the board (on average) and the manager belong to same caste, otherwise 0. *Promoter* is dummy variable which takes value 1 if manager is also one of the promoters; otherwise 0. *Age* is manager's age. Other controls are: *Size*, *Firm Age*, *Promoters Ownership*, *Institutional Ownership*, *Leverage*, *Board Size*, *Volatility*, *BIG5*, and *Number of Meetings*. To control for unobservables, firm and year fixed effects are used. All regressions include a constant. Standard errors are in parentheses. * indicates significance at 10% level; ** indicates significance at 5% level; *** indicates significance at 1% level.

VARIABLES	(1) Total Compensation	(2) Firm Value	(3) Firm Value	(4) Firm Value
Homophily Language		0.427*** (0.135)		
Homophily Place			0.231 (0.151)	
Homophily Caste				0.138 (0.223)
<i>TotalCompensation</i>		8.987*** (0.438)	9.059*** (0.438)	9.074*** (0.438)
Firm Value	0.012 (0.008)			
Age	0.003 (0.003)			
Promoter	0.272*** (0.068)			
Constant	3.573* (2.116)	-34.611*** (4.480)	-34.855*** (4.480)	-34.593*** (4.532)
Observations	3,542	3,542	3,542	3,542
R-squared	0.750	0.768	0.768	0.767
Other Controls	Yes	Yes	Yes	Yes
Manager Designation	Yes	No	No	No
Firm	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes

Table 13
Business Group vs. Non Business Group Firms

This table shows the regression results of managerial compensation and firm value (measured as price to book ratio) on group identity (measured as Homophily) for business group firms and non-business group firms. *Total Compensation* is log of total compensation paid to the manager. Here, manager is equal to CEO or CEO equivalent positions. CEO equivalent positions are chairman, president, and managing director. Firm Value is the ratio of market price of a equity to the book value of an equity. *Firm Value* is the ratio of market price of an equity to the book value of an equity. *Residual Firm Value* is the firm value condition on managerial compensation of last year. *Homophily Language* is a dummy variable whose value equals to 1 if the board (on average) and the manager speaks same Indian language, otherwise 0. *Size* is measured as log of market capitalization. *Firm Age* is the age of the firm from year of incorporation. *Promoters Ownership* is percentage of promoters equity holding to the total equity issued by the firm. *Institutional Ownership* is percentage of institutional investors equity holding to the total equity issued by the firm. *Leverage* is the ratio of total debt by total assets. *Board Size* is total number of directors on board. *Volatility* is return volatility. *BIG5* is a dummy variables which takes value 1 if the firm's auditor is BIG 5 auditor; otherwise 0. *Number of Meetings* is number of board meetings in a year. To control for unobservables, firm and year fixed effects are used. All regressions include a constant. Standard errors are in parentheses. * indicates significance at 10% level; ** indicates significance at 5% level; *** indicates significance at 1% level.

VARIABLES	Business Group			Non-Business Group		
	(1) Total Compensation	(2) Firm Value	(3) Residual Firm Value	(4) Total Compensation	(5) Firm Value	(6) Residual Firm Value
Homophily Language	0.066** (0.032)	0.188*** (0.058)	0.144*** (0.049)	0.089*** (0.028)	0.030 (0.055)	0.020 (0.047)
Size	0.064** (0.026)	0.152*** (0.048)	0.120*** (0.041)	0.126*** (0.023)	0.306*** (0.045)	0.313*** (0.038)
Firm Age	0.033 (0.021)	-0.021 (0.038)	-0.016 (0.033)	0.023 (0.033)	0.236*** (0.065)	0.184*** (0.055)
ROA	1.429*** (0.167)			0.927*** (0.119)		
Institutional Ownership	0.011*** (0.002)	0.053*** (0.004)	0.050*** (0.003)	0.018*** (0.002)	0.054*** (0.004)	0.051*** (0.003)
Promoters Ownership	0.004* (0.002)	0.058*** (0.003)	0.051*** (0.003)	0.010*** (0.002)	0.037*** (0.003)	0.034*** (0.003)
Leverage	-0.062 (0.101)	-0.837*** (0.177)	-0.701*** (0.150)	-0.231** (0.091)	-0.362** (0.175)	-0.374** (0.148)
Volatility	-8.260*** (1.507)	-5.184* (2.736)	-6.614*** (2.323)	-3.736*** (1.223)	-18.004*** (2.423)	-16.306*** (2.053)
Big5	0.038 (0.052)	0.367*** (0.095)	0.284*** (0.081)	0.178*** (0.056)	0.224** (0.110)	0.214** (0.093)
Number of Meetings	0.020*** (0.006)	0.051*** (0.010)	0.051*** (0.009)	0.012*** (0.005)	0.047*** (0.009)	0.046*** (0.008)
Board Size	-0.007 (0.006)	-0.008 (0.010)	-0.008 (0.009)	-0.003 (0.005)	0.025** (0.010)	0.022** (0.009)
Constant	8.105*** (0.958)	-3.638** (1.743)	-5.126*** (1.480)	6.934*** (1.060)	-12.563*** (2.102)	-13.062*** (1.781)
Observations	13,387	13,387	13,387	14,639	14,639	14,639
R-squared	0.769	0.577	0.061	0.752	0.581	0.071
Manager Designation	Yes	No	No	Yes	No	No
Firm	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
Business Group	Yes	Yes	Yes	No	No	No

Appendix

Appendix A

A1: Data Filtering

	Firms	Observations
Merging with surname and social identities data	24326	1473836
Total remuneration is given	13226	285159
BSE and NSE listed non- financial firms (Overall Board)	3668	208275
BSE and NSE listed non- financial firms (Managers)	3668	41507
Balanced Sample	2324	28026

A2: Yearly Distribution

Year	Unique Firms	Observations
2004	237	360
2005	365	554
2006	959	1,386
2007	1120	1,621
2008	1258	1,797
2009	1351	1,920
2010	1391	1,943
2011	1458	2,045
2012	1546	2,172
2013	1558	2,184
2014	1539	2,128
2015	1562	2,244
2016	1760	2,564
2017	1767	2,627
2018	1552	2,481
Total		28,026

A3: Business Group Distribution

Business Group	Freq.	Percent	Cum.
0	14,639	52.23	52.23
1	13,387	47.77	100
Total	28,026	100	

A4: Homophily in Different Industries

Industry	Homophily Language	Homophily Place	Homophily Caste
Abrasives	0.273	0.182	0.896
Agricultural machinery	0.364	0.424	1.000
Air transport services	0.059	0.088	0.824
Air-conditioners & refrigerators	0.120	0.120	0.978
Aluminium & aluminium products	0.216	0.279	1.000
Animation content provider	0.294	0.235	1.000
Bakery products	0.133	0.000	0.733
Beer & alcohol	0.432	0.358	0.802
Boilers & turbines	0.207	0.276	0.931
Books & cards	0.355	0.355	0.645
Business services & consultancy	0.448	0.440	0.890
Castings & forgings	0.456	0.479	0.927
Caustic soda	0.000	0.000	1.000
Cement	0.330	0.360	0.891
Ceramic products	0.545	0.625	0.670
Cloth	0.416	0.393	0.848
Coal & lignite	0.160	0.080	1.000
Cocoa products & confectionery	0.667	0.500	1.000
Coffee	0.133	0.133	1.000
Commercial complexes	0.188	0.125	0.750
Commercial vehicles	0.130	0.014	0.812
Communication equipment	0.364	0.182	0.818
Computer software	0.313	0.314	0.920
Computers, peripherals & storage devices	0.483	0.483	0.897
Consumer electronics	0.590	0.872	1.000
Conventional electricity	0.343	0.146	0.961
Copper & copper products	0.219	0.438	0.943
Cosmetics, toiletries, soaps & detergents	0.391	0.370	0.929
Cotton & blended yarn	0.423	0.440	0.936
Courier services	0.333	0.333	1.000
Crude oil & natural gas	0.265	0.193	0.976
Dairy products	0.418	0.304	0.968
Diversified	0.474	0.433	0.883
Diversified automobile	0.000	0.000	1.000
Diversified cotton textile	0.307	0.281	0.784
Diversified machinery	0.370	0.348	0.957
Diversified metal & metal products	0.720	0.560	0.980
Drugs & pharmaceuticals	0.424	0.398	0.907
Dry cells	0.433	0.500	0.417
Dyes & pigments	0.632	0.555	0.907
Education	0.365	0.394	0.865
Electricity distribution	1.000	0.364	1.000
Electricity transmission	0.091	0.091	1.000
Engines	0.360	0.260	0.780
Exhibition of films	0.629	0.571	0.657
Ferro alloys	0.571	0.578	1.000
Fertilisers	0.307	0.259	0.820
Floriculture	1.000	1.000	1.000
Footwear	0.794	0.706	0.941
Gems & jewellery	0.502	0.515	0.783
General purpose machinery	0.431	0.434	0.912
Generators, transformers & switchgears	0.353	0.335	0.909
Glass & glassware	0.198	0.189	0.468
Granite	0.589	0.589	1.000
Health services	0.642	0.572	0.994
Hotels & restaurants	0.469	0.406	0.901
Housing construction	0.561	0.585	1.000
ITES	0.403	0.368	0.924
Industrial construction	0.286	0.292	0.842
Industrial machinery	0.530	0.561	0.944
Infrastructural construction	0.458	0.470	0.863
Inorganic chemicals	0.484	0.339	0.890
Lubricants, etc.	0.467	0.417	0.983
Machine tools	0.287	0.230	0.955
Man-made filaments & fibres	0.434	0.338	0.844
Marine foods	0.000	0.000	1.000
Media-broadcasting	0.450	0.424	0.900
Media-content	0.383	0.353	0.985
Media-print	0.644	0.637	0.778

A4: Homophily in Different Industries (Contd.)

Industry	Homophily Language	Homophily Place	Homophily Caste
Milling products	1.000	1.000	1.000
Minerals	0.455	0.455	0.791
Mining & construction equipment	0.629	0.477	0.917
Miscellaneous electrical machinery	0.311	0.311	0.977
Miscellaneous manufactured articles	0.632	0.472	0.962
Natural gas trading & distribution	0.393	0.115	0.984
Online marketplaces	0.125	0.125	0.625
Organic chemicals	0.365	0.403	0.962
Other agricultural products	0.518	0.447	0.961
Other automobile ancillaries	0.382	0.344	0.911
Other chemical products	0.431	0.447	0.941
Other construction & allied activities	0.491	0.448	0.827
Other construction materials	0.280	0.326	0.970
Other consumer goods	0.602	0.637	0.832
Other domestic appliances	0.524	0.444	0.841
Other electronics	0.389	0.436	0.940
Other ferrous metal products	0.420	0.510	0.874
Other industrial machinery	0.480	0.600	0.640
Other leather & related products	0.174	0.000	0.783
Other miscellaneous services	0.538	0.487	0.952
Other non-ferrous metals & metal products	0.224	0.168	0.825
Other recreational & allied services	0.390	0.378	0.976
Other textiles	0.476	0.363	0.855
Other transport equipment & ancillaries	0.440	0.368	0.800
Paints & varnishes	0.336	0.252	0.969
Paper & newsprint	0.334	0.268	0.863
Paper products	0.600	0.600	0.900
Passenger vehicles	0.067	0.000	0.733
Pesticides	0.395	0.328	0.993
Pig iron	1.000	1.000	1.000
Plastic films & flexible packaging	0.235	0.204	0.864
Plastic furniture, floorings & miscellaneous items	0.561	0.606	0.947
Plastic packaging goods	0.486	0.646	0.913
Plastic tubes, pipes, fittings & sheets	0.390	0.444	0.912
Polymers	0.432	0.432	0.821
Poultry & meat products	0.250	0.250	0.250
Processed foods	0.483	0.510	0.867
Production & distribution of films	0.293	0.293	0.683
Railway transport services	0.095	0.143	0.857
Readymade garments	0.409	0.440	0.789
Refinery	0.388	0.321	0.993
Refractories	0.360	0.027	1.000
Renewable electricity	0.318	0.136	1.000
Retail trading	0.332	0.234	0.889
Road transport infrastructure services	0.600	0.600	1.000
Road transport services	0.769	0.808	0.885
Rubber products	0.440	0.493	0.853
Shipping transport infrastructure services	0.246	0.000	1.000
Shipping transport services	0.374	0.325	0.894
Soda ash	0.550	0.450	1.000
Sponge iron	0.404	0.372	0.862
Starches	0.091	0.000	1.000
Steel	0.391	0.351	0.899
Steel pipes & tubes	0.544	0.424	0.921
Storage & distribution	0.667	0.383	0.433
Storage batteries	0.238	0.032	0.635
Sugar	0.350	0.355	0.928
Tea	0.520	0.376	0.769
Telecommunication services	0.305	0.374	0.920
Textile processing	0.421	0.381	0.802
Tobacco products	0.316	0.316	0.908
Tourism	0.083	0.111	0.833
Transport logistics services	0.426	0.414	0.975
Two & three wheelers	0.557	0.585	0.802
Tyres & tubes	0.541	0.557	0.907
Vegetable oils & products	0.592	0.598	0.858
Wholesale trading	0.401	0.378	0.849
Wires & cables	0.593	0.586	0.947
Wood & wood products	0.548	0.444	0.976

A5: Homophily with Time



Appendix B

B1: Distribution of Manager Designation

Designation	Frequency	Percent	Cummulative
Chairperson	7,209	25.72	25.72
Chairperson & Chief Executive Officer	102	0.36	26.09
Chairperson & Chief Mentor	7	0.02	26.11
Chairperson & Director	244	0.87	26.98
Chairperson & Executive Director	549	1.96	28.94
Chairperson & Joint Managing Director	26	0.09	29.03
Chairperson & Managing Director	6,353	22.67	51.7
Chairperson & President	12	0.04	51.74
Chairperson Emeritus	112	0.4	52.14
Chairperson, Managing Director & Chief Executive Officer	136	0.49	52.63
Chief Executive Officer	545	1.94	54.57
Chief Executive Officer & Deputy Managing Director	2	0.01	54.58
Chief Executive Officer & Director	25	0.09	54.67
Chief Executive Officer & Executive Director	84	0.3	54.97
Chief Executive Officer & Manager	5	0.02	54.99
Chief Executive Officer & President	36	0.13	55.12
Director & Chief Executive Officer	73	0.26	55.38
Executive Director & Chief Executive Officer	460	1.64	57.02
Managing Director	9,701	34.61	91.63
Managing Director & Chief Executive Officer	1,057	3.77	95.4
Managing Director & Chief Operations Officer	13	0.05	95.45
Managing Director & Company Secretary	11	0.04	95.49
Managing Director & President	49	0.17	95.66
President	146	0.52	96.19
President & Chief Financial Officer	33	0.12	96.3
President & Director	45	0.16	96.46
Vice Chairperson & Managing Director	991	3.54	100
Total	28,026	100	

Appendix C

C1: Preview of Socio-Economic Caste Census (SECC) Data

head_of_hh	gender	age		
AJAY KUMAR SAH\अजय कुमार साह	M	34		
AJAY KUMAR SINGH\अजय कुमार सिंह	M	47		
AJIT KUMAR SINGH\अजित कुमार सिंह	M	33		
AKHILESH SINGH\अखिलेश सिंह	M	47		
AMERIKA BAITHA\अमेरिका बैठा	M	32		
state	district	tehsil	panchayat	
BIHAR	Purba Champaran	Phenhara	Bishunpur Basant(Village)	
BIHAR	Purba Champaran	Phenhara	Bishunpur Basant(Village)	
BIHAR	Purba Champaran	Phenhara	Bishunpur Basant(Village)	
BIHAR	Purba Champaran	Phenhara	Bishunpur Basant(Village)	
BIHAR	Purba Champaran	Phenhara	Bishunpur Basant(Village)	
social_cat	fathers_and_mothers_name	total_members	village	
Other	गोपाल जी\सीया देवी	6	Bishunpur Basant	
Other	राजमंगल सहि\मंजू कुँवर	7	Bishunpur Basant	
Other	लालबाबू सहि\श्रीकांत देवी	5	Bishunpur Basant	
Other	रामशरय सहि\फुलमत देवी	3	Bishunpur Basant	
SC	भखिरी बैठा\धनेशरी कुँवर	2	Bishunpu	

C2: Cleaning of Socio-Economic Caste Census (SECC) Data

The SECC 2011 data contains information on individual name, parents name, gender, age, caste, address (state and district) and others. This information on 474 million Indians is available in 48 CSV files. The surname data developed from SECC provide information on surname, state, district, and caste. To scrape surnames from SECCC data, the cloud computing on the Microsoft Azure is used with an instance of 32 GB RAM and 1 TB hard disk.

The steps to scrape surnames from Indian names are: firstly, the observations with missing names are removed. Second, all the special characters and names in the native language are replaced by white space. Third, the slicing of the names is done till all surnames are obtained. Fourth, if the surname length is 2 or 1, it means either names are from south India or names are long. In SECC data, south Indian names and long names are recorded differently. For example: names like "P.K. Chandrashekhar Rao" is penned as "Chandrashekhar Rao P.K." in the SECC data. Therefore, when surnames are either 1 or 2 length, it means those are not surnames but initials. These initials are replaced by second name if the name has three words such as "Rao" from "Chandrashekhar Rao P.K." or with first name if a name has two words such as "Gupta" from "Gupta S.K." Fifth, even after all these steps, if there are surnames with length 1 or 2, those surnames are removed to curb noise in the final dataset. Otherwise, final dataset would contain certain discrepancies such as surname "r" which is part of "Sahil Kuma r", but, in this name, surname is "Kumar" and not "r". All the designations such as Shri, Sh., Mr., Mr, Dr., Dr, Col., Maj., and others are removed in the beginning. Also, as Indian names are written differently in northern India and southern India. To mitigate these concerns, various books and websites are referred.

Appendix D

D1: Group Identity and Managerial Compensation (with all controls)

VARIABLES	(1) Total Compensation	(2) Total Compensation	(3) Total Compensation
Homophily Language	0.178*** (0.068)		
Homophily Place		0.033 (0.076)	
Homophily Caste			-0.718*** (0.112)
Size	-0.032 (0.064)	-0.035 (0.064)	-0.037 (0.063)
Firm Age	0.201*** (0.043)	0.200*** (0.043)	0.169*** (0.043)
Institutional Ownership	0.012*** (0.004)	0.011*** (0.004)	0.011** (0.004)
Promoters Ownership	0.016*** (0.005)	0.015*** (0.005)	0.014*** (0.005)
ROA	0.602* (0.336)	0.614* (0.336)	0.671** (0.334)
Leverage	-0.040 (0.276)	-0.025 (0.277)	-0.012 (0.275)
Volatility	-5.788 (5.267)	-5.747 (5.273)	-6.563 (5.240)
BIG5	0.061 (0.102)	0.059 (0.102)	0.072 (0.101)
Number of Meetings	0.011 (0.012)	0.011 (0.012)	0.011 (0.012)
Board Size	-0.017 (0.010)	-0.017 (0.010)	-0.014 (0.010)
Age	0.002 (0.003)	0.002 (0.003)	0.002 (0.003)
Gender	0.585*** (0.219)	0.588*** (0.219)	0.603*** (0.218)
Education	0.014 (0.039)	0.012 (0.039)	0.021 (0.039)
Constant	1.932 (2.168)	2.153 (2.170)	4.065* (2.174)
Observations	3,542	3,542	3,542
R-squared	0.750	0.750	0.753
Manager Designation	Yes	Yes	Yes
Firm	Yes	Yes	Yes
Year	Yes	Yes	Yes

D2: External Validity - Alternative Dependent Variables

VARIABLES	(1) Salary	(2) Salary	(3) Salary	(4) Sales Growth	(5) Sales Growth	(6) Sales Growth
Homophily Language	0.031* (0.019)			0.031*** (0.009)		
Homophily Place		0.057*** (0.019)			0.028*** (0.009)	
Homophily Caste			-0.128*** (0.030)			0.009 (0.014)
Size	0.119*** (0.007)	0.120*** (0.007)	0.120*** (0.007)	0.003 (0.003)	0.003 (0.003)	0.002 (0.003)
Leverage	-0.020 (0.041)	-0.019 (0.041)	-0.019 (0.041)	-0.107*** (0.019)	-0.106*** (0.019)	-0.106*** (0.019)
Firm Age	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	-0.002*** (0.000)	-0.002*** (0.000)	-0.002*** (0.000)
ROA	0.048*** (0.016)	0.047*** (0.016)	0.049*** (0.016)			
Promoters Ownership	0.007*** (0.001)	0.007*** (0.001)	0.007*** (0.001)	0.001*** (0.000)	0.001*** (0.000)	0.001*** (0.000)
Institutional Ownership	0.032*** (0.001)	0.032*** (0.001)	0.032*** (0.001)	0.001** (0.000)	0.001** (0.000)	0.001** (0.000)
Constant	8.400*** (0.124)	8.379*** (0.124)	8.530*** (0.126)	0.117** (0.058)	0.119** (0.058)	0.131** (0.059)
Observations	12,990	12,990	12,990	15,612	15,612	15,612
R-squared	0.256	0.256	0.257	0.076	0.076	0.076
Managerial Designation	Yes	Yes	Yes	No	No	No
Year	Yes	Yes	Yes	Yes	Yes	Yes

D3: Promoter as a Manager

VARIABLES	(1) Total Compensation	(2) Total Compensation	(3) Total Compensation	(4) Firm Value	(5) Firm Value	(6) Firm Value	(7) Residual Firm Value	(8) Residual Firm Value	(9) Residual Firm Value
Homophily Language	0.056 (0.038)			0.186*** (0.068)			0.146** (0.057)		
Homophily Place		0.017 (0.039)			0.156** (0.070)			0.112* (0.059)	
Homophily Caste			-0.189** (0.082)			-0.236 (0.147)			-0.207* (0.123)
Size	0.075** (0.032)	0.076** (0.032)	0.076** (0.032)	0.302*** (0.056)	0.303*** (0.056)	0.306*** (0.057)	0.302*** (0.048)	0.304*** (0.048)	0.302*** (0.048)
Firm Age	0.063** (0.025)	0.063** (0.025)	0.063** (0.025)	0.105** (0.045)	0.108** (0.045)	0.100** (0.045)	0.079** (0.038)	0.079** (0.038)	0.078** (0.038)
Institutional Ownership	0.015*** (0.003)	0.015*** (0.003)	0.015*** (0.003)	0.041*** (0.005)	0.042*** (0.005)	0.041*** (0.005)	0.040*** (0.004)	0.040*** (0.004)	0.040*** (0.004)
Promoters Holding	0.006*** (0.002)	0.006*** (0.002)	0.006*** (0.002)	0.052*** (0.004)	0.052*** (0.004)	0.052*** (0.004)	0.046*** (0.003)	0.046*** (0.003)	0.046*** (0.003)
ROA	1.558*** (0.180)	1.554*** (0.180)	1.546*** (0.180)						
Leverage	-0.210* (0.123)	-0.212* (0.123)	-0.213* (0.123)	-0.218 (0.215)	-0.276 (0.215)	-0.206 (0.216)	-0.286 (0.182)	-0.291 (0.182)	-0.291 (0.182)
Volatility	-3.582** (1.635)	-3.587** (1.635)	-3.548** (1.635)	-11.092*** (2.930)	-11.602*** (2.925)	-10.947*** (2.933)	-11.278*** (2.469)	-11.281*** (2.469)	-11.235*** (2.470)
Big5	0.101 (0.065)	0.101 (0.065)	0.100 (0.065)	0.148 (0.117)	0.129 (0.117)	0.141 (0.118)	0.092 (0.099)	0.090 (0.099)	0.093 (0.099)
Number of Meetings	0.015** (0.006)	0.015** (0.006)	0.015** (0.006)	0.049*** (0.011)	0.049*** (0.011)	0.048*** (0.011)	0.047*** (0.009)	0.047*** (0.009)	0.047*** (0.009)
Board Size	-0.005 (0.007)	-0.006 (0.007)	-0.005 (0.007)	0.014 (0.013)	0.011 (0.013)	0.018 (0.013)	0.013 (0.011)	0.013 (0.011)	0.013 (0.011)
Constant	6.829*** (1.031)	6.852*** (1.032)	7.036*** (1.033)	-10.207*** (1.849)	-10.219*** (1.847)	-9.869*** (1.857)	-10.961*** (1.558)	-10.980*** (1.559)	-10.670*** (1.561)
Observations	10,083	10,083	10,083	10,083	10,083	10,083	10,083	10,083	10,083
R-squared	0.773	0.773	0.773	0.568	0.570	0.568	0.172	0.172	0.172
Manager Designation	Yes	Yes	Yes	No	No	No	No	No	No
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

D4: Identification Test - Change in Homophily and Director's Death or Retirement

VARIABLES	(1) Total_Compensation	(2) Total_Compensation	(3) Total_Compensation	(4) Firm Value	(5) Firm Value	(6) Residual Firm Value	(7) Residual Firm Value	(8) Residual Firm Value
Homophily Language	0.076*** (0.021)					0.086** (0.034)		
DR_C_HL	-0.304*** (0.098)					-0.086 (0.158)		
DRxHomo_Lang	0.130 (0.179)					0.045 (0.288)		
Homophily Place		0.076*** (0.022)		0.061 (0.041)			0.060* (0.035)	
DR_C_HP		-0.354*** (0.096)		0.048 (0.181)			0.116 (0.154)	
DRxHomo_Place		0.210 (0.180)		-0.315 (0.342)			-0.374 (0.290)	
Homophily Caste			-0.132*** (0.041)		-0.134* (0.077)			-0.109* (0.065)
DR_C_HC			0.226 (0.269)		0.648 (0.511)			0.638 (0.433)
DRxHomo_Caste			-0.492* (0.284)		-0.576 (0.538)			-0.573 (0.457)
Size	0.094*** (0.017)	0.095*** (0.017)	0.093*** (0.017)	0.245*** (0.033)	0.245*** (0.033)	0.229*** (0.028)	0.230*** (0.028)	0.228*** (0.028)
Firm Age	0.029* (0.017)	0.028 (0.017)	0.029* (0.017)	0.050 (0.033)	0.048 (0.033)	0.039 (0.028)	0.041 (0.028)	0.041 (0.028)
Institutional Ownership	0.015*** (0.001)	0.015*** (0.001)	0.015*** (0.001)	0.053*** (0.003)	0.053*** (0.003)	0.050*** (0.002)	0.050*** (0.002)	0.050*** (0.002)
Promoters Ownership	0.007*** (0.001)	0.007*** (0.001)	0.007*** (0.001)	0.047*** (0.002)	0.047*** (0.002)	0.042*** (0.002)	0.042*** (0.002)	0.042*** (0.002)
ROA	1.133*** (0.099)	1.131*** (0.099)	1.126*** (0.099)					
Leverage	-0.146** (0.068)	-0.145** (0.068)	-0.148** (0.068)	-0.606*** (0.125)	-0.605*** (0.125)	-0.540*** (0.106)	-0.539*** (0.106)	-0.541*** (0.106)
Volatility	-5.903*** (0.960)	-5.900*** (0.960)	-5.875*** (0.960)	-11.796*** (1.818)	-11.772*** (1.818)	-11.671*** (1.542)	-11.633*** (1.542)	-11.638*** (1.542)
Big5	0.097** (0.038)	0.096** (0.038)	0.097** (0.038)	0.322*** (0.072)	0.320*** (0.072)	0.268*** (0.061)	0.267*** (0.061)	0.267*** (0.061)
Number of Meetings	0.015*** (0.004)	0.015*** (0.004)	0.015*** (0.004)	0.050*** (0.007)	0.049*** (0.007)	0.049*** (0.006)	0.049*** (0.006)	0.049*** (0.006)
Board Size	-0.005 (0.004)	-0.004 (0.004)	-0.004 (0.004)	0.009 (0.007)	0.012* (0.007)	0.008 (0.006)	0.007 (0.006)	0.007 (0.006)
Constant	7.531*** (0.668)	7.522*** (0.668)	7.675*** (0.669)	-6.905*** (1.267)	-6.740*** (1.268)	-8.223*** (1.075)	-8.284*** (1.075)	-8.109*** (1.075)
Observations	28,026	28,026	28,026	28,026	28,026	28,026	28,026	28,026
R-squared	0.762	0.762	0.762	0.577	0.578	0.062	0.062	0.062
Manager Designation	Yes	Yes	Yes	No	No	No	No	No
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

D5: Identities Index, Managerial Compensation and Firm Value

VARIABLES	(1) Total Compensation	(2) Firm Value	(3) Residual Firm Value
Homophily	0.029*** (0.008)	0.033** (0.016)	0.027** (0.013)
Size	0.095*** (0.017)	0.245*** (0.033)	0.230*** (0.028)
Firm Age	0.031* (0.017)	0.049 (0.033)	0.040 (0.028)
ROA	1.132*** (0.099)		
Institutional Ownership	0.015*** (0.001)	0.053*** (0.003)	0.050*** (0.002)
Promoters Ownership	0.007*** (0.001)	0.047*** (0.002)	0.042*** (0.002)
Leverage	-0.146** (0.068)	-0.606*** (0.125)	-0.539*** (0.106)
Volatility	-5.922*** (0.961)	-11.838*** (1.818)	-11.668*** (1.542)
Big 5	0.095** (0.038)	0.321*** (0.072)	0.267*** (0.061)
Number of Meetings	0.015*** (0.004)	0.050*** (0.007)	0.049*** (0.006)
Board Size	-0.005 (0.004)	0.009 (0.007)	0.007 (0.006)
Constant	7.455*** (0.667)	-6.867*** (1.265)	-8.219*** (1.073)
Observations	28,026	28,026	28,026
R-squared	0.762	0.577	0.062
Managerial Designation	Yes	No	No
Firm	Yes	Yes	Yes
Year	Yes	Yes	Yes

D6: Return on Equity (ROE) and Group Identity

VARIABLES	(1) ROE	(2) ROE	(3) ROE	(4) Residual ROE	(5) Residual ROE	(6) Residual ROE
Homophily Language	0.027** (0.013)			0.026** (0.012)		
Homophily Place		0.044*** (0.015)			0.036*** (0.013)	
Homophily Caste			-0.003 (0.022)			-0.001 (0.019)
Size	0.011 (0.013)	0.012 (0.013)	0.010 (0.013)	0.009 (0.011)	0.009 (0.011)	0.008 (0.011)
Firm Age	0.012 (0.009)	0.012 (0.009)	0.011 (0.009)	0.011 (0.007)	0.011 (0.007)	0.011 (0.007)
ROA	1.267*** (0.067)	1.266*** (0.067)	1.269*** (0.067)	1.180*** (0.058)	1.179*** (0.058)	1.182*** (0.058)
Institutional Ownership	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)
Promoters Ownership	-0.000 (0.001)	-0.000 (0.001)	-0.000 (0.001)	0.000 (0.001)	0.000 (0.001)	0.000 (0.001)
Leverage	0.309*** (0.055)	0.311*** (0.055)	0.312*** (0.055)	0.219*** (0.048)	0.220*** (0.048)	0.221*** (0.048)
Volatility	0.162 (1.047)	0.192 (1.046)	0.159 (1.048)	0.045 (0.911)	0.070 (0.910)	0.045 (0.912)
Big 5	0.078*** (0.020)	0.075*** (0.020)	0.079*** (0.020)	0.065*** (0.018)	0.062*** (0.018)	0.065*** (0.018)
Number of Meetings	-0.002 (0.002)	-0.002 (0.002)	-0.002 (0.002)	-0.001 (0.002)	-0.001 (0.002)	-0.001 (0.002)
Board Size	-0.001 (0.002)	-0.001 (0.002)	-0.001 (0.002)	-0.002 (0.002)	-0.002 (0.002)	-0.002 (0.002)
Age	0.001 (0.000)	0.001 (0.000)	0.001 (0.000)	0.001 (0.000)	0.001 (0.000)	0.001 (0.000)
Gender	0.036 (0.042)	0.035 (0.042)	0.037 (0.042)	0.032 (0.036)	0.032 (0.036)	0.033 (0.036)
Education	-0.008 (0.008)	-0.008 (0.008)	-0.008 (0.008)	-0.006 (0.007)	-0.006 (0.007)	-0.006 (0.007)
Constant	-0.823* (0.430)	-0.842** (0.429)	-0.774* (0.434)	-0.850** (0.374)	-0.860** (0.374)	-0.809** (0.377)
Observations	3,541	3,541	3,541	3,536	3,536	3,536
R-squared	0.421	0.422	0.421	0.215	0.216	0.214
Firm	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes

D7: Excluding Chairperson

[illegible]

D8: Profitability and Homophily

VARIABLES	(1) ROA	(2) ROA	(3) Future ROA	(4) Future ROA
Homophily Language	0.006 (0.003)		0.006 (0.003)	
Homophily Place		0.008** (0.004)		0.008** (0.004)
Size	0.002 (0.003)	0.002 (0.003)	0.002 (0.003)	0.002 (0.003)
Firm Age	0.000 (0.002)	0.000 (0.002)	0.000 (0.002)	0.000 (0.002)
Institutional Ownership	0.001*** (0.000)	0.001*** (0.000)	0.001*** (0.000)	0.001*** (0.000)
Promoters Ownership	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Leverage	-0.078*** (0.014)	-0.078*** (0.014)	-0.078*** (0.014)	-0.078*** (0.014)
Volatility	-1.678*** (0.267)	-1.666*** (0.267)	-1.678*** (0.267)	-1.666*** (0.267)
Big 5	-0.001 (0.005)	-0.002 (0.005)	-0.001 (0.005)	-0.002 (0.005)
Number of Meetings	-0.000 (0.001)	-0.000 (0.001)	-0.000 (0.001)	-0.000 (0.001)
Board Size	-0.001* (0.001)	-0.001* (0.001)	-0.001* (0.001)	-0.001* (0.001)
Age	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Gender	0.017 (0.011)	0.017 (0.011)	0.017 (0.011)	0.017 (0.011)
Education	0.002 (0.002)	0.001 (0.002)	0.002 (0.002)	0.001 (0.002)
Constant	0.016 (0.110)	0.015 (0.109)	0.016 (0.110)	0.015 (0.109)
Observations	3,388	3,388	3,388	3,388
R-squared	0.718	0.718	0.718	0.718
Firm Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes